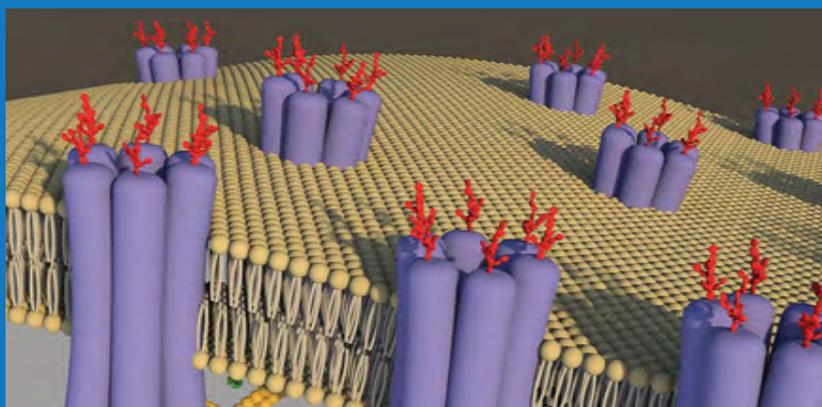
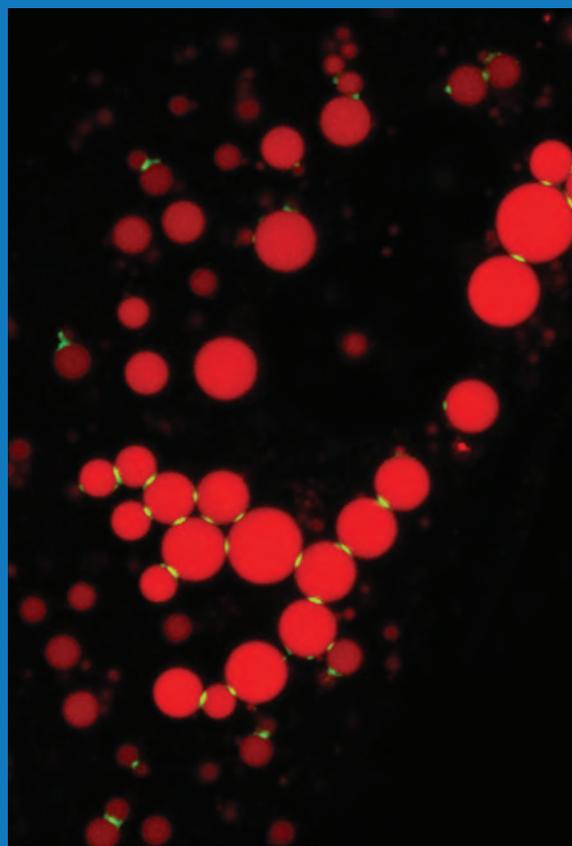
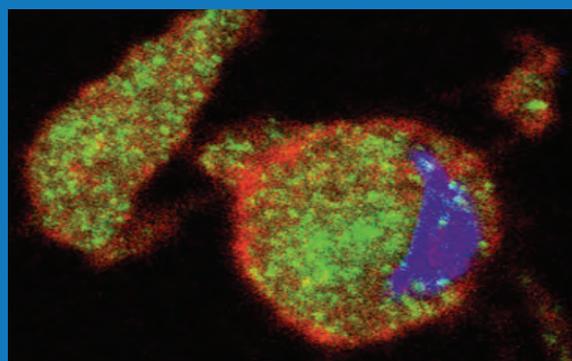
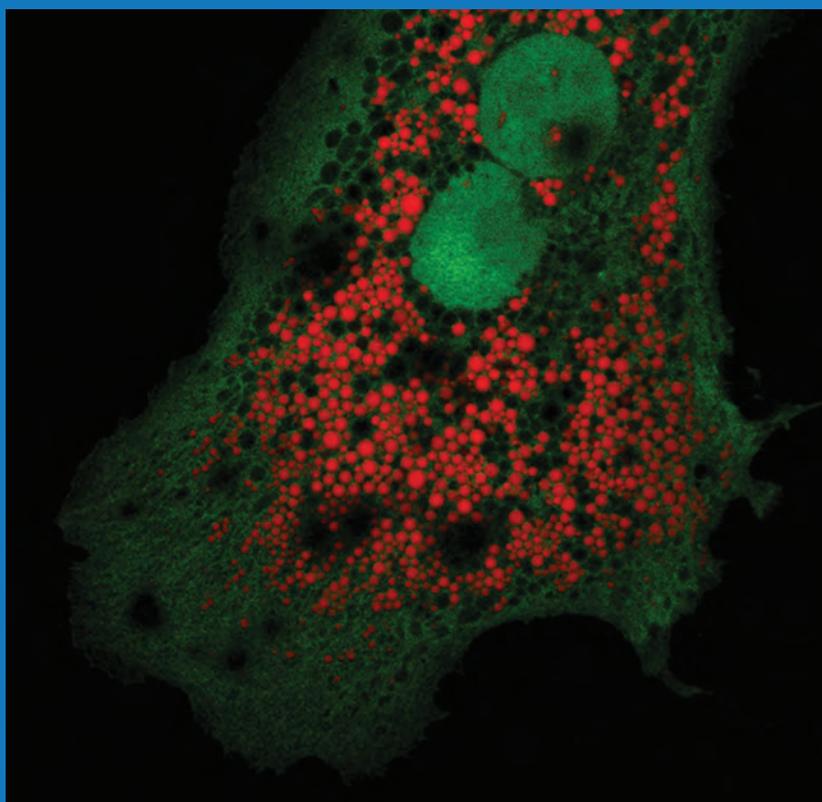


the journal of biological chemistry
jbc

Highlights of 2016



Highlights of 2016

BIOENERGETICS

- 9883 **Asymmetric Functional Conversion of Eubacterial Light-driven Ion Pumps.** *Keiichi Inoue, Yurika Nomura, and Hideki Kandori*

CELL BIOLOGY

- 6096 **YAP Nuclear Localization in the Absence of Cell-Cell Contact Is Mediated by a Filamentous Actin-dependent, Myosin II- and Phospho-YAP-independent Pathway during Extracellular Matrix Mechanosensing.** *Arupratan Das, Robert S. Fischer, Duoqia Pan, and Clare M. Waterman*

COMPUTATIONAL BIOLOGY

- 19474 **An Organic Anion Transporter 1 (OAT1)-centered Metabolic Network.** *Henry C. Liu, Neema Jamshidi, Yuchen Chen, Satish A. Eraly, Sai Yee Cho, Vibha Bhatnagar, Wei Wu, Kevin T. Bush, Ruben Abagyan, Bernhard O. Palsson, and Sanjay K. Nigam*

DEVELOPMENTAL BIOLOGY

- 27334 **RE1-silencing Transcription Factor (REST) Is Required for Nuclear Reprogramming by Inhibiting Transforming Growth Factor β Signaling Pathway.** *Qing-Ran Kong, Bing-Teng Xie, Heng Zhang, Jing-Yu Li, Tian-Qing Huang, Ren-Yue Wei, and Zhong-Hua Liu*

DNA AND CHROMOSOMES

- 21956 **Ewing Tumor-associated Antigen 1 Interacts with Replication Protein A to Promote Restart of Stalled Replication Forks.** *Sumin Feng, Yichao Zhao, Yixi Xu, Shaokai Ning, Wei Huo, Mei Hou, Ge Gao, Jianguo Ji, Rong Guo, and Dongyi Xu*

ENZYMOLGY

- 8877 **Demonstration That the Radical S-Adenosylmethionine (SAM) Enzyme PqqE Catalyzes *de Novo* Carbon-Carbon Cross-linking within a Peptide Substrate PqqA in the Presence of the Peptide Chaperone PqqD.** *Ian Barr, John A. Latham, Anthony T. Iavarone, Teera Chantarojsiri, Jennifer D. Hwang, and Judith P. Klinman*

GENE REGULATION

- 9853 **Biochemical Analysis of Yeast Suppressor of Ty 4/5 (Spt4/5) Reveals the Importance of Nucleic Acid Interactions in the Prevention of RNA Polymerase II Arrest.** *J. Brooks Crickard, Jianhua Fu, and Joseph C. Reese*

GENOMICS AND PROTEOMICS

- 15342 **Preferential Phosphorylation on Old Histones during Early Mitosis in Human Cells.** *Shu Lin, Zuo-Fei Yuan, Yumiao Han, Dylan M. Marchione, and Benjamin A. Garcia*

GLYCOBIOLOGY AND EXTRACELLULAR MATRICES

- 4268 **The E3 Ubiquitin Ligase Adaptor Protein Skp1 Is Glycosylated by an Evolutionarily Conserved Pathway That Regulates Protist Growth and Development.** *Kazi Rahman, Peng Zhao, Msano Mandalasi, Hanke van der Wel, Lance Wells, Ira J. Blader, and Christopher M. West*

IMMUNOLOGY

- 103 **A Genome-wide CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) Screen Identifies NEK7 as an Essential Component of NLRP3 Inflammation Activation.** *Jonathan L. Schmid-Burgk, Dhruv Chauhan, Tobias Schmidt, Thomas S. Ebert, Julia Reinhardt, Elmar Endl, and Veit Hornung*

LIPIDS

- 4282 **Differential Roles of Cell Death-inducing DNA Fragmentation Factor- α -like Effector (CIDE) Proteins in Promoting Lipid Droplet Fusion and Growth in Subpopulations of Hepatocytes.** *Wenyi Xu, Lizhen Wu, Miao Yu, Feng-Jung Chen, Muhammad Arshad, Xiayu Xia, Hao Ren, Jinhai Yu, Li Xu, Dijin Xu, John Zhong Li, Peng Li, and Linkang Zhou*

MEMBRANE BIOLOGY

- 1014 **Sar1 GTPase Activity Is Regulated by Membrane Curvature.** *Michael G. Hanna IV, Ioanna Mela, Lei Wang, Robert M. Henderson, Edwin R. Chapman, J. Michael Edwardson, and Anjon Audhya*

METABOLISM

- 21848 **Metformin Improves Metabolic Memory in High Fat Diet (HFD)-induced Renal Dysfunction.** *Kulbhushan Tikoo, Ekta Sharma, Venkateswara Rao Amara, Himani Pamulapati, and Vaibhav Shrirang Dhawale*

METHODS AND RESOURCES

- 24036 **Selenoprotein Gene Nomenclature.** *Vadim N. Gladyshev, Elias S. Arnér, Marla J. Berry, Regina Brigelius-Flohé, Elspeth A. Bruford, Raymond F. Burk, Bradley A. Carlson, Sergi Castellano, Laurent Chavatte, Marcus Conrad, Paul R. Copeland, Alan M. Diamond, Donna M. Driscoll, Ana Ferreiro, Leopold Flohé, Fiona R. Green, Roderic Guigó, Diane E. Handy, Dolph L. Hatfield, John Hesketh, Peter R. Hoffmann, Arne Holmgren, Robert J. Hondal, Michael T. Howard, Kaixun Huang, Hwa-Young Kim, Ick Young Kim, Josef Köhrle, Alain Krol, Gregory V. Kryukov, Byeong Jae Lee, Byung Cheon Lee, Xin Gen Lei, Qiong Liu, Alain Lescure, Alexei V. Lobanov, Joseph Loscalzo, Matilde Maiorino, Marco Mariotti, K. Sandeep Prabhu, Margaret P. Rayman, Sharon Rozovsky, Gustavo Salinas, Edward E. Schmidt, Lutz Schomburg, Ulrich Schweizer, Miljan Simonović, Roger A. Sunde, Petra A. Tsuji, Susan Tweedie, Fulvio Ursini, Philip D. Whanger, and Yan Zhang*

MICROBIOLOGY

- 22686 **Accumulation of Peptidoglycan O-Acetylation Leads to Altered Cell Wall Biochemistry and Negatively Impacts Pathogenesis Factors of *Campylobacter jejuni*.** *Reuben Ha, Emilisa Frirdich, David Sychantha, Jacob Biboy, Michael E. Taveirne, Jeremiah G. Johnson, Victor J. DiRita, Waldemar Vollmer, Anthony J. Clarke, and Erin C. Gaynor*

MOLECULAR BASES OF DISEASE

- 12432 **A Germline Variant in the PANX1 Gene Has Reduced Channel Function and Is Associated with Multisystem Dysfunction.** *Qing Shao, Kristin Lindstrom, Ruoyang Shi, John Kelly, Audrey Schroeder, Jane Juusola, Kara L. Levine, Jessica L. Esseltine, Silvia Penuela, Michael F. Jackson, and Dale W. Laird*

MOLECULAR BIOPHYSICS

- 12907 **Structure and Dynamics of Single-isoform Recombinant Neuronal Human Tubulin.** *Annapurna Vemu, Joseph Atherton, Jeffrey O. Spector, Agnieszka Szyk, Carolyn A. Moores, and Antonina Roll-Mecak*

NEUROBIOLOGY

- 9872 **Brain-derived Neurotrophic Factor in Megakaryocytes.** *Pedro Chacón-Fernández, Katharina Säuberli, Maria Colzani, Thomas Moreau, Cedric Ghevaert, and Yves-Alain Barde*

PLANT BIOLOGY

- 20270 **Structural Basis of Host Autophagy-related Protein 8 (ATG8) Binding by the Irish Potato Famine Pathogen Effector Protein PexRD54.** *Abbas Maqbool, Richard K. Hughes, Yasin F. Dagdas, Nicholas Tregidgo, Erin Zess, Khaoula Belhaj, Adam Round, Tolga O. Bozkurt, Sophien Kamoun, and Mark J. Banfield*

PROTEIN STRUCTURE AND FOLDING

- 14430 **Structural Insights into the Polyphyletic Origins of Glycyl tRNA Synthetases.** *Marco Igor Valencia-Sánchez, Annia Rodríguez-Hernández, Ruben Ferreira, Hugo Aníbal Santamaría-Suárez, Marcelino Arciniega, Anne-Catherine Dock-Bregeon, Dino Moras, Brice Beinstainer, Haydyn Mertens, Dmitri Svergun, Luis G. Brieba, Morten Grøtli, and Alfredo Torres-Larios*

PROTEIN SYNTHESIS AND DEGRADATION

- 14526 **Conserved Sequence Preferences Contribute to Substrate Recognition by the Proteasome.** *Houqing Yu, Amit K. Singh Gautam, Shameika R. Wilmington, Dennis Wylie, Kirby Martinez-Fonts, Grace Kago, Marie Warburton, Sreenivas Chavali, Tomonao Inobe, Ilya J. Finkelstein, M. Madan Babu, and Andreas Matouschek*

RNA

- 2412 **Autoinhibitory Interdomain Interactions and Subfamily-specific Extensions Redefine the Catalytic Core of the Human DEAD-box Protein DDX3.** *Stephen N. Floor, Kendall J. Condon, Deepak Sharma, Eckhard Jankowsky, and Jennifer A. Doudna*

SIGNAL TRANSDUCTION

- 5528 **Different Epidermal Growth Factor Receptor (EGFR) Agonists Produce Unique Signatures for the Recruitment of Downstream Signaling Proteins.** *Tom Ronan, Jennifer L. Macdonald-Obermann, Lorel Huelsmann, Nicholas J. Bessman, Kristen M. Naegle, and Linda J. Pike*

Asymmetric Functional Conversion of Eubacterial Light-driven Ion Pumps^{*S}

Received for publication, January 18, 2016, and in revised form, February 27, 2016. Published, JBC Papers in Press, February 29, 2016, DOI 10.1074/jbc.M116.716498

Keiichi Inoue^{‡S}, Yurika Nomura[‡], and Hideki Kandori^{‡S1}

From the [‡]Department of Frontier Materials and ^SOptoBioTechnology Research Center, Nagoya Institute of Technology, Showa-ku, Nagoya 466-8555, Japan and [¶]PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

In addition to the well-known light-driven outward proton pumps, novel ion-pumping rhodopsins functioning as outward Na⁺ and inward Cl⁻ pumps have been recently found in eubacteria. They convert light energy into transmembrane electrochemical potential difference, similar to the prototypical archaeal H⁺ pump bacteriorhodopsin (BR) and Cl⁻ pump halorhodopsin (HR). The H⁺, Na⁺, and Cl⁻ pumps possess the conserved respective DTE, NDQ, and NTQ motifs in the helix C, which likely serve as their functional determinants. To verify this hypothesis, we attempted functional interconversion between selected pumps from each category by mutagenesis. Introduction of the proton-pumping motif resulted in successful Na⁺ → H⁺ functional conversion. Introduction of the respective characteristic motifs with several additional mutations leads to successful Na⁺ → Cl⁻ and Cl⁻ → H⁺ functional conversions, whereas remaining conversions (H⁺ → Na⁺, H⁺ → Cl⁻, Cl⁻ → Na⁺) were unsuccessful when mutagenesis of 4–6 residues was used. Phylogenetic analysis suggests that a H⁺ pump is the common ancestor of all of these rhodopsins, from which Cl⁻ pumps emerged followed by Na⁺ pumps. We propose that successful functional conversions of these ion pumps are achieved exclusively when mutagenesis reverses the evolutionary amino acid sequence changes. Dependence of the observed functional conversions on the direction of evolution strongly suggests that the essential structural mechanism of an ancestral function is retained even after the gain of a new function during natural evolution, which can be evoked by a few mutations. By contrast, the gain of a new function needs accumulation of multiple mutations, which may not be easily reproduced by limited mutagenesis *in vitro*.

Animal and microbial rhodopsins are photoreceptive proteins, which use retinal as their chromophore. The main function of several groups of microbial rhodopsins is light-driven ion transport across cellular membranes (1). In the last decade, these proteins became powerful tools in neuroscience, where they are used to control neural activity of animals by light (so-

called “optogenetics”) (2, 3). Bacteriorhodopsin (BR)² and HR, the first ion-transporting rhodopsins, were discovered in 1971 and 1977, respectively, from halophilic archaea (4, 5). Since 2000, genomic and metagenomic sequencing revealed that microbial rhodopsins are widely distributed among marine and freshwater bacteria, most of which were classified as light-driven H⁺ pumps (proteorhodopsins and xanthorhodopsins, PR and XR) (Fig. 1) (6, 7). In addition, eubacterial light-driven Na⁺ and Cl⁻ pumps have been discovered recently (Fig. 1) (8–10), with the former forming a new functional class and the latter not being related to haloarchaeal HRs. These pumps are interesting not only from biophysical and biochemical point of view, but also for evolutionary and environmental microbiology and optogenetic applications. Recent structure determination of a light-driven Na⁺ pump *Krokinobacter eikastus* rhodopsin 2 (KR2) accelerated our understanding of ion selectivity, and K⁺ pump was engineered based on that structure (11, 12).

Functional interconversion of homologous proteins by site-directed mutagenesis is a powerful tool to understand molecular mechanisms of their functions. The first example of successful functional conversion of microbial rhodopsins is when BR was converted into an HR-like Cl⁻ pump by a single amino acid replacement (13, 14). The conserved functional helix C motifs for BR and HR are DTD and TSA, respectively. In BR, Asp-85 (the first Asp in the DTD motif) is the H⁺ acceptor from the protonated Schiff base during H⁺ pumping (1), and when Asp-85 is mutated to Thr, the corresponding amino acid in HR, D85T BR pumps Cl⁻ inwards, identical to HR (13, 14). This observation implied that BR and HR share a common transport mechanism, and its selectivity is determined at the position of primary proton acceptor (Asp-85 in BR). While this was a clear-cut result, the respective reverse mutation (Thr → Asp) did not convert HR into a H⁺ pump (15, 16). Ten additional mutations to make *pharaonis* HR resemble BR did not work either, indicating that functional conversion between archaeal H⁺ and Cl⁻ pumps is asymmetric (17). The vibrational analysis suggested that protein-bound water molecules may be responsible for the asymmetric functional conversion (17). While two conversions (H⁺ ⇌ Cl⁻) were tested for archaeal pumps, eubacterial rhodopsins may undergo six functional conversions (H⁺ ⇌ Cl⁻, H⁺ ⇌ Na⁺, Na⁺ ⇌ Cl⁻). In the present study, we tested

* This work was supported by grants from the Japanese Ministry of Education, Culture, Sports, Science, and Technology (to K.I.) (26708001, 26115706, 26620005) and to H.K. (25104009, 15H02391). The authors declare that they have no conflicts of interest with the contents of this article.

^S This article contains supplemental Table S1.

¹ To whom correspondence should be addressed: Dept. of Frontier Materials, Nagoya Inst. of Technology, Showa-ku, Nagoya 466-8555, Japan. Tel.: +81-52-735-5207; Fax: 81-52-735-5207; E-mail: kandori@nitech.ac.jp.

² The abbreviations used are: BR, bacteriorhodopsin; HR, halorhodopsin; KR, *Krokinobacter eikastus* rhodopsin, DDM, n-dodecyl-β-D-maltoside; GR, *Gloeobacter* rhodopsin; FR, *Fulvimarina* rhodopsin; CCCP, carbonylcyanide m-chlorophenylhydrazone.

Functional Interconversion of H^+ , Na^+ , and Cl^- Pump Rhodopsin

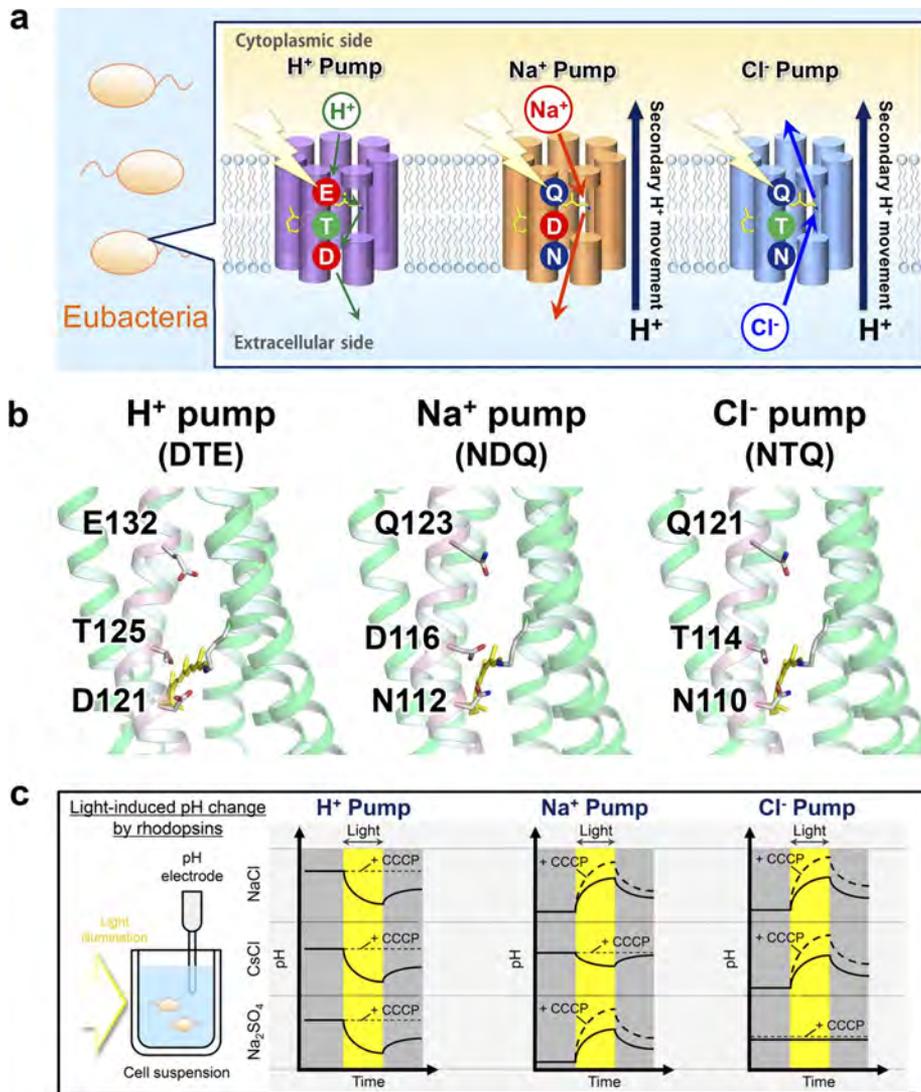


FIGURE 1. Light-driven ion-pumps in eubacteria and their transport activity. *a*, eubacterial rhodopsins contain DTE, NDQ, and NTQ motifs for H^+ , Na^+ , and Cl^- pumps, respectively. *b*, motifs in GR, KR2, and FR based on the structure of XR (left, PDB ID: 3DDL) and KR2 (middle, right, PDB ID: 3X3C). *c*, ion pumping activity assayed by pH changes. Light-driven H^+ pump shows pH decrease in the presence of any ions, which is diminished by a protonophore CCCP. In contrast, Na^+ and Cl^- pumps cause pH increase, which is enhanced by CCCP. Light-driven Na^+ and Cl^- pumps are distinguished by their cation- and anion-dependent transport, respectively.

whether a limited number of mutations can achieve functional interconversion between eubacterial pumps.

Experimental Procedures

Gene Construction, Protein Expression, and Purification—KR2 and FR genes whose codons were optimized for *Escherichia coli* expression system were synthesized (Eurofins Genomics Inc.) and subcloned into a pET21a(+)-vector with C-terminal 6×His-tag (10). A plasmid vector of GR with the additional C-terminal 6×His-tag was constructed as described previously (18). For mutagenesis, a QuikChange site-directed mutagenesis kit (Stratagene) was used according to the standard protocol. Wild-type and mutant proteins were expressed in *E. coli* C41(DE3) strain. The protein expression was induced by 0.21 mM isopropyl β -D-thiogalactopyranoside (IPTG) for 4 h at 37 °C with the supplement of 10 μ M all-*trans*-retinal (Sigma-Aldrich) to the culture. For the measurement of Cl^- dependence of absorption spectra, the expressed proteins were purified from *E. coli* cells according to the previously reported

protocols (8, 17). The cells were disrupted by French Press (Ohtake), and the membrane fraction was collected by ultracentrifugation ($125,000 \times g$, 1 h). The protein was solubilized with 2% n-dodecyl- β -D-maltoside (DDM) (Anatrace) in the presence of 300 mM NaCl, 5 mM imidazole, and 50 mM MES (pH 6.5). After Co^{2+} -NTA affinity chromatography, the collected fractions were dialyzed to the solution containing 100 mM NaCl, 50 mM Tris-HCl (pH 7.0), 0.02% DDM to remove imidazole used for elution from the column.

Assay of Light-driven Pump Activity of Rhodopsins in *E. coli* Cells—*E. coli* expressing rhodopsins were harvested by centrifugation ($3,600 \times g$, 3 min), washed three times and resuspended in the aqueous solvent containing 100 mM salt (NaCl, CsCl, or Na_2SO_4). 7.5 ml of cell suspension at $OD_{660} = 2$ was placed in the dark and then illuminated at $\lambda > 500$ nm by a 1-kW tungsten-halogen projector lamp (Rikagaku, Japan) through a glass filter (Y-52, AGC Techno Glass, Japan). The light-induced pH changes were measured by a pH electrode (HORIBA, Japan).

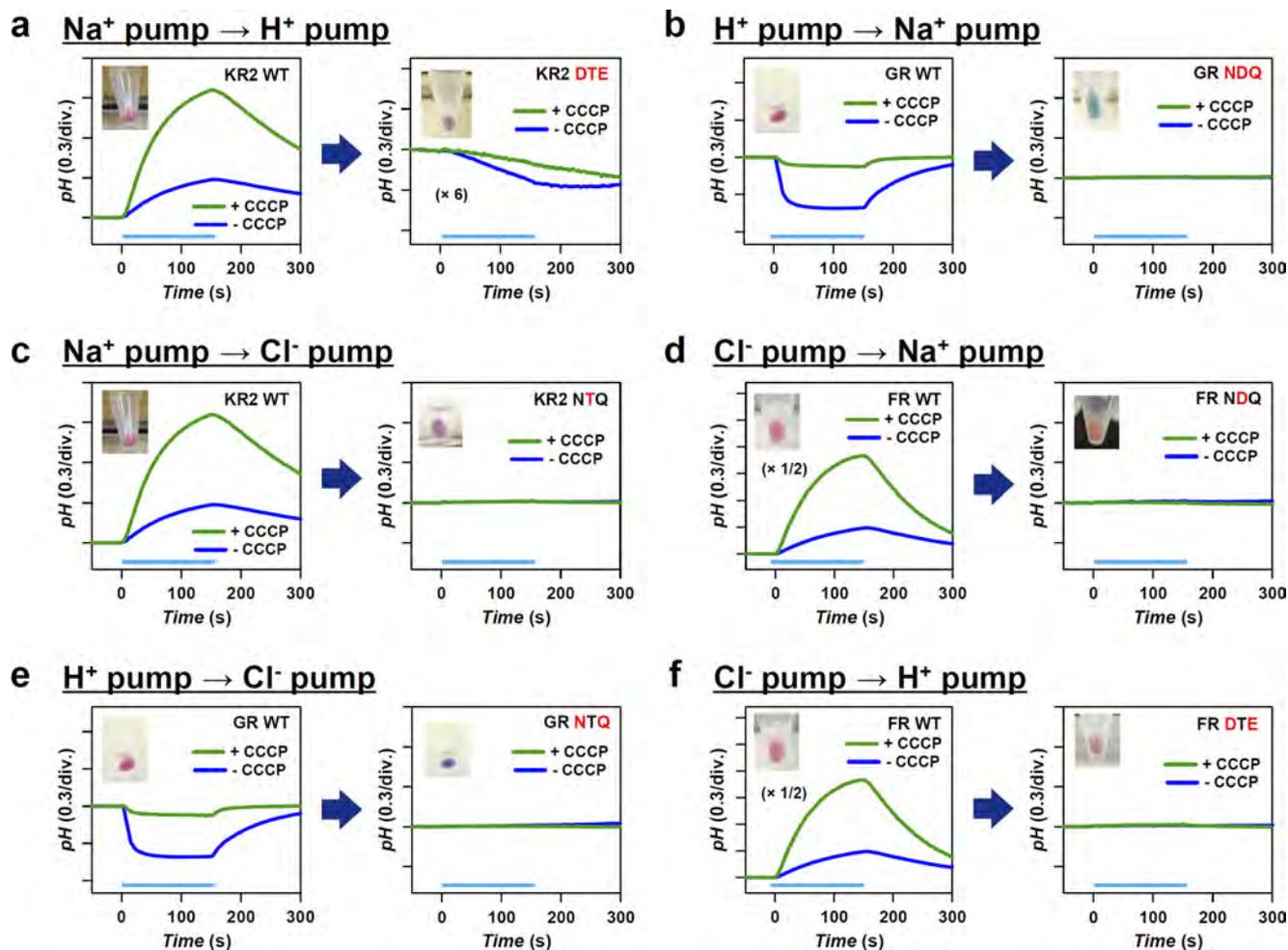


FIGURE 2. **Functional conversion of light-driven H^+ , Na^+ , and Cl^- pumps by replacing the characteristic helix C motifs.** *a–f*, three, one, and two amino acids in each motif are replaced for the functional conversions of $H^+ \rightleftharpoons Na^+$ (*a, b*), $Na^+ \rightleftharpoons Cl^-$ (*c, d*), and $H^+ \rightleftharpoons Cl^-$ (*e, f*), respectively. *Blue* and *green traces* represent pH changes upon illumination in the absence and presence of CCCP, respectively. All measurements are performed in the presence of 100 mM NaCl (pH ~ 7.0). The KR2 DTE mutant shows H^+ pumping activity, indicating a successful $Na^+ \rightarrow H^+$ pump conversion by the motif mutation (*a*). All other mutants exhibit no pumping activities (*b–f*).

Quantification of Rhodopsins Expressed in *E. coli*—To quantify the expressed rhodopsin in *E. coli* cell used for pump activity assay, the same amount of cells used for pump activity assay was collected by low-speed centrifugation at $3,600 \times g$ and $4^\circ C$ and suspended in the solution containing 100 mM NaCl, 50 mM Tris-HCl (pH 8.0), to a final volume of 3 ml. Then, 200 μ l of 1 mM lysozyme was added to the suspension and it was gently stirred at room temperature for 1 h. The *E. coli* cells were disrupted by sonication (TAITEC, Japan) and solubilized in 3.0% DDM. The absorption change, representing the bleaching of rhodopsin by hydroxylamine (HA), was measured with a UV-vis spectrometer (Shimadzu, Japan) equipped with an integrating sphere after the addition of HA (100 mM NaCl, 50 mM Tris-HCl (pH 8.0)) to a final concentration of 500 mM HA and illumination at $\lambda > 500$ nm by a 1-kW tungsten-halogen projector lamp (Rikagaku, Japan) through a glass filter (Y-52, AGC Techno Glass, Japan). The molecular extinction coefficient of rhodopsin (ϵ) was calculated from the ratio between the absorbance of rhodopsin and retinal oxime ($\epsilon = 33,600 M^{-1} cm^{-1}$ at 360 nm) produced by the reaction between retinal Schiff base and HA (19). Molar extinction coefficients and absorption maxima wavelengths of each mutant are shown in Figs. 3, 4, and

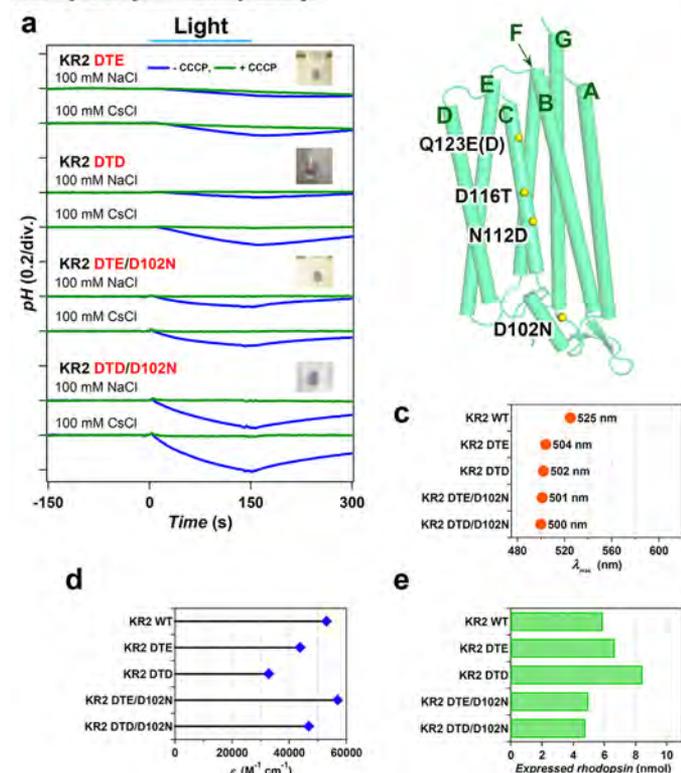
5. The amount of rhodopsin expressed in *E. coli* cells was determined by the absorbance of the bleached rhodopsin and the value of ϵ , which is shown in Figs. 3, *e, h*; 4, *e, h*; 5, *e* and *h*. The transport activity of *E. coli* cells containing each rhodopsin was quantitatively determined from the initial slope of pH change after normalizing the expression level of protein, which is summarized in Fig. 6.

Chloride Titrations—To measure the Cl^- -concentration dependence of absorption spectra of rhodopsins, the proteins were dialyzed to the solution containing 0.1% DDM and 20 mM HEPES (pH 7.0). NaCl was added to the sample and UV-visible absorption spectra were measured with a UV-vis spectrometer (JASCO, Japan) equipped with an integrating sphere at various Cl^- -concentrations.

Laser Flash Photolysis—The transient absorption change after the photo excitation of the rhodopsins were investigated by laser flash photolysis method (8, 11). The purified rhodopsin was reconstituted into the lipid bilayer of the mixture of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE) and 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (POPG) (molecular ratio: POPE:POPG = 3:1) with the protein-to-lipid molecular ratio of 1:20 at 2–4 μ M protein con-

Functional Interconversion of H^+ , Na^+ , and Cl^- Pump Rhodopsin

Na^+ pump $\rightarrow H^+$ pump



H^+ pump $\rightarrow Na^+$ pump

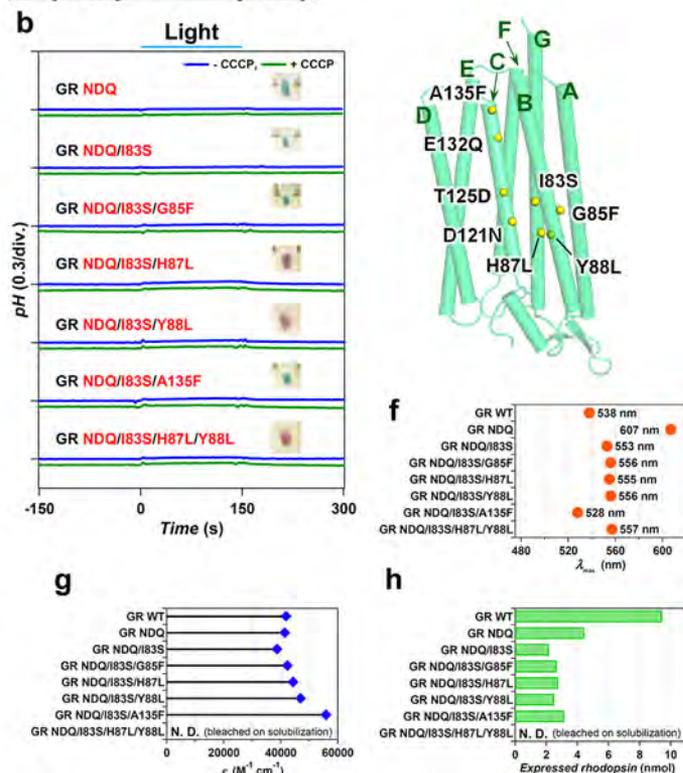


FIGURE 3. Optimization of functional conversions between light-driven Na^+ and H^+ pumps by additional mutations outside of the conserved helix C motifs. Improvement of functional conversion of $Na^+ \rightarrow H^+$ pump in KR2 (*a*) in the presence of 100 mM NaCl or CsCl (pH ~ 7.0) and unsuccessful functional conversion of $H^+ \rightarrow Na^+$ pump in GR (*b*) in the presence of 100 mM NaCl (pH ~ 7.0) in the absence (*blue*) and presence (*green*) of CCCP, respectively. The positions of mutations are indicated by yellow spheres in the modeled structure (*right*) and the names of helices are shown with green characters. The wavelength of absorption maxima (λ_{max}) (*c*, *f*), molecular extinction coefficients (*e*, *h*), and the amount of expressed rhodopsin in the *E. coli* cell used for the pump activity assay (*e*, *h*) of wild type and mutant proteins.

centration. The reconstituted sample was suspended in 100 mM NaCl, 20 mM HEPES-NaOH (pH 7). The sample solution was illuminated with a second harmonics generation of nanosecond pulsed Nd^{3+} -YAG laser ($\lambda = 532$ nm, INDI40, Spectra-Physics, CA) with the pulse energy of 3.8 mJ/cm² pulse. The transient absorption spectrum of rhodopsin after the laser excitation was obtained by measuring the intensity of white light passed through the sample before and after laser excitation at $\lambda = 350$ – 750 nm with an ICCD linear array detector (C8808–01, Hamamatsu, Japan). To increase signal-to-noise ratio, 90 identical spectra were averaged and singular-value-decomposition (SVD) analysis was applied (8).

Phylogenetic Analysis of Rhodopsin Genes—The amino acid sequences of rhodopsins were aligned using MUSCLE program (20) after the removal of weakly conserved interhelical loop, and N- and C-terminal extensions to increase the accuracy of alignment. The evolutionary history was inferred using the Neighbor-Joining method (21). The optimal tree with the sum of branch length = 23.00061278 was obtained. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) were calculated (22). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (23) and are in the units of the number of amino acid substitutions per site. The analysis involved

128 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 105 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (24).

Results and Discussion

We selected *Gloeobacter* rhodopsin (GR) (18), *Krokinobacter eikastus* rhodopsin 2 (KR2) (8), and *Fulvimarina* rhodopsin (FR) (10) as the representatives of H^+ , Na^+ , and Cl^- pumps, respectively. The amino acid identity is 26% between GR and KR2, 33% between GR and FR, and 33% between KR2 and FR (see supplemental Table S1 for all amino acid sequences). It should be noted that the eubacterial H^+ , Na^+ , and Cl^- pumps possess the highly conserved DTE, NDQ, and NTQ motifs in the helix C, respectively (Fig. 1, *a* and *b*), positions of which correspond to Asp-85, Thr-89, and Asp-96 in BR (25). As these motifs are likely to determine the function, we first examined functional conversions by altering them. Note that two, three, and one mutation(s) are needed for $H^+ \rightleftharpoons Cl^-$, $H^+ \rightleftharpoons Na^+$, and $Na^+ \rightleftharpoons Cl^-$, respectively. It should be also noted that a light-driven Na^+ pump KR2 acts as a H^+ pump in the presence of larger cations (8). Therefore, we define the $Na^+ \rightarrow H^+$ pump conversion for the KR2 mutants as successful if it pumps H^+ even in 100 mM NaCl.

To verify function of each rhodopsin mutant, the proteins were overexpressed in *E. coli* (C41 (DE3) strain) and light-in-

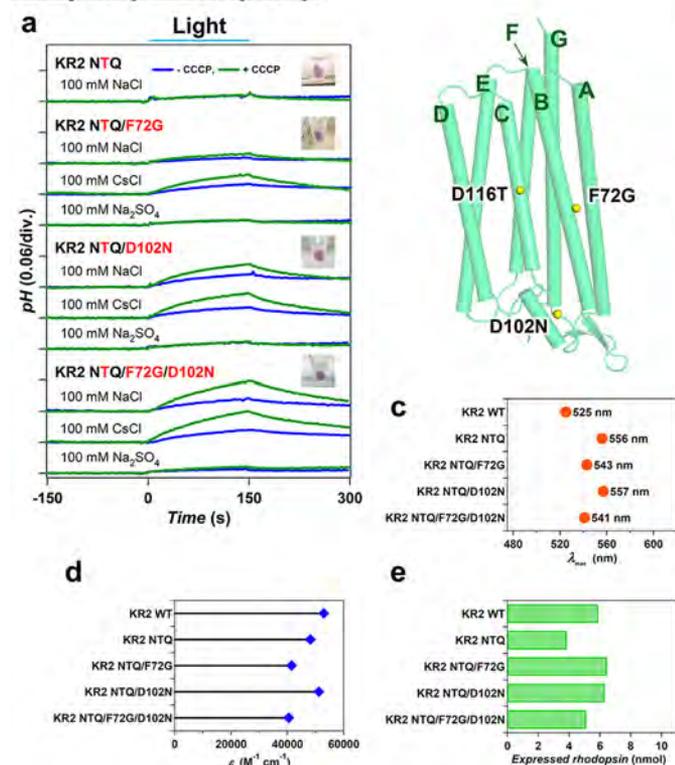
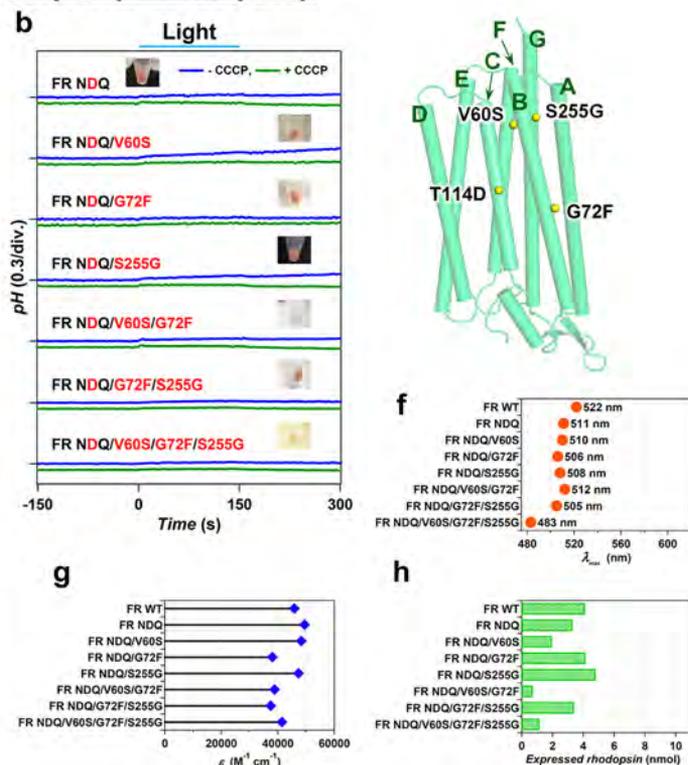
Na⁺ pump → Cl⁻ pumpCl⁻ pump → Na⁺ pump

FIGURE 4. Optimization of functional conversions between light-driven Na⁺ and Cl⁻ pumps by additional mutations outside of the conserved helix C motifs. Successful functional conversion of Na⁺ → Cl⁻ pump in KR2 (a) in the presence of 100 mM NaCl, CsCl, or Na₂SO₄ (pH ~7.0) and unsuccessful functional conversion of Cl⁻ → Na⁺ pump in FR (b) in the presence of 100 mM NaCl (pH ~7.0) in the absence (blue) and presence (green) of CCCP, respectively. The positions of mutations are indicated by yellow spheres in the modeled structure (right), and the names of helices are shown with green characters. The λ_{max} (c, f), ε (d, g) and the amount of expressed rhodopsin in the *E. coli* cell used for the pump activity assay (e, h) of wild type and mutant proteins.

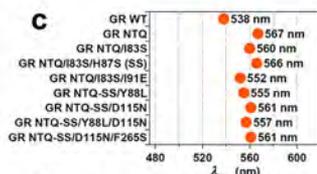
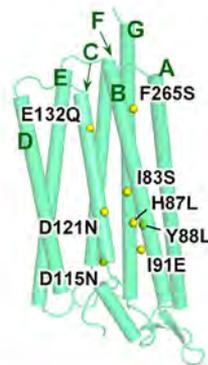
duced pH changes of *E. coli* cell suspensions containing each protein were measured. Identification of H⁺, Na⁺, and Cl⁻ pumps by the pH monitoring system is summarized in Fig. 1c. A cell suspension of *E. coli* expressing H⁺ pump shows acidification of external medium (pH decrease) upon illumination by the outward transport of H⁺ irrespective of the ionic species contained in the solvent, and the signals are diminished in the presence of a protonophore, carbonylcyanide *m*-chlorophenylhydrazine (CCCP) (Fig. 1c, left panel). In contrast, a cell suspension of *E. coli* expressing Na⁺ pump or Cl⁻ pump shows alkalinization (pH increase) upon illumination, because of secondary H⁺-uptake occurs to compensate the increased negative membrane potential inside the cells. The signals are further enhanced in the presence of CCCP, because the observed pH changes originate from secondary H⁺ movement due to cation or anion transport. Analysis of ion dependence is used to distinguish cation and anion pumps. Na⁺ pump KR2 is converted to H⁺ pump in the solvent containing the salt of larger cations such as CsCl, while it pumps Na⁺ outwards in Na₂SO₄ (Fig. 1c, middle panel). On the other hand, while Cl⁻ pump FR transports Cl⁻ ion in the solution containing NaCl or CsCl, it does not transport larger anions such as SO₄²⁻, and no alkalinization occurs in the solution of Na₂SO₄ (Fig. 1c, right panel). Thus, comparison of signals in NaCl, CsCl, and Na₂SO₄ allows distinguishing between the transported ions. Fig. 2 shows the results of functional conversion experiments performed by the replacement of the characteristic helix C motifs. Among the six

functional conversions, only one conversion was achieved, from Na⁺ to H⁺ pump (Fig. 2a). This may be consistent with the fact that Na⁺-pumping KR2 transports H⁺ in the presence of larger cations (8), suggesting that KR2 retained the molecular machinery of H⁺ pump in its structure. On the other hand, none of the other motif-altering mutations achieved functional conversion (Fig. 2, b–f). Thus, we introduced additional plausible mutations based on the sequence comparisons outside of the conserved helix C motifs.

Figs. 3, 4, and 5 summarize the results of functional conversions between H⁺, Na⁺, and Cl⁻ pumps which used such additional mutations. Fig. 3a shows the improvement of the Na⁺ → H⁺ pump conversion. Even though the DTE KR2 mutant pumps protons (Fig. 2a), the efficiency of transport estimated from the initial slope of pH change is very low (0.42% of that of GR, Fig. 6a). We found that the DTD KR2 mutant pumps better (0.79%, Fig. 6a), and the transport was more efficient in Cs⁺ than in Na⁺ (Fig. 3a). The latter may be related to the fact that KR2 possesses Na⁺ binding site at the extracellular surface (8), which inhibits outward H⁺ pumping only in Na⁺. As Asp-102 constitutes the Na⁺ binding site (11, 12), we replaced this residue, and the DTE/D102N and DTD/D102N KR2 mutants exhibited similar H⁺ pumping activities in NaCl and CsCl (3.0 and 5.6% of GR, respectively, Fig. 6a). The functional conversion of Na⁺ → H⁺ pump is thus achieved by 4 mutations. In contrast, the reverse motif mutation for H⁺ → Na⁺ pump conversion (DTE to NDQ) has not worked (Fig. 2b), which may be

Functional Interconversion of H⁺, Na⁺, and Cl⁻ Pump Rhodopsin

H⁺ pump → Cl⁻ pump



Cl⁻ pump → H⁺ pump

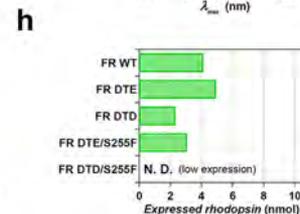
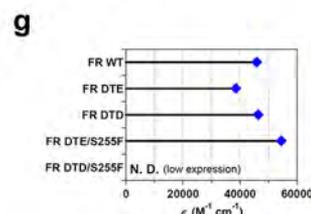
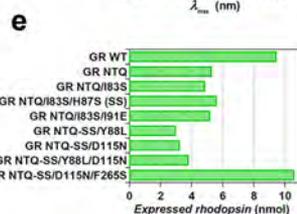
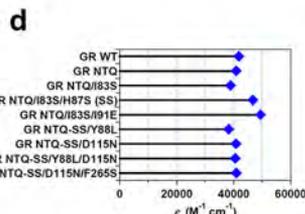
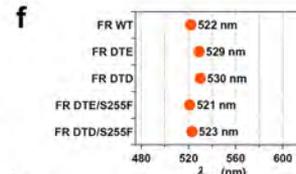
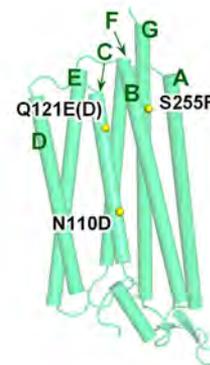
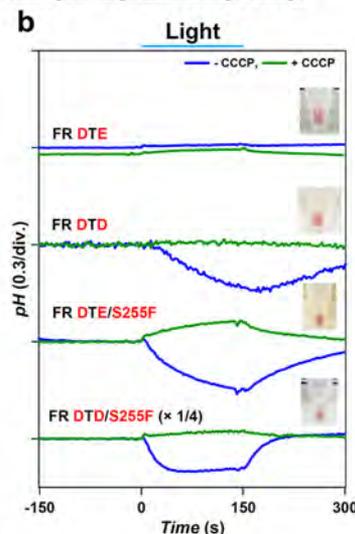


FIGURE 5. Optimization of functional conversions between light-driven H⁺ and Cl⁻ pumps by additional mutations outside of the conserved helix C motifs. Unsuccessful functional conversion of H⁺ → Cl⁻ pump in GR (a) and successful functional conversion of Cl⁻ → H⁺ pump in FR (b) in the presence of 100 mM NaCl (pH ~7.0) in the absence (blue) and presence (green) of CCCP, respectively. The positions of mutations are indicated by yellow spheres in the modeled structure (right), and the names of helices are shown with green characters. The λ_{\max} (c, f), ϵ (d, g) and the amount of expressed rhodopsin in the *E. coli* cell used for the pump activity assay (e, h) of wild type and mutant proteins.

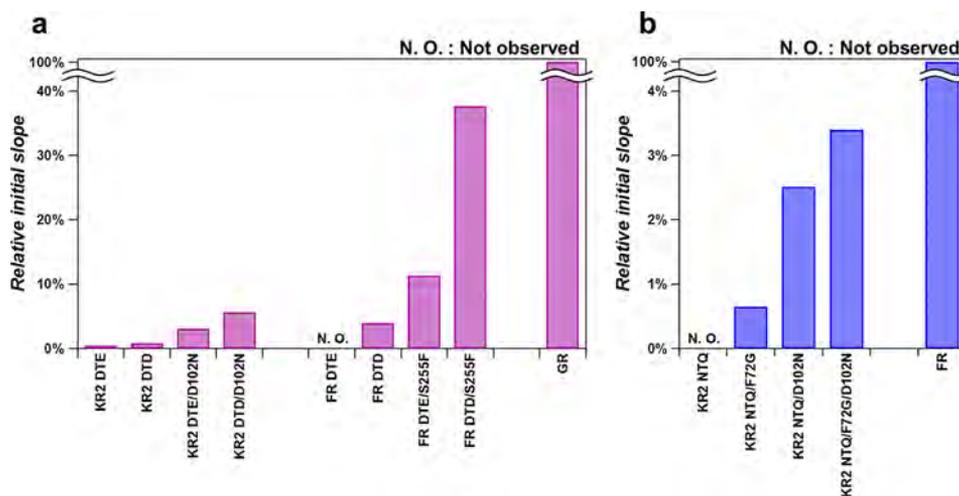


FIGURE 6. Ion pumping activities of the functionally-converted mutants. a, relative H⁺ pumping activities of KR2 and FR mutants compared with GR WT. b, relative Cl⁻ pumping activities of KR2 mutants compared with FR WT.

related to the fact that the NDQ GR mutant looks blue (λ_{\max} = 607 nm, Fig. 3f) (*i.e.* exists in the non-transporting acidic form). This suggests that Asp in the NDQ motif of this mutant is protonated and does not act as the Schiff base counterion and proton acceptor as required for sodium transport (8, 26, 27). We introduced additional mutations to the NDQ GR mutant, to change the color from blue (λ_{\max} > 600 nm) to purple (λ_{\max} < 560 nm) and make the Schiff base proton acceptor functional. Fig. 3b shows clear color

changes for the NDQ/I83S/H87L (λ_{\max} = 555 nm, Fig. 3f), NDQ/I83S/Y88L (λ_{\max} = 556 nm, Fig. 3f), and NDQ/I83S/H87L/Y88L (λ_{\max} = 557 nm, Fig. 3f) GR mutants, where mutations were introduced near the Schiff base region. However, despite the presence of the required ion-pair (Schiff base and counterion), these mutants do not pump Na⁺ (Fig. 3b). We conclude that functional interconversion for H⁺ ⇌ Na⁺ pair is asymmetric, where even 6 mutations are not sufficient for the H⁺ → Na⁺ pump conversion.

Functional Interconversion of H^+ , Na^+ , and Cl^- Pump Rhodopsin

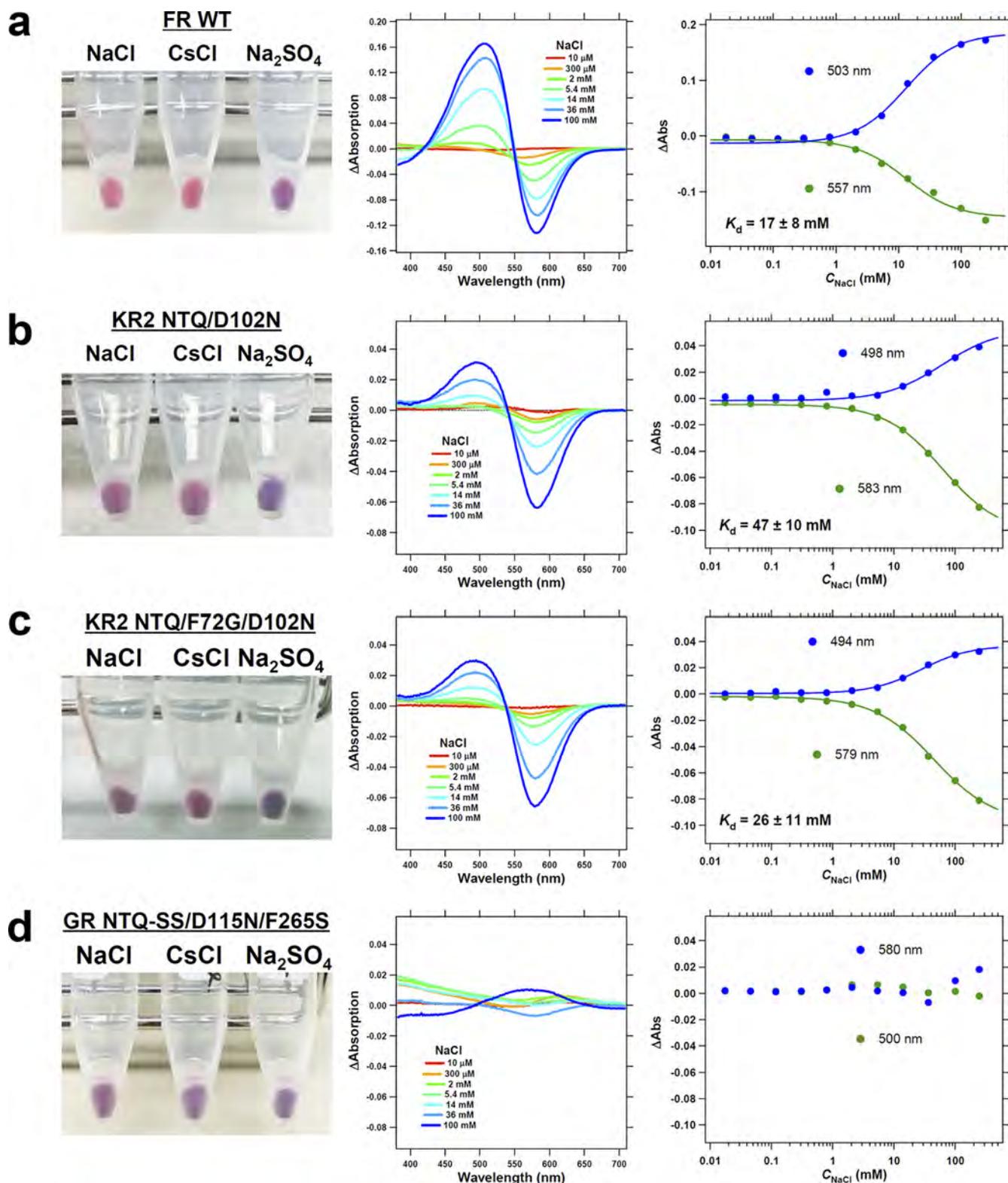


FIGURE 7. Cl^- concentration dependence of absorption spectra of rhodopsins having the NTQ motif reporting on chloride binding. The picture of *E. coli* cells in 100 mM NaCl, CsCl, and Na_2SO_4 (left), difference UV-visible absorption spectra upon the addition of NaCl of different concentrations (middle) and the titration curve for the absorption change at the peak wavelength shown in the difference spectra (right) for FR WT (a), KR2 NTQ/D102N (b), KR2 NTQ/F72G/D102N (c), GR NTQ-I83S/H87S(SS)/D115N/F265S (d). Solid lines in the titration curve indicate the fitting curves using a Hill equation and calculated dissociation constant of Cl^- (K_d) is shown in the graph.

Next, we tried to improve functional interconversion of Na^+ and Cl^- pumps (KR2 and FR) with additional mutations. As the simple motif replacement resulted in unsuccessful $Na^+ \rightarrow Cl^-$

pump conversion, we suspected that the negatively charged Na^+ binding site in KR2 (8, 11, 12) could prevent inward Cl^- pumping (Fig. 2c). It is indeed the case, as the NTQ/D102N KR2

Functional Interconversion of H^+ , Na^+ , and Cl^- Pump Rhodopsin

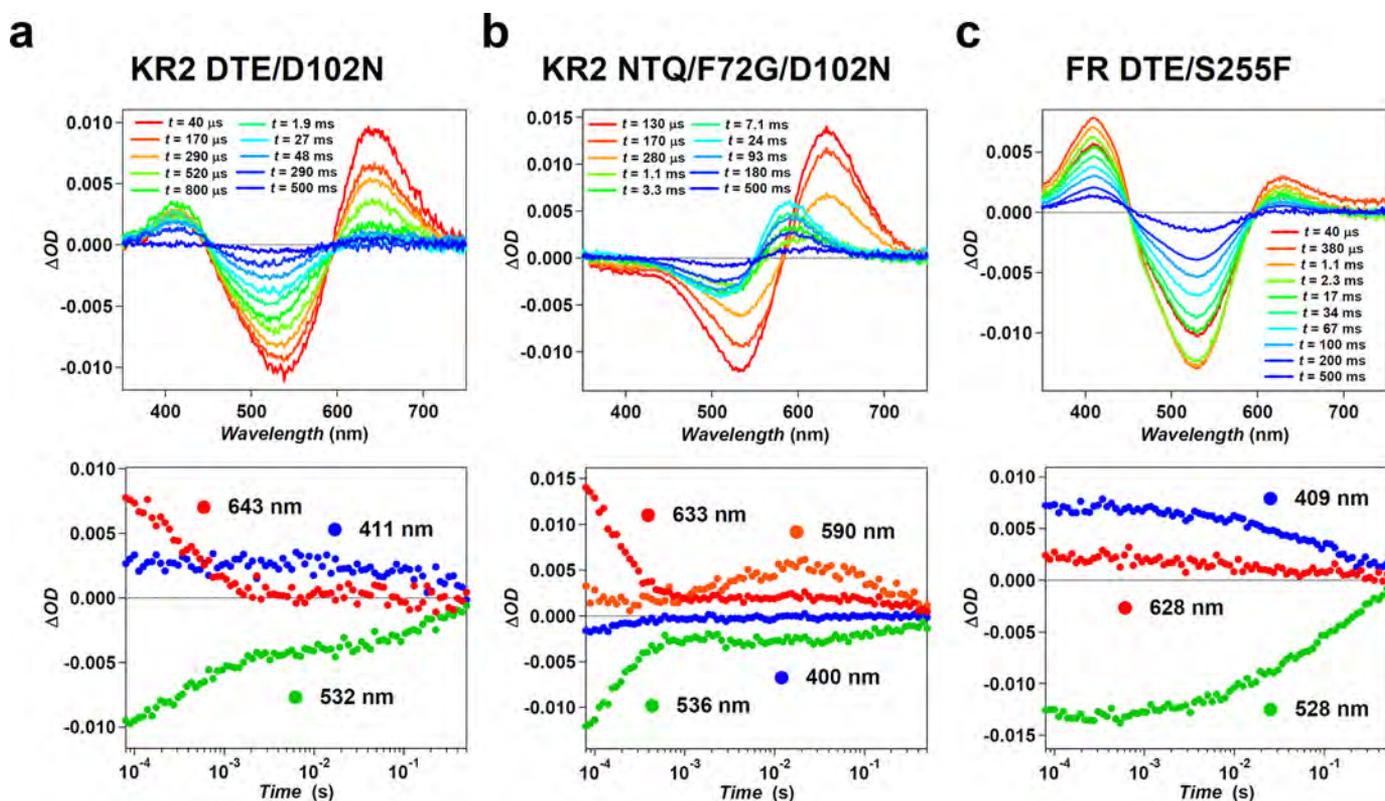


FIGURE 8. **Transient absorption change of functionally converted mutants.** Transient absorption spectra (*upper*) and the time-evolution of absorption changes (*lower*) at the wavelengths representing each photoreaction species (the K-intermediate (*red*), M-intermediate (*blue*), O-intermediate (*orange*), and initial state (*green*)) for KR2 DTE/D102N (*a*), KR2 NTQ/F72G/D102N (*b*), and FR DTE/S255F (*c*).

mutant exhibits light-induced pH increase in NaCl and CsCl, but not in Na_2SO_4 (Fig. 4*a*). This is characteristic for Cl^- pumps, and in line with this, Fig. 7*b* shows absorption changes of the NTQ/D102N KR2 mutant upon Cl^- binding. Thus, a double mutation (NTQ/D102N) is sufficient for converting KR2 into a Cl^- pump. The pumping activity is further enhanced by introducing F72G mutation, which is highly conserved near the Schiff base region, and the NTQ/F72G/D102N KR2 mutant pumps Cl^- with the 3.4% efficiency of that in FR (Figs. 4*a* and 6*b*). NTQ/F72G/D102N KR2 mutant also showed color change upon Cl^- binding (Fig. 7*c*). To make a reverse $Cl^- \rightarrow Na^+$ pump conversion possible, we searched for additional mutations outside of the conserved motif, based on the sequence differences (Fig. 4*b*) and structure of KR2. However, none of these mutants showed any ion transport (Fig. 4*b*). Thus, while $Na^+ \rightarrow Cl^-$ pump conversion was successful, the reverse $Cl^- \rightarrow Na^+$ pump conversion could not be achieved, once again showing an asymmetric interconversion.

Finally, the $H^+ \rightleftharpoons Cl^-$ interconversion was examined with additional mutations, as the motif replacement mutations (DTE \rightleftharpoons NTQ) were not sufficient for successful functional conversions (Fig. 2, *e* and *f*). The NTQ GR mutant looks blue ($\lambda_{max} = 567$ nm) (Fig. 2*e*), suggesting that the protein does not bind Cl^- in 100 mM NaCl. Therefore, we introduced additional mutations to the NTQ GR mutant to improve Cl^- binding, as assayed by change of the color from blue ($\lambda_{max} > 560$ nm) to purple ($\lambda_{max} < 560$ nm) under these conditions. Fig. 5*a* shows that several mutants exhibit color changes, suggesting FR-like Cl^- binding (10). Nevertheless, none of these mutants trans-

ported any ions (Fig. 5*a*). Using titrations of purified proteins, we found that Cl^- binding does not take place, as shown for GR NTQ/I83S/H87S/D115N/F265S, which showed only small non-systematic absorption change probably representing unspecific interaction between the protein and ions (Na^+ and/or Cl^-) (Fig. 7*d*). Thus, the color changes were caused by the mutations themselves and not by Cl^- binding, which explains the lack of ion transport. We then tried to optimize functional conversion from Cl^- to H^+ pump. Gly-263 of Na^+ pump KR2 is a key residue for Na^+ uptake, whose modification leads to K^+ pumping (11, 12). The corresponding amino acids are S and F for FR and GR, respectively, which may be important for ion specificity of each pump. Indeed, Fig. 5*b* shows that the DTE/S255F FR mutant functions as an outward H^+ pump whose relative activity is 11% of wild type GR (Fig. 6*a*). We thus conclude that functional interconversion for the $H^+ \rightleftharpoons Cl^-$ pair is asymmetric as well, where an additional mutation outside of the conserved motif converted Cl^- into H^+ pump, but four additional mutations were not sufficient for the $H^+ \rightarrow Cl^-$ pump conversion.

In this study, three types of functional conversion were succeeded in ($Na^+ \rightarrow H^+$, $Na^+ \rightarrow Cl^-$, and $Cl^- \rightarrow H^+$) with an additional mutation to motif exchange. To investigate their transport mechanism and compare it with natural pumps, the photocycles of these functionally converted mutants were studied by laser flash photolysis. The transient absorption changes of KR2 DTE/D102N, KR2 NTQ/F72G/D102N, and FR DTE/S255F were shown in Fig. 8. While wild type KR2 shows small amount of accumulation of the M-intermediate at $\lambda = 400$ nm

Functional Interconversion of H^+ , Na^+ , and Cl^- Pump Rhodopsins

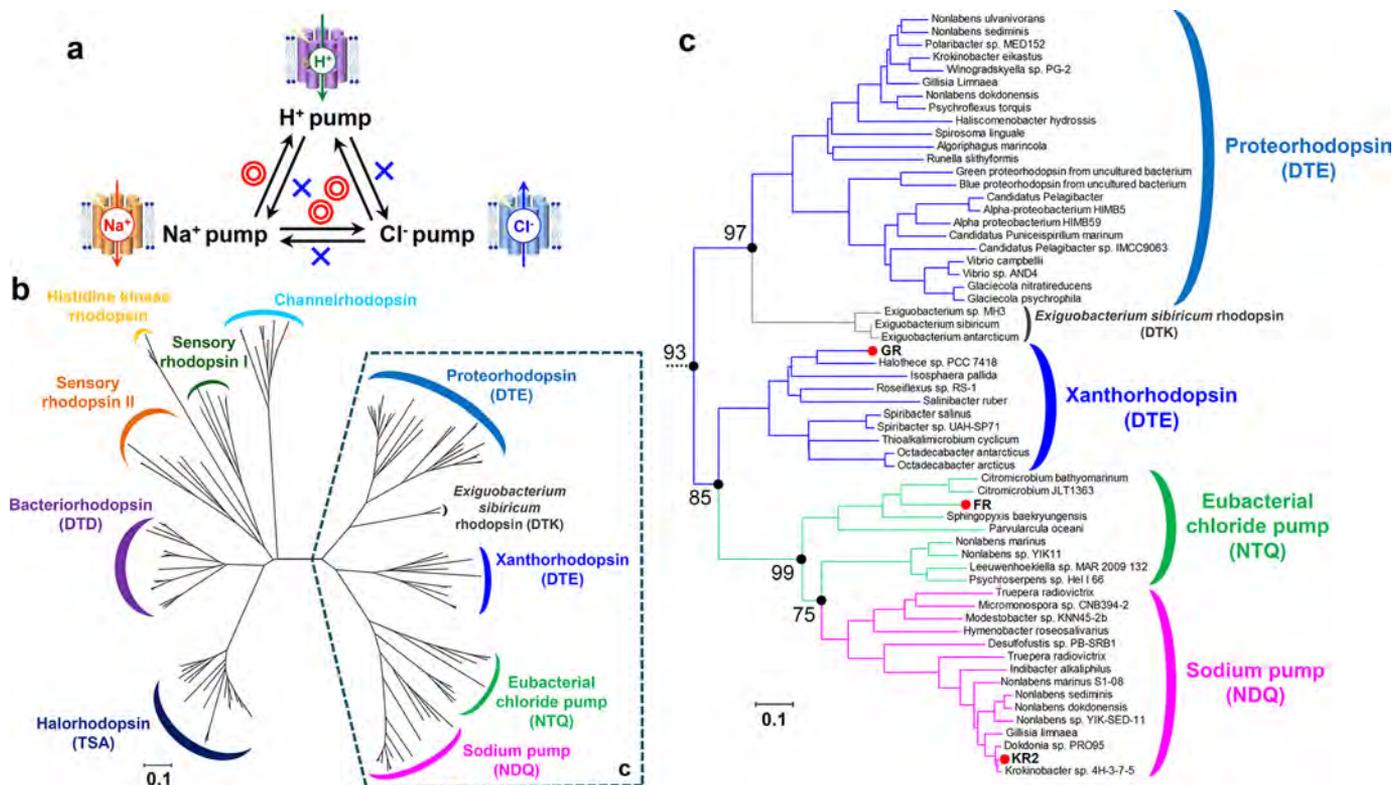


FIGURE 9. Functional conversions observed in the present study and their relationship to evolution of eubacterial ion-pumping rhodopsins. *a*, summary of functional conversions observed in the present study. *Red double-circles* and *blue crosses* represent successful and unsuccessful functional conversions, respectively. *b*, phylogenetic tree of microbial rhodopsins. The bootstrap tests are calculated for 1000 replicates. The scale bar indicates 0.1 units of the number of amino acid substitutions per site. *c*, phylogenetic subtree of eubacterial light-driven ion pumps. The corresponding branches are enclosed by *dashed line* in *b*. Bootstrap values of significantly higher score are indicated by the numbers at *filled black circles*.

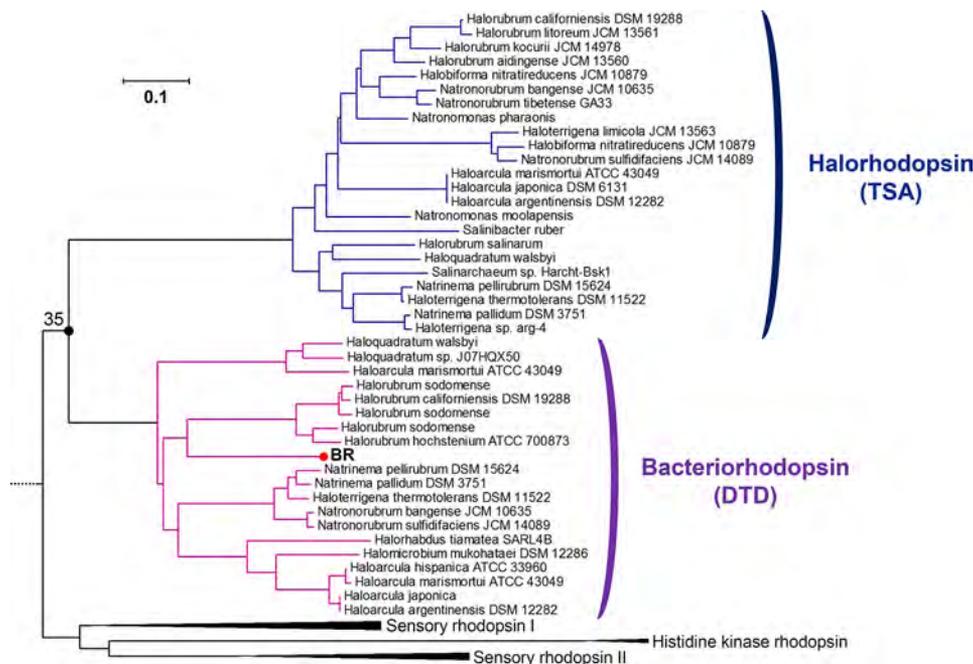


FIGURE 10. Phylogenetic subtree of bacteriorhodopsins and halorhodopsins. Subtree of the phylogenetic tree of microbial rhodopsins shown in Fig. 9*b*, including BR, HR, sensory rhodopsins, and histidine kinase rhodopsins. Subtrees of BR having DTD-motif and HR having TSA-motif are highlighted *pink* and *indigo*, respectively. The bootstrap tests are calculated for 1000 replicates. The scale bar indicates 0.1 units of the number of amino acid substitutions per site.

equilibrating with the L (8), KR2 DTE/D102N showed higher M-accumulation than wild type (Fig. 8*a*). This large accumulation of M is characteristic feature of the photocycle of many H^+ pump rhodopsins including GR (18, 28). The similar high accu-

mulation of M-intermediate was observed for FR DTE/S255F which also converted to H^+ pump (Fig. 8*c*), in contrast to the photocycle of wildtype FR in which no M accumulation occurs (10). Therefore, the mutants converted to H^+ pump transports

Functional Interconversion of H⁺, Na⁺, and Cl⁻ Pump Rhodopsin

H⁺ with the photocycle similar to that of natural H⁺ pump in which higher M accumulation occurs by the H⁺ transfer from retinal Schiff-base to the neighboring aspartate residues. On the other hand, the photocycle of KR2 NTQ/F72G/D102N showed no M-intermediate and red-shifted photo-intermediates (K and O) were dominantly observed (Fig. 8*b*). For natural Cl⁻ pump, the photocycles of HR and FR are significantly different, that is, the former and latter mainly consist of blue-shifted L-intermediate (29, 30) and red-shifted K or O-intermediates (10), respectively. The photocycle of KR2 NTQ/F72G/D102N resembles to that of wild type FR which contains the same NTQ-motif. Therefore, KR2 NTQ/F72G/D102N transports Cl⁻ with the identical mechanism to FR rather than HR with TSA-motif. The time constants of initial state recovery of these mutants (KR2 DTE/D102N: 251 ms, KR2 NTQ/F72G/D102N: 282 ms, FR DTE/S255F: 160 ms) were longer than wild-type GR (140 ms) (18) and FR (20.4 ms) (10). Therefore, a part of the reason of lower efficiencies of the mutants compared with natural pumps is considered to be its slower turnover rate (photocycle recovery) and more amino acid mutations would be required for the optimization of faster turnover rate.

The asymmetric results of functional conversions among the light-driven retinal-binding pumps are summarized in Fig. 9*a*. In principle, mutations of key residues may destabilize proteins, leading to non-functional species which lost their original function but have not gained a new one. However, we achieved three functional conversions by a very limited number of mutations. Interestingly, only one functional conversion in each pair of proteins was possible, whereas the reverse conversions did not work. We discuss the observed asymmetry of the interconversions from the evolutionary viewpoint. Fig. 9*b* shows phylogenetic tree of microbial rhodopsins. Based on the conserved helix C motif, eubacterial rhodopsins can be classified into DTX (H⁺ pumps; X is mostly E), NTQ (Cl⁻ pumps), and NDQ (Na⁺ pumps) rhodopsins, which are distinct from haloarchaeal DTD (BR; H⁺ pumps) and TSA (HR; Cl⁻ pumps) rhodopsins. Fig. 9*c* focuses on eubacterial rhodopsins, whose phylogenetic subtree strongly suggests that the origin of eubacterial rhodopsins is an H⁺ pump, from which Cl⁻ pumps emerged, followed by the appearance of Na⁺ pumps. Comparison of Fig. 9, *a* and *c* shows that successful functional conversions are attained exclusively when the mutagenesis tries to reverse the course of evolution, but not when it follows the evolutionary direction. Dependence of the observed functional conversions on the direction of evolution strongly suggests that the essential elements of an ancestral function are retained even after the gain of a new function, while the gain of a new function needs accumulation of multiple mutations, which may not be easily reproduced by limited mutagenesis *in vitro*.

It should be noted that the asymmetry of functional interconversion between eubacterial H⁺ and Cl⁻ pumps (GR and FR) is opposite to that between archaeal H⁺ and Cl⁻ pumps (BR and HR), where the H⁺ → Cl⁻ pump conversion was achieved by a single amino acid replacement (13, 14) but the reverse conversion was unsuccessful (15, 16). If success of functional conversion of ion-pumping rhodopsins depends on the direction of evolution, how can the results on the archaeal H⁺ and Cl⁻ pumps be rationalized? Fig. 9*b* suggests that phylogenetic tree

for BR and HR is less hierarchical than that for eubacterial pumps, with much older branching, as described in Fig. 10. Thus, different molecular mechanism such as hydrogen-bonding strength of protein-bound water molecules (17) underlies asymmetric functional conversion in archaeal H⁺ and Cl⁻ pumps. In this sense, each molecular mechanism of successful and unsuccessful functional conversions in the present study should be revealed by structural details in future.

Author Contributions—K. I. and H. K. contributed to the study design. K. I. contributed to construction of mutant proteins and introduction to *E. coli*. K. I. and Y. N. conducted pumping activity measurement. Y. N. quantified the amount of protein expression in *E. coli*, purified proteins and measured their absorption spectra at different NaCl concentrations. K. I. and Y. N. carried out flash photolysis experiment of purified proteins. K. I. and H. K. wrote the paper. All authors discussed and commented on the manuscript.

References

1. Ernst, O. P., Lodowski, D. T., Elstner, M., Hegemann, P., Brown, L. S., and Kandori, H. (2014) Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. *Chem. Rev.* **114**, 126–163
2. Deisseroth, K. (2011) Optogenetics. *Nat. Methods* **8**, 26–29
3. Hegemann, P., and Möglich, A. (2011) Channelrhodopsin engineering and exploration of new optogenetic tools. *Nat. Methods* **8**, 39–42
4. Oesterhelt, D., and Stoekenius, W. (1971) Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nat. New Biol.* **233**, 149–152
5. Matsuno-Yagi, A., and Mukohata, Y. (1977) Two possible roles of bacteriorhodopsin; a comparative study of strains of *Halobacterium halobium* differing in pigmentation. *Biochem. Biophys. Res. Commun.* **78**, 237–243
6. Béjà, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A., Nguyen, L. P., Jovanovich, S. B., Gates, C. M., Feldman, R. A., Spudich, J. L., Spudich, E. N., and DeLong, E. F. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906
7. Sharma, A. K., Spudich, J. L., and Doolittle, W. F. (2006) Microbial rhodopsins: functional versatility and genetic mobility. *Trends Microbiol.* **14**, 463–469
8. Inoue, K., Ono, H., Abe-Yoshizumi, R., Yoshizawa, S., Ito, H., Kogure, K., and Kandori, H. (2013) A light-driven sodium ion pump in marine bacteria. *Nat. Commun.* **4**, 1678
9. Yoshizawa, S., Kumagai, Y., Kim, H., Ogura, Y., Hayashi, T., Iwasaki, W., DeLong, E. F., and Kogure, K. (2014) Functional characterization of flavobacteria rhodopsins reveals a unique class of light-driven chloride pump in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6732–6737
10. Inoue, K., Koua, F. H., Kato, Y., Abe-Yoshizumi, R., and Kandori, H. (2014) Spectroscopic study of a light-driven chloride ion pump from marine bacteria. *J. Phys. Chem. B.* **118**, 11190–11199
11. Kato, H. E., Inoue, K., Abe-Yoshizumi, R., Kato, Y., Ono, H., Konno, M., Hososhima, S., Ishizuka, T., Hoque, M. R., Kunitomo, H., Ito, J., Yoshizawa, S., Yamashita, K., Takemoto, M., Nishizawa, T., *et al.* (2015) Structural basis for Na⁺ transport mechanism by a light-driven Na⁺ pump. *Nature* **521**, 48–53
12. Gushchin, I., Shevchenko, V., Polovinkin, V., Kovalev, K., Alekseev, A., Round, E., Borshchevskiy, V., Balandin, T., Popov, A., Gensch, T., Fahlke, C., Bamann, C., Willbold, D., Büldt, G., Bamberg, E., and Gordeliy, V. (2015) Crystal structure of a light-driven sodium pump. *Nat. Struct. Mol. Biol.* **22**, 390–395
13. Sasaki, J., Brown, L. S., Chon, Y. S., Kandori, H., Maeda, A., Needleman, R., and Lanyi, J. K. (1995) Conversion of bacteriorhodopsin into a chloride ion pump. *Science* **269**, 73–75
14. Tittor, J., Haupts, U., Haupts, C., Oesterhelt, D., Becker, A., and Bamberg, E. (1997) Chloride and proton transport in bacteriorhodopsin mutant

- D85T: different modes of ion translocation in a retinal protein. *J. Mol. Biol.* **271**, 405–416
15. Havelka, W. A., Henderson, R., and Oesterhelt, D. (1995) Three-dimensional structure of halorhodopsin at 7 Å resolution. *J. Mol. Biol.* **247**, 726–738
 16. Váró, G., Brown, L. S., Needleman, R., and Lanyi, J. K. (1996) Proton transport by halorhodopsin. *Biochemistry* **35**, 6604–6611
 17. Muroda, K., Nakashima, K., Shibata, M., Demura, M., and Kandori, H. (2012) Protein-bound water as the determinant of asymmetric functional conversion between light-driven proton and chloride pumps. *Biochemistry* **51**, 4677–4684
 18. Miranda, M. R., Choi, A. R., Shi, L., Bezerra, A. G., Jr., Jung, K. H., and Brown, L. S. (2009) The photocycle and proton translocation pathway in a cyanobacterial ion-pumping rhodopsin. *Biophys. J.* **96**, 1471–1481
 19. Scharf, B., Hess, B., and Engelhard, M. (1992) Chromophore of sensory rhodopsin II from *Halobacterium halobium*. *Biochemistry* **31**, 12486–12492
 20. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797
 21. Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425
 22. Felsenstein, J. (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783–791
 23. Zuckerkandl, E., and Pauling, L. (1965) Evolutionary divergence and convergence in proteins, in *Evolving Genes and Proteins* (Bryson, V., and Vogel, H. J., eds), pp. 97–166, Academic Press, New York
 24. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739
 25. Inoue, K., Kato, Y., and Kandori, H. (2015) Light-driven ion-translocating rhodopsins in marine bacteria. *Trends Microbiol.* **23**, 91–98
 26. Balashov, S. P., Imasheva, E. S., Dioumaev, A. K., Wang, J. M., Jung, K.-H., and Lanyi, J. K. (2014) Light-driven Na^+ pump from *Gillisia limnaea*: a high-affinity Na^+ binding site is formed transiently in the photocycle. *Biochemistry* **53**, 7549–7561
 27. da Silva, G. F., Goblirsch, B. R., Tsai, A. L., and Spudich, J. L. (2015) Cation-specific conformations in a dual-function ion-pumping microbial rhodopsin. *Biochemistry* **54**, 3950–3959
 28. Váró, G., Brown, L. S., Lakatos, M., and Lanyi, J. K. (2003) Characterization of the photochemical reaction cycle of proteorhodopsin. *Biophys. J.* **84**, 1202–1207
 29. Váró, G., Zimányi, L., Fan, X., Sun, L., Needleman, R., and Lanyi, J. K. (1995) Photocycle of halorhodopsin from *Halobacterium salinarium*. *Biophys. J.* **68**, 2062–2072
 30. Sato, M., Kubo, M., Aizawa, T., Kamo, N., Kikukawa, T., Nitta, K., and Demura, M. (2005) Role of putative anion-binding sites in cytoplasmic and extracellular channels of *Natronomonas pharaonis* halorhodopsin. *Biochemistry* **44**, 4775–4784

YAP Nuclear Localization in the Absence of Cell-Cell Contact Is Mediated by a Filamentous Actin-dependent, Myosin II- and Phospho-YAP-independent Pathway during Extracellular Matrix Mechanosensing*

Received for publication, December 4, 2015, and in revised form, January 8, 2016. Published, JBC Papers in Press, January 12, 2016, DOI 10.1074/jbc.M115.708313

Arupratan Das^{†§}, Robert S. Fischer[‡], Duoqia Pan^{§1}, and Clare M. Waterman^{‡2}

From the [‡]Cell Biology and Physiology Center, NHLBI, National Institutes of Health, Bethesda, Maryland 20892 and the

[§]Department of Molecular Biology and Genetics, Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

Cell-cell contact inhibition and the mechanical environment of cells have both been shown to regulate YAP nuclear localization to modulate cell proliferation. Changes in cellular contractility by genetic, pharmacological, and matrix stiffness perturbations regulate YAP nuclear localization. However, because contractility and F-actin organization are interconnected cytoskeletal properties, it remains unclear which of these distinctly regulates YAP localization. Here we show that in the absence of cell-cell contact, actomyosin contractility suppresses YAP phosphorylation at Ser¹¹², however, neither loss of contractility nor increase in YAP phosphorylation is sufficient for its nuclear exclusion. We find that actin cytoskeletal integrity is essential for YAP nuclear localization, and can override phosphoregulation or contractility-mediated regulation of YAP nuclear localization. This actin-mediated regulation is conserved during mechanotransduction, as substrate compliance increased YAP phosphorylation and reduced cytoskeletal integrity leading to nuclear exclusion of both YAP and Ser(P)¹¹²-YAP. These data provide evidence for two actin-mediated pathways for YAP regulation; one in which actomyosin contractility regulates YAP phosphorylation, and a second that involves cytoskeletal integrity-mediated regulation of YAP nuclear localization independent of contractility. We suggest that in non-contact inhibited cells, this latter mechanism may be important in low stiffness regimes, such as may be encountered in physiological environments.

Nuclear localization of the transcriptional co-activator YAP (and its ortholog TAZ) and activation of the TEAD family transcription factors promote cell proliferation, differentiation, stem cell fate, and organ size regulation (1–3). Deficiencies in YAP regulation lead to developmental defects (4) and tumor

formation (5). YAP nuclear localization is regulated by a broad range of biochemical and mechanical cues that include serum soluble cues (6), cell-cell contact mediated by adherens junction (AJs)³ (7), and cell mechanosensing of extracellular matrix (ECM) physical properties or stretch (8–10). These upstream inputs converge on the regulation of the core components of the Hippo signaling pathway to regulate YAP nuclear localization. Genetic studies in *Drosophila* and cell biological studies in culture have identified the epistatic relationship of the core components of the Hippo signaling pathway (3). Upstream signals activate Mst1/2 (Hippo kinase in *Drosophila*), which in turn activates Lats1/2 kinases (Wts in *Drosophila*), and Lats phosphorylates YAP (Yorkie in *Drosophila*) on serine 112, which induces it to form a complex with 14-3-3 and retains it in the cytoplasm (1, 11–13). In the absence of the activation of Hippo signaling, YAP is imported into the nucleus where it promotes TEAD-mediated transcription. However, the interplay between the inputs of serum factors, cell-cell contact, and cell mechanosensing in regulation of YAP nuclear localization is not well understood.

Hippo signaling-mediated regulation of YAP via cell-cell contact has been well characterized (14). YAP is localized to the nucleus and active in cells growing at low density, but becomes cytoplasmic in confluent cultures, and thus underlies the classic paradigm of contact inhibition of proliferation (12). YAP Ser¹¹² phosphorylation increases during contact inhibition, and overexpression of non-phosphorylatable YAP delays proliferation arrest, allowing cells to reach higher densities than controls (12). The role of cadherin- and catenin-mediated cell-cell AJs in this process is being elucidated (14). In *Drosophila*, α -catenin and the LIM protein Ajuba form a complex at AJs and recruit Wts/Lats1/2 kinase to regulate YAP phosphorylation, cytoplasmic retention, and tissue growth (15). Hippo signaling regulation via AJs occurs during mouse embryonic development, in which the protein angiominin forms a complex with Lats1/2 at the AJs to activate Hippo signaling and thus retain YAP in the cytoplasm (7).

On the other hand, cellular mechanosensation-mediated regulation of YAP localization is thought to be distinct from the

* This work was supported by the NHLBI Division of Intramural Research (to C. M. W., R. S. F., and A. D.) and Grant EY015708 from the NEI through the National Institutes of Health. The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

¹ Investigator of the Howard Hughes Medical Institute.

² To whom correspondence should be addressed: Laboratory of Cell and Tissue Morphodynamics, NHLBI, National Institutes of Health, Bldg. 50, South Dr., Rm. 4537, Bethesda, MD 20892-8019. Tel.: 301-435-2949; E-mail: watermancm@nhlbi.nih.gov.

³ The abbreviations used are: AJ, adherens junction; ECM, extracellular matrix; MEF, mouse embryonic fibroblast; DMSO, dimethyl sulfoxide; Pa, pascal; ANOVA, analysis of variance; Lat-A, latrunculin A.

cell junctional complexes that regulate the Hippo cascade (16). Cell spreading on large patches of ECM, cell attachment to stiff or stretched ECM (8, 10), or fluid shear stress (17) all induce YAP activation and nuclear localization in a range of cell types. These pathways may be Hippo independent or dominant, as LATS1/2 inactivation cannot rescue YAP inhibition in cells with reduced mechanical stress (8). Although mechanical regulation of YAP localization depends on cell attachment, it is not dependent on integrin engagement (18), but rather thought to be dependent on the maintenance of tension in an intact, contractile actomyosin cytoskeleton. Indeed, pharmacological inhibition of actin assembly or myosin II ATPase blocks YAP nuclear translocation in cells on stiff or large ECMs (8).

Although the initial evidence that YAP is regulated by the cytoskeleton came from studies of mechanosensation, more recent evidence suggests that cells need mechanical tension to sustain YAP transcriptional activities and/or nuclear localization, irrespective of the inducing input (16). In contact inhibited *Drosophila* epithelial tissue, myosin II contractility regulates the formation of the Wts kinase complex at AJs to suppress Hippo signaling, and in the absence of contractility this complex is disrupted and Hippo signaling is activated (15). Furthermore, in dense cultures *in vitro*, YAP nuclear localization is rescued by ECM stretching (9, 10). Mechanoregulation of YAP in the absence of cell-cell contact has also been shown to be dependent on ECM stiffness sensation, myosin II motor activity, and cell spreading area (8, 19).

Thus, the mechanisms by which the actomyosin cytoskeleton regulates YAP localization and phosphorylation and their relationship to cell-cell contact remain major unanswered questions in the field. Although mechanosensation is well known to alter myosin II activity (20, 21), it is not known how mechanosensation might affect F-actin levels in cells. Furthermore, recent evidence suggests that actin filaments, rather than myosin II activity, may be the key regulator of YAP. Two elegant studies in *Drosophila* showed that Hippo signaling reduces cellular F-actin (22), whereas increasing F-actin by loss of capping protein can suppress Hippo signaling (23). In addition, LATS1/2-independent regulation of YAP can be achieved in cells cultured on soft, small matrices or at high cell density by promotion of actin assembly via depletion of the actin depolymerizing factor Cofilin, or the filament capping proteins capZ or gelsolin (10). Further complicating these findings, it has also been shown that myosin inhibition by blebbistatin only partially affects YAP nuclear localization, yet it was predominantly dependent on the actin-binding protein angiomin (24).

To resolve the role of myosin contractility and F-actin in regulation of YAP nuclear localization, we utilized breast epithelial cells as a contact-inhibited model system, and mouse embryonic fibroblasts (MEF) as a contact-independent model system. We found that in the absence of cell-cell contacts in either epithelial or fibroblast cells, YAP localizes to the nucleus even in the absence of actomyosin contractility. Furthermore, actomyosin contractility suppresses YAP phosphorylation at Ser¹¹², and when contractility is inhibited, even phosphorylated YAP localized to the nucleus. Although contractility and phosphorylation is dispensable for YAP nuclear localization, we find that nuclear localization of YAP or phosphorylated YAP is

strictly dependent on the abundance of F-actin filaments. This actin-dependent regulation is also conserved during mechanotransduction when cells are grown on soft substrates, as would be expected to occur physiologically.

Experimental Procedures

Cell Culture, Treatments, and Transfections—MEFs (kindly provided by Mary C. Beckerle, Huntsman Cancer Institute, Salt Lake City, UT) were cultured in DMEM (Invitrogen) supplemented with 100 μ g/ml of penicillin/streptomycin, 2 mM L-glutamine, and 10% FBS. MCF10A mammary epithelial cells (ATCC, Manassas, VA) were cultured in mammary epithelium basal medium (Lonza, catalogue number cc-3151) supplemented with bovine pituitary extract (BPE), human epidermal growth factor (hEGF), insulin, and hydrocortisone from mammary epithelium basal medium Single Quots kit (Lonza, catalogue number cc-4136), 100 ng/ml of cholera toxin, 100 μ g/ml of penicillin/streptomycin, and 10% FBS. For immunofluorescence assays under sparse culture conditions, $\sim 3,500$ cell/cm² were plated on coverslips coated with 10 μ g/ml of fibronectin or 10 μ g/ml of poly-L-lysine, as noted, and allowed to spread overnight. For dense culture assays, $\sim 4 \times 10^5$ cells/cm² were seeded on fibronectin-coated coverslips and cultured at 37 °C in 5% CO₂ for 24 h. Cells were treated with blebbistatin (Toronto Research Chemicals), latrunculin-A (Invitrogen), Y-27632 (Sigma), manganese chloride, or DMSO vehicle control at the indicated concentrations for 2 h. Cell culture and drug treatments were performed at 37 °C in 5% CO₂.

For ECM stiffness-dependent assays, quinone-activated polyacrylamide/bisacrylamide hydrogels on coverslips or 10-cm Petri dishes were purchased from (Matrigen, CA) or were made by hand (55.0 and 0.7 kPa) in which the gel surface was cross-linked with Sulfo-SANPAH reactive amines by UV exposure. Hydrogels were coupled with 10 μ g/ml of fibronectin at room temperature for 1 h and equilibrated with media prior to cell plating. For immunofluorescence, $\sim 3,500$ cell/cm² were plated on coverslip hydrogels, for biochemical assays (whole cell lysis, F/G-actin separation) cells were plated on Petri dish hydrogels and grown overnight to $\sim 60\%$ confluence.

The FLAG-YAP and FLAG-YAP-8SA (including the following mutations: S61A, S109A, S127A, S128A, S131A, S163A, S164A, and S381A) expression vectors consisting of the cDNA encoding human YAP fused to the COOH terminus of FLAG (obtained from Kun-Liang Guan, University of California, San Diego), which were also used in Ref. 12. The YAP1-GFP expression vector consisting of the cDNA encoding mouse YAP1 fused at its COOH terminus to eGFP was purchased from Genecopoeia (catalogue number EX-Mm21312-M98). YAP1^{S112A} mutant was generated on the YAP1-GFP construct using QuikChange site-directed mutagenesis (Agilent Technologies). mApple- β -actin construct was kindly provided by M. Davidson (Florida State University, Tallahassee, FL). MEFs were transfected with these constructs (Nucleofactor solution-V, Lonza) and seeded on fibronectin-coated coverslips or hydrogels for ~ 18 h prior to fixation. FLAG-YAP constructs were transfected (Effectene, Qiagen) and cultured for 24 h then seeded on fibronectin-coated coverslips for another ~ 16 h prior to fixation.

Yap Localization during Mechanosensing Requires Filamentous Actin

Immunofluorescence—For immunolocalizing YAP and FLAG-YAP, cells on fibronectin-coated coverslips or on polyacrylamide/bisacrylamide hydrogels were fixed with 4% paraformaldehyde (Electron Microscopy Science) in cytoskeleton buffer (CB; 20 mM PIPES, pH 6.8, 138 mM KCl, 3 mM MgCl₂, and 2 mM EGTA). Cells were permeabilized with 0.5% Triton X-100 in CB for 5 min followed by a 10-min wash with 0.1 M glycine to quench free aldehydes. Next, cells were washed with TBST (Tris-buffered saline, 0.1% Tween 20) two times for 5 min each and blocked for 1 h with blocking buffer (TBST and 2% BSA). Cells were incubated with anti-mouse YAP (Santa Cruz Biotechnology, Inc., sc-101199) in blocking buffer at 1:200 dilution for 2 h at room temperature or anti-rat FLAG (BioLegend number 637301) at 1:200 overnight at 4 °C and washed three times for 5 min each with TBST. Coverslips were then incubated with fluorophore-conjugated secondary antibodies (Jackson ImmunoResearch Laboratories, Inc.) at 1:500 in blocking buffer for 1 h at room temperature. F-actin was stained with Alexa Fluor 488 or 561 phalloidin (Invitrogen) at 1:500 dilution along with the secondary antibodies. Coverslips were washed again with TBST three times for 5 min each, and DNA was stained with 1.4 μM DAPI (Sigma) with the second wash. Finally, coverslips were mounted on glass slides with mounting media (DAKO) and sealed with nail polish. To immunolocalize Ser(P)¹¹²-YAP, cells were fixed with pre-chilled (−20 °C) methanol for 20 min at −20 °C, washed three times for 5 min each with CB, and continued with the immunofluorescence protocol as above, using anti-rabbit phospho-YAP (Ser¹²⁷) (Cell Signaling, number 4911) at 1:200 dilution.

Cells expressing YAP1-GFP and mApple-actin were fixed with 4% paraformaldehyde, permeabilized with 0.5% Triton X-100 in CB, and washed with 0.1 M glycine. After a wash with TBS and TBS with DAPI, cells were mounted on a glass slide and sealed with nail polish.

F- and G-actin staining was performed as described in Ref. 25 with modifications. Cells were fixed with 4% paraformaldehyde in CB followed by a single wash with PBS and permeabilization with PBS containing 0.5% Triton X-100 for 10 min. Permeabilized cells were washed with 0.1 M glycine in CB for 10 min, then washed with PBS two times for 5 min each, and rinsed with PBS with 0.1% Triton X-100. After blocking with antibody dilution solution (PBS, 0.1% Triton X-100 and 2% BSA) for 1 h, F-actin was stained with Alexa Fluor 488 phalloidin at 1:500 dilution and G-actin with deoxyribonuclease-I conjugated with Alexa Fluor-594 (Invitrogen) at 5 μg/ml concentration for 1 h at room temperature. Coverslips were washed three times with PBS, 5 min each, and mounted on a glass slide with mounting media (DAKO) and sealed with nail polish.

Microscopy and Image Analysis—Immunolabeled samples were imaged by spinning disc confocal microscopy with an Eclipse Ti microscope body (Nikon) equipped with a Yokogawa spinning disc scanning unit as described in Ref. 26. Samples on coverslips were imaged with a 60× 1.49 NA/oil objective and on hydrogels with 60× 1.2 NA/water objective. Images were acquired with a cooled-CCD camera (HQ2 or Myo from Photometrics). For immunostained cells, Z-stacks of images were acquired for each channel, and the middle confocal slice was chosen from the images of the nucleus detected in the DAPI

channel. On the corresponding slice in the YAP channel, the average fluorescence intensity in the nucleus and just outside the nucleus (cytoplasm) was measured to determine the nuclear/cytoplasmic ratio. For cells expressing YAP1-GFP and mApple-actin, to reduce image photobleaching, instead of Z-stacks, a single image at the middle nuclear plane (determined in the DAPI channel) was acquired in the GFP channel, then the microscope was focused to the coverslip surface for acquiring the image in the mApple channel.

For F- to G-actin ratio measurements, confocal Z-stacks of images were acquired for each channel with a 400-nm step size using a stage piezo Z-stepper (Mad City Labs) integrated with an XYZ automated stage and controller (Applied Scientific Instruments). Integrated fluorescence intensity was measured on the background-subtracted sum projections for each channel and ratioed to give the F:G-actin ratio. Image analysis was performed using ImageJ software.

Nuclear/Cytoplasmic Fractionation—Nuclear/cytoplasmic fractionation was performed essentially according to Dignam *et al.* (27) but with the following modifications. All buffers used were kept on ice and centrifugations were done at 4 °C with soft braking. MEFs were grown on fibronectin-coated 15-cm culture dish to ~60% confluence. After a single wash with PBS, cells were scraped with PBS (containing 1 mM DTT and 1× protease inhibitor) and harvested by centrifugation at 1000 × *g* for 15 min. The cell pellet was gently resuspended with five times the volume of pellet with buffer A (10 mM HEPES, pH 7.9, 1.5 mM MgCl₂, 10 mM KCl, 0.5 mM DTT, and 1× protease inhibitor) and incubated on ice for 15 min, followed by homogenization (Wheaton). Cell lysis was checked by trypan blue staining of the nucleus after every 20 strokes. The cell lysate was spun at 1000 × *g* for 5 min to collect the pellet as the nuclear fraction and the supernatant as the cytoplasmic fraction. The nuclear fraction was washed by centrifugation two times with buffer A (1000 × *g* for 5 min each), resuspended with buffer A to a similar volume as the cytoplasmic fraction, and sonicated. Both fractions were boiled with sample buffer keeping an identical final volume. Each fraction was immunoblotted for YAP, Ser(P)¹¹²-YAP, and histone H3.

Fractionation of F- and G-actin by Triton Solubilization—Determination of the fraction of F- and G-actin in cells was done following the method of Rasmussen *et al.* (28) with modifications. MEFs were grown on fibronectin-coated 10-cm culture dishes to ~60% confluence. Cells were washed briefly with cold PBS, and lysed with 600 μl of actin stabilization buffer (0.1 M PIPES, pH 6.9, 30% glycerol, 5% DMSO, 1 mM MgSO₄, 1 mM EGTA, 1% Triton X-100, 1 mM ATP, and 1× protease inhibitor) for 10 min at room temperature. Cells were collected by scraping, followed by centrifugation at 90,000 × *g* for 30 min at 4 °C to separate detergent-insoluble (F-actin) and soluble (G-actin) fractions. An equal volume of 2× sample buffer was added to the soluble fraction and 1× sample buffer to the insoluble pellet to make both fractions 1× of identical volume. Samples were boiled for 10 min, and an equal volume was loaded for immunoblotting actin.

Determination of F- to G-actin from cells grown on hydrogels or dense culture was performed according to Cramer *et al.* (25) with modifications. Cells were grown to ~60% confluence

on fibronectin-coated hydrogels or to 100% confluence for dense culture in 10-cm culture dishes. Cells were washed with PBS and incubated with actin stabilization buffer for 10 min. The Triton-soluble G-actin was then collected upon cell lysis by pipetting up the solution from the dish at room temperature. After a brief wash with actin stabilization buffer, the Triton-insoluble F-actin was dissolved directly on the hydrogel surface by adding boiling F-actin solubilization buffer (0.06 M Tris, pH 6.8, 2% SDS (w/v), 1% β -mercaptoethanol (v/v)). The F-actin fraction was collected from the dish by carefully scraping and pipetting the solution. Both the Triton-soluble and insoluble fractions were concentrated by Speed-Vac and made to an identical volume with actin-stabilizing buffer. Sample buffer was added to each fraction to the identical volume and boiled for 10 min and immunoblotted for actin. For experiments to compare the F:G ratio in dense culture and sparse cultures, an extra pair of culture dishes at the two plating densities was used to count cells, and the sample loading volume was adjusted accordingly to keep the cell number similar between dense and sparse conditions.

Western Blot—For immunoblotting whole cell extracts, the same number of MEFs grown and treated with the indicated drugs for 2 h were washed with PBS and directly lysed with boiling 1 \times sample buffer (Quality Biological, Inc.) supplemented with 5% β -mercaptoethanol. Samples were collected by scraping, sonicated, and boiled for 10 min. Protein samples from whole cell lysis, nucleus/cytoplasmic, or F/G-actin fractionations were separated by SDS-PAGE and electrotransferred to Immobilon-FL (Millipore) membrane for immunoblotting. Membranes were blocked for 1 h at room temperature with blocking solution (3% skim milk in TBST), incubated with primary antibodies overnight at 4 °C in blocking solution. After reacting with primary antibodies, membranes were washed three times for 5 min each with TBST and incubated with fluorescent-conjugated secondary antibodies for 1 h 30 min at room temperature. Secondary antibodies used at 1:5,000 dilutions in blocking solutions are: anti-rabbit IRDye 680RD and anti-mouse IRDye 800CW (LI-COR, Inc.). Membranes were washed three times for 5 min each with TBST and imaged with the Odyssey infrared imaging system (LI-COR, Inc.) and analyzed with ImageJ software. Integrated fluorescence intensity for each of the Western blot bands was measured from the background-subtracted images. Primary antibodies used for Western blots were: anti-mouse actin, clone C4 (1:8,000, Millipore number MAB1501), anti-mouse YAP (1:1,000, Santa Cruz Biotechnology number sc-101199), anti-rabbit phospho-YAP (Ser¹²⁷) (1:1,000, Cell Signaling number 4911), and anti-rabbit Histone H3 (1:5,000, Cell Signaling number 9715).

Statistical Analysis—Statistical comparison done with the Student's *t* test between a pair of datasets. One-way ANOVA tests were performed on group of data containing three or more data sets, *p* values from ANOVA tests are listed in Table 1.

Results

In the Absence of Cell-Cell Contact, Nuclear Localization of YAP Requires F-actin Cytoskeletal Integrity, but Not ROCK-mediated Myosin II Contractility—We sought to examine the interplay between cytoskeletal integrity, contractility, and cell

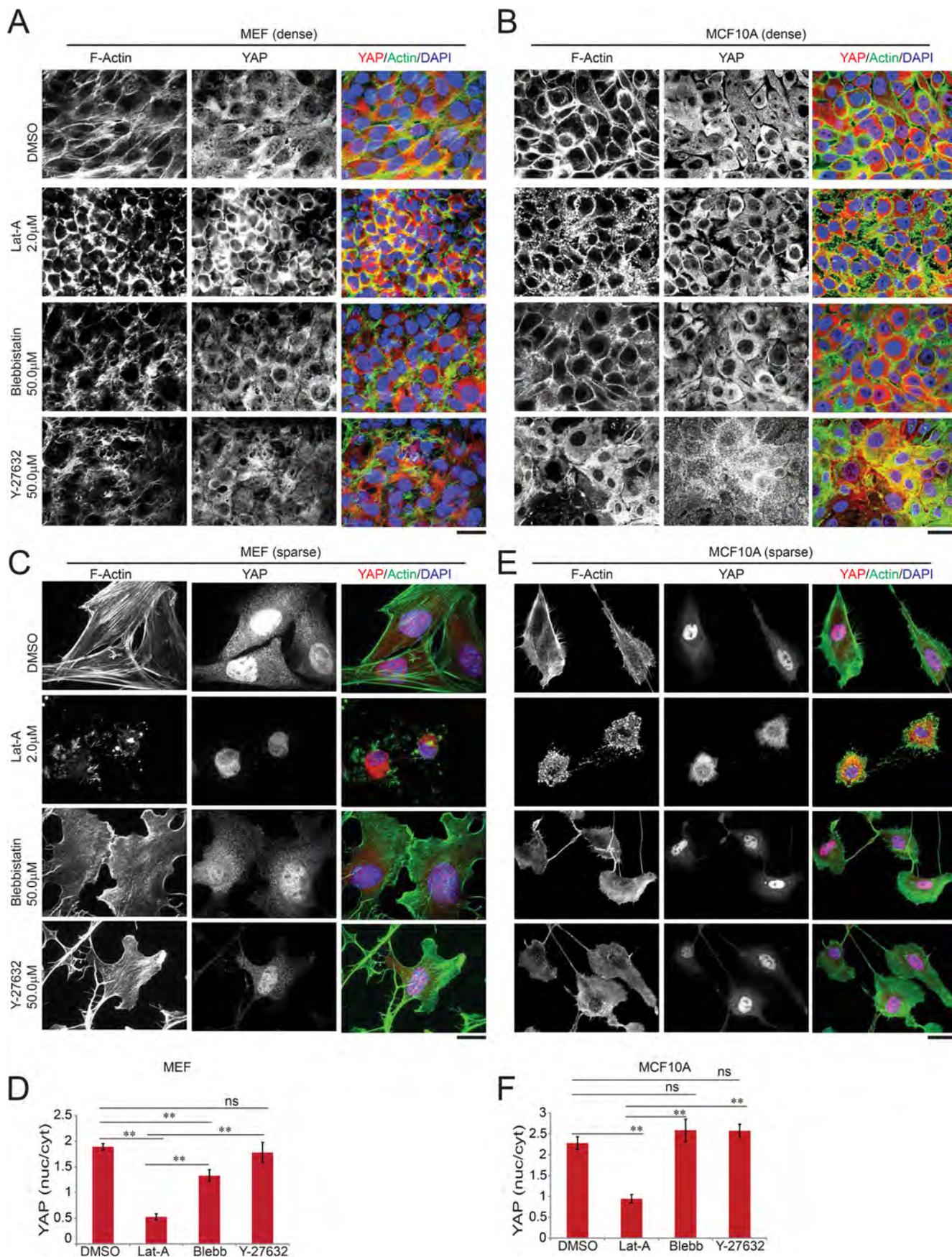
TABLE 1
One-way ANOVA test for group of three or more data sets

Figure	<i>p</i> value
Fig. 1D	4.88E-10
Fig. 1E	1.45E-07
Fig. 2D	5.95E-37
Fig. 3A	0.016866
Fig. 3C	0.000233
Fig. 5B	0.035483
Fig. 5D	1.99E-10
Fig. 5J	3.99E-16
Fig. 6C	1.04E-07
Fig. 6F	2.07E-09
Fig. 6J	6.77E-06
Fig. 6L	0.100036

contact inhibition in regulation of YAP nuclear localization and phosphorylation. We utilized MFC-10A epithelial cells that normally form tissues, and MEFs that exhibit contact inhibition of motility (29), but also function and proliferate as individual cells. We first examined the role of cytoskeletal integrity and contractility in YAP localization in confluent cultures by immunolocalizing YAP and staining actin filaments (F-actin) with fluorescent phalloidin in cells treated with pharmacological inhibitors. In confluent cultures of either MEFs or MCF10A cells plated on fibronectin, F-actin formed a band around the cell periphery at cell-cell contacts, and YAP was excluded from the nucleus (Fig. 1, A and B). To test if YAP localization was affected by cytoskeletal integrity in confluent culture, cells were treated with either vehicle (DMSO) control or 2 μ M latrunculin-A (Lat-A) for 2 h to depolymerize F-actin. This showed that treatment with Lat-A disrupted the peripheral F-actin band, which was replaced by a series of F-actin punctae, and YAP remained excluded from the nucleus (Fig. 1, A and B). To test the role of contractility, myosin II was inhibited directly with the myosin II ATPase inhibitor blebbistatin (50 μ M for 2 h) or indirectly by inhibiting the upstream regulator of myosin II, Rho-associated kinase (ROCK) with Y-27632 (50 μ M for 2 h). In fibroblasts, these treatments caused disruption of the peripheral F-actin band, whereas in MCF10-A cells the peripheral F-actin remained intact, but F-actin was additionally distributed through the cytoplasm. Despite these effects on the cytoskeleton, inhibition of contractility did not affect YAP localization, which remained excluded from the nucleus as in controls, as expected. Thus, in confluent epithelial cells and fibroblasts, YAP is retained in the cytoplasm independent of actomyosin activities.

Next, we tested the role of the actomyosin cytoskeleton in YAP localization in the absence of contact inhibition. Sparsely plated MEFs and MCF10A cells spread on fibronectin exhibited thick F-actin stress fibers, and unlike in dense cultures where YAP was retained in the cytoplasm, YAP was strongly localized to the nucleus (Fig. 1, C and E), indicating that YAP cytoplasmic retention requires cell-cell contact. Treatment of sparse cultures of both cell types with Lat-A showed that disruption of F-actin caused loss of nuclear YAP and its retention in the cytoplasm, similar to the effects of confluence (Fig. 1, C and E). Inhibition of contractility with blebbistatin or Y-27632 induced loss of stress fibers and promoted lamellipodia as expected (Fig. 1, C and E). However, surprisingly, in sparse cells with contraction inhibited, YAP remained partially localized to the nucleus

Yap Localization during Mechanosensing Requires Filamentous Actin



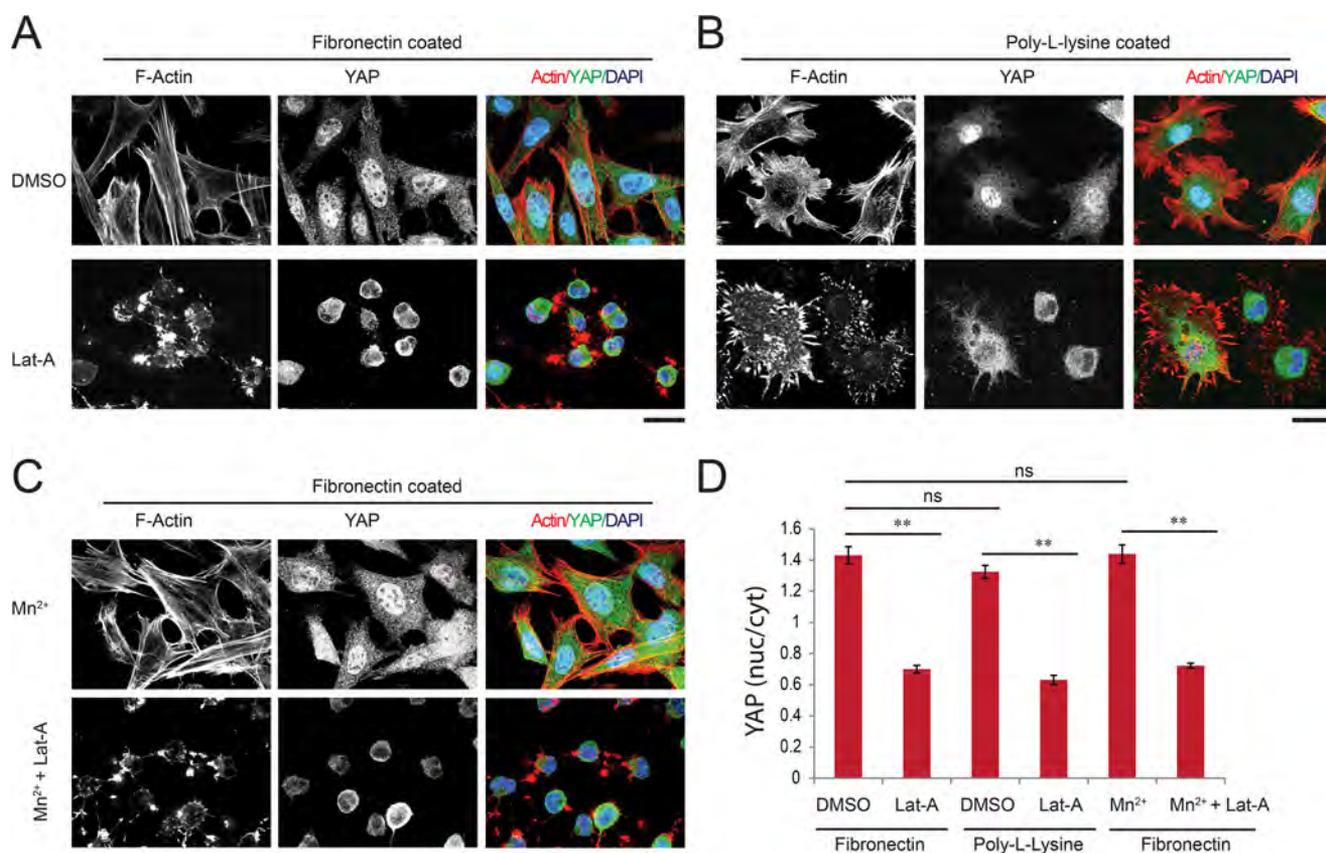


FIGURE 2. In sparse cells, F-actin cytoskeletal integrity is required for YAP nuclear localization but not integrin signaling. Maximum projections of confocal Z-stacks of sparsely plated MEFs on fibronectin (A and C) and poly-L-lysine (B)-coated coverslips and treated with Lat-A (2 μ M), manganese(II) chloride (Mn²⁺, 1 mM), and Mn²⁺ along with Lat-A (2 μ M) as indicated. F-actin was stained with phalloidin-561 (red), YAP with immunolocalization of α -YAP (green), and DNA with DAPI (blue). D, quantification of the nuclear/cytoplasmic ratio of YAP as done as described in the legend to Fig. 1. Scale bars 30 μ m. Error bars are S.E. Statistical comparison done with Student's *t* test between pair of samples connected with lines on the column plot: **, *p* value < 0.005; ns, *p* value > 0.05. One-way ANOVA test was performed on D, *p* value from ANOVA test is listed in Table 1.

in both MEFs and MCF10A cells (Fig. 1, C and E). Quantification of the nuclear:cytoplasmic ratio of YAP in immunofluorescence images confirmed these observations (Fig. 1, D and F). Together, these results indicate that in both fibroblasts and epithelial cells, YAP nuclear localization is primarily regulated by cell-cell contact. Although in the absence of cell-cell contact nuclear localization of YAP requires F-actin cytoskeletal integrity, but not ROCK-mediated myosin II contractility.

In Sparse Cells, Nuclear Localization of YAP Is Promoted by F-actin, and Is Independent of Integrin Signaling, Myosin II Contractility, and Is Not Fully Dependent on Ser¹¹² Phosphorylation—We next sought to understand the cytoskeletal mechanisms regulating YAP localization in the absence of cell-cell contact. As results were similar in fibroblasts and epithelial cells, subsequent studies were done on MEFs, because these were more easily maintained in sparse culture. Because disruption of cytoskeletal integrity by Lat-A causes dissolution of focal

adhesions, this suggests that effects of Lat-A on YAP localization could be due to changes in integrin signaling from focal adhesions. In addition, integrins have been shown to regulate Hippo signaling in MCF-10A cells under serum starvation (30). To test the role of integrin in mediating actin-dependent regulation of Yap in sparsely plated cells in the presence of serum, we cultured cells in serum-containing media and blocked integrin engagement by culturing MEFs on poly-L-lysine-coated coverslips (31) and compared YAP localization to that in cells grown on the integrin ligand fibronectin. This showed that on poly-L-lysine YAP was localized to the nucleus, whereas disrupting actin cytoskeleton by Lat-A excluded YAP from the nucleus under both plating conditions (Fig. 2, A, B, and D). Next, we constitutively activated integrin by treatment of cells with Mn²⁺ (1 mM) (32). This showed that manganese treatment failed to suppress YAP nuclear exclusion by actin cytoskeletal disruption (Fig. 2, C and D). Together, these results suggest that

FIGURE 1. In the absence of cell-cell contact, nuclear exclusion of YAP requires F-actin cytoskeletal integrity, but not ROCK-mediated myosin II contractility. Confocal immunofluorescence images of MEF (A) or human mammary epithelial cells (B, MCF10A) densely plated on fibronectin-coated coverslips and treated with the indicated agents for 2 h. F-actin was stained with phalloidin-488 (green), YAP with immunolocalization of α -YAP (red), and DNA with DAPI (blue). Maximum projections of confocal Z-stacks of sparsely plated MEFs (C) or MCF10A (E), on fibronectin-coated coverslips and treated with the indicated agents for 2 h. Immunofluorescence staining as in A and B. D and F, quantification of the nuclear/cytoplasmic ratio of YAP in the middle confocal slice of the nucleus. Scale bars 30 μ m. Error bars are S.E. Statistical comparison done with Student's *t* test between pair of samples connected with lines on the column plots: **, *p* value < 0.005; ns, *p* value > 0.05. One-way ANOVA tests were performed on group of data containing three or more than three data sets, *p* values from ANOVA tests are listed in Table 1.

Yap Localization during Mechanosensing Requires Filamentous Actin

in sparse cells in the presence of serum, YAP nuclear localization is independent of integrin signaling but dependent on cytoskeletal integrity.

The canonical pathway for regulation of YAP localization is through phosphoregulation of Ser¹¹², which promotes YAP nuclear exclusion and retention in the cytoplasm when phosphorylated, and nuclear localization when dephosphorylated (11, 12, 33, 34). To determine the effects of cytoskeletal perturbations on YAP phosphorylation, we treated sparse MEFs with cytoskeletal inhibitors followed by Western blot analysis of total YAP and YAP phosphorylated on Ser¹¹² (Ser(P)¹¹²-YAP). This showed that compared with control, both Lat-A (18) and blebbistatin treatment increased Ser(P)¹¹²-YAP (Fig. 3A). This is in agreement with effects of blebbistatin on YAP phosphorylation reported in cancer-associated fibroblasts (35) and in mammary epithelial cell (18).

Because our findings above showed that a substantial portion of YAP localized to the nucleus under myosin II inhibition in sparse cells, this suggested that phosphorylated YAP could localize in the nucleus when contractility is inhibited. To test this, we co-immunolocalized YAP and Ser(P)¹¹²-YAP in sparse MEFs under cytoskeletal perturbations. This showed that even in vehicle control, about half the Ser(P)¹¹²-YAP was localized to the nucleus (nuclear:cytoplasmic ratio of 1) (Fig. 3, B and C), in agreement with Wada *et al.* (19). However, Ser(P)¹¹²-YAP was significantly excluded from the nucleus when the actin cytoskeleton was depolymerized with Lat-A (Fig. 3, B and C). In contrast, treatment of cells with Y-27632 or a range of concentrations of blebbistatin had no effect on the nuclear:cytoplasmic ratio of Ser(P)¹¹²-YAP as compared with control (Fig. 3, B and C). Results from immunofluorescence assays were further supported by biochemical fractionation of the nucleus and cytoplasm from sparse MEFs. This showed in control cells that although YAP and Ser(P)¹¹²-YAP were concentrated in the nucleus, both were also present in the cytoplasm at low levels (Fig. 3D). Depolymerizing actin filaments with Lat-A decreased both YAP and Ser(P)¹¹²-YAP in the nuclear fraction relative to the cytoplasm (Fig. 3D). In contrast, in cells treated with blebbistatin and Y-27632, although the cytoplasmic fraction of YAP and Ser(P)¹¹²-YAP increased relative to control, there was still a substantial proportion of Ser(P)¹¹²-YAP in the nucleus (Fig. 3D). Furthermore, considerable Ser(P)¹¹²-YAP was present in the nuclear fraction across a range of blebbistatin doses (Fig. 3D). Taken together, these data show that neither lack of contractility nor YAP phosphorylation at Ser¹¹² are sufficient for YAP nuclear exclusion in sparse cells, but loss of F-actin is sufficient for nuclear exclusion of both YAP and Ser(P)¹¹²-YAP. Together our results indicate that in the absence of cell-cell contact, nuclear localization of YAP is promoted by F-actin cytoskeletal integrity, independent of integrin activation and ligand binding, myosin II contractility, or Ser¹¹²-YAP phosphorylation.

Mechanosensation of Stiff ECM Promotes Nuclear Localization of YAP and Ser(P)¹¹²-YAP in an F-actin-dependent Manner—Mechanosensation of ECM stiffness is known to modulate myosin II activity (21) and YAP nuclear localization (8) such that soft ECMs down-regulate myosin II activity, increase YAP phosphorylation, and retain YAP in the cyto-

plasm, whereas stiff ECMs promote myosin II activity, YAP dephosphorylation, and drive YAP into the nucleus (9, 36). Given our observation that Ser(P)¹¹²-YAP can reside in the nucleus during myosin II inhibition in sparse cells, we first sought to determine the effects of ECM compliance on YAP localization and phosphorylation, and then determine the role of F-actin cytoskeletal integrity in these effects. We plated cells sparsely on ECMs made of polyacrylamide substrates of defined stiffness with fibronectin coupled to their surface, and performed immunofluorescence and Western blot analysis of YAP localization and phosphorylation, and analysis of F-actin integrity. In agreement with previous reports (8, 9), in MEFs grown on stiff substrates (50 kPa), cells were well spread and exhibited stress fibers and strong nuclear localization of YAP, whereas on soft substrates (0.5 kPa), cells were poorly spread and YAP was excluded from the nucleus (Fig. 4, A and B). Western blot analysis showed that, consistent with previous studies (9), Ser(P)¹¹²-YAP was higher in cells on soft compared with that in cells on stiff ECMs (Fig. 4, C and D). Phospho-immunolocalization showed that in cells on soft ECMs, both YAP and Ser(P)¹¹²-YAP were significantly excluded from the nucleus (Fig. 4, E and F). Together, these results show that ECM compliance and myosin II inhibition both promote YAP phosphorylation at Ser¹¹², but show differential effects on YAP and Ser(P)¹¹²-YAP localization in sparse MEFs. Our results further show that the effects of substrate compliance on both YAP and cell morphology more closely resemble the effects of actin depolymerization (Figs. 1C and 4A) than they do the effects of myosin II inhibition (Fig. 1C), possibly implicating F-actin as a potential contractility-independent regulator of YAP during mechanosensation.

Whether stiffness sensing or cell-cell contacts modulate F-actin integrity to regulate YAP localization is not known. We thus sought to test whether F-actin regulates YAP nuclear localization independent of contractility during mechanosensation or cell-cell contact. To determine the effects of ECM compliance or cell-cell contact on F-actin integrity, we used either biochemical fractionation or localization-based (25) assays to quantify the F-actin to globular-actin (G-actin) ratio (F:G) in cells. We first validated the assays by determining the F:G ratio in cells treated with Lat-A. Fractionation of the detergent-soluble G-actin from the insoluble F-actin or direct staining of F-actin by fluorescent phalloidin and G-actin by fluorescent deoxyribonuclease-I (25) in individual cells both showed that Lat-A treatment significantly reduced the F:G ratio compared with vehicle control (Fig. 5, A–D), as expected. In contrast, treatment of cells with blebbistatin to inhibit myosin II contractility had no effect on the F:G ratio (Fig. 5, A–D).

We then determined how substrate compliance affected the F:G-actin ratio in cells plated sparsely on ECMs of different stiffness. Surprisingly, both detergent fractionation and localization-based assays showed that the F:G-actin ratio in cells grown on soft substrates was significantly lower compared with cells grown on stiff substrates (Fig. 5, E–H). This shows that ECM mechanosensing regulates both contractility (20, 21) and F-actin integrity. Together with our aforementioned results, this shows that the high F:G-actin ratio correlates with YAP nuclear localization, but contractility and YAP phosphoryla-

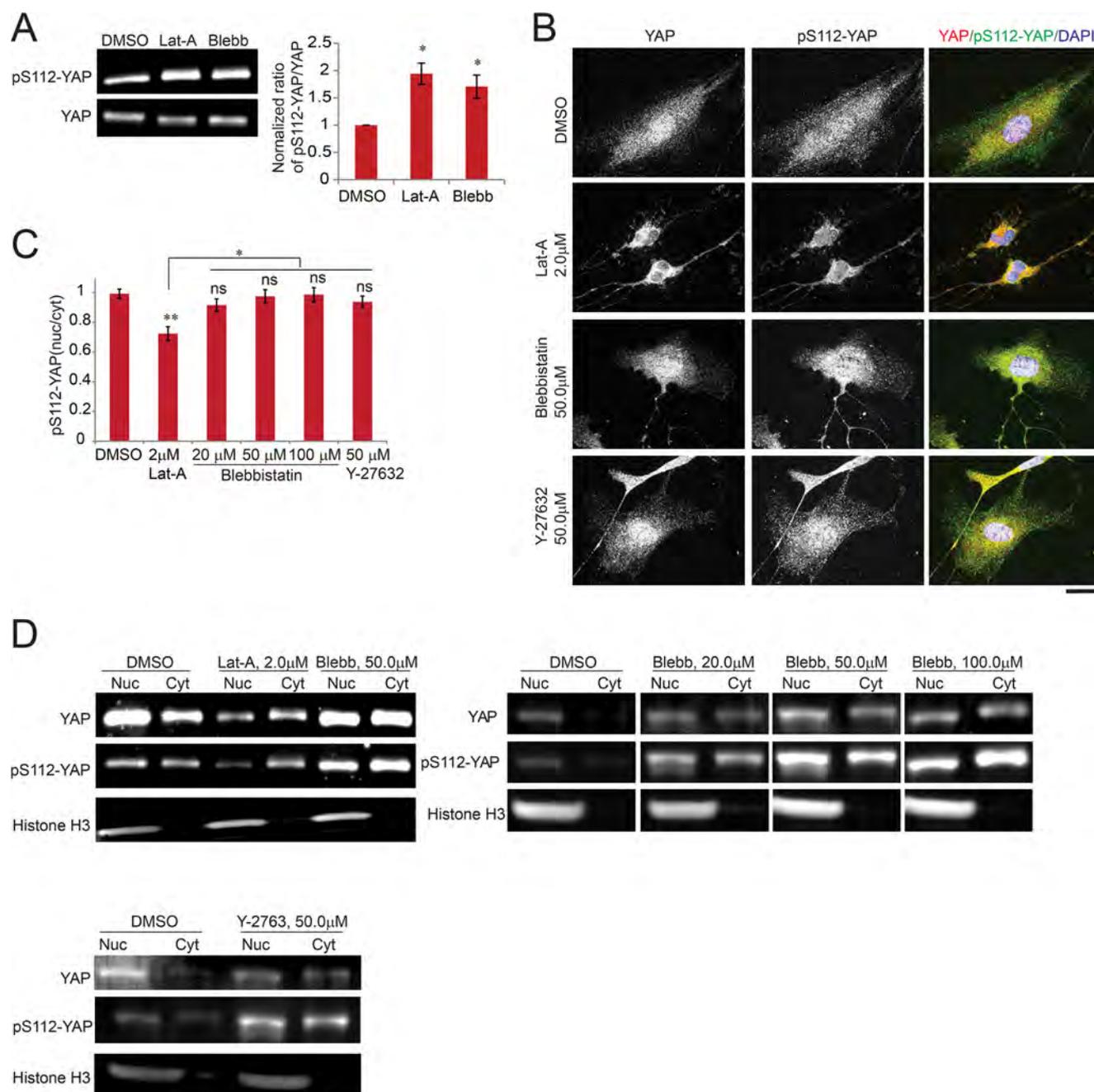


FIGURE 3. In sparse cells, nuclear localization of YAP is promoted by F-actin independent of myosin II contractility and is not fully dependent on Ser¹¹² phosphorylation. *A*, left, representative Western blots of whole cell lysates from MEFs grown on fibronectin-coated culture dishes and treated with DMSO, Lat-A (2 μM), and blebbistatin (50 μM) for 2 h and probed with antibodies to YAP or YAP phosphorylated on Ser¹¹² (pS112-YAP). Right, quantification represents the ratio of integrated band intensity between Ser(P)¹¹²-YAP and YAP from background subtracted images and normalized with respect to DMSO control ratio. *B*, maximum intensity projections of confocal Z-stack immunofluorescence images of sparsely plated MEFs treated with the indicated agents for 2 h. YAP was immunostained with α-YAP (red), α-Ser(P)¹¹²-YAP (green), and DNA with DAPI (blue). Note that Ser(P)¹¹²-YAP was better detected with methanol fixation (used here), however, YAP was better detected with paraformaldehyde fixation (used in Fig. 1). *C*, quantification of the nuclear/cytoplasmic ratio of pS112-YAP in the middle confocal slice of the nucleus. *D*, representative Western blots of the nucleus (Nuc) and cytoplasmic (Cyt) fractions from MEFs grown on fibronectin-coated culture dishes with the indicated treatments for 2 h for the indicated proteins and probed with antibodies to YAP, Ser(P)¹¹²-YAP, and Histone H3. Scale bars 30 μm. Error bars are S.E. Statistical comparison done with Student's *t* test between pair of samples connected with lines on the column plots; *, *p* value < 0.05; **, *p* value < 0.005; ns, *p* value > 0.05. One-way ANOVA tests were performed on group of data containing three or more data sets, *p* values from ANOVA tests are listed in Table 1.

tion do not. This further suggests that F-actin may be responsible for YAP nuclear localization in blebbistatin-treated cells.

To test the hypothesis that F-actin regulates YAP localization when myosin II is inhibited, we used a high dose of blebbistatin (100 μM) for 2 h to completely inhibit myosin contractility (37), then subsequently depolymerized actin filaments with Lat-A.

Similar to 50 μM blebbistatin treatment, YAP was still partially localized to the nucleus even in the presence of 100 μM blebbistatin (Fig. 5, *I* and *J*). Importantly, however, Lat-A mediated actin depolymerization in blebbistatin-treated cells caused YAP to be excluded from the nucleus (Fig. 5, *I* and *J*). Coupled with the results from low-stiffness substrates, these results sug-

Yap Localization during Mechanosensing Requires Filamentous Actin

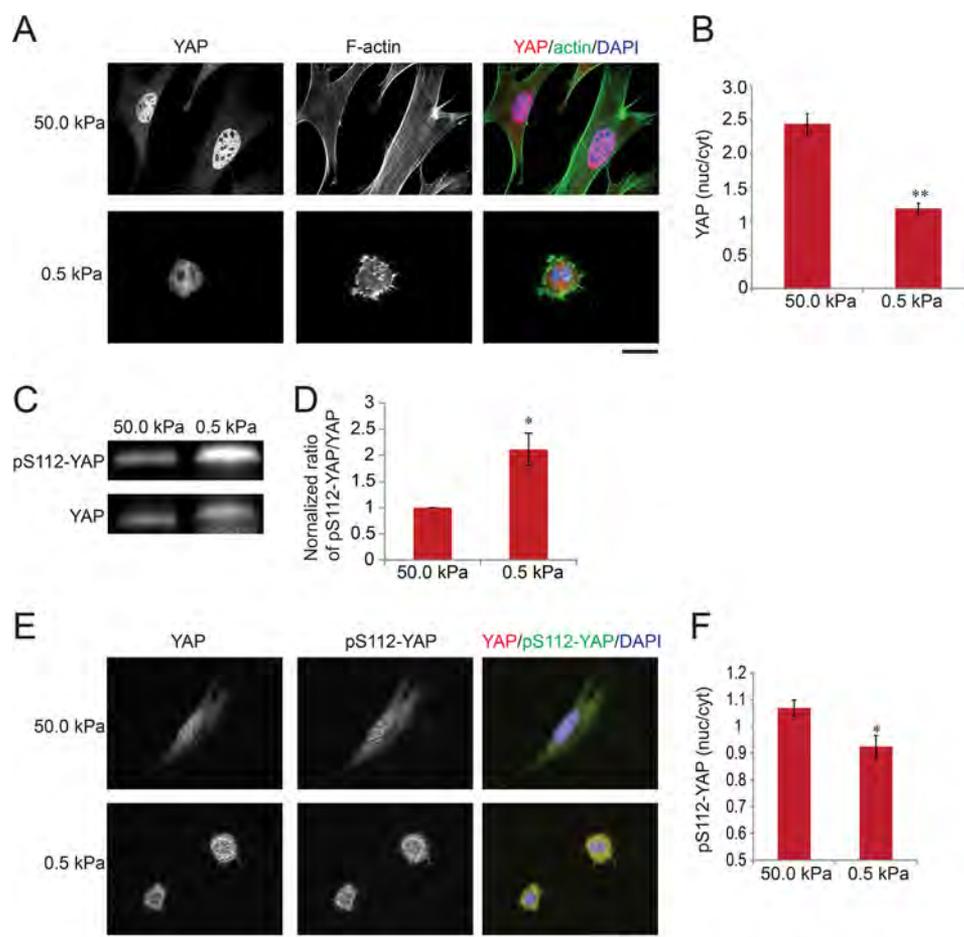


FIGURE 4. Mechanosensation of soft ECM promotes nuclear exclusion of YAP and phospho-Ser¹¹²-YAP. *A–F*, MEFs were grown on fibronectin-coupled polyacrylamide hydrogels of 50.0 (stiff) and 0.5 kPa (soft) stiffness. *A*, maximum intensity projections of confocal Z-stack immunofluorescence images of YAP (red), phalloidin staining of F-actin (green), and DAPI staining of DNA (blue). *B*, quantification of the nuclear/cytoplasmic ratio of YAP in the middle confocal slice of the nucleus. *C*, representative Western blots from whole cell lysates of MEFs grown on fibronectin-coated soft and stiff substrates and probed with antibodies to YAP and YAP phosphorylated on Ser¹¹² (pS112-YAP). *D*, quantification of the ratio of integrated band intensity between Ser(P)¹¹²-YAP and YAP from background-subtracted Western blot images and normalized with respect to 50.0 kPa. *E*, maximum intensity projections of confocal Z-stack immunofluorescence images of sparsely plated MEFs grown on fibronectin-coated stiff and soft substrates and fixed with methanol and immunostained to detect YAP (red), Ser(P)¹¹²-YAP (green), and stained with DAPI for DNA (blue). *F*, quantification of the nuclear/cytoplasmic ratio of Ser(P)¹¹²-YAP in the middle confocal slice of the nucleus. Scale bars 30 μ m. Error bars are S.E., Statistical comparison done with Student's *t* test: *, *p* value < 0.05; **, *p* value < 0.005.

gest that F-actin is responsible for YAP nuclear localization in cells under low contractility conditions.

To test whether cell-cell contact-mediated inhibition of YAP nuclear localization also occurs through actin disassembly, we compared the actin assembly state in dense and sparse cultures. This showed that the F:G-actin ratio was not effected by cell confluence (Fig. 5, *K* and *L*), suggesting that signaling from adherens junctions during contact inhibition dominates over cytoskeletal integrity in regulation of YAP nuclear localization. However, it is also possible that changes in F-actin architecture in dense culture compared with sparse culture (Fig. 1, *A–C* and *E*) may contribute to YAP regulation. Together, these data support the notion that in the absence of cell-cell contact, F-actin, and not myosin II contraction, is the critical regulator of YAP nuclear localization during mechanosensation.

Two Pathways Regulate YAP Localization in Sparse Cells: an F-actin-dependent, Phosphorylation-independent Pathway, and a Myosin II- and Phosphorylation-dependent Pathway— Our results suggest that F-actin can override the canonical phosphoregulation of YAP localization to promote nuclear

localization of phospho-YAP. We sought to test this by assaying the effects of cytoskeletal perturbation on constitutively active mutants of tagged YAP constructs (FLAG-YAP^{wt}, GFP-YAP^{wt}) that cannot be regulated by Lats1/2 kinase by using FLAG-YAP^{8SA} that has all the known Lats phosphorylation sites mutated to alanines, or GFP-YAP^{S112A}, which carries a mutation at the Lats phosphorylation site that has been shown to be the most potent regulator of YAP nuclear localization (12, 38). To first validate the functionality of the YAP constructs, we examined their localization in MEFs where phosphorylation by Lats kinase is known to be key for YAP nuclear exclusion (12, 39). This showed under sparse culture conditions where YAP phosphorylation is known to be low (12), both the FLAG-tagged and GFP-tagged YAP wild-type constructs and both FLAG-YAP^{8SA} and GFP-YAP^{S112A} localized to the nucleus (Fig. 6, *A*, *I*, and *K*), as expected. In dense culture conditions where YAP is known to become phosphorylated (12), FLAG-YAP^{wt} and GFP-YAP^{wt} were reduced in the nucleus, whereas FLAG-YAP^{8SA} and GFP-YAP^{S112A} were significantly enriched in the nucleus (Fig. 6, *D* and *F–H*). Thus, tagged YAP constructs

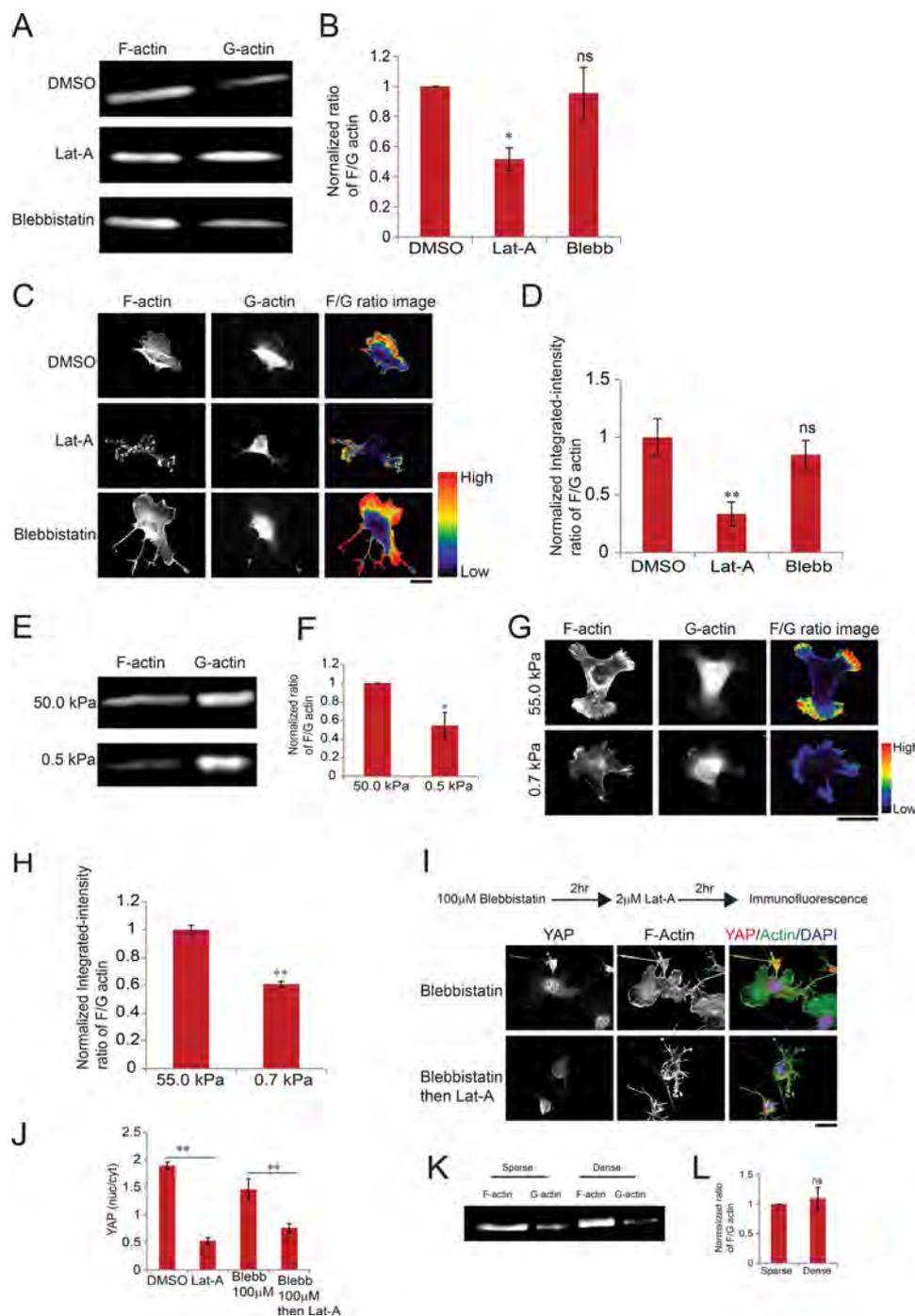
Yap Localization during Mechanosensing Requires Filamentous Actin

localize in a manner that is consistent with that described for Lats-mediated phosphoregulation of endogenous YAP in sparse and dense culture.

We next examined the localization of the phosphomutants in response to cytoskeletal perturbations. This showed that in sparse culture, treatment with Lat-A to depolymerize F-actin induced nuclear exclusion of all of the constructs (Fig. 6, *B*, *C*, and *I-L*), significantly decreasing the nuclear:cytoplasmic ratio of the YAP signal, albeit to a lesser extent for the 8SA mutant relative to the S112A mutant or for the 8SA mutant relative to the wild-types (Fig. 6, *A-C* and *K-L*). In dense culture, although actin depolymerization did not affect *FLAG-YAP^{wt}* localization,

this treatment surprisingly caused exclusion of *FLAG-YAP^{8SA}* from the nucleus (Fig. 6, *D-F*). Together, these results support the notion of a phosphorylation-independent but actin cytoskeleton-dependent pathway for YAP nuclear exclusion, even in contact-inhibited cells.

We then determined the role of YAP Ser¹¹² phosphorylation in myosin-dependent regulation of YAP localization. We found that inhibition of myosin II caused *GFP-YAP^{wt}* to behave similarly to that of the endogenous protein with a partial decrease in the nuclear localization compared with untreated cells (Figs. 6, *I* and *J*, and 1, *C* and *D*). However, the non-phosphorylatable *GFP-YAP^{S112A}* remained strongly localized to the nucleus,



Yap Localization during Mechanosensing Requires Filamentous Actin

with no significant difference in nuclear:cytoplasmic signal ratio compared with the same construct in controls (Fig. 6, *K* and *L*). These data, along with endogenous YAP data, suggest that in the absence of contact inhibition, two distinct pathways act to regulate YAP localization in MEFs, an F-actin-mediated pathway that is Ser(P)¹¹²-YAP-independent, and a myosin II-mediated pathway that requires phosphoregulation of Ser¹¹².

ECM Compliance-mediated Regulation of YAP Nuclear Localization Occurs via an F-actin-dependent, Phosphorylation-independent Pathway—We then examined the role of Ser¹¹² phosphorylation in F-actin-dependent nuclear localization of YAP during mechanosensation in MEFs. We expressed either *YAP1^{wt}-GFP* or *YAP1^{S112A}-GFP* in sparse cells grown on stiff and soft ECMs. This showed that on stiff ECMs *YAP1^{wt}-GFP* or *YAP1^{S112A}-GFP* constructs all localized to the nucleus. In contrast, on soft ECMs cells formed small clusters and could not spread effectively, and *YAP1^{wt}-GFP* or *YAP1^{S112A}-GFP* constructs were excluded from the nucleus, and exhibited a significantly reduced nuclear:cytoplasmic localization ratio (Fig. 7, *A–D*). Together with our observation of reduced F-actin content in cells plated on soft ECMs, this shows ECM compliance-mediated regulation of YAP nuclear localization occurs through an F-actin-dependent but phospho-Ser¹¹²-independent pathway.

Discussion

Our study shows that in both mouse embryonic fibroblasts and in human mammary epithelial cells, contact inhibition promotes nuclear exclusion of YAP, overriding actomyosin cytoskeletal regulation of YAP. However, in the absence of cell-cell contact, integrity of the actin filament cytoskeleton becomes the dominant regulator of YAP nuclear localization, independent of ROCK-mediated myosin II contractility. A detailed characterization in MEFs revealed that actomyosin contractility suppresses YAP phosphorylation at Ser¹¹², yet in the absence of cell-cell contact and contractility, Ser¹¹² phosphoregulation is overridden by actin cytoskeletal integrity leading to nuclear localization of both phospho- and non-phospho-YAP. These data along with YAP phosphomutant data show the existence of two mechanisms for cytoskeleton-mediated YAP regulation; one actomyosin contractility-mediated regulation

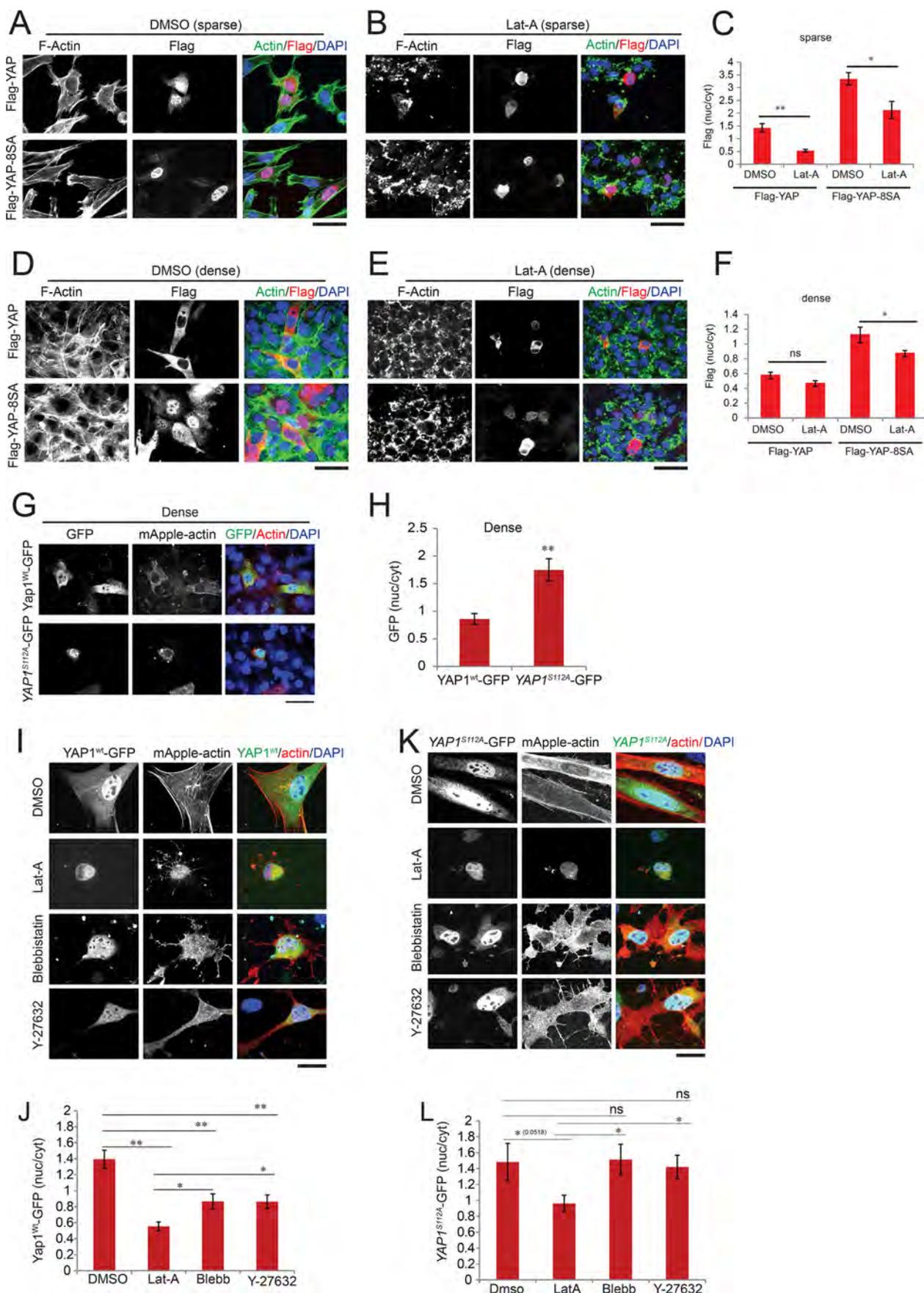
of YAP phosphorylation, and a second cytoskeletal integrity-mediated dominant mechanism that overrides phosphoregulation. This actin-mediated YAP regulation is also conserved during mechanotransduction, where we found YAP nuclear localization through ECM stiffness sensation occurs through an F-actin-dependent but phospho-Ser¹¹²-independent manner.

Hippo signaling-mediated YAP phosphorylation and nuclear exclusion have been well documented in animal tissue and *in vitro* dense cell culture systems (12, 40–43). Phosphorylation mediated YAP nuclear exclusion is largely dependent on cell-cell AJ, as YAP is sequestered at the AJ by α -catenin and *in vitro* assay showed Ser(P)¹¹²-YAP binding to 14-3-3 also depends on α -catenin (40). In the absence of α -catenin, this complex breaks apart and YAP gets into the nucleus independent of Hippo signaling (40). In sparse culture conditions there is no AJ, possibly leading to the disassembly of the α -catenin-YAP and 14-3-3-Ser(P)¹¹²-YAP- α -catenin complex, which could provide a possible mechanism for the observed YAP and Ser(P)¹¹²-YAP localization in the nucleus independent of contractility. In addition, a recent report showed that even under activation of the Hippo signaling cascade and YAP phosphorylation (Ser¹¹²), YAP localizes to the nucleus of intestinal epithelial cells via the Src family kinase Yes, causing more cell proliferation (44). This also suggests the existence of additional mechanisms in regulation of YAP nuclear localization other than phosphorylation of the Ser¹¹² residue.

Our data provides a step forward in understanding actin-mediated YAP nuclear localization beyond the requirement of actomyosin contractility and Ser¹¹² phosphorylation of YAP. A possible molecular player for such actin-mediated YAP regulation is through angiomin-like proteins (24). Angiominins bind the WW domain of YAP independent of its phosphorylation at Ser¹¹² residue, and overexpression of angiominin could sequester constitutively active YAP (*YAP^{S112A}*) in the cytoplasm (45). Angiominins competitively bind to both F-actin and YAP, and overexpression of an actin binding mutant of angiominin sequesters more YAP in the cytoplasm than overexpression of wild-type protein (24). This mechanism could underlie our observations that a high F:G-actin ratio in control and myosin-inhibited cells was associated with both YAP and

FIGURE 5. F-actin promotes YAP nuclear localization in the absence of actomyosin contractility. *A* and *E* are representative Western blots of the Triton-insoluble (*F-actin*) and Triton-soluble (*G-actin*) fractions extracted from MEFs grown on fibronectin-coated culture dish treated with DMSO, Lat-A (2 μ M), and blebbistatin (50 μ M) for 2 h (*A*) or grown on fibronectin-coupled polyacrylamide hydrogels of 50.0 (stiff) and 0.5 kPa (soft) stiffness (*E*) and probed with antibodies to actin. *B* and *F*, quantification represents the ratio of integrated band intensity between the F-actin and G-actin from background-subtracted Western blot images. The proportions of F- and G-actin were determined relative to the total actin content (total actin content = $F_s + G_s$ signal = 100%; $F = F_s/F_s + G_s$; $G = G_s/F_s + G_s$), and these values were used to determine the F:G ratio ($F/G = (F_s/F_s + G_s)/(G_s/F_s + G_s)$). The F:G ratio for the DMSO condition was then used as the normalization control to compare the effects of perturbations (*B*) or to 50-kPa control ratio (*F*). *C* and *G*, sum projections of confocal Z-stacks of MEFs grown on fibronectin-coated culture dishes and treated with DMSO, Lat-A (2 μ M), and blebbistatin (50 μ M) for 2 h (*C*) or grown on fibronectin-coupled polyacrylamide hydrogels of 55.0 kPa (stiff) and 0.7 kPa (soft) stiffness (*G*) and stained for F-actin (phalloidin-488, green) and G-actin (DNase I-594, red). The F:G-actin ratio images were generated using ImageJ as explained in (48). *D* and *H*, quantification represents the ratio of integrated fluorescence intensity measured from the background subtracted sum-projection of F- and G-actin images, normalized with respect to the average ratio of DMSO control (*D*) or 55 kPa control (*H*). S.E. and *p* values were measured from the non-normalized data set. *I*, above, experimental scheme for examining YAP localization in MEFs first treated with 100 μ M blebbistatin for 2 h followed by Lat-A (2 μ M) treatment for another 2 h (blebbistatin then Lat-A), and compared with 100 μ M blebbistatin treatment for 2 h alone. Maximum intensity projections of confocal Z-stack immunofluorescence images of sparsely plated MEFs grown on fibronectin-coated coverslips and were treated as explained above. F-actin was stained with phalloidin-488 (green), YAP with immunolocalization of α -YAP (red), and DNA with DAPI (blue). *J*, quantification of the nuclear/cytoplasmic ratio of YAP in the middle confocal slice of the nucleus. DMSO and Lat-A data are from Fig. 1D. *K*, representative Western blot of the Triton-insoluble (F-actin) and Triton-soluble (G-actin) fractions extracted from MEFs grown sparsely and densely on fibronectin coated culture dish. *L*, quantification represents the ratio of integrated band intensity between the F-actin and G-actin from background-subtracted Western blot images and normalized with respect to sparse ratio. Scale bars 30 μ m. Error bars are S.E., Statistical comparison was done with Student's *t* test between pair of samples connected with lines on the column plots: *, *p* value < 0.05; **, *p* value < 0.005; ns, *p* value > 0.05. One-way ANOVA tests were performed on a group of data containing three or more data sets, *p* values from ANOVA tests are listed in Table 1.

Yap Localization during Mechanosensing Requires Filamentous Actin



Yap Localization during Mechanosensing Requires Filamentous Actin

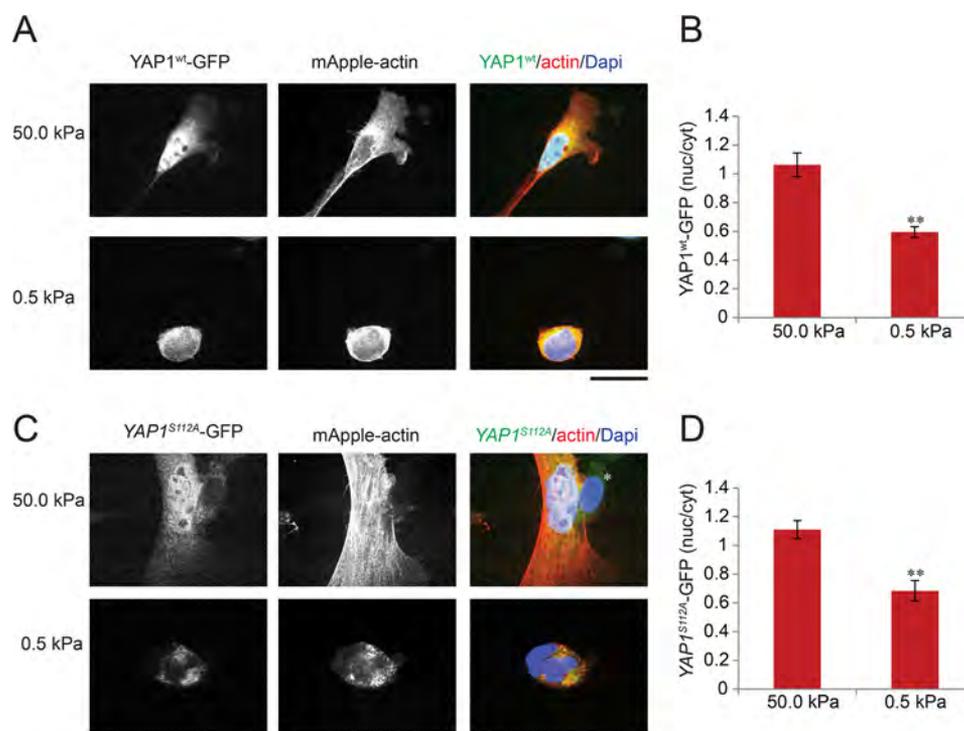


FIGURE 7. ECM compliance-mediated regulation of YAP nuclear localization is predominantly regulated by F-actin-dependent, phosphorylation-independent pathway. MEFs were grown on fibronectin-coupled polyacrylamide hydrogels of 50.0 (stiff) and 0.5 kPa (soft) stiffness. YAP1^{wt}-GFP (A, green) or non-phosphorylatable YAP1^{S112A}-GFP (C, green) was detected by the GFP signal, actin by red fluorescent protein mApple (red), and DNA stained with DAPI (blue). * in C, marks the nuclei of an adjacent untransfected cell. B and D, quantification of the nuclear/cytoplasmic ratio of YAP1-GFP constructs in the middle confocal slice of the nucleus. Scale bars 30 μ m. Error bars are S.E., Statistical comparison was done with Student's *t* test: **, *p* value < 0.005.

Ser(P)¹¹²-YAP nuclear localization. A high level of F-actin could sequester angiostatins, leaving YAP and Ser(P)¹¹²-YAP free to be imported into the nucleus. In support to this notion, it was also reported that knocking down angiostatins significantly rescued YAP nuclear localization in blebbistatin-treated cells (24). In support of our findings of an F-actin-dependent pathway of YAP nuclear localization, it was shown that knocking down the actin depolymerization factor cofilin, which is known to increase the F:G-actin ratio (28), rescued YAP transcriptional activity on soft substrates (10). Promoting more F-actin by knocking down cofilin or capping protein in animal tissue was also found to increase YAP nuclear localization and tissue growth (22, 23). In agreement with our findings, it was also reported that overexpressing actin increases F-actin, promoting cell spreading and thus overriding the small cell shape phenotype seen with cells plated on soft substrates (46). F-actin-mediated YAP regulation could have a physiologically important role as it was shown that promoting F-actin polymerization increased ovarian follicle growth mediated by YAP (47).

Our study suggests a model where in cells at low density, YAP is nuclear and has low levels of phosphorylation because of the lack of adherens junctions and the high level of myosin II contractile activity. Also at low cell density, F-actin structures promote YAP nuclear localization. Inhibition of myosin II contractility in cells in sparse culture increases YAP phosphorylation, but not sufficiently to trigger cytoplasmic accumulation of YAP. However, F-actin disruption can trigger cytoplasmic localization of YAP. At high cell density, YAP is highly phosphorylated and out of the nucleus, presumably because of signaling from adherens junctions. The observation that non-phosphorylatable YAP localizes to the nucleus in cells at high density is consistent with an essential role for phosphorylation in YAP nuclear exclusion under these conditions. One question that remains is why myosin II inhibition in cells at low density triggers YAP phosphorylation, but not nuclear exclusion, whereas at high cell density, YAP phosphorylation appears to effectively promote YAP nuclear exclusion. One possibility is that YAP is much more highly phosphorylated on Ser¹¹² and

FIGURE 6. Loss of cytoskeletal integrity significantly excludes constitutively active YAP from the nucleus. Maximum intensity projections of confocal Z-stacks of MEFs transfected with plasmids expressing FLAG-YAP and FLAG-YAP-8SA, cultured sparsely (A and B) or middle confocal slice of densely plated MEFs (D and E) on fibronectin-coated coverslips and treated with DMSO and Lat-A (2 μ M) as indicated. C and F, quantification of the nuclear/cytoplasmic ratio of FLAG in the middle confocal slice of the nucleus. F-actin stained with phalloidin-488 (green), FLAG-YAP with α -FLAG (red), and DNA with DAPI (blue). G–L, confocal images of MEFs transfected with plasmids expressing YAP1-GFP, or non-phosphorylatable YAP1^{S112A}-GFP (green) of mouse origin and mApple-actin (red) that were densely plated on fibronectin coated coverslips for 18 h and fixed with paraformaldehyde (G). MEFs were transfected with plasmids expressing YAP1-GFP (I, green), non-phosphorylatable YAP1^{S112A}-GFP (K, green) of mouse origin and mApple-actin (red) and grown on fibronectin-coated coverslips sparsely for 18 h followed by DMSO, Lat-A (2 μ M), blebbistatin (50 μ M), and Y-27632 (50 μ M) treatment for 2 h and formaldehyde fixation. YAP1 was detected by the GFP signal (green), actin by red fluorescent protein mApple (red), and DNA stained with DAPI (blue). H, J, and L, quantification of the nuclear/cytoplasmic ratio of YAP1-GFP in the middle confocal slice of the nucleus. Scale bars 30 μ m. Error bars are S.E., Statistical comparison done with Student's *t* test between pairs of samples connected with lines on the column plots: *, *p* value < 0.05; **, *p* value < 0.005; ns, *p* value > 0.05. One-way ANOVA tests were performed on group of data containing three or more data sets, *p* values from ANOVA tests are listed in Table 1.

other LATS sites in cells at high density compared with cells at low density treated with myosin inhibitors. This prediction awaits future experimentation.

Contractility- and phosphorylation (Ser¹¹²)-independent localization of YAP in the nucleus, especially in the absence of cell-cell contact, may reflect a cellular adaptation for cancer cells. Cancer cells during metastasis often lose cell-cell contact and migrate through tissues of different stiffness. Having contractility and phosphorylation-independent nuclear localization of YAP might promote cancer cell proliferation and maintenance in soft tissue.

Author Contributions—A. D. performed experiments and quantification, A. D., R. S. F., D. P., and C. M. W. interpreted data and wrote the manuscript.

Acknowledgments—We thank Mary Beckerle for the kind gift of MEFs, William Shin for maintenance of the Waterman Lab microscopes, and Schwanna Thacker for administrative assistance.

References

1. Wu, S., Huang, J., Dong, J., and Pan, D. (2003) Hippo encodes a Ste-20 family protein kinase that restricts cell proliferation and promotes apoptosis in conjunction with salvador and warts. *Cell* **114**, 445–456
2. Lian, I., Kim, J., Okazawa, H., Zhao, J., Zhao, B., Yu, J., Chinnaiyan, A., Israel, M. A., Goldstein, L. S., Abujarour, R., Ding, S., and Guan, K. L. (2010) The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. *Genes Dev.* **24**, 1106–1118
3. Pan, D. (2010) The hippo signaling pathway in development and cancer. *Dev. Cell* **19**, 491–505
4. Polesello, C., and Tapon, N. (2007) Salvador-warts-hippo signaling promotes *Drosophila* posterior follicle cell maturation downstream of notch. *Curr. Biol.* **17**, 1864–1870
5. Lu, L., Li, Y., Kim, S. M., Bossuyt, W., Liu, P., Qiu, Q., Wang, Y., Halder, G., Finegold, M. J., Lee, J. S., and Johnson, R. L. (2010) Hippo signaling is a potent *in vivo* growth and tumor suppressor pathway in the mammalian liver. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1437–1442
6. Yu, F. X., Zhao, B., Panupinthu, N., Jewell, J. L., Lian, I., Wang, L. H., Zhao, J., Yuan, H., Tumaneng, K., Li, H., Fu, X. D., Mills, G. B., and Guan, K. L. (2012) Regulation of the Hippo-YAP pathway by G-protein-coupled receptor signaling. *Cell* **150**, 780–791
7. Hirate, Y., Hirahara, S., Inoue, K., Suzuki, A., Alarcon, V. B., Akimoto, K., Hirai, T., Hara, T., Adachi, M., Chida, K., Ohno, S., Marikawa, Y., Nakao, K., Shimono, A., and Sasaki, H. (2013) Polarity-dependent distribution of angiostatin localizes Hippo signaling in preimplantation embryos. *Curr. Biol.* **23**, 1181–1194
8. Dupont, S., Morsut, L., Aragona, M., Enzo, E., Giulitti, S., Cordenonsi, M., Zanconato, F., Le Digabel, J., Forcato, M., Bicciato, S., Elvassore, N., and Piccolo, S. (2011) Role of YAP/TAZ in mechanotransduction. *Nature* **474**, 179–183
9. Codelia, V. A., Sun, G., and Irvine, K. D. (2014) Regulation of YAP by mechanical strain through Jnk and Hippo signaling. *Curr. Biol.* **24**, 2012–2017
10. Aragona, M., Panciera, T., Manfrin, A., Giulitti, S., Michielin, F., Elvassore, N., Dupont, S., and Piccolo, S. (2013) A mechanical checkpoint controls multicellular growth through YAP/TAZ regulation by actin-processing factors. *Cell* **154**, 1047–1059
11. Dong, J., Feldmann, G., Huang, J., Wu, S., Zhang, N., Comerford, S. A., Gayyed, M. F., Anders, R. A., Maitra, A., and Pan, D. (2007) Elucidation of a universal size-control mechanism in *Drosophila* and mammals. *Cell* **130**, 1120–1133
12. Zhao, B., Wei, X., Li, W., Udan, R. S., Yang, Q., Kim, J., Xie, J., Ikenoue, T., Yu, J., Li, L., Zheng, P., Ye, K., Chinnaiyan, A., Halder, G., Lai, Z. C., and Guan, K. L. (2007) Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. *Genes Dev.* **21**, 2747–2761
13. Chan, E. H., Nousiainen, M., Chalamalasetty, R. B., Schäfer, A., Nigg, E. A., and Silljé, H. H. (2005) The Ste20-like kinase Mst2 activates the human large tumor suppressor kinase Lats1. *Oncogene* **24**, 2076–2086
14. McClatchey, A. I., and Yap, A. S. (2012) Contact inhibition (of proliferation) redux. *Curr. Opin. Cell Biol.* **24**, 685–694
15. Rauskolb, C., Sun, S., Sun, G., Pan, Y., and Irvine, K. D. (2014) Cytoskeletal tension inhibits Hippo signaling through an Ajuba-Warts complex. *Cell* **158**, 143–156
16. Piccolo, S., Dupont, S., and Cordenonsi, M. (2014) The biology of YAP/TAZ: hippo signaling and beyond. *Physiol. Rev.* **94**, 1287–1312
17. Zhong, W., Tian, K., Zheng, X., Li, L., Zhang, W., Wang, S., and Qin, J. (2013) Mesenchymal stem cell and chondrocyte fates in a multishear microdevice are regulated by Yes-associated protein. *Stem Cells Dev.* **22**, 2083–2093
18. Zhao, B., Li, L., Wang, L., Wang, C. Y., Yu, J., and Guan, K. L. (2012) Cell detachment activates the Hippo pathway via cytoskeleton reorganization to induce anoikis. *Genes Dev.* **26**, 54–68
19. Wada, K., Itoga, K., Okano, T., Yonemura, S., and Sasaki, H. (2011) Hippo pathway regulation by cell morphology and stress fibers. *Development* **138**, 3907–3914
20. Discher, D. E., Janmey, P., and Wang, Y. L. (2005) Tissue cells feel and respond to the stiffness of their substrate. *Science* **310**, 1139–1143
21. Paszek, M. J., Zahir, N., Johnson, K. R., Lakins, J. N., Rozenberg, G. I., Gefen, A., Reinhart-King, C. A., Margulies, S. S., Dembo, M., Boettiger, D., Hammer, D. A., and Weaver, V. M. (2005) Tensional homeostasis and the malignant phenotype. *Cancer Cell* **8**, 241–254
22. Fernández, B. G., Gaspar, P., Brás-Pereira, C., Jezowska, B., Rebelo, S. R., and Janody, F. (2011) Actin-capping protein and the Hippo pathway regulate F-actin and tissue growth in *Drosophila*. *Development* **138**, 2337–2346
23. Sansores-Garcia, L., Bossuyt, W., Wada, K., Yonemura, S., Tao, C., Sasaki, H., and Halder, G. (2011) Modulating F-actin organization induces organ growth by affecting the Hippo pathway. *EMBO J.* **30**, 2325–2335
24. Mana-Capelli, S., Paramasivam, M., Dutta, S., and McCollum, D. (2014) Angiomotins link F-actin architecture to Hippo pathway signaling. *Mol. Biol. Cell* **25**, 1676–1685
25. Cramer, L. P., Briggs, L. J., and Dawe, H. R. (2002) Use of fluorescently labeled deoxyribonuclease I to spatially measure G-actin levels in migrating and non-migrating cells. *Cell Motil Cytoskeleton* **51**, 27–38
26. Shin, W. D., Fischer, R.S., Kanchanawong, P., Kim, Y., Lim, J., Myers, K.A., Nishimura, Y., Plotnikov, S.V., Thievessen, I., Yarar, D., Sabass, B., and Waterman, C. M. (2010) *A Versatile, Multicolor Total Internal Reflection Fluorescence and Spinning-Disk Confocal Microscope System for High-Resolution Live Cell Imaging*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
27. Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475–1489
28. Rasmussen, I., Pedersen, L. H., Byg, L., Suzuki, K., Sumimoto, H., and Vilhardt, F. (2010) Effects of F/G-actin ratio and actin turn-over rate on NADPH oxidase activity in microglia. *BMC Immunol.* **11**, 44
29. Abercrombie, M. (1970) Contact inhibition in tissue culture. *In Vitro* **6**, 128–142
30. Kim, N. G., and Gumbiner, B. M. (2015) Adhesion to fibronectin regulates Hippo signaling via the FAK-Src-PI3K pathway. *J. Cell Biol.* **210**, 503–515
31. Aoudjit, F., and Vuori, K. (2001) Integrin signaling inhibits paclitaxel-induced apoptosis in breast cancer cells. *Oncogene* **20**, 4995–5004
32. Cluzel, C., Saltel, F., Lussi, J., Paulhe, F., Imhof, B. A., and Wehrle-Haller, B. (2005) The mechanisms and dynamics of (alpha)v(beta)3 integrin clustering in living cells. *J. Cell Biol.* **171**, 383–392
33. Hao, Y., Chun, A., Cheung, K., Rashidi, B., and Yang, X. (2008) Tumor suppressor LATS1 is a negative regulator of oncogene YAP. *J. Biol. Chem.* **283**, 5496–5509
34. Mo, J. S., Meng, Z., Kim, Y. C., Park, H. W., Hansen, C. G., Kim, S., Lim, D. S., and Guan, K. L. (2015) Cellular energy stress induces AMPK-mediated regulation of YAP and the Hippo pathway. *Nat. Cell. Biol.* **17**,

Yap Localization during Mechanosensing Requires Filamentous Actin

500–510

35. Calvo, F., Ege, N., Grande-Garcia, A., Hooper, S., Jenkins, R. P., Chaudhry, S. I., Harrington, K., Williamson, P., Moeendarbary, E., Charras, G., and Sahai, E. (2013) Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts. *Nat. Cell. Biol.* **15**, 637–646
36. Sun, Y., Yong, K. M., Villa-Diaz, L. G., Zhang, X., Chen, W., Philson, R., Weng, S., Xu, H., Krebsbach, P. H., and Fu, J. (2014) Hippo/YAP-mediated rigidity-dependent motor neuron differentiation of human pluripotent stem cells. *Nat. Mater.* **13**, 599–604
37. Shutova, M. S., Spessott, W. A., Giraudo, C. G., and Svitkina, T. (2014) Endogenous species of mammalian nonmuscle myosin IIA and IIB include activated monomers and heteropolymers. *Curr. Biol.* **24**, 1958–1968
38. Chen, Q., Zhang, N., Xie, R., Wang, W., Cai, J., Choi, K. S., David, K. K., Huang, B., Yabuta, N., Nojima, H., Anders, R. A., and Pan, D. (2015) Homeostatic control of Hippo signaling activity revealed by an endogenous activating mutation in YAP. *Genes Dev.* **29**, 1285–1297
39. Zhou, D., Conrad, C., Xia, F., Park, J. S., Payer, B., Yin, Y., Lauwers, G. Y., Thasler, W., Lee, J. T., Avruch, J., and Bardeesy, N. (2009) Mst1 and Mst2 maintain hepatocyte quiescence and suppress hepatocellular carcinoma development through inactivation of the Yap1 oncogene. *Cancer Cell* **16**, 425–438
40. Schlegelmilch, K., Mohseni, M., Kirak, O., Pruszk, J., Rodriguez, J. R., Zhou, D., Kreger, B. T., Vasioukhin, V., Avruch, J., Brummelkamp, T. R., and Camargo, F. D. (2011) Yap1 acts downstream of α -catenin to control epidermal proliferation. *Cell* **144**, 782–795
41. Harvey, K. F., Pflieger, C. M., and Hariharan, I. K. (2003) The *Drosophila* Mst ortholog, hippo, restricts growth and cell proliferation and promotes apoptosis. *Cell* **114**, 457–467
42. Lai, Z. C., Wei, X., Shimizu, T., Ramos, E., Rohrbaugh, M., Nikolaidis, N., Ho, L. L., and Li, Y. (2005) Control of cell proliferation and apoptosis by mob as tumor suppressor, mats. *Cell* **120**, 675–685
43. Huang, J., Wu, S., Barrera, J., Matthews, K., and Pan, D. (2005) The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the *Drosophila* Homolog of YAP. *Cell* **122**, 421–434
44. Taniguchi, K., Wu, L. W., Grivennikov, S. I., de Jong, P. R., Lian, I., Yu, F. X., Wang, K., Ho, S. B., Boland, B. S., Chang, J. T., Sandborn, W. J., Hardiman, G., Raz, E., Maehara, Y., Yoshimura, A., Zucman-Rossi, J., Guan, K. L., and Karin, M. (2015) A gp130-Src-YAP module links inflammation to epithelial regeneration. *Nature* **519**, 57–62
45. Wang, W., Huang, J., and Chen, J. (2011) Angiomotin-like proteins associate with and negatively regulate YAP1. *J. Biol. Chem.* **286**, 4364–4370
46. Engler, A., Bacakova, L., Newman, C., Hategan, A., Griffin, M., and Discher, D. (2004) Substrate compliance versus ligand density in cell on gel responses. *Biophys. J.* **86**, 617–628
47. Cheng, Y., Feng, Y., Jansson, L., Sato, Y., Deguchi, M., Kawamura, K., and Hsueh, A. J. (2015) Actin polymerization-enhancing drugs promote ovarian follicle growth mediated by the Hippo signaling effector YAP. *FASEB J.* **29**, 2423–2430
48. Kardash, E., Bandemer, J., and Raz, E. (2011) Imaging protein activity in live embryos using fluorescence resonance energy transfer biosensors. *Nat. Protoc.* **6**, 1835–1846

An Organic Anion Transporter 1 (OAT1)-centered Metabolic Network^{*[S]}

Received for publication, June 24, 2016, and in revised form, July 18, 2016. Published, JBC Papers in Press, July 20, 2016, DOI 10.1074/jbc.M116.745216

Henry C. Liu[‡], Neema Jamshidi[‡], Yuchen Chen[§], Satish A. Eraly[¶], Sai Yee Cho^{||}, Vibha Bhatnagar^{**}, Wei Wu[¶], Kevin T. Bush^{‡‡}, Ruben Abagyan^{§§}, Bernhard O. Palsson[‡], and Sanjay K. Nigam^{¶¶¶¶¶¶}

From the Departments of [‡]Bioengineering, [¶]Medicine, ^{||}Chemistry and Biochemistry, ^{**}Family and Preventative Medicine, ^{‡‡}Pediatrics, and ^{¶¶}Cellular and Molecular Medicine, [§]Bioinformatics and Systems Biology Graduate Program, ^{§§}Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California 92093

There has been a recent interest in the broader physiological importance of multispecific “drug” transporters of the SLC and ABC transporter families. Here, a novel multi-tiered systems biology approach was used to predict metabolites and signaling molecules potentially affected by the *in vivo* deletion of organic anion transporter 1 (Oat1, Slc22a6, originally NKT), a major kidney-expressed drug transporter. Validation of some predictions in wet-lab assays, together with re-evaluation of existing transport and knock-out metabolomics data, generated an experimentally validated, confidence ranked set of OAT1-interacting endogenous compounds enabling construction of an “OAT1-centered metabolic interaction network.” Pathway and enrichment analysis indicated an important role for OAT1 in metabolism involving: the TCA cycle, tryptophan and other amino acids, fatty acids, prostaglandins, cyclic nucleotides, odorants, polyamines, and vitamins. The partly validated reconstructed network is also consistent with a major role for OAT1 in modulating metabolic and signaling pathways involving uric acid, gut microbiome products, and so-called uremic toxins accumulating in chronic kidney disease. Together, the findings are compatible with the hypothesized role of drug transporters in remote inter-organ and inter-organismal communication: The Remote Sensing and Signaling Hypothesis (Nigam, S. K. (2015) *Nat. Rev. Drug Disc.* 14, 29). The fact that OAT1 can affect many systemic biological pathways suggests that drug-metabolite interactions need to be considered beyond simple competition for the drug transporter itself and may explain aspects of drug-induced metabolic syndrome. Our approach should provide novel mechanistic insights into the role of OAT1 and other drug transporters implicated in metabolic diseases like gout, diabetes, and chronic kidney disease.

A great deal of recent evidence suggests that solute carriers (SLC)² play a much broader role in physiology, including sig-

naling and metabolism, than has been previously appreciated (1, 2). Indeed, there has been a recent call for more systematic analysis of the roles of SLCs in metabolism and signaling (1). Perhaps due to heavy emphasis on their key role in pharmacokinetics, members of the SLC (and ABC as well) “drug” transporter families are not generally depicted in biochemical pathways involving the endogenous metabolites they transport (2). Such an omission could have clinical consequences, because drugs directly or indirectly affect pathways normally involved in the movement of key metabolites, pathway intermediates, and signaling molecules, thereby fundamentally affecting cell and organ physiology in normal, pathophysiological, and developmental situations (1, 2).

According to the remote sensing and signaling hypothesis, SLC and ABC drug transporters are important in regulating the movement of small endogenous molecules such as key metabolites (e.g. α -ketoglutarate, tryptophan metabolites), signaling molecules (e.g. cAMP, prostaglandins, polyamines), vitamins, antioxidants (e.g. uric acid), and certain hormones (e.g. thyroxine) between tissues, organs, and even organisms (2–11). According to this theory, the SLC and ABC transporters form an integrated network allowing remote communication between different tissues via small endogenous molecules. This integrated network functions in a manner similar to the neuroendocrine system and is, in fact, interlinked with it. The ability of SLC and ABC drug and other transporters to regulate or modulate broad aspects of systemic physiology suggests that drug-metabolite interactions might extend well beyond simple competition for transport at the binding site(s) and provide an explanation for aspects of certain drug-induced metabolic syndromes (e.g. those seen with diuretic use or chronic HIV antiviral treatment (12, 13)). Furthermore, elucidation of their physiological role is likely to be useful for further defining the roles of drug transporters in modulating common metabolic diseases, such as diabetes, gout, and chronic kidney disease (2, 4, 14, 15).

Among the aforementioned transporters, organic anion transporter 1 (OAT1/SLC22A6, originally identified as NKT (16, 17)), likely the main basolateral probenecid-sensitive organic anion drug transporter of the kidney, mediates rate-limiting steps in the renal elimination of organic anionic drugs and a few cationic drugs (7, 18–22). This drug transporter has also long been suggested to play a role in key endogenous functions (21, 23–25), whereas analysis of *Oat1* knock-out mice have provided critical information about its potential role in

* This work was supported, in whole or in part, by National Institutes of Health Grants R01-GM098449, R01GM104098, and U54-HD07160. R. Abagyan has a significant financial interest in Molsoft or its products. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

[S] This article contains supplemental Figs. S1–S4 and Tables S1–S4.

¹ To whom correspondence should be addressed: University of California San Diego, 9500 Gilman Dr., MC0693, La Jolla, CA 92093. E-mail: snigam@ucsd.edu.

² The abbreviations used are: SLC, solute carrier; OAT1, organic anion transporter 1; ABC, ATP-binding cassette; FVA, flux variability analysis; GEM, genome-scale metabolic reconstruction; APF, atomic property field; GIMME, gene inactivity moderated by metabolism and expression; CKD, chronic kidney disease; ROC, receiving operating characteristic.

basal physiology (23, 26, 27). For example, targeted and untargeted metabolomics analyses have revealed significant changes in the concentrations of a number of endogenous metabolites in the knock-out animal, including key biochemical pathway intermediates and signaling molecules (23, 27). Moreover, OAT1 (as well as other SLC22 transporters) has also been associated with metabolic abnormalities and disease (2, 3, 7, 11, 25, 26, 28, 29). Taken together, this suggests an important, if underappreciated, role for SLC drug transporters in metabolic processes and signaling.

Taking a cue from our previous systems biology efforts to analyze the physiological role of OAT1 (11), we sought to build a detailed map of metabolic and signaling pathways modulated by “drug” transporters such as OAT1. A systems biology approach involving integration of OAT1 knock-out and wild type gene expression data into a genome-scale metabolic reconstruction (GEM) together with constraint-based modeling (*i.e.* flux variability analysis (FVA)) was used to predict metabolites affected by the absence of OAT1. Pharmacophore-based virtual screening, re-evaluation of existing transport, and knock-out metabolomics data, as well as wet-lab validation were then used to constrain and rank the predicted compounds based on their potential to directly interact with OAT1. Input of this data into the Cytoscape plug-in, MetScape, enabled the generation of a largely experimentally validated, confidence-ranked OAT1-centered metabolic interaction network. Pathway and enrichment analysis supported an important role for OAT1 in several key metabolic and signaling pathways. Moreover, the results indicate the feasibility of this novel hierarchical, integrative approach in the context of generally understanding drug transporter-related metabolism, as well as other biological processes involving SLC and ABC transporters. The results appear consistent with the Remote Sensing and Signaling Hypothesis (2–11).

Results

We have previously used GEMs to identify non-obvious, novel OAT1 substrates that were experimentally validated (3, 11). However, it was also noted that the GEMs failed to detect some of the known OAT1 substrates. Thus we hypothesized that by using constraint-based modeling (to capture systems level interactions) in conjunction with pharmacophore modeling (to capture molecular substrate-receptor (transporter) binding interactions), the strengths of each methodology could be leveraged to overcome their respective weakness. We performed a loosely constrained, context-specific analysis of transcriptomic data of a global metabolic model to catch all possible metabolites potentially interacting with OAT1 (either directly or indirectly) followed by pharmacophore-based virtual screening, along with comparison to *in vivo/in vitro* databases and wet-lab assays to provide the specificity filters to differentiate those metabolites that would directly interact with OAT1 from those likely to indirectly be part of the pathway (but not an OAT1 substrate). Ultimately, this multi-tiered systems level analysis, together with metabolomics (from knock-out and wild type animals) and transport data from OAT1-expressing cells was used to build a detailed confidence-ranked OAT1-centered metabolic interaction network that includes many small mol-

ecules with well established functions in metabolic and signaling pathways (Fig. 1). We arrived at the final network by working through 5 stages:

Stage 1: Prediction of Metabolites Affected by the Deletion of OAT1—The first phase involved a systems biology approach in which transcriptomic data derived from the *Oat1-KO* mouse was integrated into the mouse GEM, iMM1415 (30), to predict metabolites potentially affected by the absence of this transporter (supplemental Fig. S1 and Table S1). Context-specific wild type (WT) and knock-out (KO) models were constructed from kidney gene expression data using the Gene Inactivity Moderated by Metabolism and Expression (GIMME) algorithm (31) with iMM1415. This systems biology approach simulates the metabolic state of a tissue/organ based on transcriptomic data and the execution of specific biological processes by converting the metabolic network into mathematical equations followed by the application of linear programming to calculate a solution of fluxes (*i.e.* measurement of rate of production or depletion of metabolites) for each of the reactions. We confirmed that the generated models could produce many of the small molecules requisite for the normal physiologic function of kidneys.

FVA (32), which calculates the boundary points of the steady state solution space, is one of most common constraint-based approaches used in investigations of the effects of gene deletion on a system (33); however, it requires an appropriate objective function (*e.g.* biomass production, ATP production, substrate uptake) that yields realistic flux distributions (34). However, the current metabolic models do not adequately encapsulate the *in vivo* metabolic alterations likely resulting from the loss of the drug transporter (OAT1) in a complex organ, and we were forced to apply approximations to the data to gain some understanding of the *in vivo* role of OAT1 in metabolism. Thus, whereas several objective functions were considered, including the biomass objective function, the ATP objective function and the renal objective function (35), we ultimately decided to use the biomass objective function because it is: 1) comprised of many metabolites involved in core cellular processes including those involved in regulating/modulating cell maintenance; and 2) applicable to non-proliferative cells, which either produce or consume these metabolites. Furthermore, many of the essential metabolites involved in regulating/modulating cell maintenance (*e.g.* Krebs cycle intermediates, prostaglandins, vitamins, uric acid, and polyamines) are transported by OAT1 and/or other closely related SLC22 family members (2, 36, 37). (Note that, in the simulations and comparative analyses between the wild type and knock-out models, we are not optimizing for growth, thus we are not assuming that biomass is being maximized, but rather only that the biomass components can be produced by the cells, which we know to be the case, biologically.)

To increase the sensitivity of model predictions, we used broad constraints at this initial stage of investigation allowing the inclusion of all possible metabolites that were then ranked in confidence based on the application of pharmacophore screens, as well as available data on their ability to directly interact with OAT1 (from either *in vivo* metabolomics data or pharmacokinetic data). The resulting WT and KO models consisted

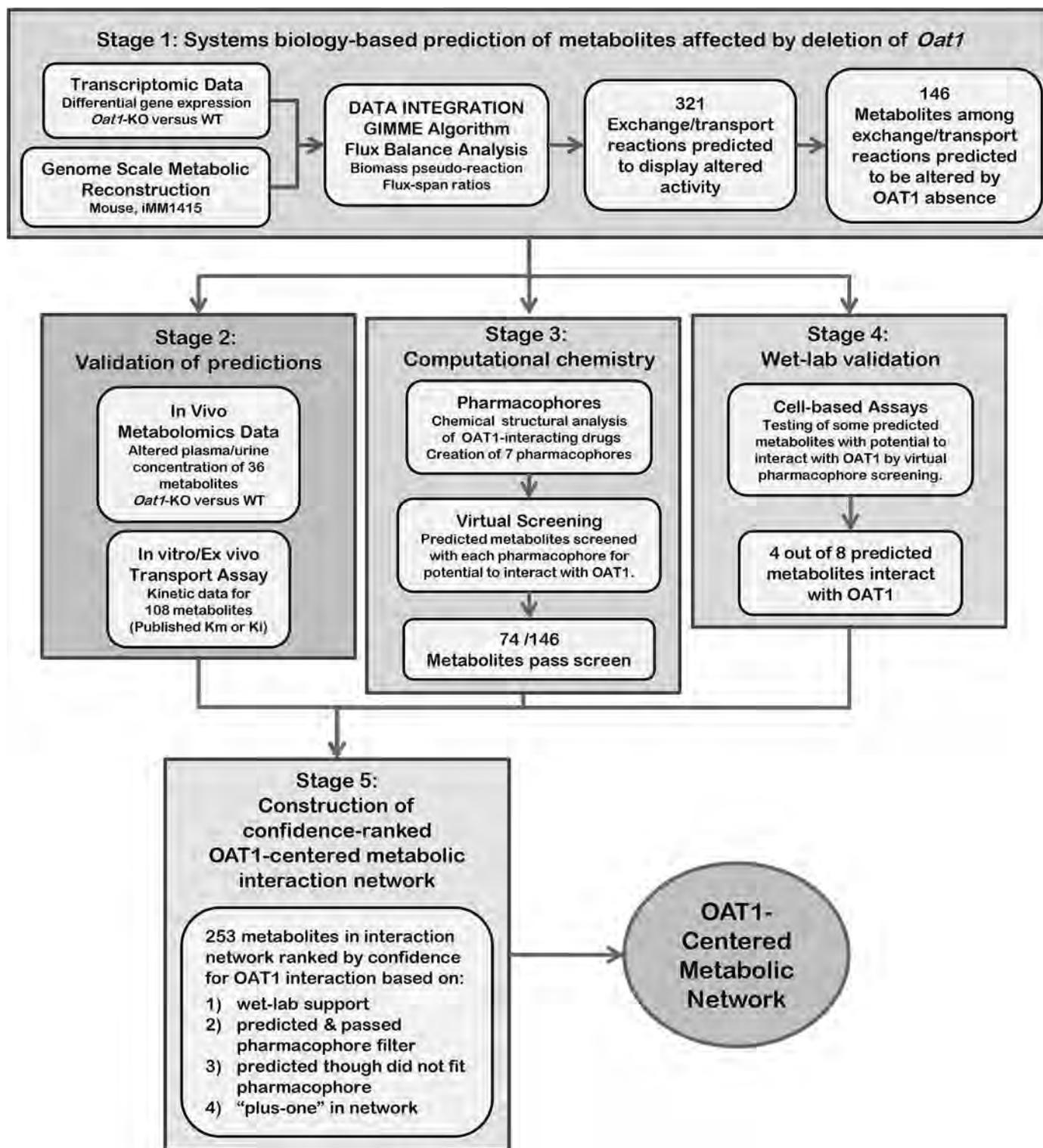


FIGURE 1. **Overall strategy employed.** A diagram of the overall multi-tiered hierarchical approach employed in this study. As described under "Results," the approach is comprised of 5 stages: 1) systems biology analysis of transcriptomic data for the prediction of metabolites altered by the absence of *OAT1* (supplemental Fig. S1 and Table S1); 2) validation of predictions using metabolomics and kinetic data (supplemental Fig. S1); 3) computational chemistry analysis, generation of pharmacophore hypotheses, and the application of pharmacophore filtering (supplemental Fig. S2 and Table S2); 4) wet-lab validation and identification of new *OAT1* metabolites/signaling molecules (Fig. 3); and 5) construction of a substantially validated *OAT1*-centered metabolic network (Fig. 4, supplemental Fig. S3).

of 2233 and 2143 reactions, respectively (supplemental Fig. S1). FVA was used to compare metabolic differences and to find the maximum and minimum flux values for each reaction in the network, enabling the calculation of flux span ratios of the KO

over WT for each reaction. Reactions with flux span ratios equal to 1 indicated no change in reaction activities due to the deletion of *Oat1*; flux span ratios less than 1 implied decreased reaction activities, whereas ratios greater than 1 indicated

increased reaction activities. It was assumed that reactions that require transport of metabolites would change in the OAT1 KO, and 1026 reactions with altered activities were identified, including 321 exchange/transport reactions, which were responsible for the handling of 177 metabolites. After excluding water and other uninformative molecules (see “Experimental Procedures”), 146 metabolites remained (supplemental Table S1), which were predicted to be linked to OAT1-mediated transport.

Stage 2: Validation of the Models Using *in Vivo* Data from the Oat1-KO, as Well as with *in Vitro/ex Vivo* Transport Data—Multiple approaches were used to validate the metabolic reconstruction, as well as to provide a measure of confidence in the ability of the predicted metabolites to interact with OAT1. Initially, metabolomics data from the Oat1-KO were interrogated to determine whether experimental observations corresponded with computational predictions. Previous metabolomics profiles identified 36 metabolites (some of which are also signaling molecules), with significantly altered plasma and/or urine concentrations between the WT and Oat1-KO (Fig. 1, supplemental Fig. S1 and Table S1) (23, 27). Among these 36 metabolites, only 19 were actually included in the iMM1415 GEM (supplemental Table S1), and direct comparison with this metabolomics data revealed that 10 of these 19 metabolites were predicted by the GEM analysis (supplemental Table S1). Applying a hypergeometric test to this sample clearly indicates statistically significant enrichment in such metabolites in our population of 146 predicted metabolites ($p < 0.01$), which suggests that the systems biology approach integrating transcriptomics data together with loosely constrained FVA modeling makes reasonable predictions of the *in vivo* metabolic differences between WT and KO.

In addition, because metabolites that interact with OAT1 would be expected to be affected by the absence of this transporter, a search of the literature was performed and a list of 108 metabolites for which kinetic data exists (*i.e.* K_m and/or K_i) indicating interaction with OAT1 was identified. Among these 108 metabolites, 56 are present in the iMM1415 GEM, of which 21 were found within the 146 GEM-predicted metabolites (supplemental Table S1). Once again, applying a hypergeometric test to this sample reveals statistically significant enrichment in OAT1-interacting metabolites in our population of 146 predicted metabolites ($p < 0.01$). Moreover, combining the metabolomics results with the kinetic data generated a list of 65 non-overlapping metabolites with wet-lab support (either *in vivo* metabolomics or *in vitro/ex vivo* kinetic data) for interaction with OAT1 in the iMM1415 GEM and out of these 65 metabolites 24 were predicted by the systems biology analysis and applying the hypergeometric test indicates significant enrichment in OAT1-interacting metabolites among the 146 GEM-predicted metabolites ($p < 0.01$). Taken together, the significant enrichment of the 146 GEM-predicted metabolites in OAT1-interacting metabolites indicates the utility of the loosely constrained systems biology approach utilized in our study.

Stage 3: Pharmacophore Analysis of Metabolites' Potential to Interact with OAT1—As described above, the broad constraints were applied in Stage 1 to identify all possible reactions and maximize the prediction of potential endogenous metabolites and signaling molecules likely affected by the absence of this

drug transporter. To generate an OAT1-centered interaction network, this broad list of metabolites was then filtered and ranked by their potential to interact with OAT1. QSAR and pharmacophore modeling have been used to analyze limited sets of OAT drugs/substrates (6, 9, 18, 38, 39). Although many drugs appear to be related to metabolites and signaling molecules (40, 41), the availability of chemical libraries and computational tools have led to more systematic comparisons of metabolites, natural compounds, and drugs (42–45). Indeed, pharmacophores based on OAT1 metabolites have previously been used to virtually screen chemical libraries and identify potential inhibitors that have been experimentally validated (27, 46). We thus reasoned that the chemical features of known OAT1-transported drugs might be used to rank the predicted metabolites for their potential to interact with OAT1. In addition, some of these metabolites could then be prioritized for later wet-lab validation to assess direct interaction with OAT1. Therefore, OAT1 pharmacophore models based on a large set of well established drug ligands were built.

To construct pharmacophore models for OAT1, 61 drugs having a published K_m or K_i less than 100 μM for OAT1 were selected as “actives” for model building and model validation; two-thirds of the drugs in this group (41 drugs) were used to build the pharmacophore models (training set), and one-third (20 drugs) of the actives was used as a validating set (supplemental Fig. S2A and Table S2). Because the drugs possess diverse chemical structures (consistent with the known multispecific nature of OAT1), they were first clustered into groups using their atomic property fields (APF) (*e.g.* hydrogen bond donors, hydrogen bond acceptors, SP2 hybridization, lipophilicity, sizes of large atoms, positive and negative charges, etc.) as discriminators (47–48). Thus, the training actives were grouped into 7 distinct clusters (Fig. 2A), and pharmacophore models were then built for each cluster based upon the alignment of its members (Fig. 2, B and C). Fig. 2B demonstrates how members of cluster 1 were first aligned, and “pharmacophore model 1” was built to represent the three-dimensional atomic properties shared among the members of that cluster. Then, the pharmacophore models were validated based on the validating set (known positives) and drugs from Drugbank database (serving as true negatives), and a ROC curve was generated (supplemental Fig. S2B). The calculated area under curve was 80.58, which supports the utility of these pharmacophores to identify OAT1-interacting compounds.

The 7 pharmacophores (Fig. 2C) were then used as three-dimensional chemical space constraints for virtual screening of the predicted molecules that revealed that 74 of the 146 predicted metabolites satisfied the constraints (supplemental Table S1) and were therefore predicted to have direct interaction with OAT1. Of these 74 metabolites, 18 are known to have direct interaction with OAT1 based on previous experimental *in vitro* observations (supplemental Table S1). Thus, compared with the original list of 146 metabolites that had 21 metabolites with kinetic data indicating interaction with OAT1 (prior to pharmacophore filtering), the percentage of metabolites known to have direct interactions was enriched about 2-fold (from 14.4% (21 of 146) to 24.3% (18 of 74) after filtering) and this

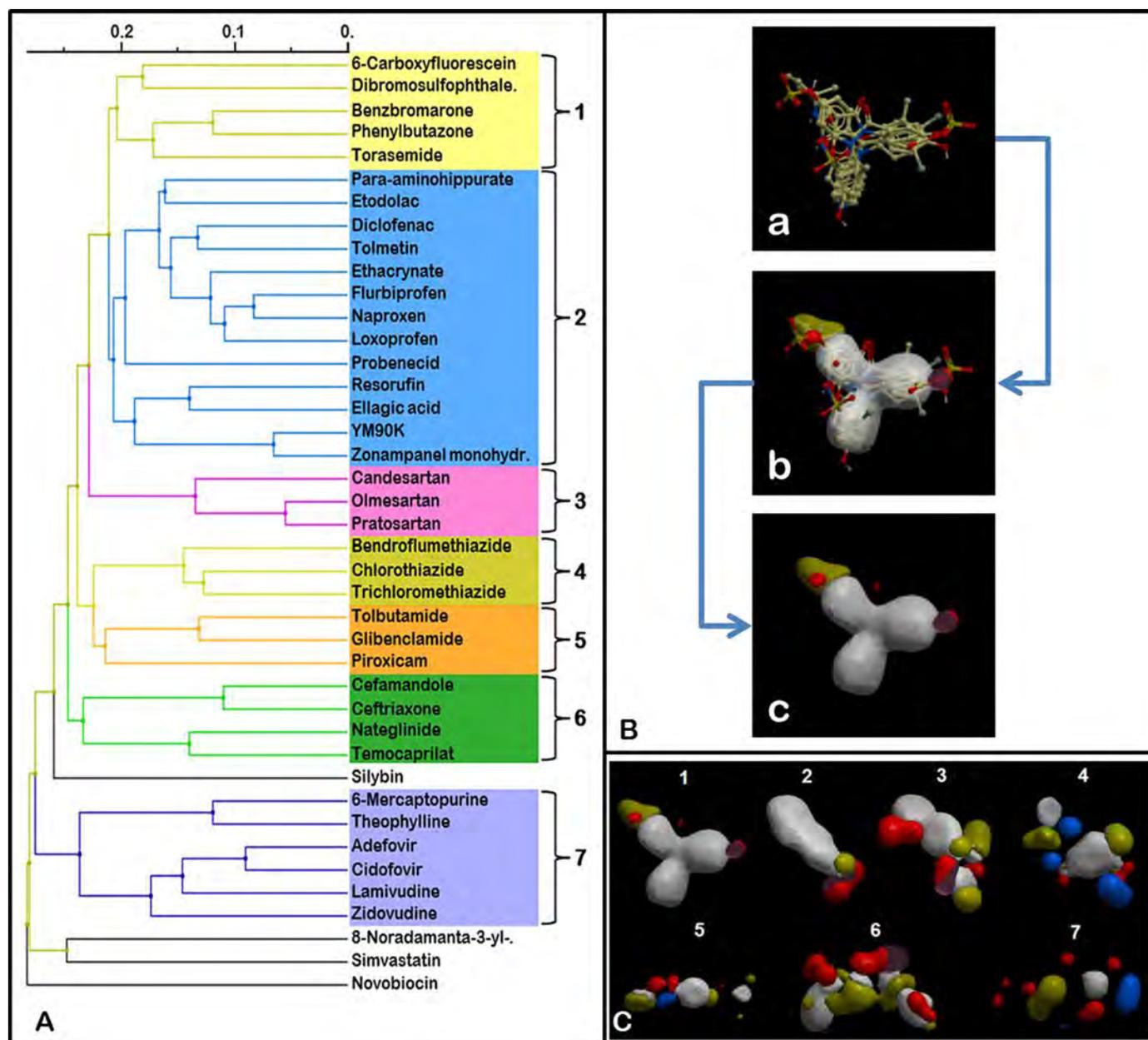


FIGURE 2. **Pharmacophore generation for characterization of metabolites.** A, clustering of 41 OAT1 drugs based on APF (supplemental Table S2 and Fig. S2). Seven clusters of 3 or more drugs were created and each cluster was used for the generation of a single pharmacophore hypothesis. B, the alignment of drugs from Cluster 1 and the creation of the pharmacophore model: (a) the chemical structures of the drugs in the cluster were aligned; (b) chemical determinants are superimposed on the cluster alignment as three-dimensional pharmacophore features; (c) three-dimensional pharmacophore model of the aligned drugs created and used for virtual screening. C, pharmacophore models that were generated from each of the 7 drug clusters; blue, hydrogen bond donor; red, hydrogen bond acceptor; white, aromaticity; yellow, hydrophobicity; light red, negative charges; light blue, positive charges.

enrichment in OAT1-interacting metabolites was found to be statistically significant by the hypergeometric test ($p < 0.01$).

Stage 4: Wet-lab Validation and Identification of Novel OAT1 Ligands—Based on their ability to fit the pharmacophore models, a subset of 8 commercially available metabolites out of the remaining 56 metabolites predicted to directly interact with OAT1 in the pharmacophore screen, but for which there is no wet-lab kinetic data indicating actual interaction with this transporter, were then randomly selected for validation in wet-lab transport assays. Of these 8 metabolites/signaling molecules, four were found to interact with OAT1 in transfected cells (Fig. 3, Supplemental Table S1). These metabolites/signal-

ing molecules were dihydrofolic acid, palmitoleic acid, 16-hydroxy-hexadecanoic acid, and prostaglandin E_1 with calculated K_i values of 93, 200, 13, and 12 μM , respectively (Fig. 3); values that are well within the documented range for many compounds shown to interact with OAT1 (supplemental Table S2). These metabolites are important in whole-body physiology, signaling, and cellular metabolism. For example, prostaglandin E_1 , an endogenous vasodilator, serves to increase peripheral blood flow (48), whereas dihydrofolic acid is required to synthesize both purines and pyrimidines. Palmitoleic acid, a long-chained fatty acid serving as a potential lipokine, is important in the regulation of lipid metabolism (49).

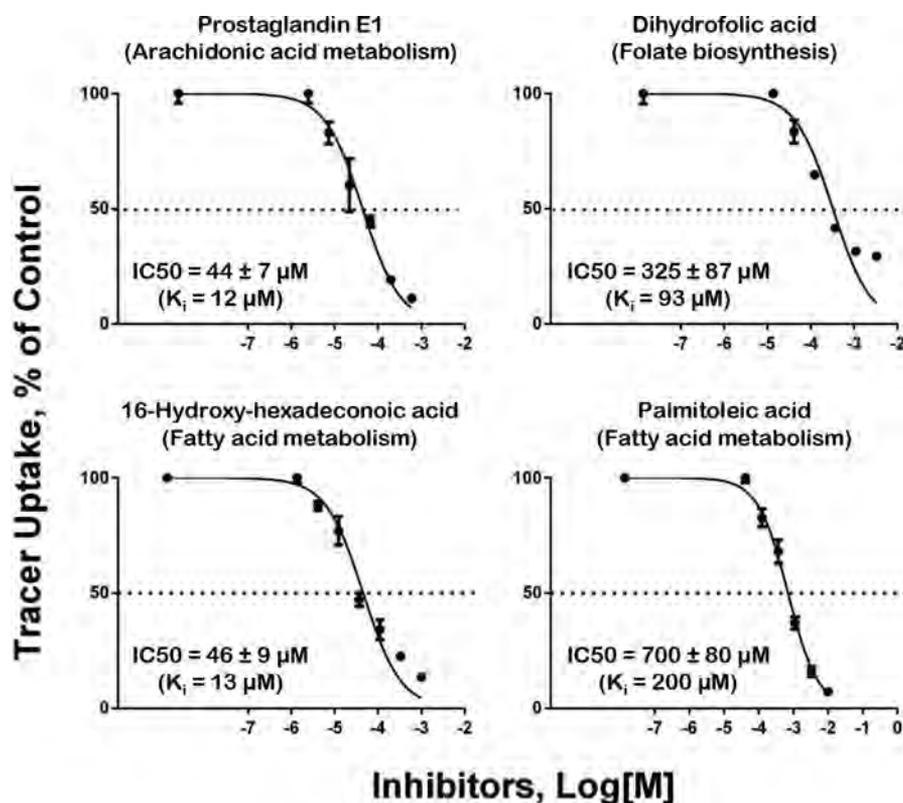


FIGURE 3. **Wet-lab validation of predicted metabolites and the identification of novel OAT1 metabolites.** The IC_{50} curves for the metabolites characterized in the *in vitro* uptake inhibition assay using OAT1-expressing CHO cells. The assay was done by testing for inhibition of uptake of 10 μM 6-carboxyfluorescein, a fluorescent tracer that is also an OAT1 substrate (35, 36) ($K_m = 3.9 \mu\text{M}$).

Taken together with the metabolomics and existing kinetic data described above, the number of wet-lab supported metabolites was increased and the application of a hypergeometric statistical test to the overall method for prediction and identification of novel OAT1 metabolites (the combined *in silico* and *in vitro* approach), indicates significant enrichment in the number of OAT1-interacting metabolites ($p < 0.01$).

Stage 5: Construction of an OAT1-centered Metabolic Interaction Network—The next phase of this study involved construction and analysis of a substantially validated, confidence ranked OAT1-centered metabolite interaction network (Figs. 4, 5, supplemental Fig. S3 and Table S3). To link OAT1 to multiple metabolic pathways, an interaction network was built based on the results of the aforementioned systems biology/pharmacophore approach and the wet-lab data (validated here or published previously), using Metscape, a Cytoscape plug-in used to construct and visualize metabolic networks based on the KEGG database (50). The resulting broader OAT1-centered metabolite network (see “Experimental Procedures”) consisted of a total of 253 metabolites, including 176 experimentally validated and/or computationally predicted and 77 “plus-one” (directly connected) metabolites (Fig. 4). Of these 176 metabolites, 73 had wet-lab support for interactions with OAT1 either by *in vivo* metabolomics from the knock-out or *in vitro* assays that were performed in this study or published. Moreover, $\sim 3\%$ of the 2272 (*i.e.* 78/2272) metabolites comprising the MetScape database at the time of analysis were among those for which OAT1-interaction data were available; however, within the 253 metabolites comprising the OAT1-interaction network more

than 28% (*i.e.* 73/253) have been shown to interact with OAT1. This represented a significant enrichment for OAT1-interacting metabolites in the OAT1-centered interaction network ($p < 0.01$). These metabolites were thus placed in the group termed “wet-lab” support (supplemental Table S3) and had the highest level of confidence for being part of an OAT1-centered metabolic network based on their ability to interact with OAT1.

Three other groups of metabolites were included, in order of level of confidence (supplemental Table S3). The metabolites with the next level of confidence were those first predicted by GEM and which also passed pharmacophore filters; these were termed “metabolites with high confidence of interacting with OAT1” (supplemental Table S3). Metabolites only predicted by GEM (but having structures such that they did not pass the drug-based pharmacophore filters with high confidence) were classified as “metabolites likely to be affected indirectly.” Finally, the remaining plus-one metabolites were termed “OAT1-first neighbor compounds” (Fig. 4). The network revealed that, in the revised OAT1-centered metabolite pathways, the majority of metabolites were interconnected to constitute a main component, and there were also a number of small self-connected components (Fig. 4); network parameters were measured and some are shown in supplemental Fig. S4. The metabolites within the network participated in more than 20 different canonical metabolic pathways, suggesting the broad importance of OAT1 in metabolism. Pathway enrichment analysis of these 253 metabolites was performed using the online bioinformatics resource, Metaboanalyst (51). This over-representation analysis provides statistical information on the

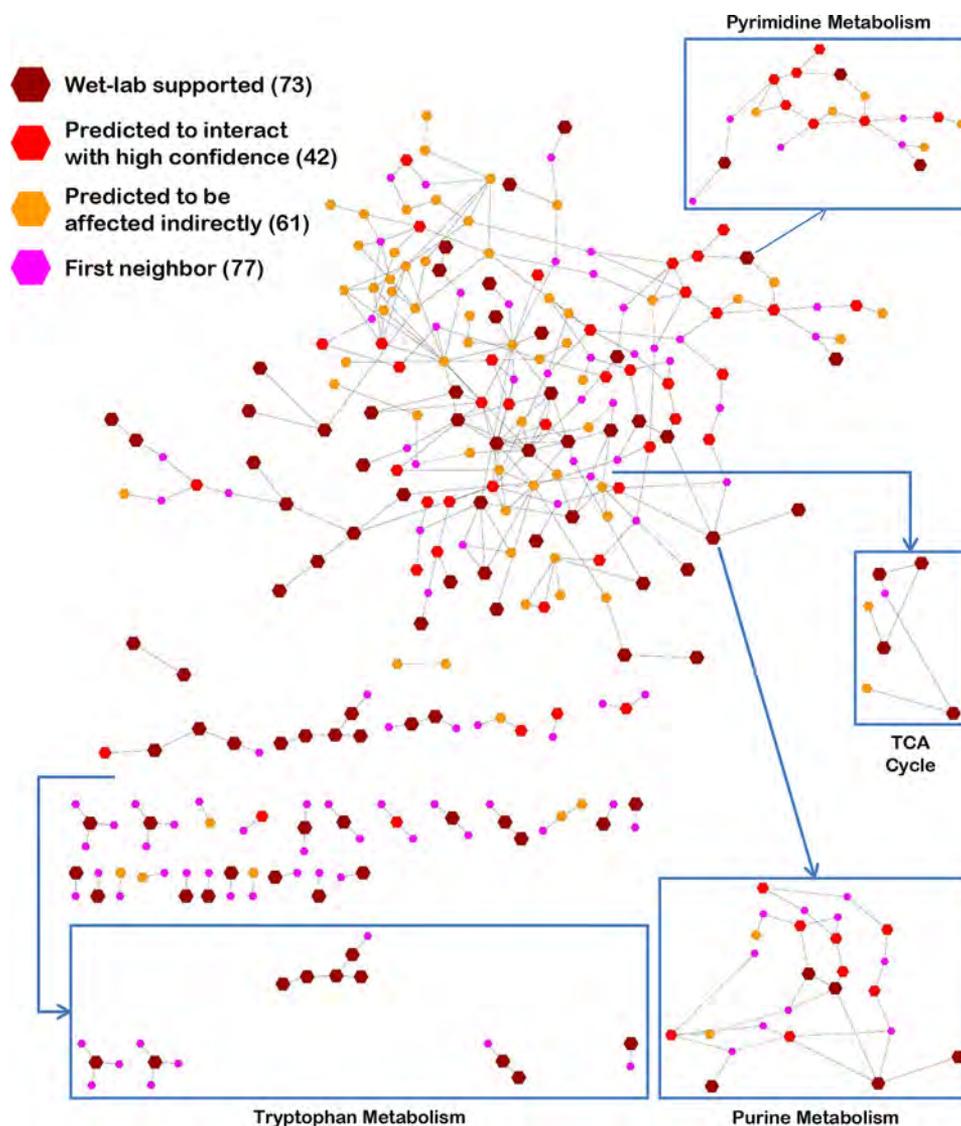


FIGURE 4. **An OAT1-centered metabolic interaction network contains several essential biochemical pathways.** Interaction network consisting of metabolites classified according to the level of confidence for OAT1 interaction: 1) wet-lab supported; 2) predicted to interact with high confidence; 3) predicted to be affected indirectly; and 4) first neighbor (plus-one) (supplemental Table S3). The level of confidence of metabolites is reflected by the size and color of nodes (larger and darker nodes have a higher ranking). In the network, some well represented metabolic pathways are shown. For example, purine metabolism is shown as 25 nodes: 5 wet-lab supported (urate, GMP, hypoxanthine, inosine, 3',5'-cyclic GMP); 7 predicted to interact with high confidence (guanosine, deoxyguanosine, deoxyinosine, dGDP, deoxyadenosine, dADP, L-glutamine); 2 predicted to be affected indirectly (3', 5'-cyclic AMP, L-aspartate); and 10 first neighbors (guanine, dGMP, reduced glutaredoxin, dAMP, adenine, AMP, IMP, xanthosine 5'-phosphate, GMP, 5-phospho-alpha-D-ribose 1-diphosphate).

impact of the metabolites on various pathways (Table 1 and Fig. 5). The affected metabolite/signaling pathways included carbohydrate (e.g. Krebs's cycle, galactose metabolism, etc.), lipid (glycosphingolipid metabolism, bile acid biosynthesis, etc.), amino acid (alanine, aspartate and glutamate, etc.), nucleotide (purine and pyrimidine), and cofactor and vitamin (vitamin A, B2, B3, B5, B6, and B9) metabolism (Table 1; Fig. 5).

The most highly represented metabolic pathways having at least 7 metabolites are shown and ranked according to the validation percentage in Table 1 (validation percentage is equal to the number of "wet-lab supported" metabolites of the pathway divided by the total number of metabolites in the pathway). Among these pathways, the two with the highest "hits with wet-lab support" were tyrosine metabolism and TCA cycle (58.8 and 71.4%, respectively). The TCA cycle is noteworthy, because it includes metabolites known to be classical substrates

of OAT1, such as α -ketoglutarate, citrate, fumarate, and succinate (2, 9, 22, 52). Also included among these top-ranked pathways was tryptophan metabolism, which had considerable wet-lab support for 8 of 15 metabolites (53.3%), including anthranilate, xanthurenic acid, kynurenine, and indole-acetic acid, which are also putative uremic toxins associated with chronic kidney disease (CKD) (27, 53, 54).

Discussion

As previously described, systemic deletion of OAT1 was accomplished using a standard homologous recombination approach and thus resulted in the generalized loss of the transporter (23). OAT1 is largely responsible for the uptake of anionic substrates from the blood, and from the viewpoint of organ physiology, any change in the function of OAT1 (e.g. *Oat1* knock-out) would concomitantly alter the concentration of

TABLE 1

Top pathways affected by OAT1 deficiency as determined by pathway analysis and ranked by the number of hits (%) with wet-lab support

The represented metabolic pathways are ranked according to their percentage of hits with wet-lab support. TCA cycle, tyrosine, and tryptophan metabolism are some of the most well validated pathways represented in the network.

Top pathways	Total	Hits	<i>p</i> Value	FDR	Hits with wet-lab support	Hits with wet-lab support %
TCA cycle	20	7	0.001972	0.011266	5	71.4
Tyrosine metabolism	76	17	0.000728	0.004852	10	58.8
Alanine, aspartate, and glutamate metabolism	24	12	5.34E-07	1.42E-05	7	58.3
Butanoate metabolism	40	12	0.000264	0.002344	7	58.3
Arginine and proline metabolism	77	13	0.032714	0.084423	7	53.8
Tryptophan metabolism	79	15	0.007898	0.033253	8	53.3
Nicotinate and nicotinamide metabolism	44	12	0.000697	0.004852	5	41.7
Valine, leucine, and isoleucine degradation	40	10	0.003935	0.018519	4	40.0
Nitrogen metabolism	39	13	4.24E-05	0.000565	5	38.5
Glyoxylate and dicarboxylate metabolism	50	12	0.002356	0.012563	4	33.3
Propanoate metabolism	35	10	0.001322	0.008134	3	30.0
Glycine, serine, and threonine metabolism	48	13	0.00045	0.003598	3	23.1
Purine metabolism	92	27	5.52E-08	2.21E-06	5	18.5
Pyrimidine metabolism	60	22	1.11E-08	8.85E-07	3	13.6

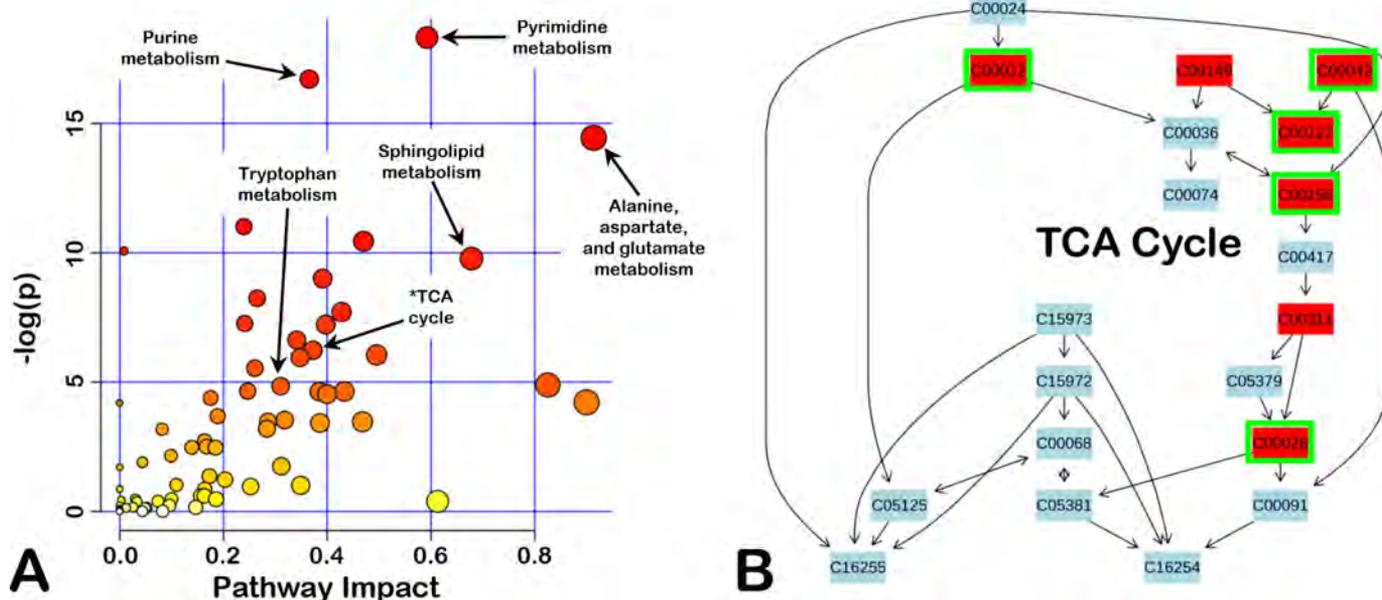


FIGURE 5. Pathway enrichment analysis of the OAT1-centered metabolic interaction network identifies several essential biochemical pathways affected by the absence of OAT1. *A*, graph of pathway enrichment analysis comparing the $-\log(p)$ to the impact on the various pathways for the network metabolites. Some of the affected pathways are indicated with circles colored based on the *p* value (darker red indicates a lower *p* value, whereas yellow indicates less significance) and sized based on the impact on the pathway (larger circles have a greater impact). Pathway impact accounts for both the number of affected nodes and its importance with the maximum importance of each pathway being 1. *B*, the TCA cycle pathway is shown as an example of an affected pathway as it has the highest percentage of hits with wet-lab support (see Table 1). Affected pathway metabolites are highlighted in red and those with wet-lab support for OAT1 interaction have a green border. The number within the rectangle is the KEGG ID for each metabolite, the seven metabolites shown in red are: C00022-pyruvate, C00149-malate, C00042-succinate, C00122-fumarate, C00158-citrate, C00311-isocitrate, C00026-2-oxoglutarate.

metabolites cleared by this transporter not only in the bloodstream, but also in the cells of the tissue in which it is expressed.

In the mouse, the kidney is the predominant site of OAT1 expression and Northern blot and immunohistochemical analyses of kidneys from OAT1 knock-out animals revealed undetectable levels of gene products (*i.e.* RNA and protein) (23). OAT1 is also found, albeit at much lower levels of expression, in some other mouse tissues, including the choroid plexus, inner ear, eye, brain, and spleen. Thus, whereas we cannot completely eliminate the possibility that these other tissues contribute to the observed metabolic alterations seen in OAT1-KO animals, based on the relatively minor contributions expected from these other tissues, it would seem likely that the observed metabolic alterations in the blood are driven mainly by the lack

of kidney-specific expression of the transporter. Moreover, despite the generalized deletion of this important drug transporter, mutant animals are born at expected sex ratios and both male and female mice are viable, appear healthy, and have a normal life expectancy (23).

The OAT1 drug transporter is a focus of regulatory agencies concerned about side effects of drugs due to interaction at the level of the transporter (55, 56). For some time, it has been clear that OAT1 and other drug transporters play key roles in regulating levels of endogenous metabolites and signaling molecules (11, 21, 27, 57). For example, the high but shifting embryonic expression of OAT1 and other SLC22 transporters (OAT3, OCT1, URAT1) in the developing nervous system and other developing tissues led to the hypothesis in 2000 that these drug

OAT1 and Metabolism: A Systems Level Analysis

transporters transported endogenous small molecules that could affect morphogenesis (21). There has recently been a call for more systematic analyses of the roles of these transporters, particularly the SLCs, in metabolism and signaling (1). There is also growing evidence showing relevance of OAT1 to metabolic disease, including chronic kidney disease (15, 25, 58). Taken together with the finding that these highly conserved drug transporters are differentially and highly expressed in various epithelial tissues lining body fluid compartments, it has been hypothesized that they potentially participate in remote communication (“remote sensing and signaling”) between organs and organisms (2–4, 7, 9, 10, 27). This may apply to other SLC families as well (1).

The remote sensing and signaling hypothesis for SLC and ABC drug transporters, which are mainly found on the apical and basolateral surfaces of epithelial cells lining body fluid compartments (e.g. blood, CSF, bile, amniotic fluid), theorizes that, together, these transporters function analogously to the endocrine system (2, 4, 7, 10). As we show here, they play a key role in regulating or modulating biochemical pathways involving a wide variety of small endogenous molecules with high informational content from the perspective of systemic or local metabolism and signaling. Our analysis thus builds upon prior work (7, 11, 23, 27) implicating OAT1 in the *in vivo* regulation of pathways involving essential metabolites (e.g. TCA cycle intermediates, tryptophan metabolites derived from the gut microbiome), key signaling molecules (e.g. prostaglandins, polyamines, cyclic nucleotides), molecules with antioxidant activity (e.g. ascorbic acid, urate), hormones (e.g. thyroxine), vitamins, and cofactors (e.g. pantothenic acid) (Fig. 5, Table 1). Crucially, from our analyses we are able to construct a partly validated OAT1-centered metabolic network (Figs. 4, 5; Table 1).

The molecules in the network are central to classical metabolism and signaling pathways, as well as organ and systemic physiology. Moreover, they are important in pathophysiological states like hyperuricemia, metabolic syndrome, diabetes, and chronic kidney disease. For example, some metabolites known to be transported by OAT1 are potential classical uremic toxins (e.g. indoxyl sulfate, kynurenate, polyamines, and uric acid), which accumulate as renal function declines in CKD (53, 54). Moreover, many of these pathophysiological states have been linked to SNPs or altered expression and/or function of members of the OAT subfamily of SLC22 (2, 4, 7, 10, 14, 15, 25, 58–61).

An important corollary has to do with the role of OATs in drug-metabolite interactions due to competition for transport. This may also be important in chronic kidney disease, where uremic toxins such as those mentioned above are transported by OAT1. For example, polyamines can inhibit transport of the drug methotrexate by OAT1 (11). If the competition is more substantial, as might occur with the drug probenecid (which binds OAT1 with high affinity), our data suggests that the metabolic changes could be quite significant, potentially resulting in many metabolic alterations that substantially overlap with the OAT1 knock-out noted here. In this regard, it is worth noting that there are a number of drug-induced metabolic syndromes that are associated with OAT1-transported drugs (12, 13).

Furthermore, the complexity of the OAT1-centered network points to the possibility of unexpected metabolic changes that could go well beyond the relatively straightforward concept of transporter-level competition for the ligand binding site. For example, in this study, we were able to separate likely direct *versus* indirect interactions of metabolites with OAT1; thus, a drug that tightly binds OAT1 may not only alter metabolites that directly compete for transport but also others in the OAT1-centered network that are not directly transported by OAT1. Because thiazide diuretics and HIV antivirals are transported by OAT1 (6, 62, 63), the OAT1-centered network may help in understanding the drug-induced metabolic syndromes associated with chronic treatment with these drugs (64–67).

In summary, we have described and validated a novel approach for systems driven discovery through integration of distinct and complementary systems biology, computational and wet-lab approaches to gain further insight into the substrate specificity and function of the OAT1 transporter in basal physiology (Fig. 1). As more wet lab transport data accumulates, and with further improvement of pathway analysis and other computational tools—including metabolic reconstruction approaches (particularly with respect to mammalian transporters), it should be possible to further refine and validate the OAT1-centered network proposed here. Furthermore, based on our results, we suggest that the types of multifaceted analyses described here for OAT1, enabling the construction of a “drug transporter”-centered metabolic network (Figs. 4, 5; Table 1), can be applied to other SLC and ABC drug transporters to generate a more comprehensive picture of the role that these transporters play in metabolism. Eventually, this approach could potentially connect cellular metabolism in different organs via molecules transported by multispecific drug transporters (as well as other transport systems such as those involving other types of transporters or channels) as proposed in the remote sensing and signaling hypothesis (2, 4, 7, 9, 10).

Importantly, once drug-transporter metabolic networks are created for other SLC and ABC multispecific transporters, the systems biology approach employed here may be useful for explicitly predicting the metabolic alterations expected for new drugs in healthy or diseased populations with globally altered metabolism (e.g. CKD, liver disease, metabolic syndrome, diabetes, multiorgan injury).

Experimental Procedures

The overall approach taken in the study is depicted in schematic flow charts (Fig. 1), which consists of the following stages: 1) systems biology analysis to predict metabolites affected by the deletion of OAT1 (supplemental Fig. S1); 2) validation of models using *in vivo* and *in vitro* data (Fig. 1); 3) computational chemistry analysis (Fig. 2, supplemental Fig. S2); 4) wet-lab validation (Fig. 3); and 5) construction of an OAT1-centered metabolic interaction network (Fig. 4, supplemental Fig. S3).

Materials—Water-soluble probenecid was purchased from Molecular Probes. The fluorescent tracer (6-carboxyfluorescein), tested metabolites, and signaling molecules (prostaglandin E₁, pristanic acid, elaidic acid, trans-vaccenic acid, dihydrofolic acid, palmitoleic acid, β -nicotinamide mononucleotide,

and 16-hydroxy-hexadecanoic acid) were purchased from Sigma.

Analysis of Transcriptomic Data with Mouse GEM Network—The transcriptomic and metabolomics data from previous studies (23, 27, 63, 68–70) were used for the context-specific analysis. The mouse GEM, iMM1415 (30), which contains the biochemical transformations for numerous tissues and cells in mice, was utilized (see [supplemental Fig. S1](#) for more details). To create context-specific models, gene expression data from wild type (WT) and *Oat1* knock-out (KO) mice were analyzed by Microarray Suite version 5.0 to assign present/absent calls. A minimum of 3 sets of microarray data for each condition (WT and KO) was analyzed separately, and for a gene to be considered present, it had to be present in at least 2 of 3 sets of data. In this way, the gene expression data were converted into a “binary” classification (*i.e.* either absent or present). This binary data without regard to actual expression values was then integrated into the genome-scale metabolic reconstruction model using the GIMME algorithm using the default setting of 90% for the percentage of the objective function needed to be met to generate the model (31).

GEM reconstructions and the application of constraint-based analysis of metabolic networks are widely used in systems biology approaches. The initial planned approach for this study described here was to apply statistical analysis or sampling of the networks with FVA. Unfortunately, the use of strictly constrained FVA models not only resulted in the prediction of a limited number of metabolites with which to work, but they also failed to capture some important interactions known to involve the transporter. Although this is not an uncommon occurrence, it is particularly problematic for this study when one takes into account the fact that our current understanding and the scope of the current metabolic models do not adequately encapsulate the metabolic interactions/alterations resulting from the deletion of SLC drug transporter. In other words, the initial approach was limited by a metabolic model that was not specifically designed to address the questions that were being asked and thus required a modified approach described below.

Sensitivity Filter Through Comparison of the Metabolic Differences between KO and WT with FVA and Model Validation—FVA was used to compare functional differences between the WT and KO mice based on maximal achievable reaction flux ranges (32). A reduced model was created by removing those reactions classified as absent based on gene expression data from the WT and KO conditions. When establishing the lower and upper bound for exchange reactions, uptake constraints for WT and KO models were set to be the same (10 $\mu\text{mol/h}$) for most metabolites, and to make the model more renal-specific, a few metabolites that were listed for the previously published renal objective function, which was used to analyze blood pressure regulation (35), were set to be either secreted or absorbed accordingly. In our analysis, a biomass maintenance function, which was defined to represent the metabolic composition necessary for the maintenance of mouse tissues, was used as the constraint to generate the context-specific models using the GIMME algorithm. We also considered the ATP objective function, as well as the so-called “renal objec-

tive function” (35). As with the biomass objective function, application of these other objective functions to the data requires some degree of approximation. Although the ATP objective function produced similar results to the biomass objective function, the application of the later produced a broader range of results. Interestingly, despite its name, the application of the renal objective function resulted in a more limited set of metabolites for subsequent analysis. This is likely due to the fact that it was designed for the analysis of the role of the kidney in regulating/modulating blood pressure.

The biomass pseudoreaction was ultimately selected to ensure that the cells could produce the complement of small molecules requisite for its function and cellular maintenance as well as metabolites/signaling molecules known to be transported by OAT1 have roles in cell maintenance and growth. Flux spans were then calculated as the differences between the maximum and minimum of reaction fluxes and the pairwise ratio between the KO and WT flux spans were calculated (*i.e.* $[(\text{max flux WT}) - (\text{min flux KO})] \div [(\text{max flux WT}) - (\text{min flux WT})]$), resulting in a set of flux span ratios for the set of reactions shared between the two models. Flux span ratios with a value of 1 were considered to be unchanged; in this loosely constrained model, any variation from 1 was considered to be affected by the absence of OAT1 and a successful model was expected to generate a list of “GEM-predicted OAT1 metabolites” that could be further evaluated ([supplemental Table S4](#)).

Specificity Filter through Pharmacophore Model Building and Validation Based on Drugs Known to Interact with OAT1—Computational chemistry analysis was performed with ICM software developed by Molsoft L.L.C (San Diego, CA). The software was used to perform clustering, alignment, and pharmacophore building based on the APF (47) of OAT1-interacting drugs or tracers (pharmaceuticals) with known K_m (substrate affinity) or K_i (inhibitory affinity) of less than 100 μM were selected. A total of 61 pharmaceuticals were selected; among which two-thirds (41 drugs) were selected for a training set for the model generation, and one-third (20 drugs) were placed in a validating set. The structure-data files ([pubchem.ncbi.nlm.nih.gov](#)) for the OAT1-interacting drugs were input into ICM, which superimposed and aligned the drugs based on using the APF superimposition method. This method takes into consideration a number of three-dimensional structural parameters, hydrogen bond donors, hydrogen bond acceptors, SP2 hybridization, lipophilicity, size of large atoms, and positive and negative charges (71, 72). Based upon these alignments, 7 pharmacophore hypotheses were generated. After the model generation, the validating set (serving as true positives) and all the drugs from the Drugbank database (serving as true negatives) were screened against the pharmacophores. Using these screening results, a ROC curve was generated ([supplemental Fig. S2](#)).

Screening the List of Metabolites Predicted by the Metabolic Networks with Pharmacophore Models—The GEM-predicted OAT1 metabolites were compared with each of the 7 pharmacophore models and ranked by how well they fit with the three-dimensional molecular space defined by the drugs known to interact with OAT1. The 30 metabolites that best fit each of the 7 pharmacophores were selected as having potential to directly

OAT1 and Metabolism: A Systems Level Analysis

interact with OAT1. Many metabolites fit more than one pharmacophore model. After eliminating overlaps, 74 metabolites remained, and these were termed “top-ranked GEM-predicted OAT1 metabolites” (passing OAT1 pharmacophore filter).

Uptake Inhibition Assay—Oat1-transfected CHO cells cultured on 96-well plates were incubated in the presence of 10 μM 6-carboxyfluorescein (6, 9, 62) with or without individual metabolites (with controls treated with the OAT1 inhibitor probenecid). The IC_{50} curves for novel ligands were plotted in Prism Software (GraphPad Inc., San Diego, CA), and the IC_{50} values were converted to K_i (inhibition affinity) using the Cheng-Prusoff equation.

$$K_i = \frac{\text{IC}_{50}}{1 + \frac{[S]}{K_m}} \quad (\text{Eq. 1})$$

Construction of an OAT1-centered Metabolic Interaction Network—The OAT1-centered metabolic interaction network was generated based on the categorization and ranking of metabolites according to the confidence for their interaction with OAT1. Those metabolites for which kinetic data exists indicating the ability to interact with OAT1 were merged together and labeled as “wet-lab supported” metabolites. Metabolites predicted by the GEM analysis, which also fit the pharmacophore filtering were designated as “metabolites predicted to interact with high confidence.” Metabolites that were predicted by the GEM analysis but which did not fit the pharmacophore screen were designated as “metabolites predicted to be affected indirectly.” The KEGG IDs for all of these metabolites were input then into Metscape, a Cytoscape plug-in (50). Metabolites present in Metscape were used as “input metabolites” to build a metabolic network. The construction of the network also introduced many “plus-one” or “first-neighbor” metabolites, and because there is currently no evidence supporting their interactions with OAT1, they were deemed to have the lowest confidence for interacting with OAT1. The network was then trimmed to eliminate uninformative nodes using the following criteria: small molecules (such as water, carbon dioxide, etc), energy-related molecules (NADH, ATP, etc.), and large peptides not known to interact with OAT1 or related transporters (somatostatin, kinetensin, etc.) were removed; unnecessary “dead-ended” and “inter-connecting” plus-ones were removed to create a more concise network (in other words, dead-ended plus-one nodes, which connected only to one node, were removed, as were inter-connecting plus-ones that did not affect connections between wet-lab validated or predicted nodes). The final network thus consisted of metabolites that fell into four categories, which in the order of level of confidence of their potential to interact with OAT1, the categories were: 1) wet-lab supported; 2) predicted to interact with high confidence; 3) predicted to be affected indirectly; and 4) plus-one after trimming.

Metabolic Pathway Analysis—Pathway and enrichment analyses were performed using MetaboAnalyst 3.0 for pathway analysis and visualization (51). Lists of the KEGG IDs for the metabolites were input into this online bioinformatics resource to either the Pathway Analysis or Enrichment Analysis func-

tionality on the MetaboAnalyst 3.0 website. For the pathway analysis, the *Homo sapiens* (human) pathway library was selected and all compounds in the selected pathway were used. The algorithms specified were the hypergeometric test for the over-representation analysis and the relative betweenness centrality for the pathway topology analysis. For enrichment analysis, the pathway-associated metabolite set was selected as the library and all compounds in the metabolite set library were used.

Statistics—To determine whether the overall *in silico* approach results in significant enrichment of metabolites known to have direct interaction with OAT1, a hypergeometric-based test was performed to calculate the various p values. The hypergeometric calculation, which is based on certain assumptions, has been used in systems biological analyses for determining the probability of a result occurring just by chance (73–75). The hypergeometric test can be used as a measure of over-representation and takes into account the overall population size, the number of successes within this population, the sample size, and the number of successes within the sample population.

Author Contributions—H. C. L., N. J., and Y. C. performed the system biology analysis of the transcriptomic data; H. C. L. and S. Y. C. performed the pharmacophore analysis and virtual screening; H. C. L. and S. Y. C. performed the wet-lab transport assays; S. A. E., V. B., W. W., and K. T. B. provided methodological assistance; V. B. and K. T. B. assisted with the statistical analysis; R. A. and B. O. P. contributed analytical tools and provided assistance with their use; S. K. N. conceived the hypothesis and supervised the experiments; H. C. L., K. T. B., W. W., and S. K. N. wrote or contributed to the writing of the manuscript.

References

1. César-Razquin, A., Snijder, B., Frappier-Brinton, T., Isserlin, R., Gyimesi, G., Bai, X., Reithmeier, R. A., Hepworth, D., Hediger, M. A., Edwards, A. M., and Superti-Furga, G. (2015) A call for systematic research on solute carriers. *Cell* **162**, 478–487
2. Nigam, S. K. (2015) What do drug transporters really do? *Nat. Rev. Drug Discov.* **14**, 29–44
3. Wu, W., Jamshidi, N., Eraly, S. A., Liu, H. C., Bush, K. T., Palsson, B. O., and Nigam, S. K. (2013) Multispecific drug transporter Slc22a8 (Oat3) regulates multiple metabolic and signaling pathways. *Drug Metab. Dispos.* **41**, 1825–1834
4. Wu, W., Dnyanmote, A. V., and Nigam, S. K. (2011) Remote communication through solute carriers and ATP binding cassette drug transporter pathways: an update on the remote sensing and signaling hypothesis. *Mol. Pharmacol.* **79**, 795–805
5. Wu, W., Bush, K. T., Liu, H. C., Zhu, C., Abagyan, R., and Nigam, S. K. (2015) Shared ligands between organic anion transporters (OAT1 and OAT6) and odorant receptors. *Drug Metab. Dispos.* **43**, 1855–1863
6. Truong, D. M., Kaler, G., Khandelwal, A., Swaan, P. W., and Nigam, S. K. (2008) Multi-level analysis of organic anion transporters 1, 3, and 6 reveals major differences in structural determinants of antiviral discrimination. *J. Biol. Chem.* **283**, 8654–8663
7. Nigam, S. K., Bush, K. T., Martovetsky, G., Ahn, S. Y., Liu, H. C., Richard, E., Bhatnagar, V., and Wu, W. (2015) The organic anion transporter (OAT) family: a systems biology perspective. *Physiol. Rev.* **95**, 83–123
8. Kaler, G., Truong, D. M., Sweeney, D. E., Logan, D. W., Nagle, M., Wu, W., Eraly, S. A., and Nigam, S. K. (2006) Olfactory mucosa-expressed organic anion transporter, Oat6, manifests high affinity interactions with odorant organic anions. *Biochem. Biophys. Res. Commun.* **351**, 872–876
9. Kaler, G., Truong, D. M., Khandelwal, A., Nagle, M., Eraly, S. A., Swaan, P. W., and Nigam, S. K. (2007) Structural variation governs substrate spec-

- ificity for organic anion transporter (OAT) homologs: potential remote sensing by OAT family members. *J. Biol. Chem.* **282**, 23841–23853
10. Ahn, S. Y., and Nigam, S. K. (2009) Toward a systems level understanding of organic anion and other multispecific drug transporters: a remote sensing and signaling hypothesis. *Mol. Pharmacol.* **76**, 481–490
 11. Ahn, S. Y., Jamshidi, N., Mo, M. L., Wu, W., Eraly, S. A., Dnyanmote, A., Bush, K. T., Gallegos, T. F., Sweet, D. H., Palsson, B. Ø., and Nigam, S. K. (2011) Linkage of organic anion transporter-1 to metabolic pathways through integrated “omics”-driven network and functional analysis. *J. Biol. Chem.* **286**, 31522–31531
 12. Hester, E. K. (2012) HIV medications: an update and review of metabolic complications. *Nutr. Clin. Pract.* **27**, 51–64
 13. Buscemi, S., Nicolucci, A., Lucisano, G., Galvano, F., Grosso, G., Massenti, F. M., Amodio, E., Bonura, A., Sprini, D., and Rini, G. B. (2013) Impact of chronic diuretic treatment on glucose homeostasis. *Diabetol. Metab. Syndr.* **5**, 80
 14. Prentice, K. J., Luu, L., Allister, E. M., Liu, Y., Jun, L. S., Sloop, K. W., Hardy, A. B., Wei, L., Jia, W., Fantus, I. G., Sweet, D. H., Sweeney, G., Retnakaran, R., Dai, F. F., and Wheeler, M. B. (2014) The furan fatty acid metabolite CMPF is elevated in diabetes and induces β cell dysfunction. *Cell Metab.* **19**, 653–666
 15. Sharma, K., Karl, B., Mathew, A. V., Gangoiti, J. A., Wassel, C. L., Saito, R., Pu, M., Sharma, S., You, Y. H., Wang, L., Diamond-Stanic, M., Lindenmeyer, M. T., Forsblom, C., Wu, W., Ix, J. H., *et al.* (2013) Metabolomics reveals signature of mitochondrial dysfunction in diabetic kidney disease. *J. Am. Soc. Nephrol.* **24**, 1901–1912
 16. Lopez-Nieto, C. E., You, G., Barros, E. J. G., Beier, D. R., and Nigam, S. K. (1996) Molecular cloning and characterization of a novel transport protein with very high expression in the kidney. *J. Am. Soc. Nephrol.* **7**, 1301
 17. Lopez-Nieto, C. E., You, G., Bush, K. T., Barros, E. J., Beier, D. R., and Nigam, S. K. (1997) Molecular cloning and characterization of NKT, a gene product related to the organic cation transporter family that is almost exclusively expressed in the kidney. *J. Biol. Chem.* **272**, 6471–6478
 18. Ahn, S. Y., Eraly, S. A., Tsigelny, I., and Nigam, S. K. (2009) Interaction of organic cations with organic anion transporters. *J. Biol. Chem.* **284**, 31422–31430
 19. Burckhardt, G., and Burckhardt, B. C. (2011) *In vitro* and *in vivo* evidence of the importance of organic anion transporters (OATs) in drug therapy. *Handb. Exp. Pharmacol.* **201**, 29–104
 20. Emami Riedmaier, A., Nies, A. T., Schaeffeler, E., and Schwab, M. (2012) Organic anion transporters and their implications in pharmacotherapy. *Pharmacol. Rev.* **64**, 421–449
 21. Pavlova, A., Sakurai, H., Leclercq, B., Beier, D. R., Yu, A. S., and Nigam, S. K. (2000) Developmentally regulated expression of organic ion transporters NKT (OAT1), OCT1, NLT (OAT2), and Roct. *Am. J. Physiol. Renal Physiol.* **278**, F635–643
 22. VanWert, A. L., Gionfriddo, M. R., and Sweet, D. H. (2010) Organic anion transporters: discovery, pharmacology, regulation and roles in pathophysiology. *Biopharm. Drug Dispos.* **31**, 1–71
 23. Eraly, S. A., Vallon, V., Vaughn, D. A., Gangoiti, J. A., Richter, K., Nagle, M., Monte, J. C., Rieg, T., Truong, D. M., Long, J. M., Barshop, B. A., Kaler, G., and Nigam, S. K. (2006) Decreased renal organic anion secretion and plasma accumulation of endogenous organic anions in OAT1 knock-out mice. *J. Biol. Chem.* **281**, 5072–5083
 24. Møller, J. V., and Sheikh, M. I. (1982) Renal organic anion transport system: pharmacological, physiological, and biochemical aspects. *Pharmacol. Rev.* **34**, 315–358
 25. Saito, H. (2010) Pathophysiological regulation of renal SLC22A organic ion transporters in acute kidney injury: pharmacological and toxicological implications. *Pharmacol. Ther.* **125**, 79–91
 26. Eraly, S. A., Vallon, V., Rieg, T., Gangoiti, J. A., Wikoff, W. R., Siuzdak, G., Barshop, B. A., and Nigam, S. K. (2008) Multiple organic anion transporters contribute to net renal excretion of uric acid. *Physiol. Genomics* **33**, 180–192
 27. Wikoff, W. R., Nagle, M. A., Kouznetsova, V. L., Tsigelny, I. F., and Nigam, S. K. (2011) Untargeted metabolomics identifies enterobiome metabolites and putative uremic toxins as substrates of organic anion transporter 1 (OAT1). *J. Proteome Res.* **10**, 2842–2851
 28. Vallon, V., Eraly, S. A., Wikoff, W. R., Rieg, T., Kaler, G., Truong, D. M., Ahn, S. Y., Mahapatra, N. R., Mahata, S. K., Gangoiti, J. A., Wu, W., Barshop, B. A., Siuzdak, G., and Nigam, S. K. (2008) Organic anion transporter 3 contributes to the regulation of blood pressure. *J. Am. Soc. Nephrol.* **19**, 1732–1740
 29. Vallon, V., Eraly, S. A., Rao, S. R., Gerasimova, M., Rose, M., Nagle, M., Anzai, N., Smith, T., Sharma, K., Nigam, S. K., and Rieg, T. (2012) A role for the organic anion transporter OAT3 in renal creatinine secretion in mice. *Am. J. Physiol. Renal Physiol.* **302**, F1293–1299
 30. Sigurdsson, M. I., Jamshidi, N., Steingrimsdottir, E., Thiele, I., and Palsson, B. Ø. (2010) A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Syst Biol* **4**, 140
 31. Becker, S. A., and Palsson, B. O. (2008) Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol.* **4**, e1000082
 32. Mahadevan, R., and Schilling, C. H. (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **5**, 264–276
 33. Raman, K., and Chandra, N. (2009) Flux balance analysis of biological systems: applications and challenges. *Brief Bioinform.* **10**, 435–449
 34. Knies, D., Wittmüß, P., Appel, S., Sawodny, O., Ederer, M., and Feuer, R. (2015) Modeling and simulation of optimal resource management during the diurnal cycle in *Emiliania huxleyi* by genome-scale reconstruction and an extended flux balance analysis approach. *Metabolites* **5**, 659–676
 35. Chang, R. L., Xie, L., Xie, L., Bourne, P. E., and Palsson, B. Ø. (2010) Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput. Biol.* **6**, e1000938
 36. Förster, J., Famili, I., Fu, P., Palsson, B. Ø., and Nielsen, J. (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* **13**, 244–253
 37. Bordbar, A., and Palsson, B. O. (2012) Using the reconstructed genome-scale human metabolic network to study physiology and pathology. *J. Intern. Med.* **271**, 131–141
 38. Duan, P., Li, S., Ai, N., Hu, L., Welsh, W. J., and You, G. (2012) Potent inhibitors of human organic anion transporters 1 and 3 from clinical drug libraries: discovery and molecular characterization. *Mol. Pharm.* **9**, 3340–3346
 39. Wolman, A. T., Gionfriddo, M. R., Heindel, G. A., Mukhija, P., Witkowski, S., Bommareddy, A., and Vanwert, A. L. (2013) Organic anion transporter 3 interacts selectively with lipophilic β -lactam antibiotics. *Drug Metab. Dispos.* **41**, 791–800
 40. Elion, G. B. (1989) Nobel lecture: the purine path to chemotherapy. *Biosci. Rep.* **9**, 509–529
 41. Hitchings, G. H., Jr. (1989) Nobel lecture in physiology or medicine—1988: selective inhibitors of dihydrofolate reductase. *In Vitro Cell Dev. Biol.* **25**, 303–310
 42. Gupta, S., and Aires-de-Sousa, J. (2007) Comparing the chemical spaces of metabolites and available chemicals: models of metabolite-likeness. *Mol. Divers.* **11**, 23–36
 43. Karakoc, E., Sahinalp, S. C., and Cherkasov, A. (2006) Comparative QSAR- and fragments distribution analysis of drugs, druglikes, metabolic substances, and antimicrobial compounds. *J. Chem. Inf. Model* **46**, 2167–2182
 44. Khanna, V., and Ranganathan, S. (2009) Physicochemical property space distribution among human metabolites, drugs and toxins. *BMC Bioinform.* **10**, S10
 45. O'Hagan, S., and Kell, D. B. (2015) Understanding the foundations of the structural similarities between marketed drugs and endogenous human metabolites. *Front Pharmacol* **6**, 105
 46. Kouznetsova, V. L., Tsigelny, I. F., Nagle, M. A., and Nigam, S. K. (2011) Elucidation of common pharmacophores from analysis of targeted metabolites transported by the multispecific drug transporter—Organic anion transporter1 (Oat1). *Bioorg. Med. Chem.* **19**, 3320–3340
 47. Totrov, M., and Abagyan, R. (2008) Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **18**, 178–184
 48. Liu, H. C., Goldenberg, A., Chen, Y., Lun, C., Wu, W., Bush, K. T., Balac, N., Rodriguez, P., Abagyan, R., and Nigam, S. K. (2016) Analysis of molecular properties of drugs interacting with SLC22 transporters, OAT1, OAT3, OCT1, OCT2: A machine-learning approach. *J. Pharmacol. Exp. Ther.* pii: jpet.116.232660

OAT1 and Metabolism: A Systems Level Analysis

49. Hodson, L., and Karpe, F. (2013) Is there something special about palmitoleate? *Curr. Opin. Clin. Nutr. Metab. Care* **16**, 225–231
50. Karnovsky, A., Weymouth, T., Hull, T., Tarcea, V. G., Scardoni, G., Laudanna, C., Sartor, M. A., Stringer, K. A., Jagadish, H. V., Burant, C., Athey, B., and Omenn, G. S. (2012) Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. *Bioinformatics* **28**, 373–380
51. Xia, J., Sinelnikov, I. V., Han, B., and Wishart, D. S. (2015) MetaboAnalyst 3.0: making metabolomics more meaningful. *Nucleic Acids Res.* **43**, W251–257
52. Pritchard, J. B. (1995) Intracellular α -ketoglutarate controls the efficacy of renal organic anion transport. *J. Pharmacol. Exp. Ther.* **274**, 1278–1284
53. Vanholder, R., De Smet, R., Glorieux, G., Argiles, A., Baurmeister, U., Brunet, P., Clark, W., Cohen, G., De Deyn, P. P., Deppisch, R., Descamps-Latscha, B., Henle, T., Jorres, A., Lemke, H. D., *et al.* (2003) Review on uremic toxins: classification, concentration, and interindividual variability. *Kidney Int.* **63**, 1934–1943
54. Duranton, F., Cohen, G., De Smet, R., Rodriguez, M., Jankowski, J., Vanholder, R., Argiles, A., and European Uremic Toxin Work Group (2012) Normal and pathologic concentrations of uremic toxins. *J. Am. Soc. Nephrol.* **23**, 1258–1270
55. Bush, K. T., Nagle, M., Truong, D. M., Bhatnagar, V., Kaler, G., Eraly, S. A., Wu, W., and Nigam, S. K. (2014) Organic anion transporters. in *Drug transporters: Molecular characterization and role in drug disposition* (You, G., and Morris, M. E., eds) 2nd Ed., pp. 25–42, John Wiley & Sons, Inc., Hoboken, NJ
56. Lee, S. C., Zhang, L., and Huang, S. M. (2014) Regulatory science perspectives on transporter studies in drug development. in *Drug Transporters: Molecular Characterization and Role in Drug Disposition* (You, G., and Morris, M. E., eds) 2nd Ed., pp. 473–487, John Wiley & Sons, Inc., Hoboken, NJ
57. Eraly, S. A., Blantz, R. C., Bhatnagar, V., and Nigam, S. K. (2003) Novel aspects of renal organic anion transporters. *Curr. Opin. Nephrol. Hypertens.* **12**, 551–558
58. Nigam, S. K., Wu, W., Bush, K. T., Hoenig, M. P., Blantz, R. C., and Bhatnagar, V. (2015) Handling of drugs, metabolites, and uremic toxins by kidney proximal tubule drug transporters. *Clin. J. Am. Soc. Nephrol.* **10**, 2039–2049
59. Bhatnagar, V., Richard, E. L., Wu, W., Nievergelt, C. M., Lipkowitz, M. S., Jeff, J., Maihofer, A. X., and Nigam, S. K. (2016) Analysis of ABCG2 and other urate transporters in uric acid homeostasis in chronic kidney disease: potential role of remote sensing and signaling. *Clin. Kidney J.* **9**, 444–453
60. Schuetz, J. D., Swaan, P. W., and Tweedie, D. J. (2014) The role of transporters in toxicity and disease. *Drug Metab. Dispos.* **42**, 541–545
61. Zhu, C., Nigam, K. B., Date, R. C., Bush, K. T., Springer, S. A., Saier, M. H., Jr, Wu, W., and Nigam, S. K. (2015) Evolutionary analysis and classification of OATs, OCTs, OCTNs, and other SLC22 transporters: structure-function implications and analysis of sequence motifs. *PLoS ONE* **10**, e0140569
62. Nagle, M. A., Truong, D. M., Dnyanmote, A. V., Ahn, S. Y., Eraly, S. A., Wu, W., and Nigam, S. K. (2011) Analysis of three-dimensional systems for developing and mature kidneys clarifies the role of OAT1 and OAT3 in antiviral handling. *J. Biol. Chem.* **286**, 243–251
63. Nagle, M. A., Wu, W., Eraly, S. A., and Nigam, S. K. (2013) Organic anion transport pathways in antiviral handling in choroid plexus in Oat1 (Slc22a6) and Oat3 (Slc22a8) deficient tissue. *Neurosci. Lett.* **534**, 133–138
64. Sinxadi, P. Z., Dave, J. A., Samuels, D. C., Heckmann, J. M., Maartens, G., Levitt, N. S., Wester, C. W., Haas, D. W., and Hulgand, T. (2013) Mitochondrial genomics and antiretroviral therapy-associated metabolic complications in HIV-infected black South Africans: a pilot study. *AIDS Res. Hum. Retroviruses* **29**, 1031–1039
65. Monier, P. L., and Wilcox, R. (2004) Metabolic complications associated with the use of highly active antiretroviral therapy in HIV-1-infected adults. *Am. J. Med. Sci.* **328**, 48–56
66. Anderson, P. L., Kakuda, T. N., and Lichtenstein, K. A. (2004) The cellular pharmacology of nucleoside- and nucleotide-analogue reverse-transcriptase inhibitors and its relationship to clinical toxicities. *Clin. Infect. Dis.* **38**, 743–753
67. Lewis, W. (2003) Mitochondrial DNA replication, nucleoside reverse-transcriptase inhibitors, and AIDS cardiomyopathy. *Prog. Cardiovasc. Dis.* **45**, 305–318
68. Vallon, V., Rieg, T., Ahn, S. Y., Wu, W., Eraly, S. A., and Nigam, S. K. (2008) Overlapping *in vitro* and *in vivo* specificities of the organic anion transporters OAT1 and OAT3 for loop and thiazide diuretics. *Am. J. Physiol. Renal Physiol.* **294**, F867–873
69. Sweeney, D. E., Vallon, V., Rieg, T., Wu, W., Gallegos, T. F., and Nigam, S. K. (2011) Functional maturation of drug transporters in the developing, neonatal, and postnatal kidney. *Mol. Pharmacol.* **80**, 147–154
70. Torres, A. M., Dnyanmote, A. V., Bush, K. T., Wu, W., and Nigam, S. K. (2011) Deletion of multispecific organic anion transporter Oat1/Slc22a6 protects against mercury-induced kidney injury. *J. Biol. Chem.* **286**, 26391–26395
71. Chen, Y. C., Totrov, M., and Abagyan, R. (2014) Docking to multiple pockets or ligand fields for screening, activity prediction and scaffold hopping. *Future Med. Chem.* **6**, 1741–1755
72. Totrov, M. (2008) Atomic property fields: generalized 3D pharmacophoric potential for automated ligand superposition, pharmacophore elucidation and 3D QSAR. *Chem. Biol. Drug Des.* **71**, 15–27
73. Mitrea, C., Taghavi, Z., Bokanizad, B., Hanoudi, S., Tagett, R., Donato, M., Voichița, C., and Drăghici, S. (2013) Methods and approaches in the topology-based analysis of biological pathways. *Front. Physiol.* **4**, 278
74. Wang, R. S., Oldham, W. M., and Loscalzo, J. (2014) Network-based association of hypoxia-responsive genes with cardiovascular diseases. *New J. Phys.* **16**, 105014
75. Zhang, Y., Fan, H., Xu, J., Xiao, Y., Xu, Y., Li, Y., and Li, X. (2013) Network analysis reveals functional cross-links between disease and inflammation genes. *Sci. Rep.* **3**, 3426

RE1-silencing Transcription Factor (REST) Is Required for Nuclear Reprogramming by Inhibiting Transforming Growth Factor β Signaling Pathway^{*[S]}

Received for publication, January 23, 2016, and in revised form, October 13, 2016. Published, JBC Papers in Press, November 7, 2016, DOI 10.1074/jbc.M116.743849

Qing-Ran Kong¹, Bing-Teng Xie¹, Heng Zhang, Jing-Yu Li, Tian-Qing Huang, Ren-Yue Wei, and Zhong-Hua Liu²

From the Key Laboratory of Animal Cell and Genetics Engineering of Heilongjiang Province, College of Life Science, Northeast Agricultural University, Harbin 150030, China

Edited by Eric R. Fearon

Differentiated cells can be reprogrammed by transcription factors, and these factors that are responsible for successful reprogramming need to be further identified. Here, we show that the neuronal repressor RE1-silencing transcription factor (REST) is rich in porcine oocytes and requires for nuclear transfer (NT)-mediated reprogramming through inhibiting TGF β signaling pathway. REST was dramatically degraded after oocyte activation, but the residual REST was incorporated into the transferred donor nuclei during reprogramming in NT embryos. Inhibition of REST function in oocytes compromised the development of NT embryos but not that of IVF and PA embryos. Bioinformation analysis of putative targets of REST indicated that REST might function on reprogramming in NT embryos by inhibiting TGF β pathway. Further results showed that the developmental failure of REST-inhibited NT embryos could be rescued by treatment of SB431542, an inhibitor of TGF β pathway. Thus, REST is a newly discovered transcription factor that is required for NT-mediated nuclear reprogramming.

Embryonic cells differentiate into all three germ layers of the body as development progresses. Once differentiated, the reversion of the differentiated state to pluripotency is strictly limited in normal development. However, experimentally the differentiated state can be returned to the pluripotent state by transcription factors (1, 2). Despite numerous attempts, the factors responsible for successful nuclear reprogramming still need to elucidate. Transcription factors maintaining the pluripotency of embryonic stem cells (ESCs)³ are called pluripotent

factors, and they have an important role in nuclear reprogramming, such as Oct4, Sox2, and Nanog (3, 4). Thus, we can identify and characterize reprogramming factors by screening the pluripotent factors.

The repressor element 1 (RE1)-silencing transcription factor (REST), as a zinc finger protein, binds 21-bp RE1 sites and functions as a key negative regulator of neurogenesis, so it is also called neuron-restrictive silencer element (5). Recently, REST has been reported to induce gene expression by recruiting TET3 to the DNA for directed 5hmC generation and Nuclear SET domain-containing protein 3-mediated H3K36 trimethylation in neurons (6). Furthermore, REST has different roles in different cellular contexts, such as oncogenic and tumor-suppressor functions and hematopoietic and cardiac differentiation (7, 8). In 2008, REST was proved to maintain self-renewal and pluripotency of mouse ESCs through suppression of microRNAs and believed to be a major pluripotent factor (9, 10). However, it has not been elaborated in nuclear reprogramming. Here, we provide evidence that REST plays a unique role in NT-mediated reprogramming as a suppressor of the TGF β signaling pathway in pig.

Results

Expression Pattern of REST—We first investigated the expression of REST in porcine oocytes, nuclear transfer (NT), and parthenogenetic activation (PA) embryos by real-time PCR and Western blotting analysis. Porcine fetal fibroblasts (PFFs) were used as donor cells to construct NT embryos, and REST was observed in the cells by Western blotting. Large amounts of REST mRNA and protein were stored in oocytes. After activation, REST mRNA was significantly decreased in NT and PA embryos ($p < 0.001$) and maintained at a low level from the four-cell to blastocyst stages (Fig. 1A), and REST protein was also degraded in one- and two-cell NT and PA embryos (Fig. 1B). We performed immunofluorescence analysis to locate REST protein in oocytes and embryos. REST was dispersed in the MII oocyte cytoplasm (Fig. 2, A and A', $n = 17$), and was incorporated into transferred donor nuclei in NT embryos when the nuclei were condensed at 2 h post-NT (Fig. 2, B and B', $n = 15$) and decondensed at 6 h post-NT (Fig. 2, C and C',

* This work was supported by National Natural Science Foundation of China Grant 31470079, National Basic Research Program of China Grant 2011CBA01006, the Fund for Fostering Talents in Basic Science of the National Natural Science Foundation of China Grant J1210069, "Academic Backbone" Project of Northeast Agricultural University Grant 15XG19, and Doctoral Research Foundation of Northeast Agricultural University Grants 2009RC56 and 2012RCB06. The authors declare that they have no conflicts of interest with the contents of this article.

[S] This article contains supplemental Table S1.

¹ Both authors contributed equally to this study.

² To whom correspondence should be addressed. Tel.: 86-045155191729; Fax: 86-045155191747; E-mail: liu086@126.com.

³ The abbreviations used are: ESC, embryonic stem cells; NT, nuclear transfer; RE1, repressor element 1; REST, RE1-silencing transcription factor; PA, parthenogenetic activation; PFF, porcine fetal fibroblasts; IVF, *in vitro* fertilization; ICSI, intracytoplasmic sperm injection; IVM, *in vitro* maturation; LNA,

locked nucleic acid; Q-PCR, quantitative PCR; iPS, induced pluripotent stem; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol.

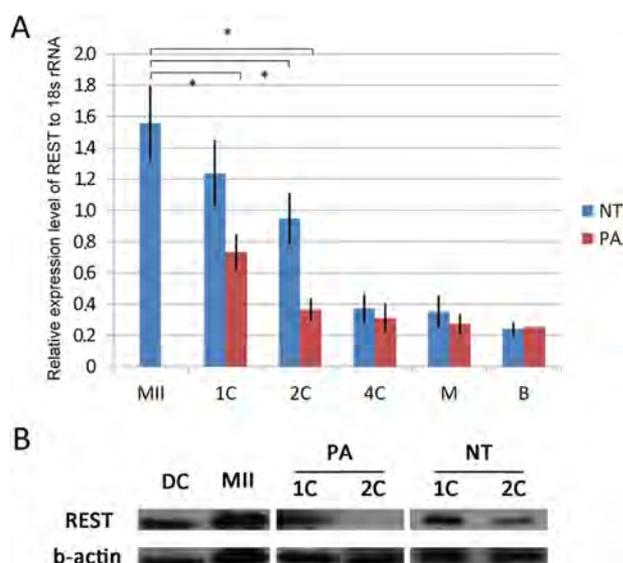


FIGURE 1. Expression pattern of REST in porcine oocytes and embryos. *A*, mRNA expressions of REST in porcine embryos detected by Q-PCR. Asterisk (*) indicates $p < 0.001$; *B*, protein expressions of REST in porcine oocytes, fibroblasts, and embryos checked by Western blotting. *MII*, MII oocytes; *DC*, donor cells; *1C*, one-cell embryos; *2C*, two-cell embryos; *4C*, four-cell embryos; *M*, morula; *B*, blastocyst.

$n = 10$). In donor cells, REST was also incorporated into the nuclei (Fig. 2, *D* and *D'*). To confirm whether maternal REST could be incorporated into the transferred donor nuclei, hREST-GFP mRNA was injected into oocytes at least 2 h before NT. In control, GFP mRNA was injected and the GFP signals were dispersed in the embryos at 2 (Fig. 2, *E* and *E'*, $n = 10$) and 6 h (Fig. 2, *G* and *G'*, $n = 12$) post-NT. But in the hREST-GFP mRNA-injected embryos GFP signals were obviously observed in the transferred donor nuclei at 2 (Fig. 2, *F* and *F'*, $n = 20$) and 6 h (Fig. 2, *H* and *H'*, $n = 16$) post-NT. In *in vitro* fertilization (IVF) and PA embryos, maternal REST was not incorporated into the nuclei when they were condensed (Fig. 2, *I* and *I'*, $n = 12$; *K* and *K'*, $n = 14$) and incorporated with the nuclei when they were decondensed (Fig. 2, *J* and *J'*, $n = 14$; *L* and *L'*, $n = 17$). These results demonstrate that maternal REST is incorporated into transferred donor nuclei, suggesting it may function in the process of NT-mediated nuclear reprogramming.

Inhibition of REST in NT Embryos—To test the role of REST in nuclear reprogramming, its function was inhibited by injection of anti-REST antibody into MII oocytes at least 2 h before NT, PA, IVF, and intracytoplasmic sperm injection (ICSI). The successful injection of the antibody used here was verified by immunofluorescence analysis (Fig. 3, *A* and *A'*, $n = 15$; *B* and *B'*, $n = 15$). In addition, no or a weak signal was detected by immunostaining the anti-REST antibody-injected oocytes at 2 h post-injection (Fig. 3, *C* and *C'*, $n = 16$), indicating that the injected antibody had been degraded and would not affect donor cell-derived and zygotic REST; and, effectively matching the anti-REST antibody to the porcine REST among whole oocyte proteins was verified by Western blotting (Fig. 3*D*). Then the *in vitro* developmental competency of porcine NT embryos, regarded as a stringent test of reprogramming efficiency, was examined. The rates of cleavage and cell numbers of blastocyst

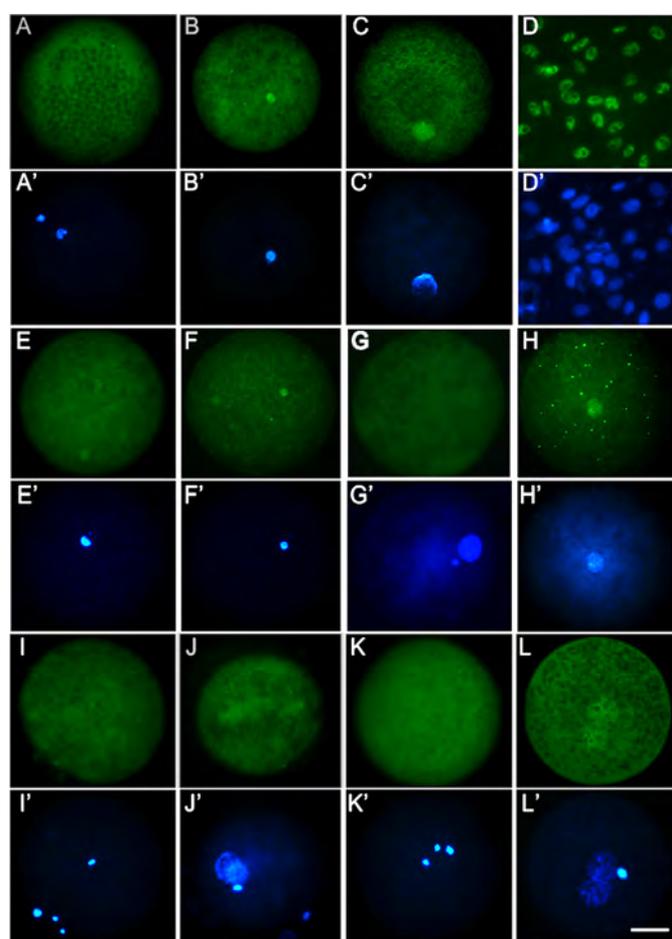


FIGURE 2. Location of REST in porcine embryos. *A* and *A'*, immunofluorescence analysis of oocyte REST; *B* and *B'*, location of REST in NT embryos at 2 h post-NT; *C* and *C'*, location of REST in NT embryos at 6 h post-NT; *D* and *D'*, expression pattern of REST in porcine fibroblasts; *E* and *E'*, detection of GFP expression pattern in NT embryos with GFP mRNA injection at 2 h post-NT; *F* and *F'*, detection of REST location in NT embryos with hREST-GFP mRNA injection at 2 h post-NT; *G* and *G'*, detection of GFP expression pattern in NT embryos with GFP mRNA injection at 6 h post-NT; *H* and *H'*, detection of REST location in NT embryos with hREST-GFP mRNA injection at 6 h post-NT; *I* and *I'*, location of REST in IVF embryos at 2 h post-fertilization; *J* and *J'*, location of REST in IVF embryos at 6 h post-fertilization; *K* and *K'*, location of REST in PA embryos at 2 h post-PA; *L* and *L'*, location of REST in PA embryos at 2 h post-PA. Green, REST or GFP; blue, DNA. Scale bar, 50 μm .

showed no significant difference among NT embryos with no injection (Con-NT), anti-REST antibody injection (anti-REST-NT), and IgG (IgG-NT) injection, but the proportion of anti-REST-NT embryos that developed to blastocysts was significantly lower than that of IgG-NT and Con-NT embryos (7.27 versus 20.8 and 21.53%, respectively; $p < 0.05$; Table 1). More anti-REST-NT embryos were arrested at the two- or four-cell stage in comparison with IgG-NT and Con-NT embryos (54.39 versus 30.33 and 28.09%, respectively; $p < 0.05$; Table 1). To further confirm the results, REST-specific locked nucleic acid (REST-LNA) was injected into oocytes at 33 h of *in vitro* maturation (IVM). Q-PCR and Western blotting analysis showed REST mRNA and protein were effectively reduced in oocytes at 42 h of IVM by REST-LNA injection ($p < 0.001$; Fig. 3, *E* and *F*). Consistent with anti-REST antibody, the proportion of NT embryos developed to the blastocyst stage was significantly decreased in the REST-LNA injection group (3.06 versus

REST Is Required for Nuclear Reprogramming

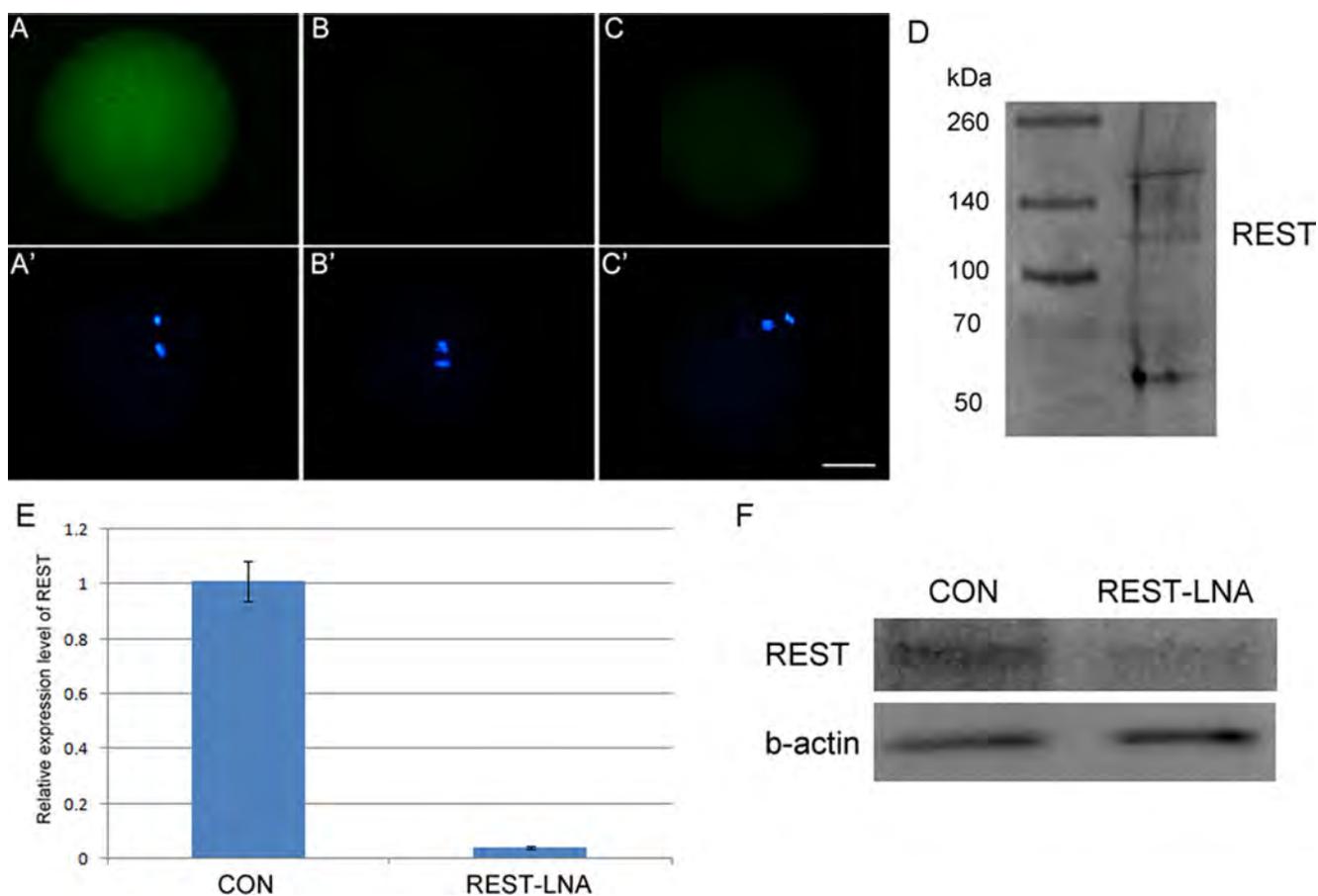


FIGURE 3. Anti-REST antibody injection and efficient knockdown of REST by REST-LNA. *A* and *A'*, immunostaining of anti-REST antibody injected oocytes only by secondary antibody; *B* and *B'*, immunostaining of non-injected oocytes only by secondary antibody. FITC-labeled donkey anti-rabbit IgG was used as secondary antibody, which was used to immunostain anti-REST antibody. *Green*, anti-REST antibody; *blue*, DNA. *Scale bar*, 50 μm . *D*, effectively match of anti-REST antibody to the porcine REST among whole oocyte proteins was verified by Western blotting. *E*, efficient knockdown of REST mRNA by REST-LNA in porcine oocytes checked by Q-PCR. Significant difference was found between the two groups by Student's *t* test ($p < 0.001$). *F*, efficient knockdown of REST protein in porcine oocytes checked by Western blotting analysis. *CON*, MII oocytes; *REST-LNA*, REST-LNA injected MII oocytes.

17.46 and 15.37%; $p < 0.05$; Table 2). We also overexpressed REST by injection of hREST-GFP mRNA into oocytes, and it had no significant effect on development of NT embryos. In contrast to that in NT embryos, injections of anti-REST antibody and hREST-GFP mRNA did not affect embryonic development of IVF, ICSI, and PA embryos (Table 1). Therefore, we suggest that REST is required for successful nuclear reprogramming.

Inhibition of REST Up-regulates TGF β Signaling Pathway in NT Embryos—REST binds RE1 sites to repress gene expression throughout the body (11). To determine how maternal REST regulates NT-mediated nuclear reprogramming, we searched REST-targeted genes in pig. A consensus RE1 based on the sequences of 32 known RE1 elements, NT(T/C)AG(A/C)(A/G)CCNN(A/G)G(A/C)(G/S)AG, was used to screen porcine in the UCSC genome sequence database (susScr3) by using a PERL script (11). The number of putative RE1s identified in the porcine genome was 1,662, and there are 324 genes that have RE1s within 10 kb from their transcriptional start site (supplemental Table S1). Pathway analysis showed that REST putative targeted genes were widely involved in the TGF β signaling pathway ($p < 0.001$, FDR = 0.1642; Fig. 4A). We therefore decided to examine expression of several key genes of the TGF β

pathway in NT embryos. By RT-PCR analysis, the expressions of TGF β R1, ACVR2A, ACVR2B, Smad2, and Smad3 were not detected in oocytes and were high in donor cell PFFs. In one-cell NT embryos, the expressions of these genes were at relatively low levels but dramatically up-regulated after REST deficiency (Fig. 4B). Consistent with RT-PCR results, Q-PCR showed that REST inhibition significantly enhanced the expressions in NT embryos from the one- to four-cell stages (Fig. 4C; $p < 0.001$). Moreover, we observed an increase of Smad3 and phosphorylated Smad3 (Smad3-p) expressions in two-cell REST-deficient NT embryos by Western blotting and immunofluorescence analysis (Fig. 4, *D* and *E*). The results show that inhibition of REST up-regulates the TGF β signaling pathway in NT embryos.

Up-regulation of TGF β Pathway by Inhibition of REST Is Involved in Reprogramming Failure—To determine whether up-regulation of the TGF β pathway in REST-deficient NT embryos leads to the failure of NT-mediated nuclear reprogramming, a specific TGF β pathway inhibitor, SB431542, was used. We found 0.1 μM SB431542 treatment for 12 h post-activation had no negative effect on development of PA embryos (Table 3), so NT embryos were treated as the method. SB431542 treatment could dramatically decrease Smad3 and

TABLE 1**Effect of maternal REST on *in vitro* development of NT, IVF, ICSI, and PA porcine embryos**Note: values with different superscripts within columns denote significant differences ($p < 0.05$).

Groups		Repeats	Embryos	Cleavage	Blastocyst	No. of blastocyst cells	Embryos arrested at two/four-cell stage
				%	%		%
NT	Con	3	122	86 (71.9 ± 6.67)	25 (21.53 ± 3.15) ^a	38.33 ± 8.33	33 (28.07 ± 3.43) ^a
	IgG	3	131	100 (78.33 ± 7.84)	26 (20.8 ± 2.27) ^a	39.43 ± 10.56	40 (30.33 ± 8.33) ^a
	Anti-REST	3	137	104 (75.52 ± 11.11)	9 (7.27 ± 3.50) ^b	32.89 ± 8.50	72 (54.39 ± 4.11) ^b
	Con + DMSO	3	169	136 (81.28 ± 4.73)	30 (18.14 ± 5.72) ^a	35.25 ± 8.38	54 (32.48 ± 5.87) ^a
	Con + SB	3	166	130 (79.31 ± 5.25)	45 (28.93 ± 4.33) ^d	41.26 ± 7.64	40 (25.74 ± 3.27) ^a
	Anti-REST + DMSO	3	164	126 (78.26 ± 6.82)	9 (6.48 ± 7.78) ^b	30.52 ± 10.65	91 (56.84 ± 5.21) ^b
	Anti-REST + SB	3	169	132 (79.10 ± 5.08)	37 (22.94 ± 5.09) ^{a,d}	37.41 ± 9.62	42 (25.25 ± 4.52) ^a
	hREST-GFP mRNA	3	96	74 (78.13 ± 10.95)	23 (23.09 ± 4.13) ^a	35.14 ± 7.50	26 (27.52 ± 3.88) ^a
	GFP mRNA	3	87	64 (75.00 ± 8.86)	16 (19.95 ± 5.86) ^a	32.25 ± 10.32	24 (29.32 ± 3.53) ^a
	Con	3	120	82 (70.08 ± 6.67)	17 (15.7 ± 4.08)	43.00 ± 7.53	54 (45.26 ± 7.92)
IVF	IgG	3	120	88 (73.78 ± 5.71)	17 (15.24 ± 4.60)	39.33 ± 8.52	61 (52.64 ± 11.33)
	Anti-REST	3	120	85 (71.92 ± 8.48)	16 (14.41 ± 6.86)	41.67 ± 10.48	55 (47.46 ± 4.27)
	hREST-GFP mRNA	3	120	78 (66.67 ± 7.86)	16 (14.17 ± 8.17)	37.33 ± 5.75	52 (43.61 ± 6.15)
	GFP mRNA	3	120	77 (65.52 ± 7.24)	15 (12.67 ± 5.70)	44.33 ± 6.67	60 (49.58 ± 8.25)
	Con	3	106	75 (70.65 ± 4.20)	22 (21.55 ± 10.92)	36.41 ± 4.21	27 (26.56 ± 5.25)
ICSI	IgG	3	99	77 (77.78 ± 2.39)	19 (20.37 ± 3.27)	34.37 ± 4.52	24 (25.45 ± 4.24)
	Anti-REST	3	116	91 (78.45 ± 6.88)	25 (21.58 ± 2.66)	33.33 ± 8.33	32 (28.24 ± 4.32)
	hREST-GFP mRNA	3	109	90 (82.57 ± 5.52)	25 (23.35 ± 8.33)	35.48 ± 5.87	29 (27.43 ± 7.92)
	GFP mRNA	3	102	74 (73.69 ± 7.05)	24 (22.67 ± 4.11)	39.39 ± 7.33	30 (30.26 ± 5.52)
	Con	3	120	105 (89.29 ± 6.04)	50 (42.86 ± 4.19)	43.22 ± 5.15	17 (15.32 ± 3.43)
PA	IgG	3	120	110 (93.33 ± 3.43)	55 (46.67 ± 4.33)	43.88 ± 7.67	20 (17.24 ± 4.67)
	Anti-REST	3	120	106 (89.10 ± 7.39)	50 (43.33 ± 5.05)	46.67 ± 7.50	21 (18.61 ± 4.34)
	hREST-GFP mRNA	3	187	173 (93.58 ± 7.46)	91 (49.73 ± 7.33)	49.48 ± 8.86	30 (16.92 ± 2.74)
	GFP mRNA	3	177	150 (85.31 ± 7.34)	77 (44.24 ± 8.81)	44.93 ± 8.30	33 (19.33 ± 3.08)

TABLE 2**Embryonic development after the microinjection of LNA**Note: values with different superscripts within columns denote significant differences ($p < 0.05$).

Groups		Repeats	Embryos	Blastocyst (%)	Embryos arrested at two/four-cell stage (%)
NT	Con	3	101	18 (17.46 ± 4.55) ^a	31 (32.76 ± 5.78) ^a
	REST-LNA	3	98	3 (3.06 ± 0.12) ^b	60 (60.39 ± 5.21) ^b
	Con-LNA	3	102	16 (15.37 ± 4.72) ^a	29 (27.39 ± 4.21) ^a
	Con	4	160	21 (12.24 ± 4.08)	74 (44.25 ± 11.78)
IVF	REST-LNA	4	160	19 (10.38 ± 4.60)	75 (46.47 ± 11.33)
	Con-LNA	4	160	26 (15.82 ± 6.86)	79 (50.35 ± 8.33)
	Con	3	98	24 (24.35 ± 5.52)	20 (22.56 ± 2.61)
ICSI	REST-LNA	3	86	19 (22.86 ± 2.98)	23 (27.32 ± 2.22)
	Con-LNA	3	92	22 (25.32 ± 4.73)	22 (24.28 ± 5.08)

Smad3-p in normal and REST-deficient NT embryos detected by Western blotting and immunofluorescence analysis (Fig. 4, *D* and *E*), indicating the TGF β pathway was efficiently inhibited by SB431542. Then, we asked whether TGF β pathway inhibition in REST-deficient NT embryos can rescue the failure of nuclear reprogramming. As expected, SB431542 treatment successfully rescued the embryonic development to blastocysts (anti-REST + DMSO-NT *versus* Con + DMSO-NT and anti-REST + SB-NT, 6.48 *versus* 18.14 and 22.94%, respectively; $p < 0.05$; Table 1). Furthermore, the development of NT embryos was significantly enhanced by SB431542 treatment (Con + DMSO-NT *versus* Con + SB-NT, 18.14 *versus* 28.93%, respectively; $p < 0.05$; Table 1). These results indicate that the failure of NT-mediated reprogramming in REST-deficient NT embryos can, at least to some extent, be attributed to TGF β pathway up-regulation and TGF β pathway may block nuclear reprogramming. In induced pluripotent stem (iPS) cells technology, inhibition of TGF β pathway promotes reprogramming through inducing Nanog (12). Here, we also found that the expressions of Nanog in one- and two-cell NT embryos were remarkably enhanced after TGF β pathway inhibition checked

by Q-PCR and Western blotting analysis ($p < 0.001$; Fig. 5, *A* and *B*). Taken together, our results indicate the REST repressing TGF β pathway regulates NT-mediated reprogramming (Fig. 5*C*).

Discussion

Differentiated cell nuclei can be reprogrammed to a pluripotent state by NT into oocytes, iPS technology, and cell fusion with ESCs (2, 3, 13, 14). NT mediated-reprogramming has been proven to be the most efficient way (15), and complex transcription factors are significant in the process. Therefore, identification and characterization of these factors will provide us important information on nuclear reprogramming. In the study, we found that REST was rich in oocytes and required for NT-mediated reprogramming through inhibiting TGF β pathway in pig.

REST is a zinc finger protein, and represses neuronal gene transcription in nonneuronal cells (16–18). Mice that lack REST exhibit malformations in the developing nervous system and die by embryonic day 11.5 of embryogenesis. However, these mice appear normal until embryonic day 9.5 (16). This is

REST Is Required for Nuclear Reprogramming

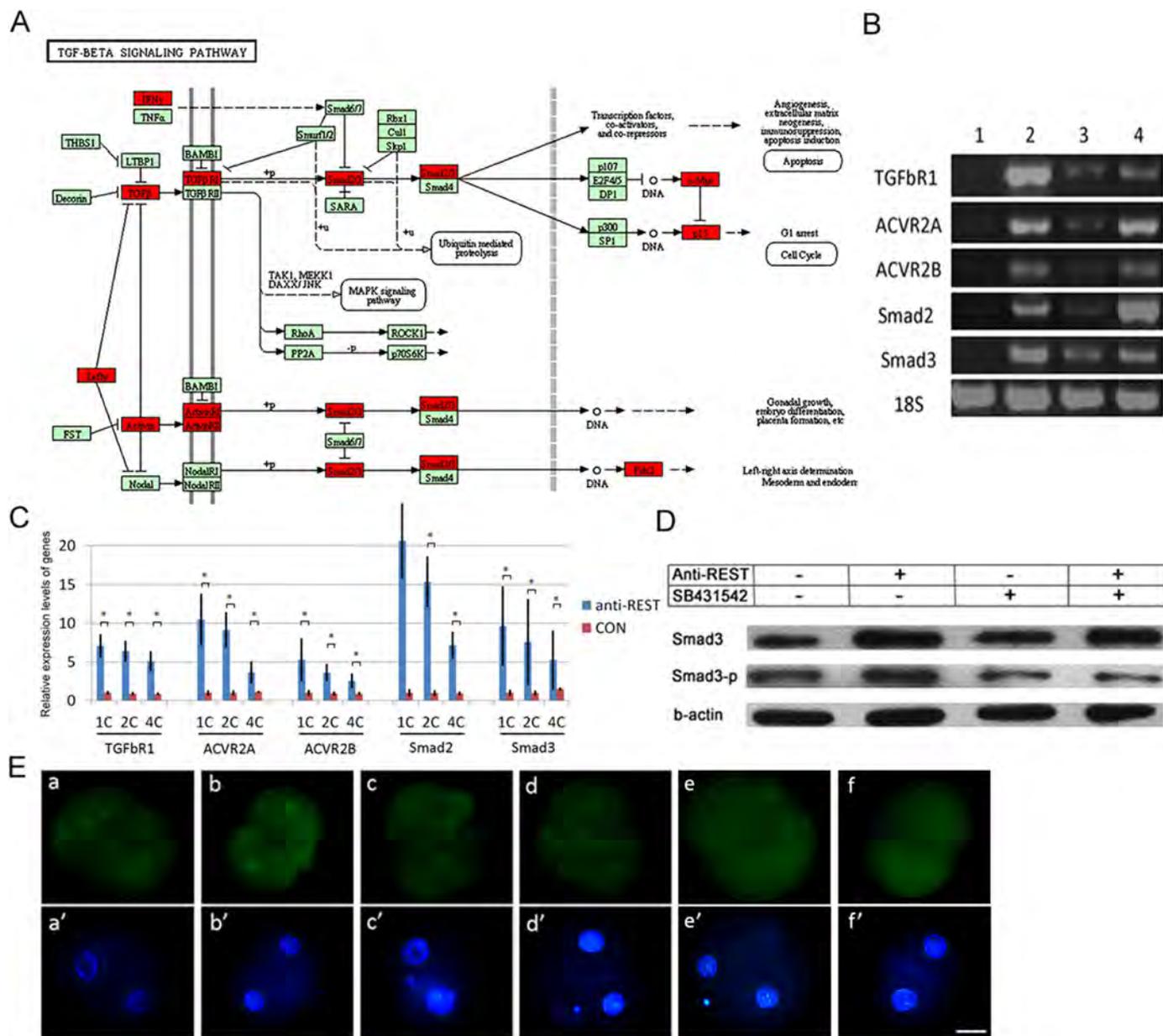


FIGURE 4. Maternal REST suppresses the TGFβ signaling pathway in porcine NT embryos. *A*, putative REST-targeted genes in the TGFβ signaling pathway. Putative REST targeted genes were marked by red. *B*, expression of REST targeted genes in TGFβ pathway checked by RT-PCR. 1, MII oocytes; 2, donor cells; 3, one-cell NT embryos; 4, anti-REST one-cell NT embryos. *C*, expression of REST targeted genes in the TGFβ pathway checked by Q-PCR. CON, NT embryos; anti-REST, anti-REST NT embryo. Asterisk (*) indicates $p < 0.001$; *D*, Western blotting analysis of Smad3 and Smad3-P in NT embryos; *E*, immunofluorescence analysis of Smad3-P in porcine two-cell NT embryos. *a* and *a'*, NT embryo ($n = 14$); *b* and *b'*, anti-REST NT embryo ($n = 15$); *c* and *c'*, NT embryo treated by SB431542 ($n = 14$); *d* and *d'*, anti-REST NT embryo treated by SB431542 ($n = 17$); *e* and *e'*, IVF embryos ($n = 8$); *f* and *f'*, PA embryos ($n = 12$). Green, Smad3-P; blue, DNA. Scale bar, 50 μm.

TABLE 3

Effect of SB431542 with different concentrations on *in vitro* development of porcine PA embryos

Note: values with different superscripts within columns denote significant differences ($p < 0.05$).

Groups	Repeats	Embryos	Cleavage	Blastocyst
Con.	3	121	98 (79.91 ± 6.36) ^a	32 (27.64 ± 7.09) ^a
1 μM	3	119	69 (56.38 ± 3.19) ^b	16 (14.53 ± 7.81) ^b
0.5 μM	3	124	84 (66.94 ± 5.48) ^c	22 (18.42 ± 4.53) ^b
0.1 μM	3	131	96 (74.48 ± 5.42) ^{a,c}	34 (26.41 ± 4.19) ^a

consistent with our observations that REST deficiency has no obvious effect on early embryonic development of IVF and PA embryos. REST was found to incorporate into the

condensed and decondensed transferred donor nuclei in one-cell NT embryos; and REST inhibition by anti-REST antibody and REST-LNA injections remarkably decreased the *in vitro* developmental competency of NT embryos. So, we believe that REST is required for NT-mediated reprogramming. REST can bind RE1 sites to repress gene expression (5). In porcine genome, 1,662 RE1s and 324 corresponding genes were identified. These numbers are comparable with human and mouse (11). Pathway analysis showed that TGFβ signaling pathway could be suppressed by REST. Confirming that, up-regulation of the TGFβ pathway was observed in REST-deficient NT embryos, revealing TGFβ pathway is suppressed by REST in NT embryos.

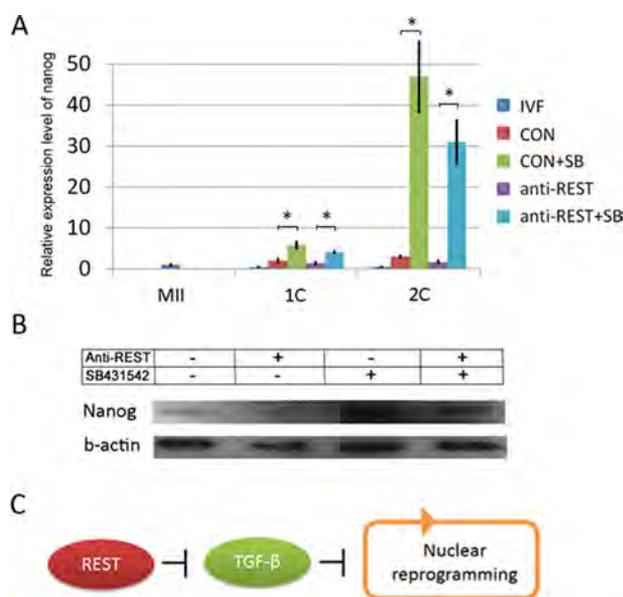


FIGURE 5. TGF β signaling pathway represses expression of Nanog in porcine NT embryos. *A*, mRNA expression pattern of Nanog in porcine embryos checked by Q-PCR. *MI*, MII oocytes; *IVF*, IVF embryos; *CON*, NT embryos; *CON + SB*, NT embryo treated by SB431542; *anti-REST*, anti-REST NT embryos; *anti-REST + SB*, anti-REST NT embryo treated by SB431542. Asterisk (*) indicates $p < 0.001$; *B*, Western blotting analysis of Nanog in porcine NT embryos at the two-cell stage. *C*, proposed model of maternal REST on nuclear reprogramming. REST promotes nuclear reprogramming by suppressing TGF β signaling pathway.

TGF β pathway has been implicated in the development and maintenance of various organs (19, 20), and is necessary for the maintenance of self-renewal and pluripotency of both human and mouse ESCs (21). During embryonic development, the pathway is believed to play critical roles in the specification of cell identities in embryonic and extra embryonic lineages of the post-implantation embryo (19, 22–24). Before implantation, embryonic phenotypes for loss-of-function mutation of the pathway are not detected (20). Transcriptome sequencing and analyzing pig embryos *in vivo* and *in vitro* also show that the TGF β pathway is not active well before maternal zygotic transition at the four- to eight-cell stages (25). Those data suggest the function of the TGF β pathway is suppressed during early embryonic development. In the study, activation of the TGF β pathway at a certain level was detected in NT embryos, but not in IVF and PA embryos, and the activation in NT embryos could be attributed to donor cell PFFs in which the TGF β pathway activated. Here, we proposed that REST was required to silence the TGF β pathway in NT embryos, and in IVF and PA embryos, the pathway was unactivated and, in the regard, REST was not needed. So, the function of REST is only necessary for NT embryos, not for IVF and PA embryos. In addition, inhibition of REST up-regulated the TGF β signaling pathway in NT embryos. So, we believe inhibition of the TGF β pathway by REST may be involved in successful nuclear reprogramming.

To test the point, SB431542 was used to treat NT embryos. SB431542 treatment could successfully rescue the developmental failure of REST-deficient NT embryos and improve developmental potential of normal NT embryos. The results reveal that

the TGF β pathway may have a negative effect on NT-mediated reprogramming. It has been demonstrated that TGF β pathway inhibition can replace Sox2 and promote the completion of iPS reprogramming through induction of the reprogramming factor Nanog (12). Coincidentally, high level Nanog expression was observed in the TGF β pathway-inhibited NT embryos. Previous reports have been shown that inhibition of the TGF β pathway by SB431542 increases Bmp signaling (26) and Bmp signaling induces Nanog expression (27). The cross-talk between TGF β and Bmp signaling may result in Nanog induction. We conclude that inhibition of the TGF β pathway improves NT-mediated reprogramming perhaps by up-regulation of Nanog.

So far, many studies have focused on identification of reprogramming factors (28–35). In the study, we demonstrate that REST acts as a repressor of the TGF β pathway and is critical for NT-mediated nuclear reprogramming, and inhibition of the TGF β pathway by SB431542 treatment promotes the reprogramming efficiency in pig. In addition to better understanding the detailed mechanism of how TGF β pathway inhibition contributes to increased reprogramming efficiency, whether or not our observation can be generally applied to other animal species warrants future investigation. The simplicity of SB431542 treatment during NT makes the testing of our approach worthwhile. If so, SB431542 treatment has the potential to enhance cloning efficiency in a broad range of mammalian species, including humans. Our method could hold great promise for human therapeutic cloning (36).

Experimental Procedures

hREST-GFP mRNA in Vitro Transcription and Plasmid Construction—pEGFP-C1 and hREST-GFP (RG211570, OriGene) plasmids were linearized before *in vitro* transcription. RNA synthesis and poly(A) tailing were carried out with a MEGA script T7 Kit (Ambion, Carlsbad, CA) according to the manufacturer’s instructions.

Oocyte and Embryo Manipulations—Before NT, IVF, ICSI, and PA, 10 picoliters of 1 mg/ml of anti-REST antibody (ab21635, Abcam), 100 ng/ μ l of GFP and hREST-GFP mRNAs solution were injected into matured oocytes. After injection, oocytes were kept for at least 2 h before manipulations, which allows the antibody to bind endogenous REST. Moreover, 10 picoliters of 10 μ M REST-LNA (Locked Nucleic Acid, Exiqon) was injected into porcine oocytes with the first polar body collected at 33 h of IVM (35), and the oocytes matured at 42 h were used for NT, ICSI, and IVF. The procedure for porcine NT, PA, and IVF has been described previously (35). After fusion, 0.1 μ M SB431542 was used to treat NT embryos for 12 h. Cumulus cell-free oocytes were directly activated by the same parameters as for the somatic cell nuclear transfer procedure to produce PA embryos.

ICSI was performed by using an inverted microscope (Olympus IX71, Olympus Optical Co. Ltd.) with a piezoactuated micromanipulator (PMAS-CT150; Prime Tech Ltd, Tsuchiura, Japan). A 100-ml drop of HEPES-M199 containing 0.5% (v/v) FBS and a 20-ml drop of 4% (w/v) polyvinylpyrrolidone (M_r 360,000; Sigma) were placed in a 35-mm dish and covered with

REST Is Required for Nuclear Reprogramming

TABLE 4
The primer list

Gene	Primer sequence (5'-3')	Length	Accession number
18S rRNA	F: TCCAAATGGATCCTCGCGGAA R: GGCTACCACATCCAAGGAAG	149	NR002170
REST	F: GAGGCGGAGTCTGAGGAGCAG R: GTGGTCGTAGCGGTTGGTGTTG	192	GU991112.1
ACVR2A	F: AGGTGCTATACTTGGTAGATCAGAAACTC R: CAGCCAACAACCTTGCTTCACTA	187	NM_001204765.1
ACVR2B	F: CACGGGAGTGCATCTACTACAACGCC R: CGTCTAGCCAGCAGCCCTTCTTCCAC	165	NM_001005350.1
ACVR1B	F: GAGTTATGAGGCGCTGCGGGTG R: GCTGAGCTGGGACAGGGTCTTCTTG	112	NM_001195322.1
TGFBR1	F: AGTAAGACATGATTGAGCCACAGATACAA R: AGCTATTTCAGAACTAAGCCCAT	172	NM_001038639.1
Smad2	F: GCTGCTCTTCTGGCTCAGTCCG R: TACTGTCTGCCTTCGGTATCTGCTC	123	NM_001256148.1
Smad3	F: CAGCGACCACAGATGAACCACAG R: CTCGTAGTAGGAGATGGAGCACCAGAAG	145	NM_214137.1
Nanog	F: CCTCCATGGATCTGCTTATTC R: CATCTGCTGGAGGCTGAGGT	118	AY596464

mineral oil. Next, 20–30 oocytes were placed in the 100-ml drop and the sperm suspension was transferred to the polyvinylpyrrolidone drop. The oocyte was positioned with a holding pipette so that the first polar body was at the 6 or 12 o'clock positions. A single sperm was injected into the cytoplasm with a micropipette. Activation of ICSI zygotes was induced with 2DC pulses of 1.2 kV/cm for 30 ms on a BTX Elector-Cell Manipulator 2001 (BTX, San Diego, CA).

The embryos were cultured in porcine zygote medium-3 at 39 °C in 5% CO₂ in air. The cleavage and blastocyst rates were assessed at 48 and 156 h after activation, and the number of blastocyst cells was examined by nuclear staining with 5 μg/ml of Hoechst 33342.

RE1 Database Construction—A search was performed of the porcine genome GenBank formatted DNA sequence flat files (downloaded from UCSC susScr3 version) by using a PERL script constrained by a core 17 nucleotide regular expression pattern. This pattern represents an RE1 consensus sequence, derived from alignment of 32 known RE1 sequences containing degeneracies reflecting known variations (12). The search output and corresponding annotations or external references (SWISS-PROT Version 40.43 and TREMBL Version 22.13 protein sequence databases) were used to assign gene description and determination of annotated genes within 100 kb on either strand.

Q-PCR Analysis—Total RNA was extracted using the PureLink™ Micro-to-Midi System (Invitrogen) according to the manufacturer's instructions, and reverse transcription was used to generate cDNAs using the Prime Script™ RT Reagent kit (TaKaRa). Real-time PCR was performed using SYBR Premix Ex Taq™ (TaKaRa) and the 7500 Real-time PCR System (Applied Biosystems). The reaction parameters were 95 °C for 30 s followed by 40 two-step cycles of 95 °C for 5 s and 60 °C for 34 s. All the primer pairs used for PCR amplification are shown in Table 4. *C_t* values were calculated using Sequence Detection System software (Applied Biosystems), and the amount of target sequence normalized to the reference sequence was calculated as $2^{-\Delta\Delta C_t}$.

Western Blotting—The procedure for Western blotting analysis has been described previously (11). Oocytes or embryos

without zona pellucida were transferred to cold 40 mM sodium phosphate, pH 7.6, containing 50 mM NaCl, 50 μM sodium orthovanadate, 10 mM sodium fluoride, 20 μM MG132, 2 μM matrix metalloprotease inhibitor III (444264, Calbiochem), and 1% protease inhibitor mixture III (539134, Calbiochem). Homogenization was carried out with a Tekmar homogenizer by three 15-s bursts with a minute cooling between. Homogenates were centrifuged for 1 h at 100,000 × *g*. The supernatant solutions are referred to as “soluble” fractions. The pellets were suspended in 0.2 ml of complete buffer containing 1% ASB-14 and were mixed every 15 min for 2 h with Radnoti glass pestles (Unitek, Monrovia, CA). After centrifugation for 1 h at 100,000 × *g*, the supernatants, referred to as “membrane extracts” were removed, and the pellets were discarded. About 50 embryos of each soluble and membrane extract for each gene were separated by lithium dodecyl sulfate-polyacrylamide gel electrophoresis on 4–12% BisTris NuPAGE gels (the gels have been run under the same experimental conditions) and transferred to PVDF membranes (Invitrogen); nonspecific binding was blocked by overnight incubation in 1% casein in PBS at room temperature. Antibodies against REST (ab21635, Abcam), Smad3 (ab40854, Abcam), Smad3-p (ab118825, Abcam), and Nanog (500-P236, Peprotech) were used, and β-actin (A1978, Sigma) served as a loading control. After a 2-h incubation at room temperature with secondary antibodies, protein bands were detected by enhanced chemiluminescence with the RPN2108 kit (Amersham Biosciences) and BioMax Light film (Eastman Kodak Co.).

Immunofluorescence Analysis—Oocytes and embryos without zonae pellucidae were washed twice in PBS, then fixed in freshly prepared 4% paraformaldehyde in PBS, permeabilized in 1% Triton X-100 in PBS, and left in blocking solution (1% BSA in PBS) for 1 h. For immunolabeling, the embryos were incubated overnight with anti-REST (ab21635, Abcam), anti-Nanog (500-P236, Peprotech), or anti-Smad3-p (ab118825, Abcam) antibodies; washed three times, and incubated for 1 h with secondary antibody FITC-labeled donkey anti-mouse IgG (A21202, Invitrogen) diluted 1:1000 with blocking solution. Immunofluorescence of injected oocytes and one-cell NT embryos without anti-REST antibody (only secondary anti-

body) was used to analyze REST antibody injection and degradation. Samples were washed and counterstained with 5 $\mu\text{g}/\text{ml}$ of Hoechst 33342. Fluorescence was detected and imaged using a Nikon fluorescence microscope.

Statistical Analysis—Statistical analysis was performed using SPSS 13.0 for MicroSoft™ Windows. Data are shown as the mean \pm S.D. One-way analysis of variance was used to assess any differences between groups. The Duncan method was used for pairwise comparisons followed by a Bonferroni correction. $p < 0.05$ (two-tailed) was considered statistically significant.

Author Contributions—Z. H. L. and Q. R. K. designed and conceived the experiments; B. T. X., Q. R. K., and H. Z. conducted the oocyte and embryo manipulations; Q. R. K. and T. Q. H. conducted molecular experiments; J. Y. L. contributed to bioinformatics analysis; R. Y. W. conducted the cell manipulations. Z. H. L. and Q. R. K. wrote and all authors reviewed the manuscript.

Acknowledgment—We gratefully acknowledge Dr. Jilong Liu in Huanan Agricultural University for proteomic analysis.

References

- Gurdon, J. B., Elsdale, T. R., and Fischberg, M. (1958) Sexually mature individuals of *Xenopus laevis* from the transplantation of single somatic nuclei. *Nature* **182**, 64–65
- Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J., and Campbell, K. H. (1997) Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813
- Takahashi, K., and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676
- Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., and Thomson, J. A. (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920
- Ooi, L., and Wood, I. C. (2007) Chromatin crosstalk in development and disease: lessons from REST. *Nat. Rev. Genet.* **8**, 544–554
- Perera, A., Eisen, D., Wagner, M., Laube, S. K., Künzel, A. F., Koch, S., Steinbacher, J., Schulze, E., Splith, V., Mittermeier, N., Müller, M., Biel, M., Carell, T., and Michalakis, S. (2015) TET3 is recruited by REST for context-specific hydroxymethylation and induction of gene expression. *Cell Rep.* **11**, 283–294
- Negrini, S., Prada, I., D'Alessandro, R., and Meldolesi, J. (2013) REST: an oncogene or a tumor suppressor? *Trends Cell Biol.* **23**, 289–295
- Tokoyoda, K., and Radbruch, A. (2012) Signals controlling rest and reactivation of T helper memory lymphocytes in bone marrow. *Cell. Mol. Life Sci.* **69**, 1609–1613
- Kagalwala, M. N., Singh, S. K., and Majumder, S. (2008) Stemness is only a state of the cell. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 227–234
- Ding, N., Zhou, H., Esteve, P. O., Chin, H. G., Kim, S., Xu, X., Joseph, S. M., Friez, M. J., Schwartz, C. E., Pradhan, S., and Boyer, T. G. (2008) Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation. *Mol. Cell* **31**, 347–359
- Bruce, A. W., Donaldson, I. J., Wood, I. C., Yerbury, S. A., Sadowski, M. I., Chapman, M., Göttgens, B., and Buckley, N. J. (2004) Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10458–10463
- Ichida, J. K., Blanchard, J., Lam, K., Son, E. Y., Chung, J. E., Egli, D., Loh, K. M., Carter, A. C., Di Giorgio, F. P., Koszka, K., Huangfu, D., Akutsu, H., Liu, D. R., Rubin, L. L., and Eggan, K. (2009) A small-molecule inhibitor of TGF- β signaling replaces Sox2 in reprogramming by inducing nanog. *Cell Stem Cell* **5**, 491–503
- Tada, M., Morizane, A., Kimura, H., Kawasaki, H., Ainscough, J. F., Sasai, Y., Nakatsuji, N., and Tada, T. (2003) Pluripotency of reprogrammed somatic genomes in embryonic stem hybrid cells. *Dev. Dyn.* **227**, 504–510
- Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., and Tada, T. (2001) Nuclear reprogramming of somatic cells by *in vitro* hybridization with ES cells. *Curr. Biol.* **11**, 1553–1558
- Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M. J., Ji, H., Ehrlich, L. I., Yabuuchi, A., Takeuchi, A., Cunniff, K. C., Hongguang, H., McKinney-Freeman, S., et al. (2010) Epigenetic memory in induced pluripotent stem cells. *Nature* **467**, 285–290
- Chen, Z. F., Paquette, A. J., and Anderson, D. J. (1998) NRSF/REST is required *in vivo* for repression of multiple neuronal target genes during embryogenesis. *Nat. Genet.* **20**, 136–142
- Qureshi, I. A., and Mehler, M. F. (2010) Impact of nuclear organization and dynamics on epigenetic regulation in the central nervous system: implications for neurological disease states. *Ann. N.Y. Acad. Sci.* **1204**, E20–E37
- Qureshi, I. A., Gokhan, S., and Mehler, M. F. (2010) REST and CoREST are transcriptional and epigenetic regulators of seminal neural fate decisions. *Cell Cycle* **9**, 4477–4486
- Robertson, E. J. (2014) Dose-dependent Nodal/Smad signals pattern the early mouse embryo. *Semin. Cell Dev. Biol.* **32**, 73–79
- Papanayotou, C., and Collignon, J. (2014) Activin/Nodal signalling before implantation: setting the stage for embryo patterning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, pii. 21030539
- Park, K. S. (2011) Tgf- β family signaling in embryonic stem cells. *Int. J. Stem Cells* **4**, 18–23
- Rossant, J., and Tam, P. P. (2009) Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development* **136**, 701–713
- Shen, M. M. (2007) Nodal signaling: developmental roles and regulation. *Development* **134**, 1023–1034
- Takaoka, K., Yamamoto, M., and Hamada, H. (2011) Origin and role of distal visceral endoderm, a group of cells that determines anterior-posterior polarity of the mouse embryo. *Nat. Cell Biol.* **13**, 743–752
- Cao, S., Han, J., Wu, J., Li, Q., Liu, S., Zhang, W., Pei, Y., Ruan, X., Liu, Z., Wang, X., Lim, B., and Li, N. (2014) Specific gene-regulation networks during the pre-implantation development of the pig embryo as revealed by deep sequencing. *BMC Genomics* **15**, 4
- Xu, R. H., Sampsel-Barron, T. L., Gu, F., Root, S., Peck, R. M., Pan, G., Yu, J., Antosiewicz-Bourget, J., Tian, S., Stewart, R., and Thomson, J. A. (2008) NANOG is a direct target of TGF β /activin-mediated SMAD signaling in human ESCs. *Cell Stem Cell* **3**, 196–206
- Suzuki, A., Raya, A., Kawakami, Y., Morita, M., Matsui, T., Nakashima, K., Gage, F. H., Rodríguez-Esteban, C., and Izpisua Belmonte, J. C. (2006) Nanog binds to Smad1 and blocks bone morphogenetic protein-induced differentiation of embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10294–10299
- Kikyo, N., Wade, P. A., Guschin, D., Ge, H., and Wolffe, A. P. (2000) Active remodeling of somatic nuclei in egg cytoplasm by the nucleosomal ATPase ISWI. *Science* **289**, 2360–2362
- Jullien, J., Astrand, C., Halley-Stott, R. P., Garrett, N., and Gurdon, J. B. (2010) Characterization of somatic cell nuclear reprogramming by oocytes in which a linker histone is required for pluripotency gene reactivation. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5483–5488
- Tamada, H., Van Thuan, N., Reed, P., Nelson, D., Katoku-Kikyo, N., Wudel, J., Wakayama, T., and Kikyo, N. (2006) Chromatin decondensation and nuclear reprogramming by nucleoplasmin. *Mol. Cell Biol.* **26**, 1259–1271
- Wen, D., Banaszynski, L. A., Liu, Y., Geng, F., Noh, K. M., Xiang, J., Elemento, O., Rosenwaks, Z., Allis, C. D., and Rafii, S. (2014) Histone variant H3.3 is an essential maternal factor for oocyte reprogramming. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7325–7330
- Miyamoto, K., Nagai, K., Kitamura, N., Nishikawa, T., Ikegami, H., Binh, N. T., Tsukamoto, S., Matsumoto, M., Tsukiyama, T., Minami, N., Yamada, M., Ariga, H., Miyake, M., Kawarasaki, T., Matsumoto, K., and

REST Is Required for Nuclear Reprogramming

- Imai, H. (2011) Identification and characterization of an oocyte factor required for development of porcine nuclear transfer embryos. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7040–7045
33. Hansis, C., Barreto, G., Maltry, N., and Niehrs, C. (2004) Nuclear reprogramming of human somatic cells by *Xenopus* egg extract requires BRG1. *Curr. Biol.* **14**, 1475–1480
34. Gonda, K., Fowler, J., Katoku-Kikyo, N., Haroldson, J., Wudel, J., and Kikyo, N. (2003) Reversible disassembly of somatic nucleoli by the germ cell proteins FRGY2a and FRGY2b. *Nat. Cell. Biol.* **5**, 205–210
35. Kong, Q., Xie, B., Li, J., Huan, Y., Huang, T., Wei, R., Lv, J., Liu, S., and Liu, Z. (2014) Identification and characterization of an oocyte factor required for porcine nuclear reprogramming. *J. Biol. Chem.* **289**, 6960–6968
36. Tachibana, M., Amato, P., Sparman, M., Gutierrez, N. M., Tippner-Hedges, R., Ma, H., Kang, E., Fulati, A., Lee, H. S., Sritanandomchai, H., Masterson, K., Larson, J., Eaton, D., Sadler-Fredd, K., Battaglia, D., *et al.* (2013) Human embryonic stem cells derived by somatic cell nuclear transfer. *Cell* **153**, 1228–1238

Ewing Tumor-associated Antigen 1 Interacts with Replication Protein A to Promote Restart of Stalled Replication Forks*

Received for publication, July 11, 2016, and in revised form, August 27, 2016
 Published, JBC Papers in Press, September 6, 2016, DOI 10.1074/jbc.C116.747758

Sumin Feng, Yichao Zhao, Yixi Xu, Shaokai Ning, Wei Huo, Mei Hou, Ge Gao, Jianguo Ji, Rong Guo¹, and Dongyi Xu²

From the State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China

The replication protein A (RPA) complex binds single-stranded DNA generated at stalled replication forks and recruits other DNA repair proteins to promote recovery of these forks. Here, we identify Ewing tumor-associated antigen 1 (ETAA1), which has been linked to susceptibility to pancreatic cancer, as a new repair protein that is recruited to stalled forks by RPA. We demonstrate that ETAA1 interacts with RPA through two regions, each of which resembles two previously identified RPA-binding domains, RPA70N-binding motif and RPA32C-binding motif, respectively. In response to replication stress, ETAA1 is recruited to stalled forks where it colocalizes with RPA, and this recruitment is diminished when RPA is depleted. Notably, inactivation of the *ETAA1* gene increases the collapse level of the stalled replication forks and decreases the recovery efficiency of these forks. Moreover, epistasis analysis shows that ETAA1 stabilizes stalled replication forks in an ataxia telangiectasia and Rad3-related protein (ATR)-independent manner. Thus, our results reveal that ETAA1 is a novel RPA-interacting protein that promotes restart of stalled replication forks.

The faithful replication of DNA is essential for the maintenance of genomic stability and the prevention of cancer-promoting mutations. Replication forks can be stalled by numerous obstacles on the DNA template, including unrepaired DNA damage, DNA-bound proteins, and secondary structures (1). Stalled replication forks are able to restart once the obstacles are removed or become broken (collapse) into DNA double strand breaks, which pose the most serious threat to genome integrity when fork protection fails (2, 3). However, how stalled replication forks are protected is not well understood.

* This work was supported in part by the National Basic Research Program of China Grant 2013CB911002 and the National Natural Science Foundation of China Grants 31271435 and 31370836. The authors declare that they have no conflicts of interest with the contents of this article.

¹ To whom correspondence may be addressed. Tel.: 86-10-62767767; E-mail: guorong@pku.edu.cn.

² To whom correspondence may be addressed. Tel.: 86-10-62759816; E-mail: xudongyi@pku.edu.cn.

The replication protein A (RPA)³ complex, which consists of RPA1 (RPA70), RPA2 (RPA32), and RPA3 (RPA14), plays crucial roles in a variety of DNA metabolic pathways, including DNA replication, recombination, repair, and DNA damage checkpoint (4–6). When replication forks stalled, single-stranded DNA (ssDNA) is generated and extended by minichromosome maintenance protein complex helicases (7, 8). The ssDNA is bound by RPA, which protects ssDNA from cleavage by nucleases and recruits repair proteins to initiate DNA damage responses. The RPA-ssDNA complex recruits and activates ATR/ATRIP thereby eliciting checkpoint signaling (9). In addition, RPA-ssDNA complex also recruits factors necessary for the stabilization and resumption of stalled replication forks, such as RAD51 (10, 11) and SMARCAL1 (12–15). Recently, several studies have also revealed a physical and functional interaction between RPA and the ubiquitin E3 ligases RFD3 (16–18) and PRP19 (19), which ubiquitinate RPA and facilitate replication fork restart. Here, we identified a new RPA interaction protein, ETAA1, whose gene variation has been associated with susceptibility to pancreatic cancer (20). ETAA1 is recruited to stalled replication forks in response to replication stress, and the disruption of ETAA1 leads to fork collapse in an ATR-independent manner. These results suggest that ETAA1 is a new player involved in the stabilization of stalled replication forks.

Results

ETAA1 Is a Novel RPA-associated Protein—We transiently expressed FLAG-tagged RPA1 in HEK293 cells and immunoprecipitated the complexes with an anti-FLAG antibody (Fig. 1A). Mass spectrometry analysis revealed that, in addition to RPA2, RPA3, and the BLM complex, a novel protein, ETAA1, also immunoprecipitated with RPA1 (Fig. 1A). Immunoblotting confirmed this finding (Fig. 1B). To verify that ETAA1 indeed associates with the RPA complex, we performed a reciprocal immunoprecipitation using HEK293 cells expressing FLAG-tagged ETAA1; immunoblotting revealed that RPA was present in the ETAA1-associated complexes (Fig. 1C). To further confirm this interaction, we also performed immunoprecipitations with anti-RPA2 and anti-ETAA1 antibodies, and we found that endogenous ETAA1 strongly associated with the RPA complex *in vivo* (Fig. 1, D and E). Together, these data suggest that ETAA1 is a *bona fide* RPA-associated protein.

To examine the interaction during the cell cycle, cells were synchronized at the G₁-S boundary by a double thymidine treatment or at metaphase by a nocodazole treatment and then released. ETAA1 showed the strongest interaction with RPA at S phase and the lowest at G₂/M phase (Fig. 1, F and G), suggesting that ETAA1 may function with RPA in DNA replication.

³ The abbreviations used are: RPA, replication protein A; HU, hydroxyurea; MBP, maltose-binding protein; ATR, ataxia telangiectasia and Rad3-related protein; ssDNA, single-stranded DNA; IdU, 5-iodo-2'-deoxyuridine; QIBC, quantitative image-based cytometry; WB, Western blot; IF, immunofluorescence; CldU, 5-chloro-2'-deoxyuridine; CPT, camptothecin.

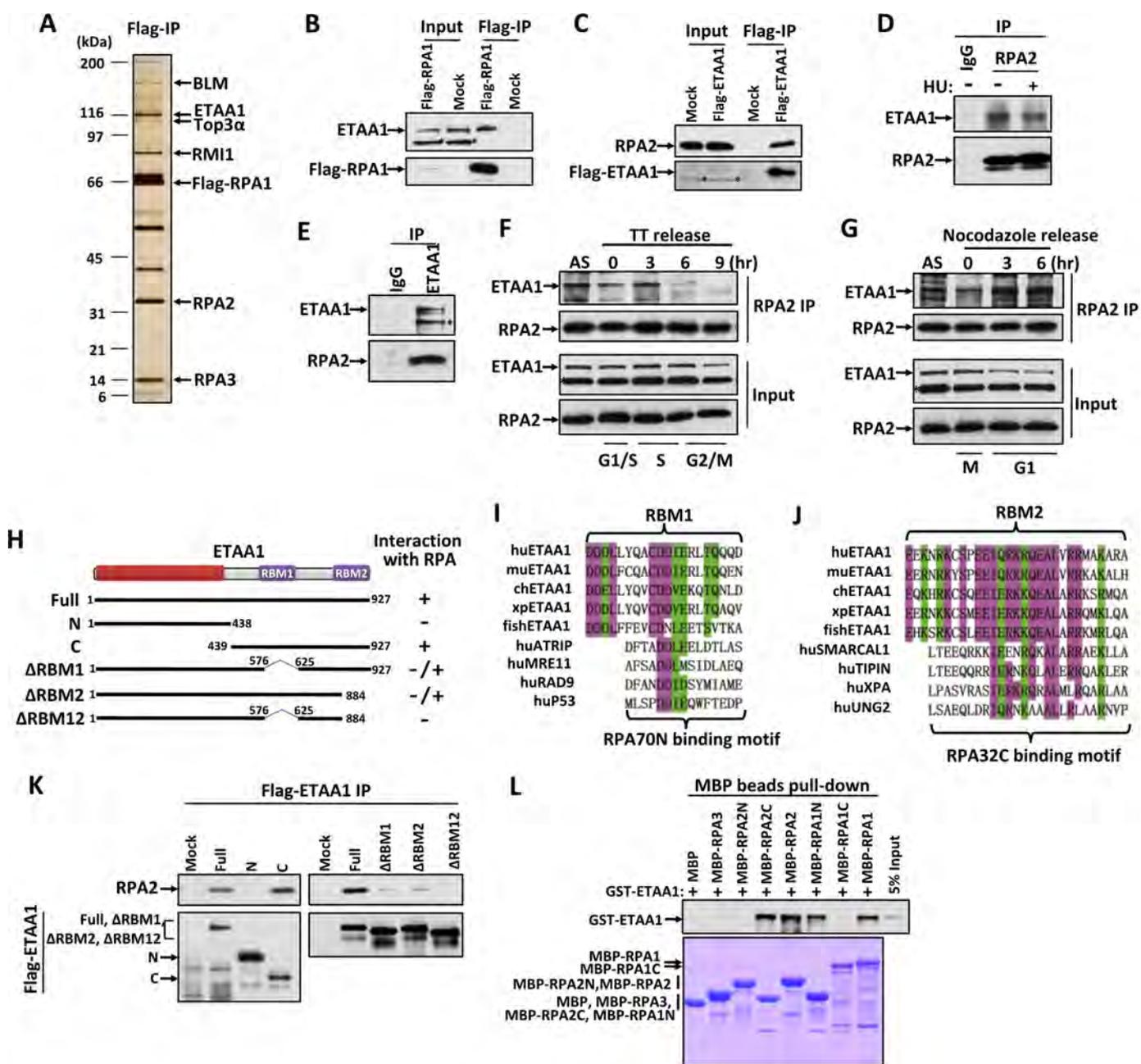


FIGURE 1. ETAA1 associates with the RPA complex. *A*, silver-stained SDS-polyacrylamide gel showing the polypeptides that were immunoprecipitated from extracts of HEK293 cells expressing FLAG-tagged RPA1 using the anti-FLAG antibody. The major polypeptides on the gel (arrows) were identified by mass spectrometry. *B* and *C*, immunoblot showing the immunoprecipitation (IP) of FLAG-tagged RPA1 (*B*) and ETAA1 (*C*). Asterisks indicate cross-reactive polypeptides. *D* and *E*, immunoblot showing the endogenous RPA2 (*D*) and ETAA1 (*E*) immunoprecipitation. HEK293 cells were treated with or without 4 mM HU for 3 h before harvest in the RPA2 immunoprecipitation. *F* and *G*, interaction of ETAA1 with RPA during the cell cycle. HEK293 cells were synchronized at the G₁-S boundary by a double thymidine treatment (TT), released into fresh medium, and collected at the indicated times (*F*). Alternatively, cells were synchronized at prometaphase by a nocodazole block, released into fresh medium, and harvested at the indicated times (*G*). The cell cycle profile was analyzed by flow cytometry with the DNA content determined propidium iodide-staining (data not shown). Cell lysates and the anti-RPA immunoprecipitates (IP) were analyzed by Western blotting. AS, asynchronous cells. The asterisk indicates a cross-reactive polypeptide. *H*, schematic representation of the different ETAA1 deletion mutants (left) and their ability to coimmunoprecipitate with RPA from HEK293 extracts (right). *I* and *J*, sequence alignment of RBM1 and the RPA70N-binding motif (*I*) or RBM2 and the RPA32C-binding motif (*J*). *K*, immunoprecipitation and Western blotting to assess whether the various deletion mutants of ETAA1 described in *H* coimmunoprecipitated with RPA. *L*, direct binding between recombinant GST-tagged ETAA1 and MBP-tagged RPA. Upper panel, GST-ETAA1 was detected by immunoblotting with anti-GST antibodies. Lower panel, purified MBP-fused proteins was visualized by Coomassie staining.

Additionally, the binding of RPA2 to ETAA1 was not changed after cells were exposed to HU (Fig. 1D).

ETAA1 Contains Two RPA-binding Motifs—ETAA1 is expressed only in vertebrates. Sequence analyses showed that ETAA1 contains three conserved regions (Fig. 1H) as follows. The N-terminal conserved region contains no well known

domains; the middle conserved region, RBM1, shows weak similarity to the RPA70N-binding motifs of ATRIP, MRE11, RAD9, and P53 (Fig. 1I) (21); and the C-terminal conserved region, RBM2, is similar to the RPA32C-terminal binding motifs of SMARCAL1, TIPIN, XPA, and UNG2 (Fig. 1J) (22, 23). To identify the region(s) of ETAA1 responsible for its interaction

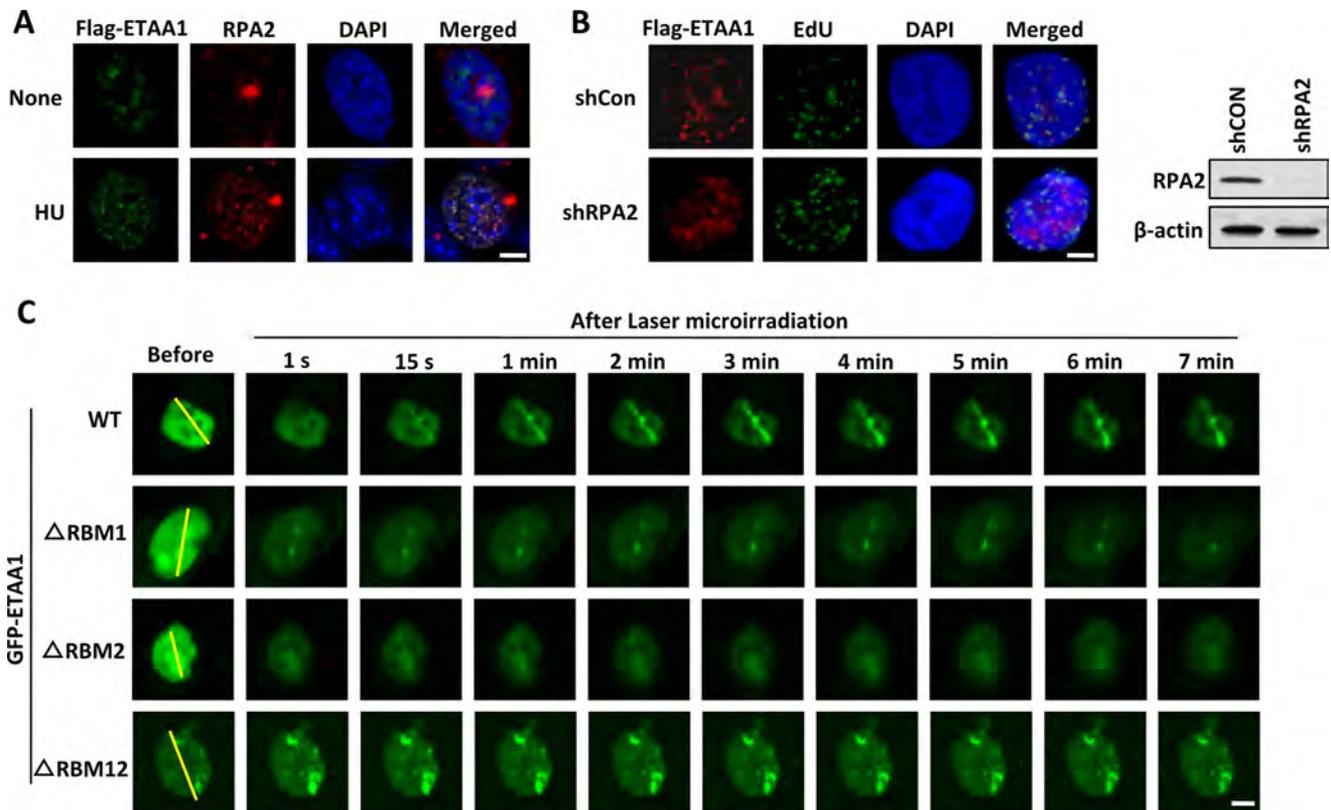


FIGURE 2. **ETAA1** colocalizes with RPA at stalled replication forks. **A**, ETAA1 localizes to stalled replication forks in response to replication stress. U2OS cells expressing FLAG-ETAA1 were mock-treated or treated with 5 mM HU for 6 h. Immunostaining was performed using anti-FLAG and anti-RPA2 antibodies. (Scale bar, 5 μ m.) **B**, localization of ETAA1 to stalled replication forks requires RPA. U2OS cells expressing FLAG-ETAA1 were transfected with control or RPA2 shRNA. The cells were pulse-labeled with 20 μ M EdU for 30 min and then treated with 5 mM HU for 6 h. Scale bar, 5 μ m. Immunoblotting shows knockdown efficiency in the right panel. **C**, cells expressing wild-type or mutant GFP-ETAA1 were treated with laser microirradiation and monitored via live-cell imaging. The yellow lines indicate the positions for laser microirradiation. (Scale bar, 5 μ m.)

with RPA, we generated a series of ETAA1 truncation mutants (Fig. 1H). As shown in Fig. 1, H and K, the N-terminal region (residues 1–438, containing the first conserved region) of ETAA1 is dispensable for its interaction with RPA. The mutant lacking the RBM1 motif (Δ RBM1) or the RBM2 motif (Δ RBM2) dramatically decreased the ETAA1-RPA interaction, thus indicating that both RBM1 and RBM2 regions of ETAA1 are important for binding to RPA. Moreover, deletion of both RBM1 and RBM2 motifs (Δ RBM12) completely lost its interaction with RPA. These results are consistent with the sequence analyses that ETAA1 contains two RPA-binding motifs, RBM1 and RBM2.

To determine whether the interaction between RPA and ETAA1 is direct, we expressed and purified recombinant MBP-tagged RPA1, RPA2, RPA3, and GST-tagged ETAA1 from *Escherichia coli*. Pulldown experiments revealed that ETAA1 binds strongly with RPA1 and RPA2 but not with RPA3 (Fig. 1L), indicating that two independent ETAA1-binding sites exist on RPA1 and RPA2. Moreover, mapping analysis revealed that the RPA1 N-terminal region (residues 1–120) and RPA2 C-terminal region (residues 204–270) directly interacted with ETAA1 (Fig. 1L), thus agreeing that RBM1 and RBM2 of ETAA1 are RPA70N-terminal and RPA32C-terminal binding motifs, respectively.

ETAA1 Is Recruited to DNA Damage Sites by RPA—The RPA complex binds to ssDNA generated at stalled replication forks

or DNA damage sites and forms an RPA-ssDNA platform, which facilitates the recruitment of many repair proteins. The interaction of ETAA1 and RPA suggests that ETAA1 may colocalize with RPA at ssDNA regions *in vivo*. As shown in Fig. 2A, FLAG-ETAA1 rarely formed foci in untreated cells. After HU treatment, ETAA1 was recruited to nuclear foci, where it colocalized with RPA, thus suggesting that ETAA1 localizes to stalled replication forks and may play a role in cellular responses to replication stress.

We then determined whether the localization of ETAA1 at stalled replication forks was dependent on the RPA complex. As shown in Fig. 2B, ETAA1 exhibited dramatically decreased foci formation in RPA2-depleted cells but not in control cells after HU treatment, thereby indicating that ETAA1 is mainly recruited to stalled replication forks by the RPA complex.

We then examined the dynamic recruitment of GFP-ETAA1 at microirradiation-induced DNA damage sites in time-lapse experiments. GFP-ETAA1 accumulated at laser tracks very quickly, within 1 s, and the signal peaked at \sim 2 min and then persisted for more than 1 h (Fig. 2C and data not shown). To determine whether this recruitment was also dependent on RPA, we tested ETAA1 mutants lacking the RBM1 or RBM2 motif. Deletion of RBM1 or RBM2 did not affect ETAA1 recruitment at the early stage (<1 s) but significantly reduced its recruitment to and (or) retention on DNA damage sites at the late stage (Fig. 2C). The deletion of both RBM1 and RBM2

completely abolished the localization of ETAA1 to DNA damage sites (Fig. 2C). These results suggest that the localization of ETAA1 at DNA damage sites is dependent on its interaction with RPA through the RBM1 and RBM2 motifs.

ETAA1 Is Required for Stalled Replication Fork Restart—The RPA complex plays a key role in replication and in stalled replication restart. Thus, we generated ETAA1-knock-out HCT116 cells using two different CRISPR target sites (Fig. 3A) and performed single DNA fiber analysis. Cells were pulse-labeled with CldU for 30 min and then incubated with HU (5 mM) and aphidicolin (5 μ M) for 5 h to arrest replication forks. After a washout step, cells were incubated with IdU for 20 min. Restarted replication forks were visualized as tracks of CldU incorporation followed by tracks of IdU incorporation, whereas stalled or collapsed replication forks appeared as tracks of only CldU incorporation. We found that two *ETAA1*^{-/-} clones both exhibited significantly decreased fork restart, as indicated by an approximate 2-fold reduction (Fig. 3B). These results suggest that ETAA1 promotes stalled fork restart upon replication stress.

We also examined replication rate under normal conditions by measuring IdU track length (Fig. 3C). The *ETAA1*^{-/-} cells showed similar track length as that of wild-type cells, indicating that ETAA1 is not required for normal replication.

ETAA1 Stabilizes Stalled Replication Forks—Combined treatment with HU and an ATR inhibitor exhausts RPA and triggers fork breakage, and it is accompanied by hallmarks of ataxia telangiectasia-mutated (ATM) activity such as H2AX hyperphosphorylation (24). We examined fork stability during replication stress in ETAA1-defective cells by measuring H2AX hyperphosphorylation using a previously described quantitative image-based cytometry (QIBC) method (24). HU treatment induced H2AX hyperphosphorylation in *ETAA1*^{-/-} cells but not in wild-type cells, a result similar to that induced in wild-type cells after treatment with both HU and an ATR inhibitor (Fig. 3D). These results suggest that ETAA1 has a similar function to that of ATR in preventing stalled fork collapse. Indeed, the loss of ETAA1 also caused hypersensitivity to HU and CPT (Fig. 3E), a topoisomerase I inhibitor that induces topoisomerase I-DNA adducts and blocks replication, thus further supporting a role for ETAA1 in fork protection upon replication stress. Although the expression of the wild-type ETAA1 completely rescued stabilization of stalled replication fork and CPT resistance in the *ETAA1*^{-/-} cells, reconstitution with ETAA1 mutants lacking RBM1, RBM2, or both failed to do so (Fig. 3, F–H). This result suggests that the interaction of ETAA1 with RPA is important for its function in response to replication stress.

Moreover, the combination of the ATR inhibitor and HU treatment induced a higher γ H2AX signal in *ETAA1*^{-/-} cells than in wild-type cells, or in *ETAA1*^{-/-} cells treated with HU alone (Fig. 3D), suggesting that ETAA1 has an ATR-independent role in protecting the stalled replication forks.

Discussion

The RPA complex is a key player in initiating DNA-damage checkpoint signaling, replication fork stabilization, and DNA repair. The binding of RPA to ssDNA not only protects ssDNA

from degradation by nucleases but also forms a platform facilitating the recruitment of many binding partners for diverse functions. Here, we identified a novel RPA-binding protein, ETAA1, which is required for the stability of the stalled replication fork.

ETAA1 contains two RPA-binding motifs, RBM1 and RBM2. RBM1 shows weak similarity to the RPA70N-binding motif of ATRIP, MRE11, RAD9, and p53, whereas RBM2 is similar to the RPA32C-binding motifs of SMARCAL1, TIPIN, XPA, and UNG2. The presence of either RBM1 or RBM2 is sufficient for ETAA1 to bind RPA at ssDNA regions, although the signal is weaker, thus allowing one RPA complex to bind other proteins together with ETAA1. This implies that ETAA1 can be orchestrated with other RPA-binding proteins on the RPA-ssDNA platform, which coordinates different DNA damage responses.

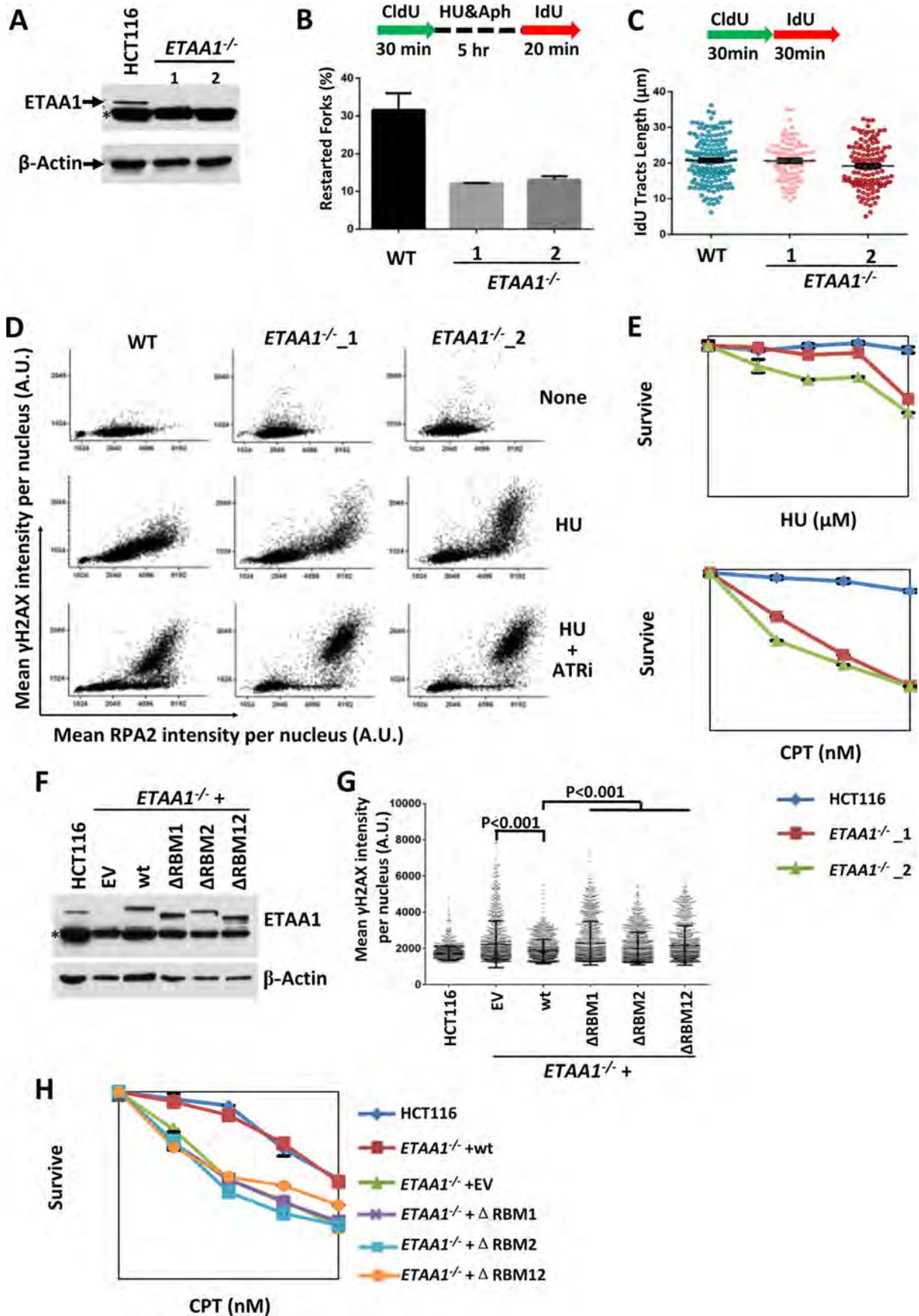
RPA complexes protect stalled replication forks through at least three mechanisms. First, RPA directly binds and protects ssDNA generated at stalled replication forks from nucleolytic digestion. ETAA1 probably does not contribute to this step because ETAA1 did not affect the ssDNA binding activity of RPA, and *ETAA1*^{-/-} cells even showed more ssDNA-bound RPA under replication stress than did wild-type cells (Fig. 3D). Second, RPA activates the ATR-dependent DNA replication checkpoint to stabilize stalled replication forks. However, our epistasis analysis using an ATR inhibitor in *ETAA1*-null cells showed that ETAA1 and ATR function in two parallel pathways that protect stalled replication forks. Third, RPA recruits repair proteins, which remodel and protect stalled replication forks. Whether ETAA1 facilitates remodeling of stalled replication forks, thereby promoting fork stabilization, remains to be determined.

Moreover, GFP-ETAA1 was recruited to laser-induced DNA damage sites in both G₁ and G₂/S phase cells (data not shown), thus suggesting that ETAA1 functions not only in stalled replication fork restart but also in other DNA repair pathways. Repair of the double strand breaks, which are commonly induced by laser treatment, also recruits RPA after end resection, which normally takes more than a few seconds. ETAA1 was recruited quickly to DNA damage sites (within 1 s), a result suggesting that ETAA1 may be recruited by RPA-independent mechanisms in the early stage after laser-induced DNA damage. The RPA-dependent and (or) independent functions of ETAA1 in other DNA damage repair processes require further study.

In summary, we identified ETAA1 as a novel RPA-binding protein, which protects stalled replication forks and maintains genomic stability. Our results provide insights into the function of ETAA1 in preventing pancreatic cancer.

Experimental Procedures

Cell Culture—HeLa, U2OS, and HEK293T cells were cultured in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum (FBS; Invitrogen). HCT116 cells were cultured in RPMI 1640 medium with 10% FBS (Invitrogen). HEK293 suspension cells were cultured in Freestyle medium (Invitrogen) supplemented with 1% FBS (Gibco) and 1% glutamine in an incubator with shaking at 130 rpm.



Coimmunoprecipitation—The immunoprecipitation of the complexes was performed as described previously (25).

MBP Pulldown Assay—MBP-tagged RPA proteins were expressed in *E. coli*. Cells were harvested and resuspended in lysis buffer (20 mM Tris-HCl, pH 7.0, 300 mM NaCl, 1% Triton X-100, and 1 μ g/ml each of leupeptin, aprotinin, and pepstatin). After sonicating, the extract was centrifuged at 18,000 rpm for 40 min. The supernatant was collected and incubated with amylose resins for 2 h at 4 °C. After washing the beads with washing buffer (20 mM Tris-HCl, 500 mM NaCl, 0.5% Nonidet P-40, 1 mM DTT), the protein-bound beads were incubated with the extract of *E. coli* cells expressing GST-ETAA1 for 2 h at 4 °C. After washing with binding buffer (20 mM Tris-HCl, pH 7.0, 150 mM NaCl, 0.5% Nonidet P-40, 1 mM DTT), the proteins were eluted with sample buffer (63 mM Tris-HCl, pH 6.8, 10% glycerol, 2% SDS, 0.0025% bromophenol blue) and analyzed with SDS-PAGE.

Laser Microirradiation—U2OS cells expressing GFP-ETAA1 were cultured at 37 °C in CO₂-independent medium (Invitrogen) containing 10% FBS in a temperature-controlled container in glass-bottom dishes (MatTek). Laser microirradiation was carried out with the MicroPoint Laser Illumination and Ablation System coupled to a Nikon eclipse Ti microscope with a plan fluor 60 \times 0.5–1.25 oil iris immersion objective. Time-lapse images were acquired with ANDOR IQ3 software through an ANDOR IXON camera.

Generation of ETAA1 Knock-out Cells—ETAA1-deficient HCT116 cells were generated using CRISPR. Briefly, two guide sequences, AGGAAACACCAAGATATCTG and GCTACAAAAAGCCAATCAA, targeting two different sites of the human *ETAA1* gene were inserted into the pX330 vector (26). The guide sequence containing pX330 plasmids were transfected into HCT116 cells. Single colonies were picked after 8–10 days of incubation. The genomic fragments of the *ETAA1* gene were amplified by PCR using the following primers: GAGCTAGATGTGATTCAAGAGC and CTGTCCGCTAC-ATTTCTGAG. The products were digested with EcoRV and BslI, respectively. Colonies containing the expected PCR fragments were then sequenced and examined by Western blotting.

DNA Fiber Assay—The restart efficiency of stalled replication forks was determined by using DNA fiber assays as described previously (27). Cells were first labeled with CldU (100 μ M) for 30 min and then treated with HU (5 mM) and aphidicolin (5 μ M) for 5 h. After being washed with PBS, cells were recovered in fresh medium with IdU (20 μ M) for 20 min. Cells were then trypsinized and resuspended in PBS to a concentration of 2.5×10^5 cells/ml. Then the cells were diluted 1:4 with unlabeled cells at the same concentration, and 2.5 μ l of cells was mixed with 7.5 μ l of lysis buffer (200 mM Tris-HCl, pH

7.5, 50 mM EDTA, and 0.5% SDS) on a clean glass slide. After the edges were dried for 3–5 min, the slides were tilted at 15° to horizontal, allowing the DNA to slowly flow down along the slide. The slides were then air-dried, fixed in 3:1 methanol/acetic acid, and refrigerated overnight. The slides were treated with 2.5 M HCl for 1 h, neutralized in 0.1 M Na₃B₄O₇, pH 8.5, and rinsed three times in PBST (PBS buffer with 0.1% Tween 20). The slides were then blocked in blocking buffer (PBST buffer containing 1% BSA) for 20 min and incubated with rat anti-BrdU antibody (Abcam BU1/75, 1:200) in blocking buffer at 37 °C for 1 h. After three washes, with PBST, slides were incubated with Alexa Fluor 488-conjugated anti-rat (Molecular Probes, 1:200 dilution) for 45 min. After additional washes, the slides were incubated with mouse anti-BrdU (BD Biosciences, B44, 1:40) for 1 h and then washed once with high-salt PBST (0.5 M NaCl) and three times with PBST. Then the slides were incubated with Alexa Fluor 549-conjugated anti-mouse (Molecular Probes, 1:200 dilution) for 45 min. After three washes with PBST, the slides were mounted in SlowFade Gold antifade reagent (Invitrogen). The slides were imaged on a Zeiss Axiovert microscope with a 100 \times objective.

Immunostaining and Immunoblotting—U2OS or HCT116 cells were cultured on polylysine-coated coverslips 24 h before the experiments. After washing with PBS, the cells were pre-extracted with 0.5% Triton X-100 in CSK buffer (20 mM HEPES, pH 7.0, 100 mM NaCl, 300 mM sucrose, and 3 mM MgCl₂). The cells were then washed three times with PBS and fixed with 3% paraformaldehyde for 10 min at room temperature. Before staining, the cells were permeabilized for 10 min with PBS, 0.5% Triton X-100, and washed three times. For EdU staining, the cells were incubated with Click-iT reaction buffer (PBS with 50 μ M Alexa Fluor® 488 azide, 10 mM sodium ascorbate, and 2 mM CuSO₄) for 30 min at room temperature. After washing, the cells were blocked with 5% BSA (Sigma) in PBS for 15 min. The primary antibodies were diluted in PBS containing 1% BSA and incubated with the cells for 90 min. After washing, secondary antibodies diluted in PBS containing 1% BSA were added to the cells for 30 min. The cells were washed three times and mounted with ProLong Gold antifade reagent with DAPI (Invitrogen). Images were acquired with an LSM710 confocal microscope (Zeiss) using a 100 \times /1.4 NA objective. For immunoblotting, primary antibodies were incubated for 1.5 h at room temperature in PBST containing 5% powder milk. Secondary peroxidase-coupled antibodies (Jackson ImmunoResearch) were incubated at room temperature for 45 min. ECL-based chemiluminescence was detected by using film. Primary antibodies were used at the following dilutions: RPA2/RPA32 (Bethyl, A300-244A, WB, 1:2000; IF, 1:500), ETAA1 (Abcam,

FIGURE 3. **ETAA1 stabilizes stalled replication forks.** *A*, immunoblot shows that ETAA1 protein is absent in two *ETAA1*^{-/-} HCT116 cell clones. β -Actin is included as a control. The asterisk indicates a cross-reactive polypeptide. *B*, ETAA1 is required for stalled replication restart. Cells were pulse-labeled and treated as outlined in the top panel. DNA fibers were stained with antibodies recognizing IdU (red) and CldU (green). Restarted forks are indicated by green tracts followed by red tracts. Data represent the mean and S.D. *C*, ETAA1 is not required for normal replication. Cells were pulse-labeled as outlined in the top panel. The lengths of the IdU tracks were measured and are presented in the graph. *D*, QIBC of immunolabeled wild-type or *ETAA1*^{-/-} HCT116 cells. Cells were treated with/without HU (2 mM) and the ATR inhibitor VE821 (10 μ M) for 3 h, pre-extracted, and immunostained with the indicated antibodies. Mean nuclear intensities for RPA2 and γ -H2AX were determined for each of >5000 individual cells and are plotted in a scatter diagram. *E*, cell survival assay of the wild-type and *ETAA1*^{-/-} HCT116 cells to HU and CPT. *F–H*, complementation experiments show that RPA-binding motifs of ETAA1 are required for its function in response to replication stress. The expression levels of ETAA1 in different cells were determined by immunoblotting (*F*). γ H2AX intensity per nucleus was determined by QIBC (*G*). CPT sensitivity was measured by colony formation assay (*H*). EV, empty vector.

ab192402, WB, 1:1000), H2AX-pS139 (Millipore, 05-636, IF, 1:5000), FLAG (MBL, M185-3L, WB, 1:2500; IF, 1:250).

QIBC—QIBC was performed as described previously (24). Briefly, images were acquired in an unbiased fashion with an Image Xpress Micro XL microscope (Molecular Devices) with a 10×/0.3 NA objective and a scientific CMOS camera. For every sample, 30–50 images were acquired, containing a total of 5000 to 10,000 cells per condition. After acquisition, the images were processed for automated analysis with the MetaXpress High Content Image Acquisition and Analysis software.

Cell Survival Assay—Cell survival curves for HCT116 cells treated with HU and CPT were generated as described previously (28). An appropriate number of cells were plated in 6-well plates and cultured for 24 h, and then the indicated dose of HU or CPT was added to the medium. After an additional 9–14 days of incubation, the colonies were stained with methylene blue and counted.

Author Contributions—S. F., Y. Z., Y. X., S. N., W. H., and M. H. conducted the experiments. G. G., J. J., R. G., and D. X. supervised the project. S. F., Y. X., S. N., W. H., G. G., J. J., R. G., and D. X. analyzed the data. S. F., R. G., and D. X. wrote the manuscript with input from the other authors.

Acknowledgments—We thank Weidong Wang and Lee Zou for their advice and revisions to the manuscript, Jun Huang for providing MBP-tagged RPA2- and RPA3-expressing plasmids, and Xingzhi Xu for providing the laser microirradiation system. We thank the mass spectrometry facility at the National Center for Protein Sciences, Peking University, for assistance with the identification of the proteins, and the Core Facility of Life Sciences, Peking University, for assistance with the imaging.

References

1. Petermann, E., and Helleday, T. (2010) Pathways of mammalian replication fork restart. *Nat. Rev. Mol. Cell Biol.* **11**, 683–687
2. Tercero, J. A., Longhese, M. P., and Diffley, J. F. (2003) A central role for DNA replication forks in checkpoint activation and response. *Mol. Cell* **11**, 1323–1336
3. Lambert, S., and Carr, A. M. (2005) Checkpoint responses to replication fork barriers. *Biochimie* **87**, 591–602
4. Wold, M. S. (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* **66**, 61–92
5. Iftode, C., Daniely, Y., and Borowiec, J. A. (1999) Replication protein A (RPA): the eukaryotic SSB. *Crit. Rev. Biochem. Mol. Biol.* **34**, 141–180
6. Fanning, E., Klimovich, V., and Nager, A. R. (2006) A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res.* **34**, 4126–4137
7. Sogo, J. M., Lopes, M., and Foiani, M. (2002) Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science* **297**, 599–602
8. Byun, T. S., Pacek, M., Yee, M. C., Walter, J. C., and Cimprich, K. A. (2005) Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes Dev.* **19**, 1040–1052
9. Zou, L., and Elledge, S. J. (2003) Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* **300**, 1542–1548
10. Jackson, S. P. (2002) Sensing and repairing DNA double-strand breaks. *Carcinogenesis* **23**, 687–696

11. Stauffer, M. E., and Chazin, W. J. (2004) Physical interaction between replication protein A and Rad51 promotes exchange on single-stranded DNA. *J. Biol. Chem.* **279**, 25638–25645
12. Bansbach, C. E., Bétous, R., Lovejoy, C. A., Glick, G. G., and Cortez, D. (2009) The annealing helicase SMARCAL1 maintains genome integrity at stalled replication forks. *Genes Dev.* **23**, 2405–2414
13. Ciccía, A., Bredemeyer, A. L., Sowa, M. E., Terret, M. E., Jallepalli, P. V., Harper, J. W., and Elledge, S. J. (2009) The SIOD disorder protein SMARCAL1 is an RPA-interacting protein involved in replication fork restart. *Genes Dev.* **23**, 2415–2425
14. Driscoll, R., and Cimprich, K. A. (2009) HARPing on about the DNA damage response during replication. *Genes Dev.* **23**, 2359–2365
15. Yuan, J., Ghosal, G., and Chen, J. (2009) The annealing helicase HARP protects stalled replication forks. *Genes Dev.* **23**, 2394–2399
16. Gong, Z., and Chen, J. (2011) E3 ligase RFD3 participates in replication checkpoint control. *J. Biol. Chem.* **286**, 22308–22313
17. Liu, S., Chu, J., Yucer, N., Leng, M., Wang, S. Y., Chen, B. P., Hittelman, W. N., and Wang, Y. (2011) RING finger and WD repeat domain 3 (RFD3) associates with replication protein A (RPA) and facilitates RPA-mediated DNA damage response. *J. Biol. Chem.* **286**, 22314–22322
18. Elia, A. E., Wang, D. C., Willis, N. A., Boardman, A. P., Hajdu, I., Adeyemi, R. O., Lowry, E., Gygi, S. P., Scully, R., and Elledge, S. J. (2015) RFD3-dependent ubiquitination of RPA regulates repair at stalled replication forks. *Mol. Cell* **60**, 280–293
19. Maréchal, A., Li, J. M., Ji, X. Y., Wu, C. S., Yazinski, S. A., Nguyen, H. D., Liu, S., Jiménez, A. E., Jin, J., and Zou, L. (2014) PRP19 transforms into a sensor of RPA-ssDNA after DNA damage and drives ATR activation via a ubiquitin-mediated circuitry. *Mol. Cell* **53**, 235–246
20. Childs, E. J., Mocci, E., Campa, D., Bracci, P. M., Gallinger, S., Goggins, M., Li, D., Neale, R. E., Olson, S. H., Scelo, G., Amundadottir, L. T., Bamlet, W. R., Bijlsma, M. F., Blackford, A., Borges, M., et al. (2015) Common variation at 2p13.3, 3q29, 7p13, and 17q25.1 associated with susceptibility to pancreatic cancer. *Nat. Genet.* **47**, 911–916
21. Xu, X., Vaithiyalingam, S., Glick, G. G., Mordes, D. A., Chazin, W. J., and Cortez, D. (2008) The basic cleft of RPA70N binds multiple checkpoint proteins, including RAD9, to regulate ATR signaling. *Mol. Cell Biol.* **28**, 7345–7353
22. Feldkamp, M. D., Mason, A. C., Eichman, B. F., and Chazin, W. J. (2014) Structural analysis of replication protein A recruitment of the DNA damage response protein SMARCAL1. *Biochemistry* **53**, 3052–3061
23. Xie, S., Lu, Y., Jakoncic, J., Sun, H., Xia, J., and Qian, C. (2014) Structure of RPA32 bound to the N terminus of SMARCAL1 redefines the binding interface between RPA32 and its interacting proteins. *FEBS J.* **281**, 3382–3396
24. Toledo, L. I., Altmeyer, M., Rask, M. B., Lukas, C., Larsen, D. H., Povlsen, L. K., Bekker-Jensen, S., Mailand, N., Bartek, J., and Lukas, J. (2013) ATR prohibits replication catastrophe by preventing global exhaustion of RPA. *Cell* **155**, 1088–1103
25. Xing, M., Yang, M., Huo, W., Feng, F., Wei, L., Jiang, W., Ning, S., Yan, Z., Li, W., Wang, Q., Hou, M., Dong, C., Guo, R., Gao, G., Ji, J., Zha, S., Lan, L., Liang, H., and Xu, D. (2015) Interactome analysis identifies a new paralog of XRCC4 in non-homologous end joining DNA repair pathway. *Nat. Commun.* **6**, 6233
26. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823
27. Davies, S. L., North, P. S., and Hickson, I. D. (2007) Role for BLM in replication-fork restart and suppression of origin firing after replicative stress. *Nat. Struct. Mol. Biol.* **14**, 677–679
28. Katsube, T., Mori, M., Tsuji, H., Shiomi, T., Shiomi, N., and Onoda, M. (2011) Differences in sensitivity to DNA-damaging agents between XRCC4- and Artemis-deficient human cells. *J. Radiat. Res.* **52**, 415–424

Demonstration That the Radical *S*-Adenosylmethionine (SAM) Enzyme PqqE Catalyzes *de Novo* Carbon-Carbon Cross-linking within a Peptide Substrate PqqA in the Presence of the Peptide Chaperone PqqD*[‡]

Received for publication, October 21, 2015, and in revised form, February 18, 2016
 Published, JBC Papers in Press, March 8, 2016, DOI 10.1074/jbc.C115.699918

Ian Barr^{‡§¶}, John A. Latham^{‡¶}, Anthony T. Iavarone^{‡¶},
 Teera Chantarosiri[‡], Jennifer D. Hwang[§],
 and Judith P. Klinman^{‡§¶1}

From the [‡]Department of Chemistry, [§]Department of Molecular and Cell Biology, and [¶]California Institute for Quantitative Biosciences, University of California, Berkeley, California 94720

The radical *S*-adenosylmethionine (SAM) protein PqqE is predicted to function in the pyrroloquinoline quinone (PQQ) biosynthetic pathway via catalysis of carbon-carbon bond formation between a glutamate and tyrosine side chain within the small peptide substrate PqqA. We report here that PqqE activity is dependent on the accessory protein PqqD, which was recently shown to bind PqqA tightly. In addition, PqqE activity *in vitro* requires the presence of a flavodoxin- and flavodoxin reductase-based reduction system, with other reductants leading to an uncoupled cleavage of the co-substrate SAM. These results indicate that PqqE, in conjunction with PqqD, carries out the first step in PQQ biosynthesis: a radical-mediated formation of a new carbon-carbon bond between two amino acid side chains on PqqA.

Pyrroloquinoline quinone (PQQ)² is employed by a wide variety of bacteria, where it functions as a redox cofactor in substrate oxidation via an alternate (non-glycolytic) pathway for production of cellular ATP (1). Hundreds of bacterial species are thought to produce PQQ, based on the presence in their genomes of the strongly conserved *pqq* operon (2) (Fig. 1A). Although the existence of this important redox cofactor has been known for decades, knowledge of the biosynthetic pathway for PQQ has remained scant (3, 4). The PQQ molecule derives from the evolutionarily conserved glutamate and tyro-

sine side chains within a ribosomally produced peptide substrate PqqA (Fig. 1, B and C). To produce PQQ, a large number of chemical modifications are required: (i) the formation of a new carbon-carbon bond between the γ -carbon of glutamate and the 3-position of tyrosine (labeled in *green* in Fig. 1C); (ii) the addition of two additional hydroxyl groups to the tyrosine ring; (iii) a condensation between the 4-OH of tyrosine and the backbone amine of glutamate to produce a heterocyclic ring; and (iv) an 8-electron oxidation catalyzed by PqqC, a cofactor-less enzyme that converts 3a-(2-amino-2-carboxyethyl)-4,5-dioxo-4,5,6,7,8,9-hexahydroquinoline-7,9-dicarboxylic acid (AHQQ) to PQQ in the final step of the pathway (5–7).

The radical SAM enzyme PqqE has been the primary candidate as the catalyst for the formation of the new carbon-carbon bond between glutamate and tyrosine. Radical SAM proteins use the reductive cleavage of *S*-adenosyl methionine to initiate free radical chemistry, and can accomplish a wide variety of reactions, including carbon-carbon bond formation (8, 9). PqqE is a founding member of the SPASM domain-containing radical SAM proteins, which contain one or more auxiliary clusters in their C-terminal regions, and of which several are known to modify small peptides or proteins (10, 11). Recently, a small (~10 kDa) protein, PqqD, has been shown to form a strong, sub-micromolar K_D complex with the peptide substrate PqqA. This complex was further demonstrated to associate with PqqE (12). These findings suggested that PqqD would serve as a chaperone to deliver PqqA to PqqE, functioning as a necessary and heretofore missing component of the subsequent radical-based chemistry. This connects PqqD and the PQQ biosynthesis pathway to a growing class of proteins involved in binding and modifying small, ribosomally produced peptides, of which many are radical SAM proteins (13–15). We now show that PqqE, in association with the complex, PqqA/D, generates a functional catalyst for the cross-linking of glutamate and tyrosine within PqqA.

Experimental Procedures

Methylobacterium extorquens AM1 PqqD, a C-terminal domain of the longer PqqCD gene fusion (12), was cloned into the pGEX-6p-1 vector using the NotI and BamHI restriction sites. Purification was carried out over a glutathione-Sepharose column, followed by cleavage of the GST tag with PreScission protease overnight at 4 °C. The GST tag was removed by passing cleaved protein through a glutathione-Sepharose column. PqqA was made synthetically by CPC Scientific. A non-conserved cysteine was mutated to serine to increase the stability of the peptide. PqqE was expressed in the presence of oxygen, in TB medium supplemented with 100 μ M ferric citrate. The incubation was started at 31 °C for 4 h, and then the temperature was decreased to 20 °C. An hour later, isopropyl-1-thio- β -D-galactopyranoside was added to a final concentration of 100 μ M, and cysteine was added to a concentration of 50 μ M. Cells were harvested 18 h later. UV-visible absorbance spectroscopy of PqqE was performed on a Cary 50 spectrophotometer (spectral bandwidth 1.5 nm), in a septum-sealed cuvette. Flavodoxin A

* This work was supported by National Institutes of Health Grant GM 039296 (to J. P. K.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

[‡] This article was selected as a Paper of the Week.

¹ To whom correspondence should be addressed. E-mail: klinman@berkeley.edu.

² The abbreviations used are: PQQ, pyrroloquinoline quinone; SAM, *S*-adenosylmethionine; FldA, flavodoxin A; FNR, flavodoxin reductase; dA, deoxyadenosine; CID, collision-induced dissociation; MS/MS, tandem mass spectrometry.

REPORT: Carbon-Carbon Bond Formation in PqqA by PqqE Protein

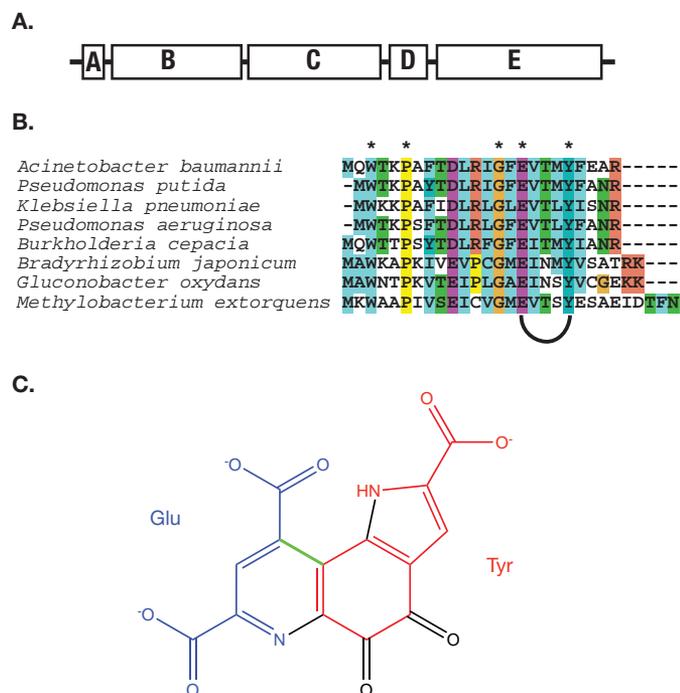


FIGURE 1. A, the *pqq* operon contains, at a minimum, the genes PqqA, PqqB, PqqC, PqqD, and PqqE. PqqF, a protease, is often present as well, but is not uniformly conserved. The gene order is strictly conserved. In *M. extorquens* AM1, there is a gene fusion between PqqC and PqqD. B, alignment of PqqA from PQQ producing organisms. Asterisks denote conserved residues. The conserved glutamate and tyrosine near the C terminus are modified to form PQQ. C, PQQ, showing the origin of the atoms that make its backbone. The bond highlighted in green is the carbon-carbon bond hypothesized to be made by PqqE.

(FldA) and flavodoxin reductase (FNR) from *Escherichia coli* were grown essentially as described previously (16, 17). Purification employed a DEAE-Sephacel column at pH 7.5, 50 mM Tris, with an elution gradient from 50 to 500 mM NaCl over 10 column volumes. The protein was then passed over a preparatory gel filtration column to remove high molecular weight contaminants, exchanged into oxygen-free buffer in the glove box, and frozen in small aliquots. FNR from *M. extorquens* AM1 was cloned from genomic DNA (ATCC 14718) using the primers CCATTCATATGCCCAAGATCACCTTCGT and CCATTCTCGAGTCAGCCCTGGCGGG, and then ligated into the plasmid pET-28a between the NdeI and XhoI restriction sites. This plasmid was used to transform BL21 (DE3) cells, which were grown overnight in LB medium supplemented with 100 μ M riboflavin at 31 $^{\circ}$ C. Protein was purified over a nickel-nitrilotriacetic acid column. *Azotobacter vinelandii* (ATCC 478) FldA was cloned from genomic DNA using the primers GATTGGATCCATGAGCGTAACCATTGTTTATGGGTCC and ATTGCGGCCGCTACATGAGCTGAGCAAGCCATGC and ligated into the pGEX-6p-1 plasmid between the BamHI and NotI restriction sites. This was grown as described for FNR, with the exception that 50 μ M FMN was added to the medium instead of riboflavin. Protein was purified over a glutathione-Sephacel column and cleaved with PreScission protease, and the final product was separated on a size-exclusion column.

For ^{57}Fe labeling of PqqE, the protein was expressed in M9 medium supplemented with 5 g/liter casamino acids and 20 μ M

iron-57 citrate. The casamino acids were passed through a Chelex column prior to addition to broth. Iron-57 was purchased as the iron oxide (Cambridge Isotopes) and reacted with 10 M HCl at 80 $^{\circ}$ C to form $^{57}\text{FeCl}_3$. This was diluted 10-fold and mixed with a 2-fold excess of citric acid. Trace minerals were added, and the medium was supplemented with 10 μ M pyridoxine HCl and 100 μ M cysteine after the addition of isopropyl-1-thio- β -D-galactopyranoside. Purification and reconstitution were carried out as with the native form of PqqE.

Iron and Sulfide Quantification of PqqE—The method of Crack *et al.* (18) was used for both, with ferrozine substituting for ferene in the iron assay. The sample absorbance was compared with an iron standard, and then determined in triplicate. The ferrozine assay was found to be unperturbed by the presence of protein through the method of standard addition. For the sulfide quantification, the standard was generated from a sodium sulfide solution dissolved in 10 mM NaOH. In this case, the method of standard addition was used to correct for interference from the protein (19).

Protein Quantification Assay—PqqE was mixed with 6 M guanidinium HCl plus 0.1 M citric acid and allowed to react for 1 h to fully denature the protein. The citric acid was added to destroy the Fe-S clusters and chelate free iron. The protein was then exchanged into 6 M guanidinium HCl using a PD-10 column, and the concentration of protein was determined by absorbance at 280 nm using the method of Pace *et al.* (20). Protein concentration of the same sample was determined using a Bradford assay where the BSA standard curve was linearized by dividing the 590 nm absorbance by the 450 nm absorbance, according to the method of Zor and Selinger (21). The comparison of these two values provides a conversion factor for PqqE of 0.855 (*i.e.* the true protein concentration is the Bradford-determined value multiplied by 0.855) that was used to correct for systematic error in the Bradford assay. This conversion factor was then used to determine the concentration of subsequent PqqE preparations.

Mössbauer Spectroscopy—Zero-field, ^{57}Fe Mössbauer spectra were recorded in a constant acceleration spectrometer (SEE Co., Edina, MN) at 4 K using a Janis Research Co. cryostat (Woburn, MA). Collected spectra were analyzed with the WMOSS software package (See Co., Edina, MN). Isomer shifts are reported relative to α -iron (27- μ m foil) at room temperature. Samples of PqqE were prepared by freezing solutions in a Teflon sample holder (thickness 0.2 inch) under an inert atmosphere. The sample holder was placed snugly in the sample rod holder and wrapped in Kapton[®] tape prior to introduction into the spectrometer.

Reconstitution of PqqE—The as-isolated PqqE was diluted to 100–200 μ M in a buffer containing 2 mM DTT, 50 mM Tris, pH 7.9, 300 mM KCl, and 10% (v/v) glycerol in a glovebox containing <5 ppm of O_2 . While stirring, aliquots of Na_2S and ammonium iron(II) sulfate were added to the sample in 100 μ M doses every 30 min. The sample was kept at room temperature. The protein was then passed over a PD-10 column to remove excess iron and sulfide, and the protein was diluted 1:1 with water and bound to a DEAE column. Using 50 mM Tris, pH 7.9, 500 mM KCl, 10% glycerol as eluent, a black aggregate remained on the column, while soluble protein eluted as a brown fraction.

Peptide Modification Assay—Assays contained 50 μM PqqE, 50 μM PqqD, 50 μM PqqA, and 500 μM SAM, and were carried out in the glove box. The buffering system was 50 mM Tris, pH 7.9, 200 mM KCl, 10% (v/v) glycerol. For reducing reagents, the assay had either 500 μM sodium dithionite, 500 μM titanium(III) citrate, or a mixture of 1 mM NADPH, 5 μM FNR, and 20 μM FldA. FNR concentration was maintained at a concentration less than FldA concentration to ensure that FldA binding to FNR did not out-compete association of FldA with PqqE. The reaction was quenched with 5% formic acid.

Mass Spectrometric Assay for Conversion of PqqA into Its Cross-linked Product—Samples were quenched with formic acid (5%) and analyzed by an Agilent 1200 LC that was connected in-line with a Thermo LTQ-Orbitrap-XL mass spectrometer equipped with an electrospray ionization source and operated in the positive ion mode. The LC was equipped with a C4 column (Restek), and analytes were eluted using a linear water/acetonitrile gradient. Data acquisition and analysis were performed using the Xcalibur software (version 2.0.7, Thermo). The extent of cross-linking of the PqqA peptide was determined by integrating extracted ion chromatograms for the $[\text{M}+2\text{H}]^{2+}$ ions of cross-linked and unmodified forms of PqqA (occurring at $m/z = 1536.7$ and 1537.7 , respectively). Tandem mass spectrometry (MS/MS) measurements were performed using collision-induced dissociation (CID). Mass-to-charge ratios (m/z) of theoretical fragment ions of a given amino acid sequence were calculated using the MS-Product tool of the ProteinProspector software. Mass envelopes were simulated using the MassXpert software.

Results

As-isolated PqqE Contains Incomplete Fe-S Centers—As reported previously, we have expressed *M. extorquens* AM1 PqqE as a His₆ tag fusion in *E. coli*, and the protein is found to be very soluble and to express in high yield (12). Quantification of iron shows that there are roughly 7–10 iron atoms per polypeptide, which, although varying among enzyme preparations, suggested multiple 4Fe-4S clusters (22). Due to a persistent lack of activity in the as-purified PqqE, we turned to Mössbauer spectroscopy to assess the status of its iron-sulfur centers. The zero-field spectrum of oxidized PqqE shows only two significant doublets: one being a $[\text{4Fe-4S}]^{2+}$ signal ($\delta = 0.478$, $\Delta E_q = 1.214$), comprising 66% of the total iron, and the other being a $[\text{2Fe-2S}]^{2+}$ or $[\text{3Fe-4S}]^+$ signal ($\delta = 0.324$, $\Delta E_q = 0.499$), comprising 33% of the total iron (Fig. 2A). Taken together, this suggests that protein as-isolated contains a mixture of 2Fe-2S and 4Fe-4S clusters as recently reported by Saichana *et al.* (23).

In an effort to obtain active forms of PqqE, we turned to reconstitution of the as-purified *M. extorquens* PqqE, initially adding in 4 eq of iron and sulfide. Following dialysis to remove any adventitiously bound iron, an increase in both the 4Fe-4S Mössbauer signal and the total iron loading is observed, yielding 13.0 ± 0.1 total irons per polypeptide (average of four reconstituted enzyme preparations), whereas sulfide quantification shows 12.2 ± 0.5 sulfurs per polypeptide. This result, together with the fact that PqqE contains seven conserved cysteines in its C-terminal region, suggested the presence of three 4Fe-4S clusters, analogous to the SPASM domain-containing

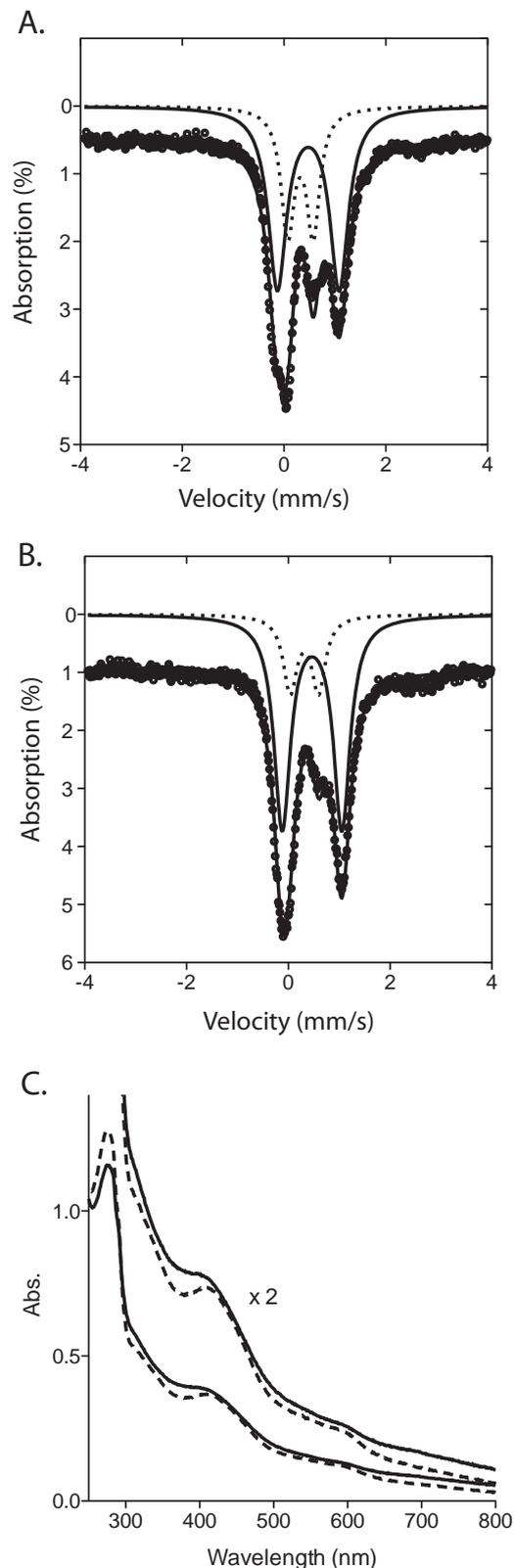


FIGURE 2. Reconstitution of *M. extorquens* PqqE shows a 4Fe-4S protein. A, zero-field Mössbauer spectrum of PqqE at 4 K, as recombinantly expressed in the presence of ^{57}Fe and purified from *E. coli*. Black circles show the raw data, which have been offset from the axis by 0.5% for ease of interpretation. These data are decomposed into 4Fe-4S (solid line) and 2Fe-2S (dotted line) components. B, Mössbauer spectrum of as-isolated PqqE reconstituted with additional ^{57}Fe and sulfide; the 4Fe-4S component comprises 80% of all iron in the sample. C, absorbance (Abs.) spectrum of PqqE before (dashed lines) and after (solid lines) reconstitution of the as-isolated protein.

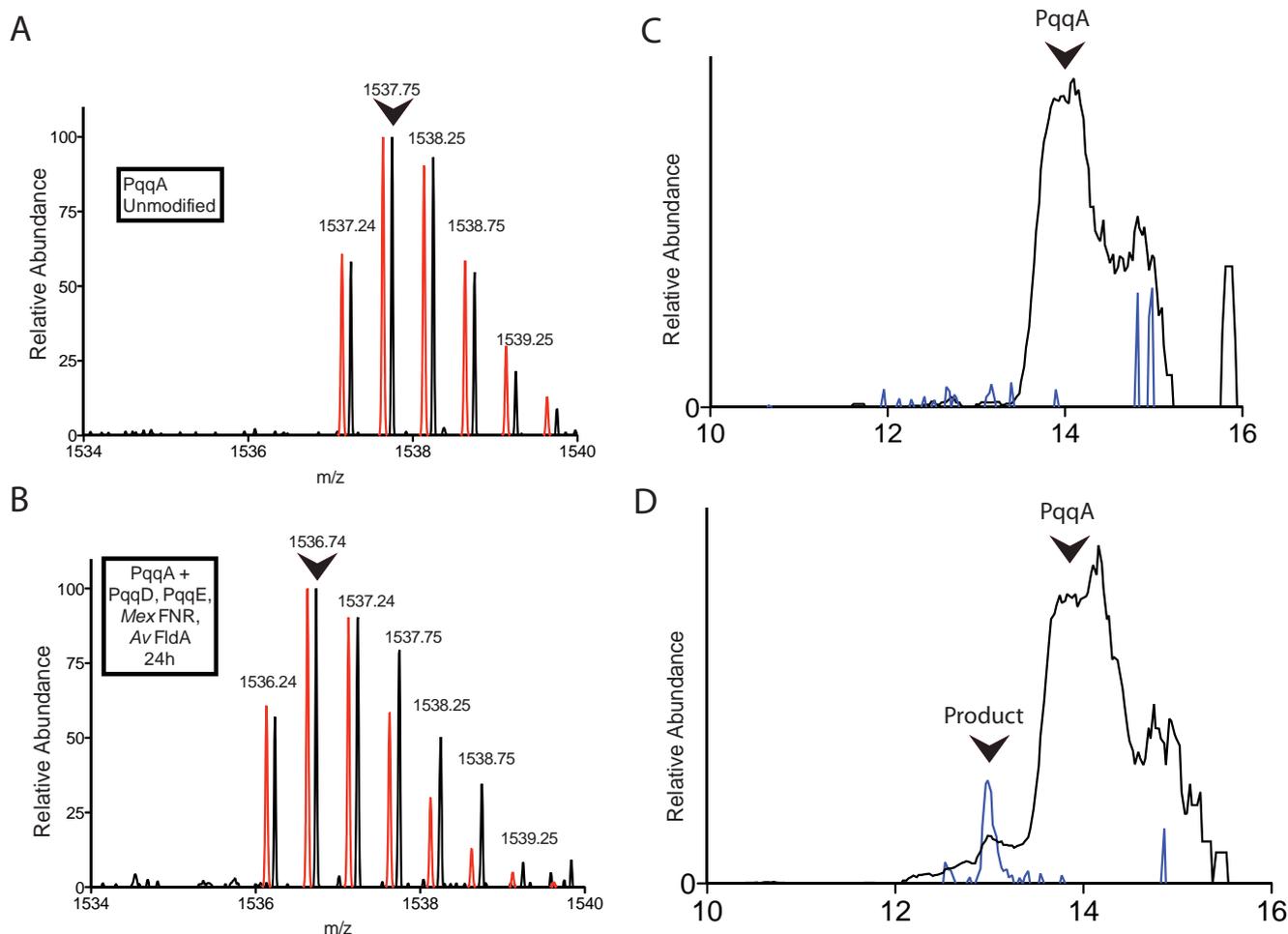


FIGURE 3. Modification of PqqA by PqqE and PqqD monitored by LC-MS. *A*, the initial 2+ ion mass envelope of unmodified PqqA. *B*, the 2+ mass envelope of a minor peak, eluting 1 min earlier, seen following a 24-h reaction under anaerobic conditions. A noticeable shift in mass by 2 Da is observed, consistent with cross-linking of residues in PqqA according to Fig. 5. In both *A* and *B*, the red spectra are calculated mass envelopes for PqqA (*A*) and modified PqqA (*B*), offset slightly for ease of comparison. Arrows indicate the most abundant ions, which were used to quantify the relative amount of modified PqqA. *C*, chromatograph showing the elution profile of 1537.7 (black) and 1536.7 (blue) ions in an unreacted PqqA sample. *D*, chromatograph showing the elution profile of 1537.7 (black) and 1536.7 (blue) ions in a 24-h reaction mixture; a small peak containing cross-linked PqqA is seen to elute earlier than the unreacted PqqA.

radical SAM enzyme anSME described by Drennan and co-workers (25). We then turned to the analysis of Mössbauer spectra of such reconstituted PqqE at 4 K and zero field (Fig. 2*B*). These data show a 4Fe-4S signal that represents ~80% of the total iron detected. The Mössbauer spectrum of the remaining 20% irons implicates either [2Fe-2S]²⁺ or [3Fe-4S]⁺ states; the latter, which has a spectroscopic signature similar to [2Fe-2S]²⁺, could result from loss of a labile iron at one (or more) [4Fe-4S] clusters (24). UV-visible absorbance spectra of the reconstituted PqqE (Fig. 2*C*) display a spectrum more in line with known 4Fe-4S clusters: a broad, flat absorbance shoulder at ~405 nm without the distinct absorbance peaks seen in the as-purified form (Fig. 2*C*). Based on the combined iron binding and spectroscopic data, we conclude that functional PqqE contains three Fe-S centers, at least two of which are present as [4Fe-4S]. Further structural and spectroscopic analyses of PqqE, and their relationship to function, are needed before the structure of the third site can be unambiguously assigned.

A Flavin-containing Reductase Is Necessary to Obtain Peptide Modification by PqqE—It has been observed in several instances that radical SAM enzymes are unable to be activated

using strong reducing reagents such as sodium dithionite as an electron source. The presence of such reagents leads instead to an uncoupled reaction that produces 5'-deoxyadenosine and methionine, but no modification of the substrate. We have previously reported the uncoupled production of 5'-deoxyadenosine (dA) by *Klebsiella pneumoniae* PqqE using sodium dithionite as the reducing reagent (22). As an alternative to such a non-physiological reductant, researchers have employed the flavodoxin (FldA)/flavodoxin reductase (FNR) system from *E. coli*. In this system, NADPH functions as a source of electrons: FNR binds to NADPH and its FAD cofactor is reduced by two electrons; these electrons are then transferred to the FMN cofactor of FldA, and finally to the radical SAM enzyme itself. For the enzyme ThiH, this reduction method was found to be the sole method for productive cleavage of SAM (26), with dithionite leading to uncoupled SAM cleavage as the sole pathway. Bruender *et al.* (27) found that flavodoxin homologues increased the activity of 7-carboxy-7-deazaguanine (CDG) synthase 75-fold, and in this case, flavodoxin was able to accept electrons from sodium dithionite and increase the reaction rate relative to dithionite alone. Notably, the most effective reduc-

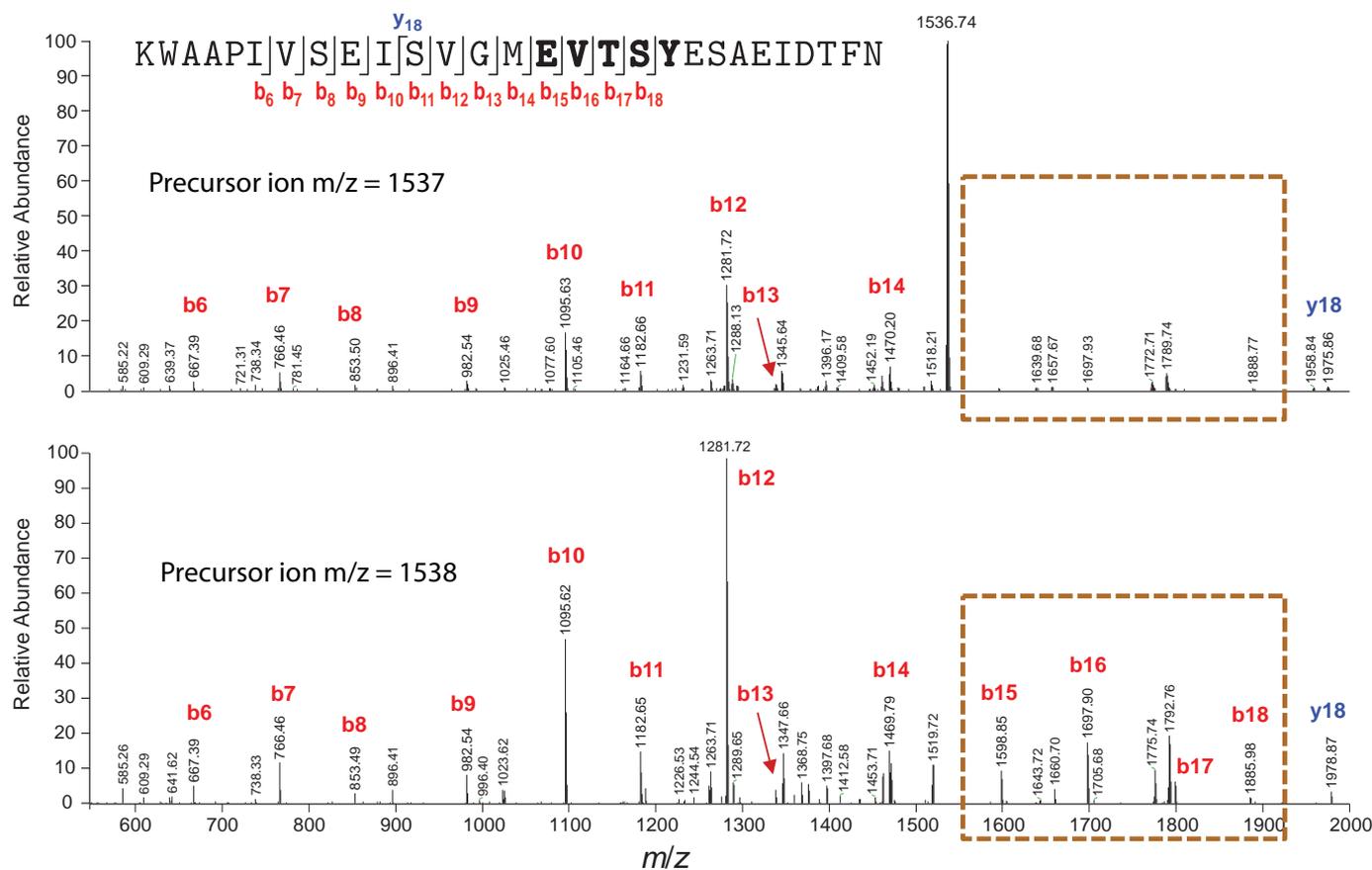


FIGURE 4. MS-MS analysis of modified (top) and unmodified (bottom) PqqA. The modified peptide shows cleavage protection between E and Y; few or none of the fragments detected in the parent PqqA are observed. Mass fragments are labeled according to the sequence of PqqA shown in the top panel. b ions that are not present in the cross-linked sample are outlined in the dashed box.

tion rate came from using the native flavodoxins, suggesting that there may be specific interactions between the radical SAM enzyme and flavodoxin, or that a highly tuned redox potential is required for productive turnover. With these results in mind, we set up reactions between PqqE, PqqD, and PqqA using different reducing reagents that utilize FNR/FldA, sodium dithionite, or titanium(III) citrate as the sole electron source. We examined both the *E. coli* FldA/FNR pair and a second reducing system composed of an FNR from *M. extorquens* FNR and an FldA from *A. vinelandii*. Although a flavodoxin from *M. extorquens* would have been preferable, we were unable to locate an open reading frame for such a gene product in the genome of *M. extorquens*; further, the FNR of *A. vinelandii* is 67% identical to that from *M. extorquens*. LC-MS measurements were used to analyze the products of the enzymatic reactions, to identify any mass changes that occur as a result of modification of PqqA. The most abundant peak of unmodified PqqA appears at $m/z = 1537.7$, with $z = +2$ (Fig. 3A). The elution profile of the unmodified PqqA appears as a broad peak beginning at around 13 min in the chromatogram (Fig. 3C). Upon reaction at room temperature overnight, a new, broader fraction eluted from the LC, which contained a mixture of unmodified and modified PqqA, with the latter observed in the leading edge of the fraction (Fig. 3D), and exhibiting a -2 -Da molecular mass shift relative to unmodified PqqA (Fig. 3B). This mass shift is as expected for a cross-linking reaction

between glutamate and tyrosine in PqqA, in which formation of the new carbon-carbon bond is followed by re-aromatization of the tyrosine ring and concomitant elimination of two hydrogen atoms (see below). Use of either titanium(III) citrate or sodium dithionite as reductant failed to lead to modified PqqA. Use of sodium dithionite resulted in complete formation of dA through an uncoupling reaction, both in the absence and in the presence of PqqA and PqqD. Surprisingly, despite the low yield of modified PqqA (estimated as $\sim 4\%$ at 24 h), no uncoupled production of dA could be detected using FldA/FNR/NADPH as reductant. Reaction mixtures in which PqqD is omitted are unreactive and show no change in PqqA mass relative to starting material. This was also recently found in another radical SAM protein that contains a PqqD-like domain and modifies a small peptide (28). The slow turnover of PqqE may be related to the use of a less than optimal flavodoxin as electron carrier or PqqE inactivation in the course of reaction, and there is clearly room for further optimization. Importantly, these results indicate for the first time that an intramolecular cross-linking reaction occurs between PqqA residues, and that this process is critically dependent on PqqE, PqqD, and the flavodoxin reductase system.

MS/MS Supports Cross-linking of Glu¹⁶ to a C-terminal Side Chain of PqqA—The $[M+2H]^{2+}$ ions of unmodified PqqA and modified PqqA were selected for analysis by MS/MS using CID. CID of peptide ions typically results in cleavage of peptide

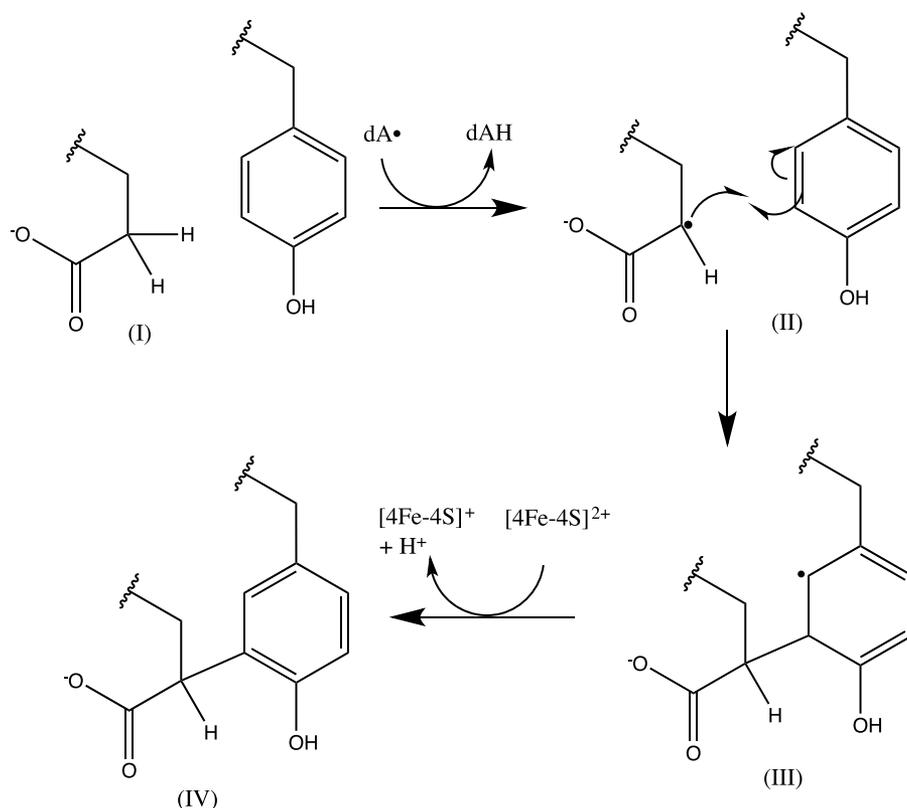


FIGURE 5. Scheme for the formation of PQQ from glutamate and tyrosine of PqqA. PqqA's conserved glutamate undergoes a hydrogen abstraction on the γ -carbon, forming a carbon radical that reacts with the 3-position of tyrosine to form a carbon-carbon bond. At this point, re-aromatization is favored, and the modified tyrosine loses a proton and electron, potentially to either the radical SAM cluster or an auxiliary cluster.

bonds to form b-type and/or γ -type fragment ions (29). The particular fragment ions that are observed in an MS/MS spectrum depend on several factors, including the primary sequence, the precursor ion charge state, the locations of charges (*i.e.* ionizing protons) in the peptide precursor ions, the internal energy of the precursor ions, etc. The MS/MS spectrum of unmodified PqqA exhibits a series of consecutive b-type fragment ions, b_6 through b_{18} (Fig. 4, lower panel). For comparison, in the MS/MS spectrum of modified PqqA, fragment ions b_{15} through b_{18} are observed at markedly lower abundance (highlighted in the orange dashed box in the upper panel of Fig. 4). Formation of fragment ions b_{15} through b_{18} results from cleavage of peptide bonds in the region of PqqA encompassed by Glu¹⁵ and Tyr¹⁹, and separation of the resulting, complementary N-terminal and C-terminal fragments. In the case of a modified PqqA, formation of fragment ions b_{15} through b_{18} is hindered by the covalent, intramolecular cross-link between Glu¹⁵ and Tyr¹⁹, which prevents detection of fragments resulting from peptide bond cleavage within this region of the peptide. Moreover, in the MS/MS spectrum of unmodified PqqA, the b_{12} fragment ion ($m/z = 1281.72$) is the base peak (*i.e.* the peak of highest abundance) and residual precursor ion is not observed, indicating that the precursor ion of unmodified PqqA was completely fragmented under CID. In contrast, the base peak in the MS/MS spectrum of the modified PqqA is residual precursor ion ($m/z = 1536.74$), indicating incomplete fragmentation of the modified PqqA under CID. The collision energy settings used for CID of unmodified and modified PqqA were identical. The incomplete fragmentation of modified PqqA is

consistent with a modification that stabilizes the precursor peptide ions against fragmentation. In summary, the incomplete precursor ion fragmentation and lowered abundance of fragment ions resulting from peptide backbone cleavage within the region between Glu¹⁵ and Tyr¹⁹ strongly support the presence of a covalent cross-link between Glu¹⁵ and Tyr¹⁹ in the modified PqqA.

Discussion

Among the three major classes of enzymatic redox cofactors (flavins, nicotinamides, and quinones), there has been a singular lack of understanding regarding the biosynthetic pathway for PQQ production. The operon encoding PQQ biosynthesis (Fig. 1A) was identified more than 25 years ago (4, 30), and more recent bioinformatics studies demonstrated a conservation of gene order that includes an occasional gene fusion between PqqD with either the C terminus of PqqC or the N terminus of PqqE (2). However, the description of function for each gene product from this operon has lagged considerably, with PqqC representing the only enzyme within the pathway successfully characterized with regard to structural, spectroscopic, and kinetic properties (5, 6, 31). Although PqqE had been shown to be a radical SAM enzyme (22), it had not been possible to demonstrate a reaction between the peptide substrate PqqA and PqqE, suggesting either that PqqA must be modified before it can interact with PqqE or that key components of the PqqE system were missing (7). In a recent breakthrough, PqqD was shown to bind tightly to PqqA and to enter into a ternary complex with PqqE (12). The resulting suggestion that PqqD was a

missing component for productive PqqE activity is verified in the present study. Two additional features of a functional PqqE not previously defined are the requirements for reconstitution of enzyme with additional iron equivalents, as well as the uncovering of suitable reductant systems that can support coupling between SAM cleavage and the subsequent functionalization of PqqA.

With this study, both the initial step in PQQ production, catalyzed by PqqE, and the ultimate step, catalyzed by PqqC, have been identified. Further, with a function assigned to PqqD, PqqB remains the sole gene product without any clearly established catalytic role. Although hints to the function of PqqE came with its demonstration as a radical SAM enzyme, it was uncertain whether the reaction would involve simply the formation of a new carbon-carbon bond formation or, perhaps, more drastic modifications. Based on the present work and its relationship to the recently published work on the StrB protein, a radical SAM enzyme that catalyzes carbon-carbon bond formation between lysine and tryptophan residues in a small, ribosomally produced peptide (32), we conclude that the principal role of PqqE is to catalyze formation of a new carbon-carbon bond between the γ -carbon of a glutamate and the ring of a tyrosine residue (Fig. 5). A great deal of fascinating biochemistry remains to be determined that includes determining the manner in which PqqD controls the interaction between PqqA and PqqE, the catalytic function of PqqB, and the process that liberates the cross-linked Glu-Tyr product from its parent peptide. The data presented herein, demonstrating that the initial step in PQQ biosynthesis is carried out by the combined action of PqqE and PqqD, provide a critical clarification of the mechanistic underpinnings of PQQ production as well as a defined direction for future studies of this unique member of the family of ribosomally encoded, post-translationally modified peptides (13–15).

Author Contributions—I. B. and J. P. K. designed the study and the experiments. I. B., J. A. L., A. T. I., T. C., and J. D. H. performed and analyzed the experiments. I. B. and J. P. K. wrote the manuscript, and A. T. I. and T. C. revised the manuscript and added experimental details. All authors analyzed the results and approved the final version of the manuscript.

Acknowledgments—We thank Prof. Michael Waterman (Vanderbilt University) for providing the *E. coli* FNR/FLD plasmids and Prof. Chris Chang (University of California, Berkeley) for use of the Mössbauer spectrometer. A mass spectrometer was acquired with National Institutes of Health support (1S10RR022393-01).

References

- Duine, J. A. (1999) The PQQ story. *J. Biosci. Bioeng.* **88**, 231–236
- Shen, Y.-Q., Bonnot, F., Imsand, E. M., RoseFigura, J. M., Sjölander, K., and Klinman, J. P. (2012) Distribution and properties of the genes encoding the biosynthesis of the bacterial cofactor, pyrroloquinoline quinone. *Biochemistry* **51**, 2265–2275
- van Kleef, M. A., and Duine, J. A. (1988) A search for intermediates in the bacterial biosynthesis of PQQ. *Biofactors* **1**, 297–302
- Goosen, N., Vermaas, D. A., and van de Putte, P. (1987) Cloning of the genes involved in synthesis of coenzyme pyrrolo-quinoline-quinone from *Acinetobacter calcoaceticus*. *J. Bacteriol.* **169**, 303–307
- Magnusson, O. T., Toyama, H., Saeki, M., Rojas, A., Reed, J. C., Liddington, R. C., Klinman, J. P., and Schwarzenbacher, R. (2004) Quinone biogenesis: structure and mechanism of PqqC, the final catalyst in the production of pyrroloquinoline quinone. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7913–7918
- Bonnot, F., Iavarone, A. T., and Klinman, J. P. (2013) Multistep, eight-electron oxidation catalyzed by the cofactorless oxidase, PqqC: identification of chemical intermediates and their dependence on molecular oxygen. *Biochemistry* **52**, 4667–4675
- Klinman, J. P., and Bonnot, F. (2014) The intrigues and intricacies of the biosynthetic pathways for the enzymatic quinocofactors: PQQ, TTQ, CTQ, TPQ and LTQ. *Chem. Rev.* **114**, 4343–4365
- Shisler, K. A., and Broderick, J. B. (2012) Emerging themes in radical SAM chemistry. *Curr. Opin. Struct. Biol.* **22**, 701–710
- Broderick, J. B., Duffus, B. R., Duschene, K. S., and Shepard, E. M. (2014) Radical S-adenosylmethionine enzymes. *Chem. Rev.* **114**, 4229–4317
- Haft, D. H., and Basu, M. K. (2011) Biological systems discovery *in silico*: radical S-adenosylmethionine protein families and their target peptides for posttranslational modification. *J. Bacteriol.* **193**, 2745–2755
- Grell, T. A. J., Goldman, P. J., and Drennan, C. L. (2015) SPASM and Twitch domains in AdoMet radical enzyme structures. *J. Biol. Chem.* **290**, 3964–3971
- Latham, J. A., Iavarone, A. T., Barr, I., Juthani, P. V., and Klinman, J. P. (2015) PqqD is a novel peptide chaperone that forms a ternary complex with the radical S-adenosylmethionine protein PqqE in the pyrroloquinoline quinone biosynthetic pathway. *J. Biol. Chem.* **290**, 12908–12918
- Burkhart, B. J., Hudson, G. A., Dunbar, K. L., and Mitchell, D. A. (2015) A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nat. Chem. Biol.* **11**, 564–570
- Link, A. J. (2015) Biosynthesis: leading the way to RiPPs. *Nat. Chem. Biol.* **11**, 551–552
- Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., *et al.* (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160
- Jenkins, C. M., Pikuleva, I., Kagawa, N., and Waterman, M. R. (1997) *Escherichia coli* flavodoxin Sepharose as an affinity resin for cytochromes P450 and use to identify a putative cytochrome P450c17/3 β -hydroxysteroid dehydrogenase interaction. *Arch. Biochem. Biophys.* **347**, 93–102
- Jenkins, C. M., and Waterman, M. R. (1998) NADPH-flavodoxin reductase and flavodoxin from *Escherichia coli*: characteristics as a soluble microsomal P450 reductase. *Biochemistry* **37**, 6106–6113
- Crack, J. C., Green, J., Thomson, A. J., and Le Brun, N. E. (2014) Techniques for the production, isolation, and analysis of iron-sulfur proteins. *Methods Mol Biol.* **1122**, 33–48
- Bader, M. (1980) A systematic approach to standard addition methods in instrumental analysis. *J. Chem. Educ.* **57**, 703
- Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**, 2411–2423
- Zor, T., and Selinger, Z. (1996) Linearization of the Bradford protein assay increases its sensitivity: theoretical and experimental studies. *Anal. Biochem.* **236**, 302–308
- Weckslers, S. R., Stoll, S., Tran, H., Magnusson, O. T., Wu, S.-P., King, D., Britt, R. D., and Klinman, J. P. (2009) Pyrroloquinoline quinone biogenesis: demonstration that PqqE from *Klebsiella pneumoniae* is a radical S-adenosyl-L-methionine enzyme. *Biochemistry* **48**, 10151–10161
- Saichana, N., Tanizawa, K., Pechoušek, J., Novák, P., Yakushi, T., Toyama, H., and Frébortová, J. (2016) PqqE from *Methylobacterium extorquens* AM1: a radical S-adenosyl-L-methionine enzyme with an unusual tolerance to oxygen. *J. Biochem.* **159**, 87–99
- Pandelia, M.-E., Lanz, N. D., Booker, S. J., and Krebs, C. (2015) Mössbauer spectroscopy of Fe/S proteins. *Biochim. Biophys. Acta* **1853**, 1395–1405
- Goldman, P. J., Grove, T. L., Sites, L. A., McLaughlin, M. I., Booker, S. J., and Drennan, C. L. (2013) X-ray structure of an AdoMet radical activase

REPORT: Carbon-Carbon Bond Formation in PqqA by PqqE Protein

- reveals an anaerobic solution for formylglycine posttranslational modification. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8519–8524
26. Chatterjee, A., Hazra, A. B., Abdelwahed, S., Hilmey, D. G., and Begley, T. P. (2010) A “radical dance” in thiamin biosynthesis: mechanistic analysis of the bacterial hydroxymethylpyrimidine phosphate synthase. *Angew. Chem. Int. Ed. Engl.* **49**, 8653–8656
27. Bruender, N. A., Young, A. P., and Bandarian, V. (2015) Chemical and biological reduction of the radical SAM enzyme CPH₄ synthase. *Biochemistry* **54**, 2903–2910
28. Wieckowski, B. M., Hegemann, J. D., Mielcarek, A., Boss, L., Burghaus, O., and Marahiel, M. A. (2015) The PqqD homologous domain of the radical SAM enzyme ThnB is required for thioether bond formation during thurincin H maturation. *FEBS Lett.* **589**, 1802–1806
29. Roepstorff, P., and Fohlman, J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **11**, 601
30. Meulenber, J. J. M., Loenen, W. A. M., Sellink, E., and Postma, P. W. (1989) The role of PQQ in *K. aerogenes* and cloning of *pqq* genes. in *PQQ and Quinoproteins: Proceedings of the First International Symposium on PQQ and Quinoproteins, Delft, The Netherlands, 1988* (Jongejan, J. A., and Duine, J. A. eds), pp. 187–189, Springer Netherlands, Dordrecht, Netherlands, 10.1007/978-94-009-0957-1_28
31. Schwarzenbacher, R., Stenner-Liewen, F., Liewen, H., Reed, J. C., and Lidington, R. C. (2004) Crystal structure of PqqC from *Klebsiella pneumoniae* at 2.1 Å resolution. *Proteins* **56**, 401–403
32. Schramma, K. R., Bushin, L. B., and Seyedsayamdost, M. R. (2015) Structure and biosynthesis of a macrocyclic peptide containing an unprecedented lysine-to-tryptophan crosslink. *Nat. Chem.* **7**, 431–437

Biochemical Analysis of Yeast Suppressor of Ty 4/5 (Spt4/5) Reveals the Importance of Nucleic Acid Interactions in the Prevention of RNA Polymerase II Arrest*[♦]

Received for publication, January 15, 2016, and in revised form, February 26, 2016. Published, JBC Papers in Press, March 4, 2016, DOI 10.1074/jbc.M116.716001

J. Brooks Crickard[‡], Jianhua Fu[§], and Joseph C. Reese^{†1}

From the [‡]Department of Biochemistry and Molecular Biology, Center for Eukaryotic Gene Regulation, Penn State University, University Park, Pennsylvania 16802 and the [§]Department of Biochemistry, Medical College of Wisconsin, Milwaukee, Wisconsin 53226.

RNA polymerase II (RNAPII) undergoes structural changes during the transitions from initiation, elongation, and termination, which are aided by a collection of proteins called elongation factors. NusG/Spt5 is the only elongation factor conserved in all domains of life. Although much information exists about the interactions between NusG/Spt5 and RNA polymerase in prokaryotes, little is known about how the binding of eukaryotic Spt4/5 affects the biochemical activities of RNAPII. We characterized the activities of Spt4/5 and interrogated the structural features of Spt5 required for it to interact with elongation complexes, bind nucleic acids, and promote transcription elongation. The eukaryotic specific regions of Spt5 containing the Kyrpides, Ouzounis, Woese domains are involved in stabilizing the association with the RNAPII elongation complex, which also requires the presence of the nascent transcript. Interestingly, we identify a region within the conserved NusG N-terminal (NGN) domain of Spt5 that contacts the non-template strand of DNA both upstream of RNAPII and in the transcription bubble. Mutating charged residues in this region of Spt5 did not prevent Spt4/5 binding to elongation complexes, but abrogated the cross-linking of Spt5 to DNA and the anti-arrest properties of Spt4/5, thus suggesting that contact between Spt5 (NGN) and DNA is required for Spt4/5 to promote elongation. We propose that the mechanism of how Spt5/NGN promotes elongation is fundamentally conserved; however, the eukaryotic specific regions of the protein evolved so that it can serve as a platform for other elongation factors and maintain its association with RNAPII as it navigates genomes packaged into chromatin.

The conversion of DNA to RNA is a fundamental aspect of all life, and this process is carried out by RNA polymerases (RNAPs).² These enzymatic powerhouses must maintain both

high levels of fidelity and processivity over long distances to ensure that RNAs are accurately produced on a time scale amenable to life. Families of proteins called elongation factors have evolved to assist RNA polymerases during transcription elongation. The oldest and most conserved of these factors is the NusG/suppressor of Ty element (Spt) 5 family (1, 2). NusG is the eubacterial version of Spt5 and functions as a single polypeptide; however, archaea and eukaryotic Spt5 associate with an additional small protein, Spt4. In yeast, *SPT5* is essential, but *SPT4* is not. Deleting the gene encoding Spt4 impairs elongation, transcription-coupled repair, and mRNA processing (2–5). Some of the functions of Spt4 may be partially dependent on its ability to prevent degradation of Spt5 in cells (4).

The NusG/Spt5 family of proteins has been shown to enhance RNA polymerase transcription elongation in all domains of life (6–10). NusG regulates RNAP activity by stabilizing the post-translocated state thereby inhibiting backtracking and reducing long lifetime pauses (6, 11). The NusG homolog RfaH has also been implicated in regulating movement of the RNAP bridge helix suggesting that NusG and RfaH may function to alter RNAP conformational dynamics (6, 12). In fact, the movement of the trigger loop and bridge helix in the active site is a fundamental process in nucleotide incorporation and regulates arrest of active elongation complexes in both prokaryotes and eukaryotes (12–15). In prokaryotes, the movement of the trigger loop and bridge helix is linked to the formation of RNA hairpins, which regulate RNAP pausing (13, 16, 17). Although this method of pausing is not known to exist in eukaryotes, x-ray crystal structures of yeast RNAPII in different stages of elongation have generated a model in which the movement of the bridge helix and trigger loop can be coupled to translocation through the non-template strand of DNA (18). This information implies that the nucleic acid scaffold is critical in maintaining active RNAP during elongation.

Using a combination of crystal structures, cryo-EM, and model building of archaeal Spt4/5 bound to RNAP, it has been proposed that Spt4/5 closes the “crab claw”-like clamp of RNAP (8, 10) by binding across the jaws and interacting with a coiled-coil domain of RpoA (7). This binding may prevent the dissoci-

* This work was supported by National Institutes of Health Grants GM058672 (to J. C. R.) and GM100997 (to J. F.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

[♦] This article was selected as a Paper of the Week.

¹ To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biology, Center for Eukaryotic Gene Regulation, Penn State University, 463A North Frear, University Park, PA 16802. Tel.: 814-865-1976; Fax: 814-863-7024; E-mail: Jcr8@psu.edu.

² The abbreviations used are: RNAP, RNA polymerase; RNAPII, RNA polymerase II; RNAPI, RNA polymerase I; CTD, C-terminal region of Rpb1; TFIIIS, transcription factor II S; EC, elongation complex; CTR, C-terminal region of

Spt5; KOW, Kyrpides, Ouzounis, Woese; NGN, NusG N terminus; DSIF, 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole sensitivity-inducing factor; NTS, non-template strand; FRB, FKBP12-rapamycin-binding domain of human mTor; NusG, N utilization substance G; H2B, histone 2B; ExoIII, exonuclease III; oligo, oligonucleotide; PDB, Protein Data Bank.

Spt4/5 Interaction with DNA Prevents Arrest of RNAPII

ation of the RNA polymerase from the template by encircling DNA and enhancing processivity (7, 8, 10). “Bridging” of the two lobes of RNAP occurs through the universally conserved NGN domain (NusG N-terminal region), and all known biomechanical properties of Spt5 are linked to this domain (1, 7, 19, 20).

Much of what we know about the biochemical activities of eukaryotic Spt4/5 arose from its identification as the DRB-sensitive inducing factor (DSIF) in HeLa cell extracts and was later found to be required to stably pause RNAPII in promoter proximal regions (9, 21). In this latter case, DSIF acts as a negative elongation factor working with the negative elongation factor (22). Conversion to a positive elongation factor requires the phosphorylation of Spt5 by positive transcription elongation factor \bar{h} (P-TEFb) (23). Human Spt5 suppresses the arrest of RNAPII at poly(A) tracts (24), and the zebrafish version of Spt5 stimulated transcription elongation in extracts (25). In these two examples, eukaryotic Spt4/5 has been shown to function in a positive manner to support transcription elongation. Because the ability to induce promoter proximal pausing is unique to higher eukaryotes, much of the focus to understand DSIF function has been carried out in metazoans using depleted extracts or crude fractions. What is lacking is a highly defined biochemical reconstitution system to study eukaryotic Spt4/5 that can provide a deeper mechanistic understanding of how it affects RNAPII activity and promotes elongation. Studying yeast Spt4/5 provides an opportunity to understand the positive effect of Spt4/5 on transcription elongation, as yeast lacks promoter proximal pausing. Furthermore, genetic assays indicate that Spt4/5 functions in a purely positive function (3, 26, 27).

In prokaryotes, NusG of eubacteria and Spt4/5 from archaea bind to RNAP (7, 20). However, the association between fly and human DSIF and RNAPII is facilitated by the emergence of the nascent transcript (28, 29). Eukaryotic Spt5 contains an extended C terminus containing five Kyrpides, Ouzounis, Woese (KOW) domains that theoretically can reach anywhere on the surface of RNAPII. KOW domains are generally part of larger Tudor domains that are known to interact with nucleic acids (30, 31). In fact, a crystal structure of an isolated KOW1-linker (K1L1) fragment of Spt5 and biochemical analysis using isolated K1L1 provided evidence that it binds to nucleic acids (32). Furthermore, *in vivo* cross-linking data conducted on yeast Spt5 revealed that KOW domains 4 and 5 interact with surfaces on RNAPII, including the Rpb4/7 sub-module. Although more information is being gained on the functions of these KOW domains, how they affect RNAPII binding and activity is not clear. Eukaryotic Spt5 also contains an extended intrinsically disordered modifiable C-terminal region (CTR) that has been shown to be phosphorylated during transcription, which in turn recruits elongation and RNA-processing factors (4, 33–37). This raised the possibility that the elongation activity associated with Spt5 may have been delegated to other elongation factors during the course of evolution. The CTR of Spt5 does not contribute to the essential function of the protein because a CTR-less mutant of Spt5 is viable and displays minor phenotypes (36, 37); however, systematic deletion of KOW domains and the linker regions of Spt5 in yeast are lethal and result in reduced affinity for both RNAPI and RNAPII (38).

Spt4/5 likely associates with several regions of RNAPII and nucleic acids in the elongation complex. The consequences of these interactions with polymerase are not well understood. We show that the emerging RNA transcript is critical for the binding of yeast Spt4/5 to the elongation complex. Additionally, eukaryotic specific regions of Spt5 are required for stable binding to RNAPII elongation complexes. Furthermore, we provide direct evidence that interaction of the NGN domain of Spt5 with the non-template strand of DNA in the elongation complex is required for Spt4/5 to promote elongation. Together, these findings suggest a eukaryotic specific mode of binding to polymerase but a universally conserved mechanism of action to promote elongation. It also suggests an elongation promoting activity for Spt4/5 different from its ability to encircle the DNA and retain RNAP on DNA.

Experimental Procedures

Plasmid Construction—The coding sequences of *SPT4* and *SPT5* were amplified by PCR from genomic DNA with primers that introduced BamHI and BsrGI or BamHI and NgoMIV sites into the ends, respectively. The products were then ligated into the shuttle vectors pST50-N-term-Strep tag (39) and pST66-C-terminal His₆ tag. The expression modules were then excised using unique restriction sites and ligated into the pST69 co-expression plasmid (a kind gift from Song Tan). Details are available upon request. The SPT5-HA₃ sequence was amplified by PCR from pHQ1494 (a kind gift from Allen Hinnebusch) and then inserted into pRS313 plasmids using the In-fusion cloning method (Clontech). Mutations were constructed using the In-fusion mutagenesis procedure. Constructs were confirmed by sequencing.

Yeast Strains—Homologous recombination was used to fuse the FRB sequences in-frame with the endogenous *SPT5* coding sequence in the anchor-away strain Y40343 (Euroscarf) as described previously (40). The strain, JR1734, genotype is *MAT α tor1-1 fpr1::NAT RPL13A-2xFKBP12::TRP1 SPT5xFRB::KANMX*. Spotting of strains was performed by growing yeast cells transformed with pRS313 containing SPT5-HA₃, and its mutant derivatives, to an OD of 1.0 and then spotting 10-fold serial dilutions onto synthetic media (–HIS) plates \pm 1 μ g/ml rapamycin. For preparing extracts for Western blotting, cells were grown to an A_{600} of 0.7–0.8; rapamycin was added to 1 μ g/ml, and after 1.5 h the cells were harvested, and whole cell extracts were prepared using the trichloroacetic acid method (41). Whole cell extract was then run on SDS-PAGE and subjected to Western blotting with anti-HA (Pierce) or anti-H2B (Active Motif) antibodies. Signals were detected using chemiluminescence detected on x-ray film.

Purification of Recombinant Spt4/5 and Yeast RNAPII—BL21 codon plus cells transformed with a plasmid co-expressing Strep-Spt4 and Spt5-His₆ were grown to an OD of 0.4. Spt4/5 was induced by the addition of 0.1 mM isopropyl 1-thio- β -D-galactopyranoside and 0.3 mM ZnSO₄ overnight at 18 °C. Cells were then harvested, washed in lysis buffer (20 mM Tris-Cl, pH 8.0, 500 mM KCl, 0.010 mM ZnCl₂, 5 mM imidazole, and 10% glycerol), lysed by adding lysozyme, and sonicated. The lysate was cleared by centrifugation at 14,000 rpm and incubated in Talon resin (Clontech) for 1 h at 4 °C. Resin was

washed with lysis buffer and then eluted in lysis buffer containing 200 mM imidazole. The peak fractions were pooled and diluted to 200 mM salt in heparin buffer (20 mM Tris-Cl, pH 6.8, 100 mM NaCl, 8% glycerol). Protein was purified on heparin-agarose using step elution fractions of 0.5, 1.0, 1.5, and 2.0 M NaCl. Peak fractions were pooled and diluted to 100 mM NaCl and then passed over a Q-Sepharose column. Protein was eluted with step gradients from 200 to 800 mM KCl in 100 mM steps in Q buffer (20 mM HEPES, pH 7.5, 5 mM MgCl₂, 1 mM β-mercaptoethanol, 0.01 mM ZnCl₂, 10% glycerol). Peak fractions were pooled and dialyzed against 20 mM HEPES, pH 7.5, 200 mM KCl, 5 mM MgCl₂, 1 mM β-mercaptoethanol, 0.01 mM ZnCl₂, 10% glycerol. Protein was stored at -80 °C. Yeast RNA-Pol II was purified as described previously (42).

Elongation Complex Reconstitution and Runoff Experiments—DNA templates for EC42 runoff experiments were prepared by annealing two chemically synthesized oligos corresponding to the non-template and template strands, producing a template with a 3' overhang. The DNA template was gel-purified. The sequences are as follows: EC42 template strand, 5'-GCCACCGCGGTCTAGAGGATCCCCGGGAGTGG-AATGAGAAATGAGTGTGAAGATAGAGGAGAGATCA-AAAAAATTA-3', and EC42 non-template strand, 5'-CTC-CTCTATCTTCACACTCATTTCTCATTTCCACTCCCCGGG-GATCCTCTAGACCGCGGTGGC-3'. EC70 templates were prepared using a PCR and tail ligation strategy (28, 43, 44)

Elongation complexes (ECs) were formed as described previously (28, 43, 44). Briefly, reactions were formed in 15 μl of transcription buffer (50 mM HEPES, pH 7.5, 100 mM KCl, 1 mM MnCl₂, and 0.5 mM DTT, 10% glycerol, 0.5 mM UpG, 20 units of RNasin (Promega), 100 ng/μl BSA). Approximately 150 fmol of RNAPII was pre-incubated with DNA template, and then 5 μl of 0.1 mM ATP, 0.1 mM CTP, 0.005 mM UTP, and 8 μCi of [α -³²P]UTP (6000 Ci/mmol) was added. For the anti-arrest assays, ECs were formed on EC42 templates in the presence of 0.1 mM ATP, 0.1 mM CTP, 0.005 mM UTP, and 8 μCi of [α -³²P]UTP (6000 Ci/mmol) for 2.5 min, and then Spt4/5 was added to the reaction along with 100 μM UTP and 250 ng/μl salmon sperm DNA. ECs were then incubated for 5, 7.5, 10, and 20 min. After each time point, ECs were chased with 100 μM ATP, CTP, and GTP for 10 min. Reactions were terminated by transcription stop mix (20 mM EDTA, pH 8.0, 200 mM NaCl, 1% SDS, 0.5 μg/ml yeast total RNA). Samples were treated with 1 μg/ml proteinase K, and the RNA was purified by phenol/chloroform/isoamyl alcohol (25:24:1) extraction and ethanol precipitation. RNA was resuspended in formamide dye solution (90% formamide, 1 mM EDTA, bromophenol blue, and xylene cyanol), heated, separated on 10% denaturing polyacrylamide gels, dried, and exposed to a phosphorimaging screen. Images were recorded using a Typhoon system (GE Healthcare) and analyzed using ImageJ. Percent (%) active complex was calculated by the following formula: % RNA runoff = (full-length RNA + EC RNA), plotted on the y axis versus time (x axis), and the data were then fit to an exponential decay curve for comparison using Excel.

RNA Cross-linking and Footprinting—RNA cross-linking was performed as described previously (28, 43), except that 3-O-MeGTP was omitted. Briefly, ECs were formed on DNA tem-

plates using ~500 fmol of RNAPII in transcription buffer with 0.03 mM bromo-UTP, 0.03 mM ATP, 0.005 mM CTP, and 4 μCi of [α -³²P]CTP (6000 Ci/mmol). ECs were formed for 20 min, and then Spt4/5 was added with 500 ng of competitor yeast total RNA. Samples were then UV-irradiated for 10 min at 308 nm, and 1 μg of RNase A and 10 units of DNase I (Roche Applied Science) were added and the samples incubated at 37 °C for 1 h. Samples were resolved on SDS-PAGE, dried, and exposed to a phosphorimaging screen.

For RNase protection, footprinting experiments were conducted on ECs formed on DNA templates immobilized on streptavidin M-280 Dynabeads. ECs were formed in transcription buffer using ~100 fmol of RNAPII plus 0.1 mM ATP, 0.1 mM CTP, 0.1 mM for 20 min. Nucleotides and free protein were removed by magnetic collection and three washes in wash buffer (20 mM HEPES, pH 7.5, 100 mM KCl, 1 mM DTT, 0.02% Nonidet P-40, 100 ng/μl BSA). Afterward, the beads were resuspended in reaction buffer (20 mM HEPES, pH 7.5, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT, 0.01 mM ZnCl₂, 100 ng/μl BSA, 10% glycerol), and [α -³²P]GTP (6000 Ci/mmol) was added and incubated for 5 min to end label the RNA, and then cold GTP was added to 100 μM and incubated for an additional 5 min. The complexes were then collected and washed three times with wash buffer and resuspended in reaction buffer. Spt4/5 was then titrated into the binding buffer, and RNase I (New England Biolabs) was added to 1 unit/μl and incubated for 3 min. The reaction was terminated by the addition of transcription stop mixture, and then phenol/chloroform/isoamyl alcohol (25:24:1) was extracted and ethanol-precipitated. Samples were resuspended in formamide dye and run on 15% denaturing polyacrylamide gels.

DNA Gel Shift and ExoIII Footprinting—Transcription templates for EMSA and ExoIII nuclease footprinting were generated by annealing complementary gel-purified oligos. The top strand was end-labeled with [³²P]ATP using polynucleotide kinase. EC formation was performed as described above, except no radiolabeled nucleotides were added, and 10 nM RNAPII for ExoIII footprinting the scale of the reaction was increased to 100 nM RNAPII. ECs (±Spt4/5) were loaded on a 4.5% native PAGE in 0.5× TBE and electrophoresed at room temperature for 2 h. For ExoIII footprinting, ExoIII was added to a final concentration of 10 units/μl and allowed to digest for varying amounts of time. Samples were quenched with 20 mM Tris-Cl, pH 7.5, 200 mM NaCl, 1% SDS, and 100 ng/μl single-stranded DNA. Samples were then treated with proteinase K; phenol/chloroform/isoamyl (25:24:1) was extracted, ethanol-precipitated, resuspended, and run on a 10% denaturing polyacrylamide gel. Quantification of the ExoIII footprinting was performed by taking the sum total of the bands comprising the RNAPII footprint at nucleotides +18, +17, +16, and +15. The signal in the free DNA digestion samples was considered background and was subtracted out. The sum of these four bands was then calculated, and each individual band was divided by the total (+18/Σ+18,+17,+16,+15), and this generated a fraction of the total population.

Immobilization of RNAPII and KMnO₄ Footprinting—Protein A-Sepharose magnetic beads (GE Healthcare) were blocked with 20 mM Na-HEPES, pH 7.8, 150 mM NaCl, 10%

Spt4/5 Interaction with DNA Prevents Arrest of RNAPII

glycerol, 250 ng/ μ l BSA and then bound to 8WG16 IgG for 1 h at room temperature. Beads were then washed three times with 10 bed volumes of bead wash buffer (50 mM Tris-HCl, pH 7.4, 250 mM KCl, and 100 ng/ μ l BSA) and then three times with IgG Wash buffer (50 mM Tris-HCl, pH 7.4, 100 mM NaCl). RNAPII was then immobilized through the CTD of Rpb1 on protein A magnetic beads in transcription buffer for 1 h at 4 °C, washed three times with IgG wash buffer, and finally resuspended in transcription buffer. End-labeled DNA was added to form ECs for 20 min in the presence of 100 μ M CTP, ATP, and UTP. Beads were washed two times with 20 mM HEPES, pH 7.5, 500 mM KCl, 5 mM MgCl₂, 0.001 mM ZnCl₂, 10% glycerol, and 100 ng/ μ l BSA and then two times with the same buffer except that the KCl was reduced to 100 mM. The immobilized ECs were resuspended in transcription reaction buffer supplemented with 2 mM CaCl₂. Competitor salmon sperm DNA was added to a concentration of 100 ng/ μ l. For RNAPII walking experiments, complexes were eluted using a GST-CTD recombinant protein. Spt4/5 was either added or not, and either 100 μ M CTP, GTP, or 100 μ M CTP ATP was added back. 3-Fold molar excess of TFIIS was added back. Complexes were then treated with 5 mM KMnO₄ for 2 min at room temperature and quenched with 20 mM EDTA, 250 mM β -mercaptoethanol, 300 ng/ μ l salmon sperm DNA, and 10% piperidine. Samples were then heated to 90 °C for 10 min and extracted once with water-saturated *n*-butanol, precipitated with ethanol, resuspended in 90% formamide buffer, and run on 10% denaturing polyacrylamide gels.

DNase I Footprinting—Elongation complexes were formed on immobilized RNAPII as described above. After washing, elongation complexes were eluted from the magnetic beads using 1 μ g of recombinant GST-CTD in DNase I reaction buffer (20 mM HEPES, pH 7.5, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT, 0.01 mM ZnCl₂, 100 ng/ μ l BSA, 1 mM CaCl₂, 10% glycerol). Spt4/5 or buffer was added for 5 min, and then 200 ng of salmon sperm DNA was added, and samples were treated with 0.0006 units of DNase I for 1 min at room temperature. Samples were then quenched with 10 mM EDTA and immediately loaded on 0.5 \times TBE, 4.5% native PAGE. After separation, the gel was exposed to x-ray film, and the bands corresponding to free DNA, ECs, and EC-Spt4/5 complexes were excised, and the DNA was recovered. The DNA was then purified and separated on 10% denaturing PAGE. Quantification of the DNase I footprint was performed using ImageQuant software (GE Healthcare). A trace of each individual lane was taken and adjusted for the total counts in each lane using Excel. Then the traces were generated in Excel and overlaid.

Generation of Photo-probes and 5-Iodo-CTP Photo-cross-linking—Photoreactive nucleotide was incorporated at unique positions by hybridizing an oligo to a long template strand immediately upstream of CA pairs. 10 μ M 5-iodo-dCTP (Sigma) and 5 μ Ci of [α -³²P]dATP (6000 μ Ci/mmol) were added along with Klenow *exo*- (New England Biolabs) and incubated for 10 min at 37 °C. Afterward, all dNTPs were added to a final concentration of 100 μ M, and the reaction was allowed to extend for 30 min. Double-stranded modified probes were gel-purified on 10% native PAGE. ECs were formed as described above (EC42), and then Spt4/5 was added followed by 200 ng of carrier salmon sperm DNA. Samples were then irra-

diated using a 308 nm UV light source for 10 min. After irradiation, the samples were digested with DNase I (10 units) and RNase A (2.5 μ g) at 37 °C for 60 min. SDS-loading buffer was then added, and samples were separated on SDS-polyacrylamide gels, dried, and exposed to phosphorimaging screens.

Results

Characterization of Elongation Complexes Formed from End-initiated Templates—We use a minimal system for assembling ECs from highly purified components (28, 43–46). RNAPII elongation was initiated from a 3' single strand extension on a duplex template to assemble an EC with a defined transcript produced from a G-less cassette (Fig. 1A). ECs were formed with RNA transcripts between 35 and 70 nucleotides in length, where indicated. These lengths of RNA are sufficient to bind Spt4/5 (see below) (28, 29). To produce ECs with different length transcripts, the DNA sequence in the template between the initiation site and the first G was shortened. The sequences immediately upstream of the arrest site and the single strand-double strand junction were always identical. Previous studies have shown that initiating transcription from some 3'-extended templates produced aberrant ECs with extended RNA-DNA hybrids caused by the hybridization of the transcript to the template strand and a displacement of the non-template strand (NTS) from DNA (47–49). However, it was later found that the extent to which this occurs is dependent on a number of factors, including the source of RNAPII, the sequences of the 3' tail, and the region encompassing the single strand-double strand junction on the template (47). Previously, it has been shown that templates using the same 3' tail and duplex region as that described here produced transcripts that are resistant to RNase H (cuts RNA-DNA hybrids) and sensitive to RNase A, suggesting extended DNA-RNA hybrids are not produced (44). Nonetheless, because one of the goals of this study was to characterize interaction between Spt4/5 and the nucleic acid scaffold, we further characterized ECs built from this template to validate that a proper nucleic acid scaffold was forming.

We first tested whether the ECs produced here had a properly formed transcription bubble. Potassium permanganate (KMnO₄) was used to interrogate the structure and location of the transcription bubble in RNAP (50, 51). The thymidine (T) bases in single-stranded DNA are prone to oxidation by KMnO₄, whereas those located in double-stranded nucleic acids were resistant. If the NTS was being displaced and the transcription bubble failed to close, thymidines in the NTS behind RNAPII would be highly reactive to permanganate. This was not observed. ECs formed by this method resulted in strong permanganate reactivity immediately upstream of the G-tract (+1) at T-4, T-9, and T-10, which are expected to be in the transcription bubble when RNAPII arrests at the first G (Fig. 1B, compare lanes 2 and 3). Weaker activity is also observed at T+5 and T+7, which lie just on the downstream edge of the expected location of the transcription bubble. The weaker reactivity just ahead of the bubble of RNAPII ECs may be caused by bending of the template and enhanced base flipping in this region. Importantly, thymidines in the NTS immediately behind RNAPII are resistant to KnMO₄ indicating bubble closure. Weak RNAPII-dependent reactivity of T-27, T-35, and

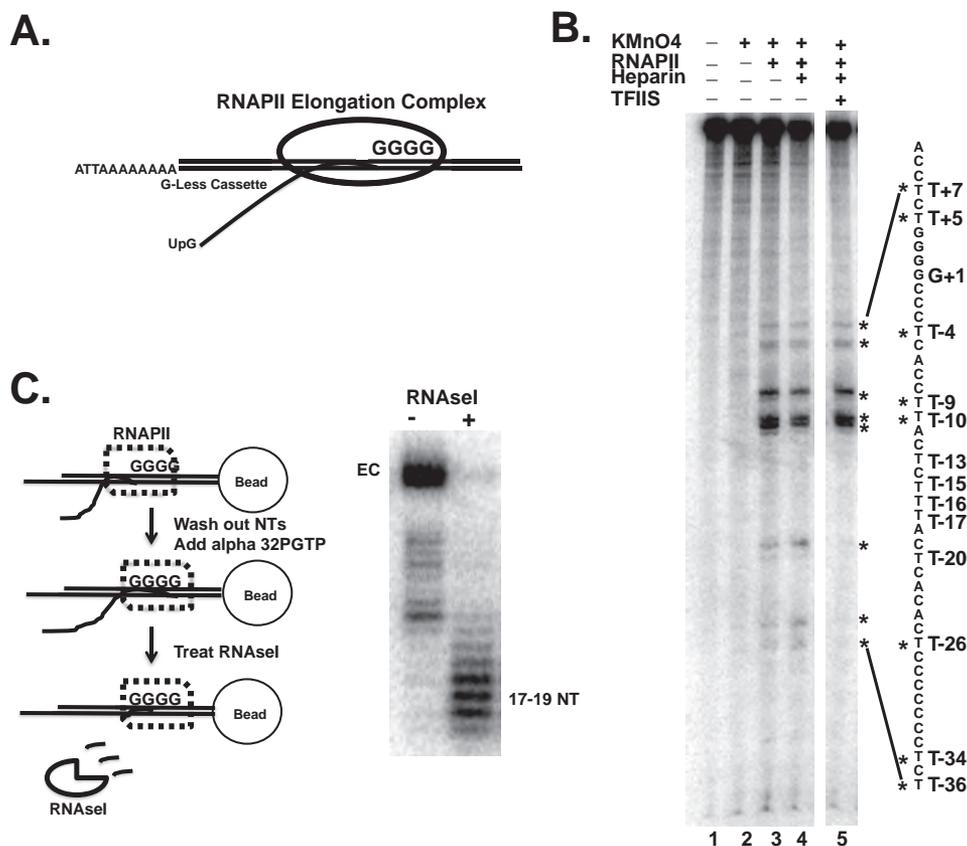


FIGURE 1. Characterization of ECs. *A*, schematic diagram of the formation of RNAPII ECs using 3' end initiated templates. Radiolabeled UTP is incorporated into the transcript during the transcription of the G-less cassette. *B*, KMnO₄ footprint of the non-template strand. ECs (EC70) were prepared on immobilized RNAPII (see under "Experimental Procedures"). DNA sequence (+1 to +4) are equivalent to the G-tract shown in *A*. The reactive Ts are marked with an asterisk. Heparin and TFIIS was added to the samples shown in lanes 4 and 5, respectively. All samples were run on the same gel, but irrelevant lanes were removed between lanes 4 and 5. *C*, RNase I footprinting of the transcript in ECs. ECs were formed on a template that produced a 35-nucleotide (NT) transcript with radiolabeled GTP incorporated into the 3' end of the transcript in the active site of RNAPII. *Left*, schematic representation of RNA protection experiment. *Right*, gel image of the labeled RNA produced with and without RNase I digestion.

T-36 was observed. However, these Ts are flanked by a C-tract found in the template. RNAPII stalls or arrests at polynucleotide tracts (52), and this reactivity may be caused by a fraction of RNAPII becoming arrested over this region due to backtracking. To test this hypothesis, the elongation factor TFIIS was added to the reaction. TFIIS would rescue the arrested complexes over the C-tract and lead to a loss of reactivity at these T bases. This was indeed the case; adding TFIIS eliminated the permanganate reactivity at these thymidines (Fig. 1*B*, compare lanes 3 and 5). Finally, we demonstrated the KMnO₄ reactivity was resistant to heparin, suggesting that this was not due to a nonspecific interaction between RNAPII and the DNA (Fig. 1*B*, lanes 3 versus 4). Thus, KMnO₄ reactivity indicates that a transcription bubble is produced and that the NTS is re-annealing with the template strand behind RNAPII.

Previous EC complex mapping revealed that RNAP protects 16–19 nucleotides of RNA from digestion by single-stranded RNases (53, 54). The structure and location of the emerging transcript in the ECs prepared by this method were examined. RNAPII ECs were formed on immobilized DNA templates by transcribing up to the G-tract using "cold" nucleotides lacking GTP; the complexes were washed to remove nucleotides, and then [α -³²P]GTP was added to produce a 3' end-labeled transcript. RNase I digested the transcript down to 17–19 nucle-

tides, a length expected if the RNA was threaded through the RNA exit channel (Fig. 1*C*). If it emerged aberrantly, such as through the main channel, the size of the protected transcript should be different. Additionally, the sensitivity of the transcript to single-stranded specific RNase indicates that extended RNA-DNA hybrids were not formed on the vast majority of the templates. All of the data described here, and elsewhere (28, 43–46), indicate that this template produces ECs with the expected nucleic acid scaffold.

Spt4/5 Reduces or Prevents RNAPII Arrest—After validating the structure of the ECs, we tested how Spt4/5 bound to ECs and affected the activity of RNAPII. Until now, it has been shown that human Spt4/5 (DSIF) stimulates elongation and mediates 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole sensitivity of RNAPII in extracts (9, 24, 55, 56), but how Spt4/5 affects RNAPII activity is still unclear. We produced recombinant yeast Spt4/5 complex in *Escherichia coli* using a polycistronic co-expression system (39). Spt4/5 complexes containing a full-length Spt5 and another lacking the C-terminal region (Δ CTR) were isolated to high purity (Fig. 2*A*). Removing the disordered CTR increased expression of the complex and allowed for its separation from Rpb2 in gels in subsequent experiments (see below). Because the CTR is not required for viability in yeast or binding to RNAPII *in vitro* (36, 38), we

Spt4/5 Interaction with DNA Prevents Arrest of RNAPII

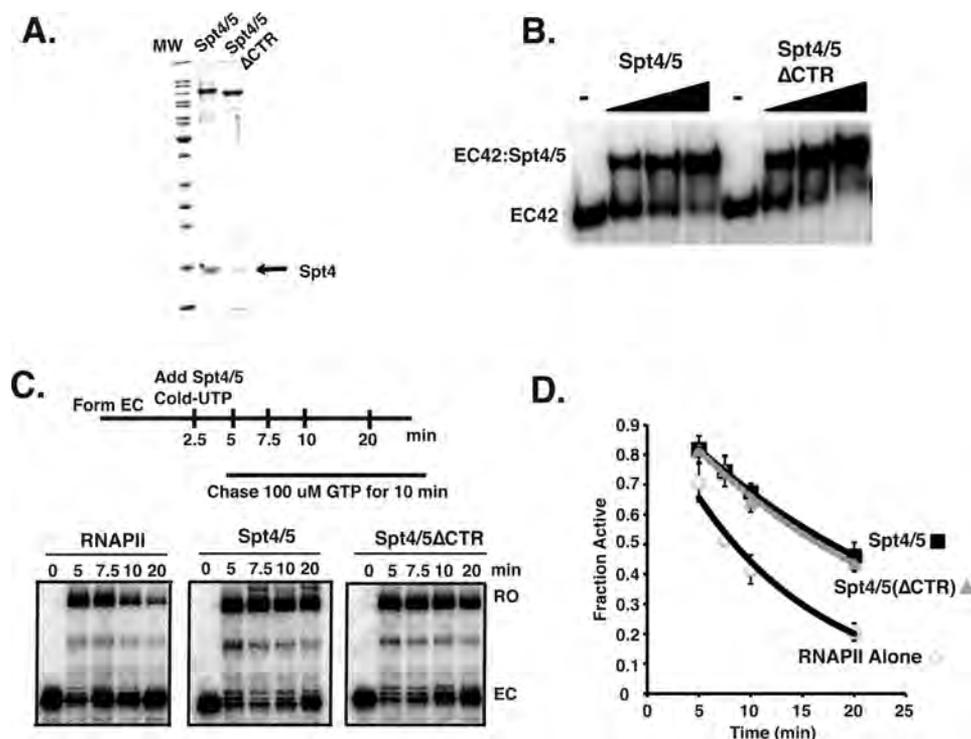


FIGURE 2. Biochemical analysis of recombinant Spt4/5. *A*, Coomassie Blue-stained SDS-PAGE of recombinant of Spt4/5 and Spt4/5(Δ CTR). An *arrow* shows Spt4. The *asterisk* indicates a contaminating protein migrating just above Spt4 found in the preparation of full-length Spt4/5. *B*, EMSA of Spt4/5 and Spt4/5(Δ CTR) binding to RNAPII EC prepared on end-labeled DNA templates. The ratio of Spt4/5 to RNAPII was 0.75, 1.5, and 3.0. Free DNA migrated at the bottom of the gel and was cropped out of the image. *C*, RNAPII arrest assay comparing intact Spt4/5 and Spt4/5(Δ CTR). Saturating amounts of Spt4/5 were used in the assay, which was determined by shifting all of the EC in EMSA. *D*, RNAPII arrest assays were quantified as described under "Experimental Procedures" and are plotted as a fraction of active RNAPII as a function of time. The data were fit to an exponential decay curve. The data represent the averages and standard deviations of three independent assays.

suspected it might be dispensable for the elongation promoting activities of the complex, but this was not known. So the biochemical activities of these two forms were compared.

We tested the ability of intact Spt4/5 and the CTR-less derivative to bind to ECs. Although it was known that the CTR-less version of Spt5 interacts with RNAPI and RNAPII (38), it was not known whether the CTR played a role in the binding of Spt4/5 to ECs. ECs (EC42) were assembled onto an end-labeled DNA template; Spt4/5 and Spt4/5(Δ CTR) were titrated into the binding assay, and the complexes were resolved on native gels. The results shown in Fig. 2*B* indicate that Spt4/5 complexes containing intact and CTR-less Spt5 bound to ECs to the same extent, indicating that the CTR region of Spt5 is not required for Spt4/5 to bind RNAPII ECs *in vitro*.

The bacterial homolog of Spt5, NusG, has been shown to reduce long lifetime pauses of RNAP and promote elongation (6, 57). Archaea Spt4/5 has been shown to enhance general processivity through interactions with the coiled-coil domain of RNAP (7). To determine whether yeast Spt4/5 promotes elongation, we established a transcription assay to measure the activity of Spt4/5 that relies on the propensity of RNAPII to become paused or arrested during prolonged periods of nucleotide depletion. In this assay, radiolabeled ECs are formed for 2.5 min in the absence of GTP, allowing RNAPII to transcribe up to the G-tract in the template (pulse). At that point, cold UTP and Spt4/5 or buffer was added; the excess cold UTP prevented the detection of newly formed ECs during the chase. The ECs were then incubated for increasing amounts of time

before GTP was added to initiate the chase (Fig. 2*C*, *upper panel*). The fraction of full-length transcript (active RNAPII) was plotted as a function of time. As expected, the longer GTP was withheld from the reaction, the smaller the fraction of ECs converted into active complexes that produced runoff product (Fig. 2, *C* and *D*). When intact Spt4/5 (WT) or the CTR-less Spt4/5 (Δ CTR) was added, the amount of arrested complexes detected at the 5-, 7.5-, 10-, and 20-min time points was reduced compared when Spt4/5 was omitted. The percentage of runoff over time was fit to an exponential decay curve, and the slope of these lines was compared. This analysis indicated Spt4/5 reduced arrest by about half, and the CTR-less version showed the same level of activity (Fig. 2*D*). Complete runoff was not observed, however, even in the presence of Spt4/5. The fraction of ECs refractory to Spt4/5 activity (\sim 30%) likely represents RNAPII that reached the G-tract and arrested during the 2.5-min pulse before Spt4/5 was added, suggesting that Spt4/5 prevented arrest but cannot rescue pre-arrested complexes. In support of this hypothesis, if Spt4/5 was added after 20 min of GTP starvation, it could not rescue the pre-arrested RNAPII complexes.³ We conclude from these results that Spt4/5 reduces or prevents arrest of RNAPII, and the CTR of Spt5 is not required for its biochemical activities. Because we could generate higher quality complexes from the CTR-less Spt4/5, it was used for the remaining experiments unless noted

³ J. B. Crickard and J. C. Reese, unpublished results.

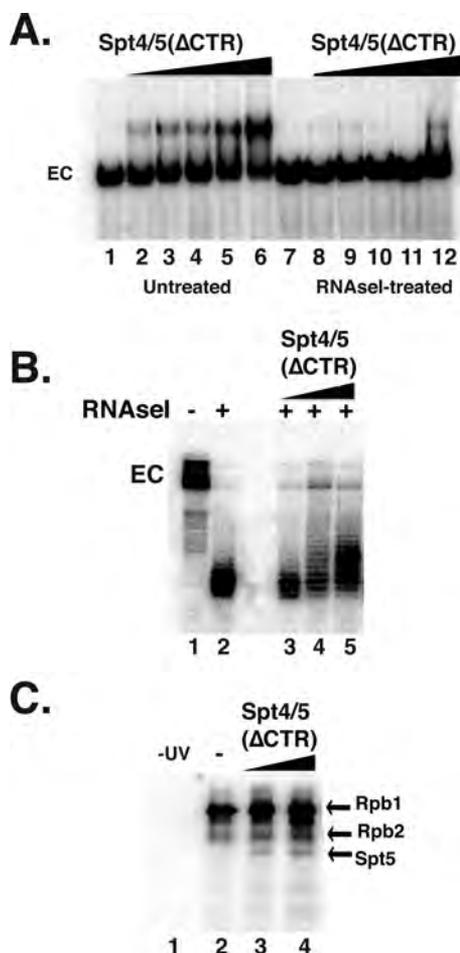


FIGURE 3. Spt4/5 contacts the emerging transcript. *A*, EMSA analysis of EC2 treated with (lanes 7–12) and without RNase I (lanes 1–6). Spt4/5 Δ CTR was titrated in and complexes were resolved on native polyacrylamide gels. *B*, RNase I footprint of the emerging transcript. Spt4/5 (Δ CTR) was titrated in prior to digestion. The experiment was conducted as described in the legend of Fig. 1C. *C*, ECs were prepared with body-labeled transcripts using [32 P]CTP and bromo-UTP. After exposure to UV, where indicated, samples were digested with DNase I and RNase A and separated on SDS-PAGE. Labeled proteins were visualized by exposing dried gels to a phosphorimager screen. Two different amounts (1- and 3-fold molar excess) of Spt4/5 were added. The locations of the bands corresponding to Rpb1, Rpb2, and Spt5(Δ CTR) are indicated on the right.

otherwise. Following this section, the Spt4/5(Δ CTR) version will be referred to as Spt4/5 in the text for simplicity.

Nascent RNA Transcript Is Required for Yeast Spt4/5 to Bind to ECs—Previous studies indicated that the stable association of metazoan DSIF (Spt4/5) with ECs required RNA to emerge from RNAPII (28, 29). To test whether this was also the case for yeast Spt4/5, we measured the binding of Spt4/5 to ECs before and after digesting the transcript with RNase I. RNAPII ECs were formed on end-labeled DNA templates and then split into two pools; one pool was treated with RNase I to digest the emerging transcript, and the other was left untreated. Spt4/5 bound to the untreated ECs but not to those treated with RNase I (Fig. 3A). Even the small amount of binding observed in some lanes in this experiment could be attributed to a small fraction of ECs not digested with RNase.

It was estimated that PDB DSIF requires \sim 4 nucleotides of RNA to associate with RNAPII ECs, suggesting that Spt4/5 contacts the transcript near the RNA exit channel (28). We con-

ducted RNase I protection assay on ECs containing 3' end-labeled transcripts in the presence and absence of Spt4/5 (Fig. 3B). As described above, RNAPII protected \sim 17–19 nucleotides of the transcript from digestion (Fig. 3B); however, adding Spt4/5 to the assay extended the protection by \sim 2–4 nucleotides (Fig. 3B). This result suggests that, like its *Drosophila* counterpart, yeast Spt4/5 interacts with RNA as it emerges from RNAPII.

UV cross-linking was performed to determine whether the RNase I protection is caused by direct contact between Spt5 and RNA. ECs were prepared in the presence of radiolabeled CTP and a photoreactive UTP analog, bromo-UTP, to prepare body-labeled transcripts. Complexes were then UV-irradiated and digested with nucleases, and the transfer of labeled RNA to proteins was detected by SDS-PAGE followed by autoradiography. The strongest cross-linking occurred between Rpb1 and RNA, with lower levels detected for Rpb2 (Fig. 3C). This is in agreement with other studies (28, 43, 58). Adding Spt4/5 resulted in the appearance of an additional radiolabeled band migrating at the expected position of Spt5 (Fig. 3C).

C-terminal KOW Domains of Spt5 Are Required for Spt4/5 Binding to RNAPII—Prokaryotic NusG or Spt4/5 contains a minimal NGN-KOW1 region, and this is sufficient for it to associate with RNAP (7, 20). A previous study using glutathione *S*-transferase-Spt5 fusion proteins (GST-Spt5) determined that deleting several of the KOW domains resulted in a loss of Spt5 binding to RNAPI and RNAPII (38). Furthermore, *in vivo* cross-linking studies suggest that the region around KOW domains 4 and 5 cross-link to RNAPII (19). When these cross-link sites were modeled onto existing structures of RNAPII elongation complexes, KOW domains 4 and 5 fit between the Rpb4/7 subunits of RNAPII and the base of the Rpb1 clamp domain (19). Importantly, this model would position Spt5 to protect and physically interact with the emerging transcript. We next wanted to determine how removal of these eukaryotic specific domains would affect the association of Spt4/5 with ECs. Mutants lacking the eukaryotic specific regions of the C terminus of Spt5(Δ 419–1063) and another that removed linkers 3 and 4 and KOW domains 4 and 5 (Δ 666–901) (Fig. 4A) were produced. A gel showing the purified complexes is displayed as Fig. 4B. The Δ KOW4 and 5 mutant can only be expressed if the CTR region was included,³ so this construct contained the CTR. Because the effect of deleting these regions on the interaction between Spt4/5 and ECs is unknown, this was examined by EMSA. In these experiments, a range of Spt4/5 concentrations was used so that a K_d value for the association of Spt4/5 with RNAPII ECs can be estimated. Spt4/5 bound to RNAPII ECs with high affinity (K_d 8.9 ± 2.5 nM). However, deleting most of the C terminus of the Spt4/5(Δ 419–1063) abolished the association of Spt4/5 with RNAPII ECs, and deleting KOW4 and 5 (Δ 666–901) reduced the binding of the complex more than 10-fold (Fig. 4, C and D). When the RNA cross-linking was conducted on these binding mutants, no obvious bands arising from Spt5 were observed (Fig. 4E). These results confirm that the eukaryotic specific regions of Spt5 are required for Spt4/5 to bind to ECs. Additionally, to further confirm that the labeled protein detected in the RNA cross-linking experiments was indeed Spt5, we produced an Spt4/5 complex con-

Spt4/5 Interaction with DNA Prevents Arrest of RNAPII

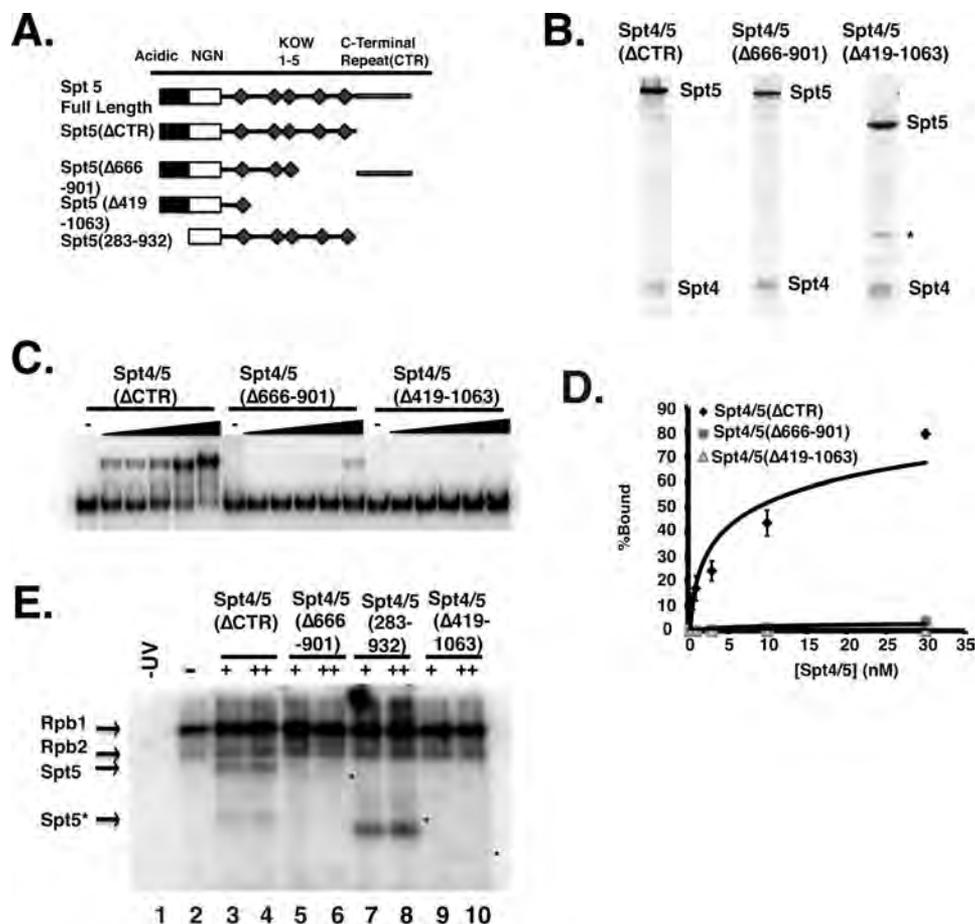


FIGURE 4. C-terminal KOW domains of Spt5 are required for the binding to RNAPII ECs. *A*, schematic representation of Spt5 deletion mutants that were used in the figure. The area shaded in black is an acidic region in the N terminus; the white box is the NGN domain, and the diamonds are the KOW domains. *B*, Coomassie Blue-stained gel of the mutant complexes. *C*, EMSA comparing the binding of Spt4/5(Δ CTR), Spt4/5(Δ 666-901), and Spt4/5(Δ 419-1063) to ECs formed from 10 nM RNAPII. Increasing amounts of complex were titrated in and then resolved by native PAGE. *D*, quantification of EMSA experiments. Percent binding was plotted on the y axis as a function of Spt4/5 concentration. The data were then fit to a logarithmic binding curve, and a K_d value was estimated. Values represent the averages and standard deviations of three independent experiments. *E*, RNA cross-linking assay. The assay was conducted as described in Fig. 3C. A sample of RNAPII alone was untreated with UV (lane 1) (–UV) and all others were exposed to UV (lanes 2–10). Two amounts of each of the Spt4/5 complexes were titrated in. “+” indicates the amount of Spt4/5 required to shift ECs in EMSA assays and “++” is twice that amount (3- and 6-fold molar excess relative to RNAPII). Lane 1, –UV; lane 2, +UV; lanes 3 and 4, Spt4/5(Δ CTR); lanes 5 and 6, Spt4/5(Δ 666-901); lanes 7 and 8, Spt4/5(Δ 284-931); and lanes 9 and 10, Spt4/5(Δ 419-1063). The asterisk (Spt5*) marks a breakdown product in the Spt4/5 preparation. The dots mark the migration of the mutants in the gel.

taining Spt5 lacking the N terminus (283–932). Cross-linking assays using this mutant resulted in a faster migrating band at the size expected of the Spt5 mutant lacking the N terminus, confirming that the labeled protein is Spt5 (Fig. 4E).

Probing the Nucleic Acid Scaffold in Spt4/5-containing ECs—Next, we analyzed interactions between Spt4/5 and DNA in the EC. Current structural models generated from x-ray crystallographic data of archaeal Spt4/5 fused to the clamp domain of RNAP suggest that eukaryotic Spt4/5 may interact with upstream DNA exiting RNAPII. Additionally, two other structures, one of RNAPII including both strands of DNA (18) and another of the isolated KOW1-Linker1 region of Spt5(372–508, KIL1) (32), also suggest that Spt4/5 binds to upstream DNA. In addition, structural modeling of Spt4/5 in ECs suggests that Spt5 could contact the non-template strand of DNA. Although this function has not been documented for Spt5, it was recently reported that *Bacillus subtilis* NusG interacts with the non-template strand of DNA to promote sequence-specific pausing (59). Additionally, the recruitment of RfaH requires exposure of the operon polarity suppressor element within the non-tem-

plate strand of DNA (60). Although it has been widely speculated that Spt4/5 binds DNA in the ECs, direct evidence for this is lacking.

To identify contacts between Spt4/5 and DNA, DNase I footprinting of RNAPII ECs was conducted in the presence and absence of Spt4/5. Because only a fraction of the DNA templates will form ECs, ECs were formed on RNAPII immobilized on protein-A magnetic beads coated with 8WG16 antibody that binds the CTD of Rpb1. This allowed enrichment for RNA-P-II-engaged templates from the more abundant free templates. Complexes were then eluted from the beads with the recombinant glutathione *S*-transferase C-terminal domain (GST-CTD), treated with DNase I, and then resolved on preparative native gels to separate free DNA, ECs, and Spt4/5-containing ECs (Fig. 5A, left). The DNA contained in each band was purified and run on a denaturing gel (Fig. 5A, right). Consistent with published DNase I footprints of RNAPII, elongation complexes showed a region of protection of ~35–40 bases (Fig. 5A, lanes 3 versus 4) (51, 61). Importantly, Spt4/5 extended the footprint upstream of RNAPII, and protection was extended to bases

Spt4/5 Interaction with DNA Prevents Arrest of RNAPII

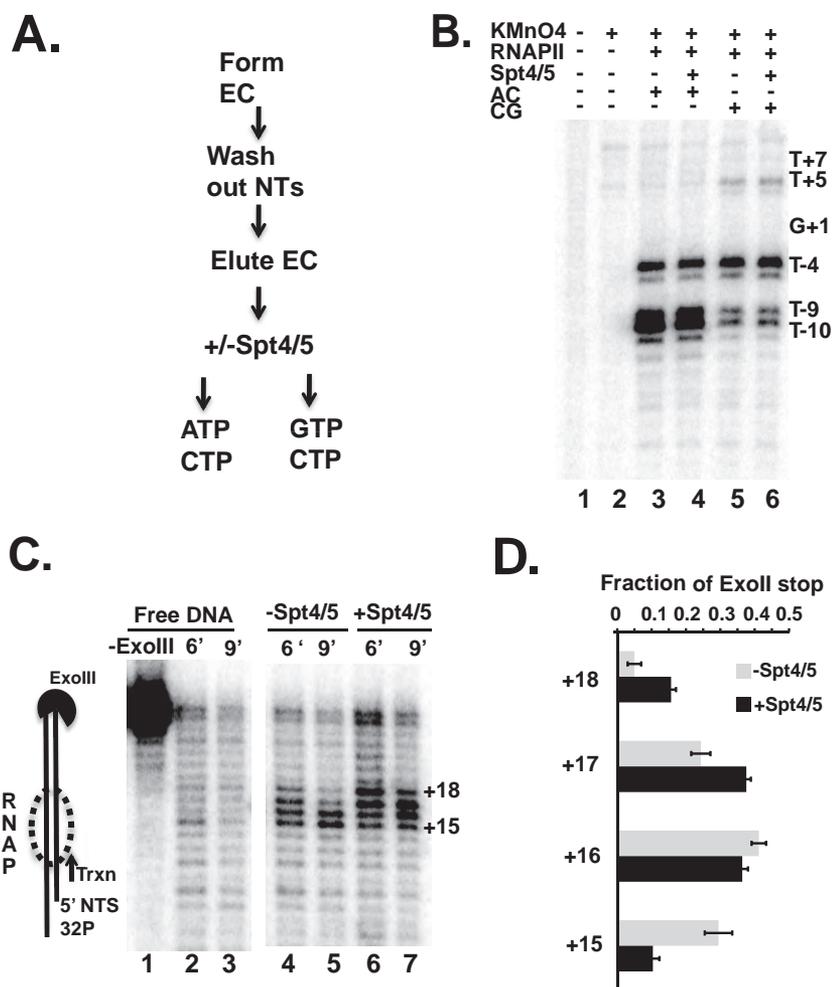


FIGURE 6. Footprinting elongation complexes with permanganate and exonuclease III. *A*, flow diagram explaining the experimental design. Details are given under "Experimental Procedures." Briefly, immobilized RNAPII ECs were provided ATP and CTP (AC, lanes 3 and 4 in *B*) or provided CTP and GTP to allow transcription through the 4-base G-tract (CG, lanes 5 and 6 in *B*). Spt4/5(Δ CTD) was added where indicated. TFIIIS was added after nucleotide addition to reduce the potential effects of backtracking on bubble progression. *B*, denaturing polyacrylamide gel showing the end-labeled non-template strand. Ts in the sequence are indicated on the right of the gel and labeled relative to the first G. *C*, ExoIII protection experiment to map the leading edge of RNAPII along the DNA. Denaturing PAGE comparing 6 and 9 min of cleavage by ExoIII. Free DNA (lanes 1–3) was compared with DNA + RNAPII-Spt4/5 (lanes 4 and 5) and DNA + RNAPII with Spt4/5(Δ CTD) (lanes 6 and 7). Positions of the bands are labeled to the right of the gel and represent the distance downstream from G+1. *D*, quantification of the ExoIII data from four independent experiments presented as the averages and S.E. The numbers on the left indicate the size of the band in nucleotides and correspond to those indicated in *C*.

These changes were observed (Fig. 6*B*, compare lanes 3 versus 5), although it appears that not all of the RNAPII walked to the new location, because the increase in the reactivity of T+5 and reduction in reactivity of the -9 and -10 positions were not quantitative. Importantly, there were no differences in the reactivity of these locations in the presence of Spt4/5 (Fig. 6*B*, compare lanes 5 and 6). Although we cannot rule out that Spt4/5 may kinetically affect the opening or closing of the transcription bubble, we conclude that under equilibrium conditions Spt4/5 does not alter the size of the transcription bubble.

The results above suggest that Spt4/5 protects bases upstream of RNAPII. It is possible that Spt4/5 contacts DNA downstream (ahead of polymerase). Furthermore, nothing is known about how Spt4/5 affects the tracking of RNAPII along DNA. NusG causes the lateral movement of RNAP downstream by one nucleotide, which may represent the post-translocated state of RNAP (11). Either of these two changes brought

about by Spt4/5 can be revealed by ExoIII footprinting of the leading edge of RNAPII. RNAPII ECs were formed on end-labeled DNA templates and incubated with or without Spt4/5, followed by digestion with ExoIII. Probing ECs with ExoIII generated an RNAPII-dependent footprint, as shown by the protection of the DNA compared with digestion of naked DNA (Fig. 6*C*, lanes 2 and 3 versus lanes 4 and 5). Adding Spt4/5 extended the footprint compared with RNAPII alone by approximately one or two nucleotides downstream (Fig. 6*C*, compare lanes 4 and 5 versus lanes 6 and 7). The shift in the distribution of ExoIII stops was quantified, and the data from multiple experiments are presented in Fig. 6*D*. This footprinting result argues against extensive protection of downstream DNA by Spt4/5, but it is strikingly similar to the lateral movement of RNAP caused by NusG (11). A conservative interpretation of this result is that Spt4/5 extends the leading edge of RNAPII downstream, which may be caused by a structural change in RNAPII.

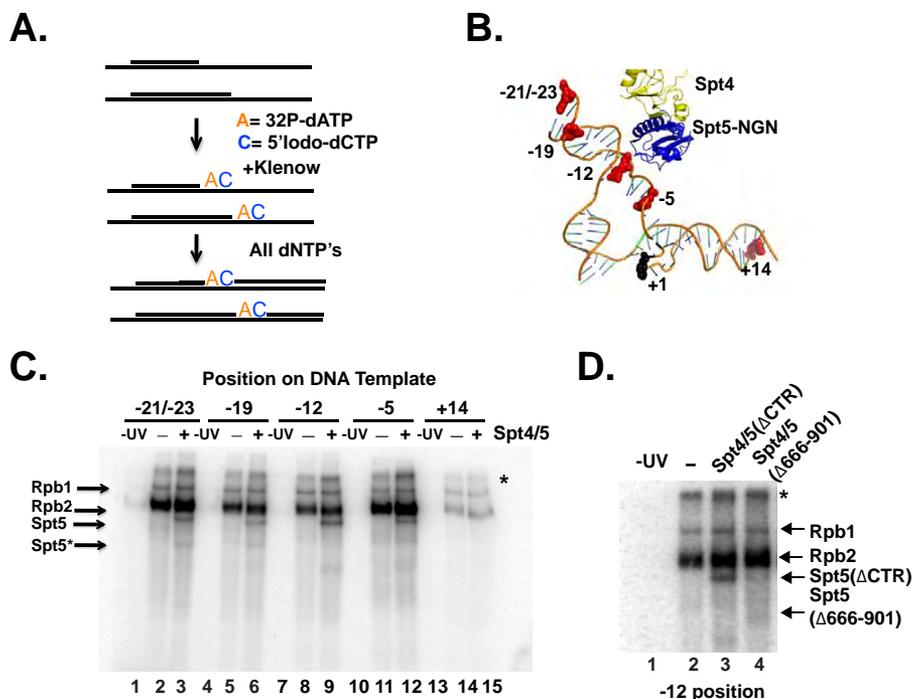


FIGURE 7. Spt5 cross-links to the non-template strand in ECs. *A*, schematic outline of the strategy used to generate site-specific labeled DNA probes on the non-template strand of the EC42 DNA template. A short oligo complementary to the template strand was annealed and used to prime synthesis of the NTS. In the first step, [³²P]dATP and 5-iodo-dCTP were incorporated using the Klenow (–exo) fragment of DNA polymerase. In a second step, cold dNTPs were added to complete the top strand. *B*, A model derived from overlaid crystal structures of RNAPII with the non-template strand, yeast Spt4/5, and archaeal RpoA fused to Spt5 (PDB codes 5C4X, 2EXU, and 3QQC) using PyMOL (version 1.7.4 Schrödinger, LLC). Highlighted in red are bases representing the position of photoreactive 5-iodo-dCTP nucleotide analog in the non-template strand of RNAPII. +1 is the site of NTP incorporation. *C*, SDS-PAGE analysis of cross-linked products. The location of the photoreactive nucleotide in each template is indicated above. The presence of Spt4/5(ΔCTD) (3-fold molar excess relative to RNAPII) is indicated in the panel. Arrows on the left indicate the migration of Rpb1, Rpb2, and Spt5(ΔCTD). Spt5* indicates the position of a breakdown product of Spt5 that cross-links to DNA. The band above Rpb1 marked with an asterisk may result from incomplete digestion of the nucleic acid template or products formed by protein-protein cross-links. *D*, SDS-PAGE comparing cross-linking of RNAPII only (lane 2), Spt4/5(ΔCTR) (lane 3), and Spt4/5(Δ666–901) mutant to the –12 probe.

Spt5 Contacts the Non-template Strand in the Transcription Bubble and DNA Upstream of RNAPII—To provide further evidence that Spt4/5 contacts DNA within the nucleic acid scaffold, we turned to site-specific cross-linking. Using the findings from our DNase I footprinting experiments (Fig. 5) and the model of the Spt5(NGN)-Spt4 in the RNAPII EC (10), we selected locations to incorporate the photoreactive nucleotide analog 5-iodo-2'-dCTP adjacent to radiolabeled adenine nucleotide in the NTS (Fig. 7A). The photoreactive probes were placed into the NTS because structural models predicted that the NTS would be in close proximity to Spt4/5 and because incorporating the probes in the template strand could impede transcription of RNAPII up to the G-tract. Photoreactive nucleotides were incorporated at positions –21/23, –19, –12, –5, and +14 (+1 is the first G in the G-tract) (Fig. 7B). –21/23 and –19 are located upstream of RNAPII (14, 51); –12 is located near the expected site of strand re-annealing; –5 is located in the transcription bubble, and +14 is located downstream of RNAPII and is a control because we would expect lower levels of cross-linking of RNAPII at this location. The strategy to incorporate the radiolabeled and photoactive nucleotides produced templates with photoreactive nucleotides at both the –21 and –23 positions because of the DNA sequence. Rpb2 cross-linked to the DNA with the highest efficiency, with lower levels of Rpb1 cross-linking detected (Fig. 7C). The strong cross-linking of the NTS to Rpb2 is consistent with the recently

published crystal structure of RNAPII with an intact transcription bubble, which revealed that the NTS passes along the surface of Rpb2 (18). A diffuse band above Rpb1 was also observed. It is unclear what this band is, but it may represent undigested DNA-protein products or UV-induced aggregates. As expected, low levels of Rpb1 and Rpb2 cross-linking were observed at the +14 site, which should be positioned just ahead of RNAPII. Spt5 cross-linked at –21/23, –19, –12, and –5. No cross-linking of Spt5 was detected at position +14 (Fig. 7C), suggesting that the cross-linking of Spt5 to DNA requires it to be properly positioned in the RNAPII EC. Weaker bands migrating faster than Spt5 were also observed in the lanes containing Spt4/5. These bands may be truncated forms of Spt5 contained in the preparations or breakdown of Spt5 during cross-linking. Interestingly, even though Spt4 is in close proximity to DNA in the model (Fig. 7B), we failed to conclusively identify a labeled band of the size of Spt4 in high percentage gels.³ However, this negative result does not definitely rule out the possibility that Spt4 interacts with another region of the nucleic acid scaffold. Furthermore, as a control, we performed cross-linking at the –12 position using a mutant form of Spt4/5(Δ666–901) that binds poorly to RNAPII ECs (see Fig. 4C), and we found that no cross-linking was observed (Fig. 7D). These data are the first evidence that eukaryotic Spt5 physically interacts with DNA upstream of RNAPII and contacts the non-template strand of the transcription bubble.

Spt4/5 Interaction with DNA Prevents Arrest of RNAPII

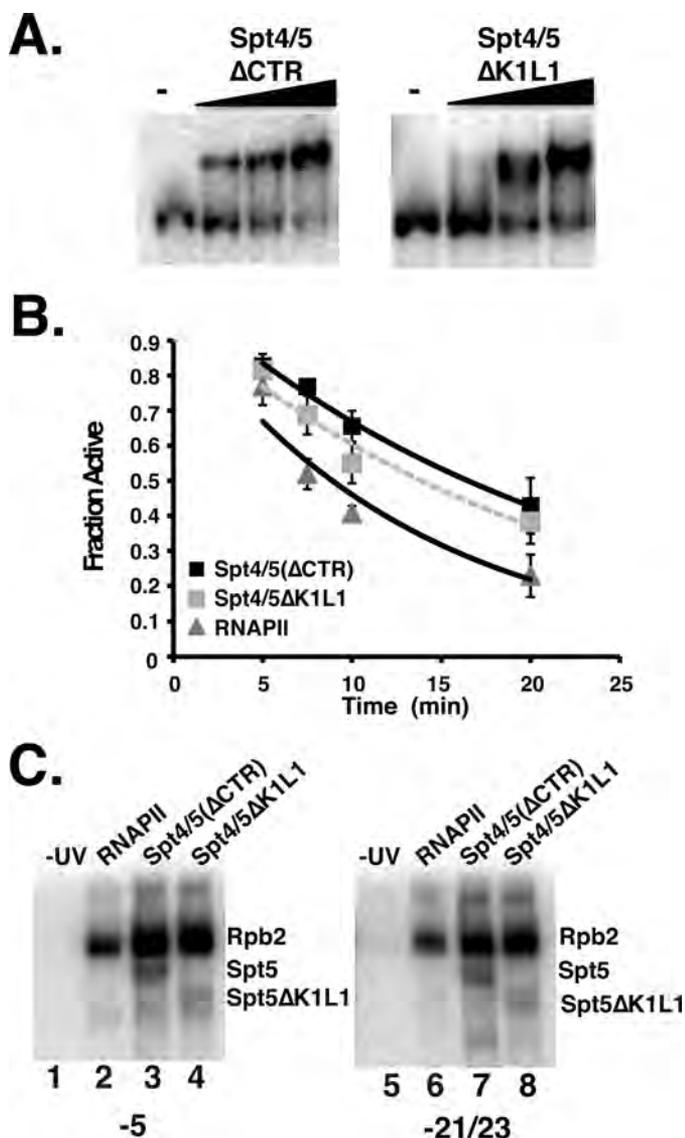


FIGURE 8. Conserved KOW1-linker region is dispensable for the biochemical activities of Spt4/5. *A*, EMSA comparing the binding of Spt4/5(Δ CTR) and Spt4/5(Δ K1L1) to ECs. Assays were conducted as described in the legend of Fig. 2B. The Spt4/5(Δ K1L1) mutants also have the CTR removed, but it is labeled more simply in the figure. *B*, RNAPII arrest assay comparing the activity of RNAPII alone or RNAPII plus Spt4/5(Δ CTR) or Spt4/5(Δ K1L1). Data are plotted as the fraction of active ECs as a function of time. The data were fit to an exponential decay curve. Error bars represent the standard deviation of three independent experiments. *C*, DNA photo-cross-linking of Spt4/5 to the -5 (lanes 1–4) and $-21/-23$ (lanes 5–8) positions.

KOW1 Linker of Spt5 Is Dispensable for the Biochemical Activities of Spt4/5—A high resolution x-ray crystal structure of the KOW1-linker (K1L1) region of Spt5 was solved, and this fragment of the protein bound single- and double-stranded nucleic acids (32). However, the importance of this region, which is believed to be the counterpart of the KOW domain in prokaryotic Spt5 homologs, for the biochemical activities of the complex is not known. We examined the association of an Spt4/5 complex containing Spt5 lacking the K1L1 region (Δ K1L1) with ECs. Fig. 8A shows that the mutant complex bound to ECs, and a titration of different amounts of the complex suggests that it has a slightly reduced affinity compared with the intact complex (about 2-fold). Thus, the K1L1 region

plays a minor role in the binding of Spt4/5 to RNAPII elongation complexes *in vitro*.

The anti-arrest activity of the K1L1 mutant was examined using the EC arrest assay described above. The results show that the Spt5 K1L1 mutant can prevent arrest but may not be as effective as the wild type complex (Fig. 8B). The slight reduction in activity, which is reproducible, may be related to the small decrease in binding observed in the EMSA experiments. Nonetheless, although the K1L1 region may be required for optimal interaction with RNAPII ECs, it is not essential for the biochemical activities of Spt4/5 examined here.

One potential function of the K1L1 region in Spt5 is binding to DNA in the transcription scaffold. DNA cross-linking assays were conducted using the Spt4/5(Δ K1L1) mutant under saturated binding conditions. We examined cross-linking within the transcription bubble (-5) and upstream DNA ($-21/-23$). It was proposed that the K1L1 of Spt5, together with Spt4, “cradles” upstream DNA (32). However, the results show that the K1L1 mutant cross-linked at the -5 and $-21/-23$ positions (Fig. 8C). This does not rule out the possibility that K1L1 binds to DNA, but deleting it does not affect Spt5 cross-linking in the regions tested here.

Highly Conserved Basic Surface on the NGN Domain of Spt5 Contacts the Non-template Strand and Is Required for Spt4/5 to Prevent RNAPII Arrest—The recent publication of the structure of an RNAPII EC with the complete transcription bubble provided an opportunity to more precisely position Spt4/5 within the RNAPII EC (18). Based on this model, we identified two regions within the NGN domain of Spt5 that contained stretches of positively charged amino acids that could form a DNA-binding surface. One of these surfaces (302–308) is highly conserved in all Spt5 homologs and is predicted to face the transcription bubble (Fig. 9A, green residues), whereas the other (316–321) is not universally conserved and projects away from the NTS (Fig. 9A, orange). Spt4/5 complexes containing Spt5 with alanine substitutions in the basic residues on these surfaces, 302–308A and 316–321A, were purified and analyzed.

The mutants were subjected to site-specific DNA cross-linking on ECs with 5-iodo-2'-dC at the -5 and -12 positions in the NTS. Consistent with our modeling results, mutating the highly conserved patch of basic residues (302–308) prevented the cross-linking of Spt5 to the NTS, whereas the mutant with substitutions in the non-conserved patch facing away from the transcription bubble (316–321) cross-linked to both positions (Fig. 9B). These results indicate that the NTS binds to this basic surface of the NGN domain of Spt5 and provides stronger evidence that eukaryotic Spt5 binds to DNA in the EC. The 302–308A Spt5 mutant was able to form a complex with Spt4,³ suggesting the mutations did not cause gross structural changes to the NGN domain. However, a trivial explanation for why the Spt4/5(302–308A) derivative failed to cross-link to the NTS is that the mutations prevented the binding of the mutant complex to ECs. However, this was not the case as the mutant complex bound as well as the wild type version (Fig. 9C).

We next wanted to understand the significance of the binding of the NTS to the NGN domain. Based on work on NGN homologs from prokaryotes (60), interaction between the NGN

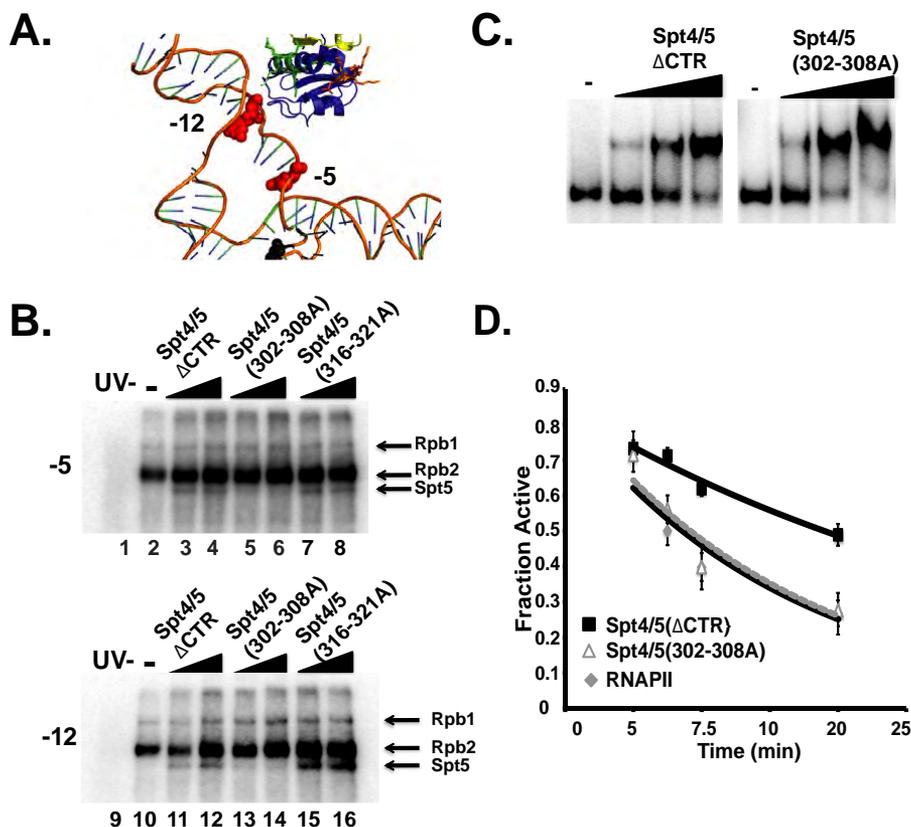


FIGURE 9. Charged surface on the NGN domain is required for the anti-arrest activity of Spt4/5. *A*, model derived as in Fig. 7*B* of Spt4/5(NGN) with the transcription bubble. Spt4 is colored in yellow and Spt5-NGN in blue. The region of Spt5-NGN encompassing the conserved basic patch (302–308) is colored in green, and the region Spt5-NGN containing the non-conserved patch (316–321) is colored in orange. *B*, SDS-PAGE of DNA cross-linking products using a template with a probe at positions -5 (lanes 1–8) and -12 (lanes 9–16). Note, each of the point mutations contained the CTR removed like the “wild type” Spt4/5 but is labeled only to indicate the amino acid changes in the NGN. *C*, EMSA testing binding of Spt4/5(Δ CTR) and Spt4/5(302–308A) to RNAPII ECs. *D*, RNAPII arrest assay comparing RNAPII alone and RNAPII plus the Spt4/5 derivatives indicated in the panel. The assay was conducted with saturating amounts of Spt4/5, estimated by the results of the EMSA. Data were plotted as the fraction of active RNAPII as a function of time and fit to an exponential decay curve for comparison. Error bars represent the standard deviation of four independent experiments.

domain and DNA may be necessary for Spt4/5 to prevent arrest of RNAPII. As described above, Spt4/5 prevented the pausing or arrest of RNAPII over the G-tract during prolonged GTP omission (Fig. 9*D*). However, the Spt4/5(302–308A) mutant lost this activity. Because this mutant associated with RNAPII ECs (Fig. 9*C*), we conclude that contact between this conserved surface of Spt5 and the NTS is required for Spt4/5 to prevent arrest of RNAPII. The significance of this interaction will be addressed under the “Discussion.”

Phenotypic Effects of 302–308A and K1L1 Mutations—SPT5 is an essential gene (27). The ability of the K1L1 deletion and the NGN 302–308A mutant to support viability was examined. To do so, we utilized the “anchor-away” approach (62) to conditionally deplete Spt5 from the nucleus. A strain whose only copy of *SPT5* contains an in-frame fusion with the FRB fragment (*SPT5-FRB*) and expressing an FKBP12 fusion of the ribosomal protein Rpl13 was constructed. Rapamycin induced dimerization of Spt5-FRB, and FKBP12-Rpl13 conditionally sequesters Spt5 in the cytoplasm, and cells transformed with the empty vector (pRS313) were not viable on rapamycin plates (Fig. 10*A*). However, a plasmid containing a wild type copy *SPT5* fully complemented the growth defect (Fig. 10*A*). The *spt5* K1L1 Δ mutant did not support viability (Fig. 10*A*). Interestingly, expression of the K1L1 mutant caused slightly slowed

growth of the strain even in the absence of rapamycin. This can be explained by competition between the mutant and the Spt5-FRB protein that supported the functions. Thus, even though this mutant displayed only mild biochemical defects, it could not complement the loss of Spt5 from the nucleus. This suggests it has additional functions in the cell. In contrast, the *spt5*(302–308A) mutant was viable but displayed temperature-sensitive growth. The viability of this mutant suggests that the essential functions of Spt5 are not impacted by mutation of these residues (see under “Discussion”).

The plasmid copy of the *SPT5* derivatives contained an epitope tag (HA) at the C terminus to allow for the measurement of protein levels by Western blotting. When Spt4/5 is recruited to genes, the CTR of Spt5 is phosphorylated by Kin28 and Bur1, thereby changing its migration in gels (36). Analyzing the phosphorylation state of the Spt5 derivatives can provide information on its incorporation into elongation complexes *in vivo*. Spt5 migrated as two species, a lower un-phosphorylated form and a slower migrating phosphorylated form (Fig. 10*B*). Previous studies confirmed that the upper species is phosphorylated Spt5 (36, 37, 63). As a control, extracts from cells expressing a version of Spt5 with all the phosphorylation sites in the CTR (CTR Ser \rightarrow Ala) mutated were included to help identify the un-phosphorylated form of Spt5 in gels (37). A few

Spt4/5 Interaction with DNA Prevents Arrest of RNAPII

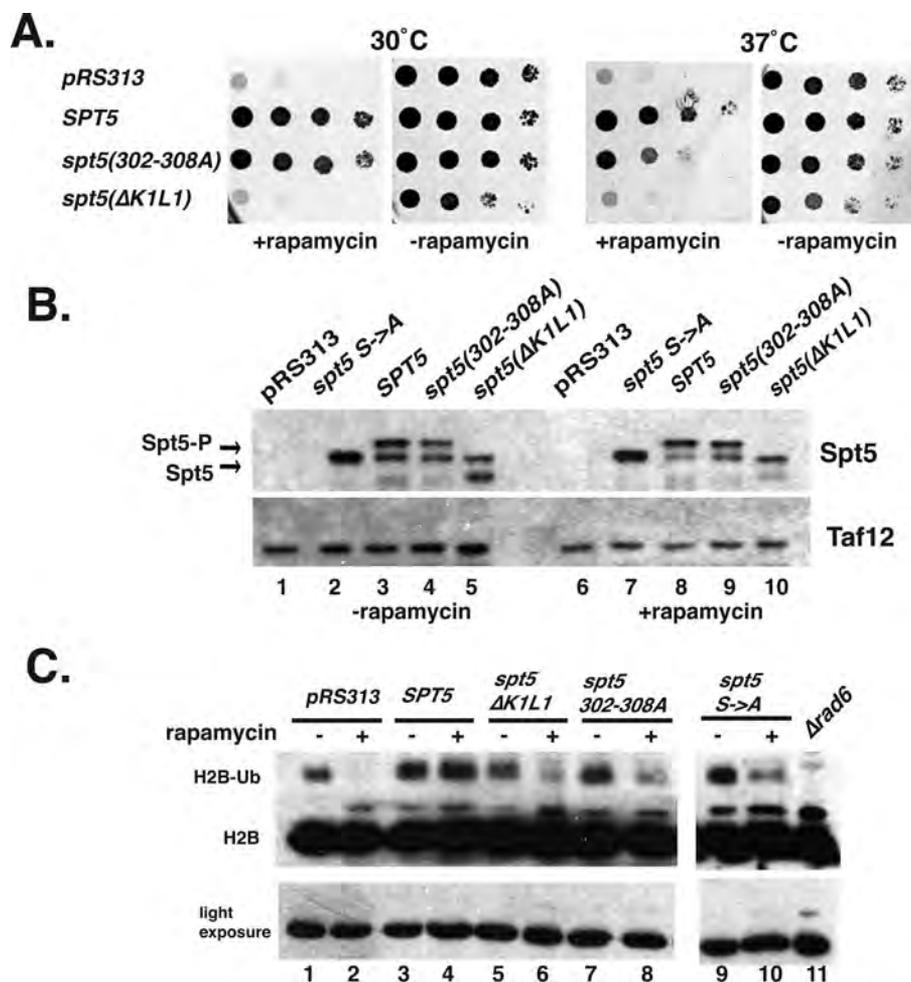


FIGURE 10. Analysis of the growth properties of Spt5 mutants. *A*, anchor-away yeast strains (*SPT5*-FRB) containing either pRS313, pRS313-*SPT5*-HA₃ (*WT*), pRS313-*spt5*(302-308A)-HA₃, or pRS313-*spt5*(ΔK1L1)-HA₃ were spotted onto synthetic media-His ± rapamycin (1 μg/ml) and incubated at either 30 or 37 °C for 3 days. *B*, Western blot analysis of yeast extracts from cells grown in the absence or presence of 1 μg/ml rapamycin. Cells were grown to an A₆₀₀ of 0.7, and then rapamycin was added for 1.5 h. Yeast bearing a plasmid containing an *SPT5* mutant with the phosphorylated serine residues changed to alanines (*spt5*-CTR Ser → Ala) was used to discriminate the phosphorylated from the un-phosphorylated forms of Spt5 in the blot. *C*, analysis of H2B ubiquitylation. Western blot of cell extracts using an antibody to yeast H2B. A light exposure panel (*below*) serves as a control for total histone levels. Extracts from a *rad6*Δ mutant was run on the gel as a control to identify the ubiquitylated form of H2B.

interesting things were noted. First, when rapamycin was added the proportion of the phosphorylated species increased. This makes sense because as the Spt5-FRB leaves the nucleus, it no longer competes with the plasmid copy of Spt5. This increases the proportion of Spt5-HA recruited to genes where it can become phosphorylated. Second, the mutants were expressed to similar levels as the wild type cells. Finally, both mutants were phosphorylated. This suggests that these mutants are incorporated into ECs *in vivo*. This is consistent with the data showing that both mutants bound to RNAPII ECs *in vitro* (Figs. 8 and 9).

The phosphorylation of the CTR of Spt5 recruits Paf1c (RNA polymerase-associated factor 1), which in turn leads to the co-transcriptional ubiquitylation of histone H2B by Rad6 (35, 63–65). H2B ubiquitylation is a mark of active transcription elongation and is used as an indicator of ongoing transcription elongation on a global scale. Ubiquitylated H2B appears as a slower migrating species in Western blots that is dependent on the *RAD6* gene. The Spt5 anchor-away strain containing plasmid copies of wild type or Spt5 mutants were grown to log

phase and then rapamycin was added to half of the culture. Cell extracts were prepared after 90 min and analyzed by Western blotting using an antibody to H2B (Fig. 10C). Treating cells transformed with an empty vector (pRS313) with rapamycin caused a dramatic reduction of H2B ubiquitylation (Fig. 10C, lanes 1 versus 2), but introducing a plasmid containing a wild type copy of *SPT5* fully restored H2B ubiquitylation levels. However, the 302–308A and the ΔK1L1 mutants only partially complemented the loss of H2B ubiquitylation (Fig. 10C, lanes 5 versus 6 and lanes 7 versus 8). The reduced H2B ubiquitylation suggests that transcription elongation is impaired in these *SPT5* mutants and that the basic patch in the NGN of Spt5(302–308) contributes to elongation *in vivo*. As controls, a strain expressing an Spt5 derivative where all of the serines in the CTR were changed to alanines (*spt5* Ser → Ala) was analyzed. As expected, this mutant caused reduced H2B ubiquitylation (35, 63).

Discussion

Changes occur to RNAPII as it transitions from the pre-initiation complex into a productive elongation complex. One

such change is its association with elongation factors, including the primordial elongation factor Spt4/5 (66). The recruitment of DSIF (metazoan Spt4/5 complex) requires a minimal length of mRNA emerging from RNAPII, and this event is thought to be among the earliest changes in the elongation complex during the transition from initiation to elongation into the body of genes. Although it has been shown that Spt5 can bind across the jaw and the clamp of archaeal RNAP, which may close the clamp to encircle DNA to enhance processivity (67), little was known about how the binding of Spt4/5 to the nucleic acid scaffold in ECs affects elongation. The interaction between Spt5 and upstream DNA was predicted from structural modeling studies based on x-ray crystal structures and high resolution cryo-EM of Spt5 homologs (8, 10); however, prior to our study no direct evidence for the binding of Spt5 to the DNA has been presented nor has the significance of these interactions been tested. Here, we used a defined, highly purified system to demonstrate that a surface on the NGN domain of Spt5 contacts the non-template strand in the transcription bubble and that this interaction is important for Spt4/5 to promote elongation by reducing arrest or pausing of RNAPII.

Importance of Spt5-DNA Interactions—It has been shown that RfaH, a NusG homolog, and NusG from eubacteria interact with the NTS of DNA in a sequence-specific manner, where it controls pausing of RNAP (59, 60). Whether or not the interaction between NGN domains of Spt5 and the non-template strand is important for function was unknown, especially because sequence-specific pausing is not known to occur in eukaryotes. We used site-specific photoreactive DNA cross-linking to show for the first time that a highly conserved basic patch on the NGN domain of Spt5 contacts the NTS in the transcription bubble. Contact between the NGN domain of eukaryotic Spt5 and the NTS strongly suggests that this function is conserved throughout evolution, and we propose that it may be a fundamental requirement for Spt5 homologs to regulate RNAP elongation. Recently, Spt4/5 from the archaea *Methanocaldococcus jannaschii* was shown to bind to double-stranded DNA. They described a region of the NGN domain as containing an alkaline surface that is responsible for binding to free DNA that overlaps with residues 302–308 of yeast Spt5 (68). Additionally, mutations of basic amino acids in this region affect the activity of RfaH (69). Thus, the DNA binding properties of the NGN domain and that of its paralog appear to be conserved.

It has been speculated that Spt4/5 plays a role in stabilizing the upstream duplex, assisting in bubble closure (8, 10, 18, 32). Until recently, DNA upstream of the RNAPII elongation complex was a virtual unknown, which lacked structural definition. This led many to view it as a semi-stable region of duplexed DNA, which could be stabilized by Spt4/5. Recent crystallization of the RNAPII complete transcription bubble seems to indicate all of the structural requirements to promote strand re-annealing is located within Rpb1 and Rpb2. In particular, the wedge/hairpin loop within Rpb2 was shown to contact the upstream duplex and likely works with “the arch,” composed of the rudder and FL1 residues, to close the bubble. Mutations in the tip of the loop reduce transcription elongation rates *in vitro*, and the mutants display a number of phenotypes consistent

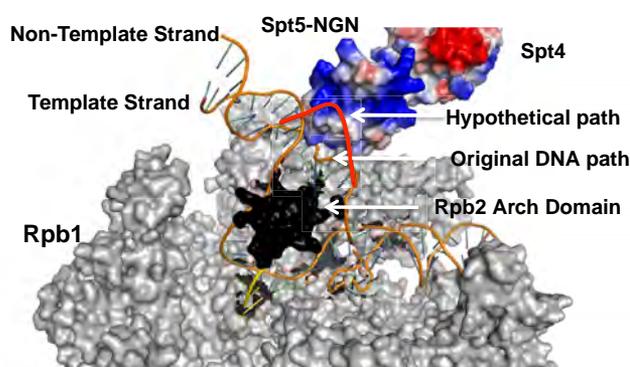


FIGURE 11. Bubble chaperone model of Spt4/5 action on ECs. A model illustrating the bubble chaperone mechanism. Model was generated by overlaying crystal structures of the archaeal clamp fused to archaeal Spt5 (data not shown, PDB code 3QQC), RNAPII with template and non-template strand (PDB code 5C4X) (18), and yeast Spt4/5 (PDB code 2EXU) (74). Spt4 and Spt5 are shown in electrostatic space-filling mode. Gray indicates all subunits of RNAPII except for Rpb2. Rpb2 was omitted from this structure except the Arch domain, which is shown in black. The template and non-template strand are shown as orange, and the RNA is in yellow. Hypothetical DNA path in the presence of Spt4/5 was drawn manually in red over the existing structure.

with impaired elongation in cells (18). Clearly, RNAPII can close the bubble on its own without Spt4/5, and our permanganate footprinting experiment failed to detect changes in the size and/or closure of the transcription bubble in the presence of Spt4/5 (Fig. 6B). However, as stated under the “Results,” such experiments can only provide a static picture of the bubble. Whether or not Spt4/5 stabilizes upstream of DNA during the dynamic act of transcription remains to be tested.

So what is the role of the interaction between the basic patch on the NGN domain of Spt5 and the non-template strand? We propose a modification of the bubble closure models proposed by others, called the bubble chaperone mechanism. In RNAPII, the NTS passes over the arch (rudder and FL1) and rejoins the template strand behind it (18). The basic patch on the NGN provides an additional surface to guide the NTS around the arch, in essence “taming” the NTS to ensure it proceeds along the most productive path to rejoin the NTS (Fig. 11). Additionally, the binding of the NGN domain to the NTS could also prevent the association of the NTS with surfaces on the wall or arch that stabilize a paused/arrested state, thus reducing the pausing/arrest of RNAPII. This could explain why mutating the residues on the basic patch (302–308) of the NGN abolished the anti-arrest properties of Spt4/5. This model also suggests a role for Spt4/5 beyond encircling the DNA and closing the clamp and suggests that Spt4/5 plays a more extensive role in promoting elongation.

Modeling, footprinting, and cross-linking data suggest Spt4/5 contacts DNA as it exits RNAPII, so binding to the NTS strand in the transcription bubble is one part of the equation. Contact between Spt5 and upstream DNA may stabilize the association of RNAPII with DNA. A recent crystal structure of the KOW1-Linker1 (K1L1) region of yeast Spt5 identified a nucleic acid interaction surface, called the positively charged patch (32). The authors proposed, based on genetic evidence that showed mutations in the positively charged patch of K1L1 are synthetically lethal when Spt4 was absent, that the K1L1 wraps up along the DNA on the opposite side of the helix where Spt4 is positioned to act like pinchers in a claw to stabilize the

Spt4/5 Interaction with DNA Prevents Arrest of RNAPII

interaction of Spt4/5 in the EC (32). We found that deleting the entire K1L1 region had a modest ~2-fold effect on the binding of the mutant Spt4/5 complex to ECs (Fig. 8A), which generally supports a role for this domain in stabilizing the interaction of Spt4/5 with elongation complexes. Surprisingly, this mutant displayed very little reduction in anti-pausing/arrest activity and still interacted with the NTS in the transcription bubble. Thus, it appears that the K1L1 is dispensable for most known biochemical activities of the Spt4/5 complex. Interestingly, the Δ K1L1 mutant is expressed and phosphorylated *in vivo* but is unable to support viability (Fig. 10A). These results suggest that the K1L1 has functions in cells beyond preventing RNAPII arrest. Further characterization of the Δ K1L1 mutant *in vivo* would be interesting but is beyond the scope of our study. Compare that with the NGN(302–308A) mutant, which could support viability. The ability of the SPT5(302–308A) mutant to support viability seems at odds with its inability to prevent arrest *in vitro*. However, most *bona fide* elongation factors are non-essential in yeast; SPT5 is the exception because it participates in a number of essential functions such as splicing, capping, and ribosomal RNA production (5, 34, 38, 70, 71). The elongation promoting activity of Spt4/5 may not be an essential function for cell survival under idealized growth conditions, but it imparts a competitive advantage to the organism or is more important during stress. The latter explanation would be consistent with the temperature-sensitive phenotype of the SPT5(302–308A) mutant and the H2B ubiquitylation defects observed in both Spt4/5 mutants.

Extended Regions of Eukaryotic Spt5 May Coordinate Changes in RNAPII Structure—We provide evidence that contact between the NGN and the NTS is important for the activity of Spt4/5. However, this is not likely the only way Spt4/5 promotes elongation. NusG and archaeal Spt5 completely bridge the two lobes of RNAP, interacting with the clamp coil of Rpb1/ β' and the β gate loop of β and RpoB. Mutation of these surfaces compromises the activity of NusG/Spt5 (7, 20). From this study and other data, a model has been proposed that NusG/Spt5 acts to reduce “fluttering” of the two subunits, which in turn controls the movement of the RNAP trigger loop (12, 13, 17). NusG acts to stabilize RNAP in a post-translocated state (11). Likewise, the binding of Spt4/5 to the clamp could affect structures such as the trigger loop or increase processivity by binding across the jaws of RNAPII.

In the case of prokaryotes, the NGN domain is sufficient for association with the elongation complex and controlling elongation activity (7, 72). Our data, and that of others (38), suggest that this is not the case in eukaryotes and that additional domains in Spt5 are required for its association with RNAPII. Elegant *in vivo* site-specific cross-linking studies found that a region of Spt5 containing KOW domains 4 and 5 cross-linked to Rpb4/7, the stalk, of RNAPII (19, 58, 73). Our biochemical data support the important role that KOW domains 4 and 5, and other regions in the C terminus, play in the binding of Spt4/5 to RNAPII (Fig. 4). It is interesting that KOW domains 4 and 5 contact Rpb4/7, because Spt5 cross-links to RNA near the exit channel (this study), and the binding of the emerging transcript to Rpb7 has been observed (58, 73). Interactions between Spt4/5 and ECs may be stabilized through contact with Rpb4/7

and the emerging transcript. Because these regions are physically linked to the NGN domain, this could allow “communication” between the clamp, the stalk of RNAPII, and the emerging transcript.

Spt5 Links Different Features of the EC Complex to the Elongation Process—Our data show that Spt4/5 makes contact with several different structural features of the RNAPII elongation complex required for it to bind to ECs and prevent arrest. It could simply be that Spt4/5 requires additional binding surfaces contributed by the eukaryotic specific regions of Spt5 to stabilize its interaction with elongating RNAPII as it navigates the chromatin barrier or as different co-transcriptional RNA and chromatin-modifying factors are exchanged during elongation. A more interesting idea is that Spt5 serves as a sensor of multiple steps in the transcription process. Spt5 contacts upstream DNA, the transcription bubble, the clamp, the Rpb4/7 module, and the emerging transcript. Spt4/5 may coordinate the translocation of RNAPII with the growth of the emerging transcript, the movement of the Rpb4/7 and clamp, and the tracking of the DNA in the transcription bubble. Conceptually, the ability of the Spt4/5 complex to interact with all of the “moving parts” of the ECs (nucleic acids and RNAPII) provides a mechanical rationale for Spt4/5 to stabilize the RNAPII ground state after each translocation event during elongation. This could prevent arrest by preventing non-productive, uncoordinated movements among the different structures that would cause arrest.

Author Contributions—J. B. C. and J. C. R. designed the studies and wrote the manuscript. J. B. C. performed the experiments and analyzed the data. J. F. provided RNAPII, edited the manuscript, and shared unpublished data.

Acknowledgments—Alan Hinnebusch, Steve Hahn, and Song Tan are recognized for providing plasmids used in this study. Dave Gilmour is acknowledged for input during the course of the work and reading the manuscript. The members of the Center for Eukaryotic Gene Regulation are recognized for their comments and feedback during the completion of this work.

References

1. Werner, F. (2012) A nexus for gene expression—molecular mechanisms of Spt5 and NusG in the three domains of life. *J. Mol. Biol.* **417**, 13–27
2. Hartzog, G. A., and Fu, J. (2013) The Spt4–Spt5 complex: a multi-faceted regulator of transcription elongation. *Biochim. Biophys. Acta* **1829**, 105–115
3. Rondón, A. G., García-Rubio, M., González-Barrera, S., and Aguilera, A. (2003) Molecular evidence for a positive role of Spt4 in transcription elongation. *EMBO J.* **22**, 612–620
4. Ding, B., LeJeune, D., and Li, S. (2010) The C-terminal repeat domain of Spt5 plays an important role in suppression of Rad26-independent transcription coupled repair. *J. Biol. Chem.* **285**, 5317–5326
5. Lindstrom, D. L., Squazzo, S. L., Muster, N., Burckin, T. A., Wachter, K. C., Emigh, C. A., McCleery, J. A., Yates, J. R., 3rd, and Hartzog, G. A. (2003) Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol. Cell. Biol.* **23**, 1368–1378
6. Herbert, K. M., Zhou, J., Mooney, R. A., Porta, A. L., Landick, R., and Block, S. M. (2010) *E. coli* NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *J. Mol. Biol.* **399**, 17–30

7. Hirtreiter, A., Damsma, G. E., Cheung, A. C., Klose, D., Grohmann, D., Vojnic, E., Martin, A. C., Cramer, P., and Werner, F. (2010) Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.* **38**, 4040–4051
8. Klein, B. J., Bose, D., Baker, K. J., Yusoff, Z. M., Zhang, X., and Murakami, K. S. (2011) RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 546–550
9. Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., Buratowski, S., and Handa, H. (1998) DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343–356
10. Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C., and Cramer, P. (2011) Architecture of the RNA polymerase–Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* **30**, 1302–1310
11. Bar-Nahum, G., Epshtein, V., Ruckenstein, A. E., Rafikov, R., Mustaev, A., and Nudler, E. (2005) A ratchet mechanism of transcription elongation and its control. *Cell* **120**, 183–193
12. Nayak, D., Voss, M., Windgassen, T., Mooney, R. A., and Landick, R. (2013) Cys-pair reporters detect a constrained trigger loop in a paused RNA polymerase. *Mol. Cell* **50**, 882–893
13. Touloukhonov, I., Zhang, J., Palangat, M., and Landick, R. (2007) A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol. Cell* **27**, 406–419
14. Cheung, A. C., and Cramer, P. (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249–253
15. Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D., and Kornberg, R. D. (2006) Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**, 941–954
16. Hein, P. P., Kolb, K. E., Windgassen, T., Bellecourt, M. J., Darst, S. A., Mooney, R. A., and Landick, R. (2014) RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat. Struct. Mol. Biol.* **21**, 794–802
17. Sekine, S., Murayama, Y., Svetlov, V., Nudler, E., and Yokoyama, S. (2015) The ratcheted and ratchetable structural states of RNA polymerase underlie multiple transcriptional functions. *Mol. Cell* **57**, 408–421
18. Barnes, C. O., Calero, M., Malik, I., Graham, B. W., Spahr, H., Lin, G., Cohen, A. E., Brown, I. S., Zhang, Q., Pullara, F., Trakselis, M. A., Kaplan, C. D., and Calero, G. (2015) Crystal structure of a transcribing RNA polymerase II complex reveals a complete transcription bubble. *Mol. Cell* **59**, 258–269
19. Li, W., Giles, C., and Li, S. (2014) Insights into how Spt5 functions in transcription elongation and repressing transcription coupled DNA repair. *Nucleic Acids Res.* **42**, 7069–7083
20. Sevostyanova, A., Belogurov, G. A., Mooney, R. A., Landick, R., and Artsimovitch, I. (2011) The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Mol. Cell* **43**, 253–262
21. Yamaguchi, Y., Wada, T., Watanabe, D., Takagi, T., Hasegawa, J., and Handa, H. (1999) Structure and function of the human transcription elongation factor DSIF. *J. Biol. Chem.* **274**, 8085–8092
22. Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999) NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41–51
23. Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D., and Handa, H. (1998) Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription *in vitro*. *EMBO J.* **17**, 7395–7403
24. Bourgeois, C. F., Kim, Y. K., Churcher, M. J., West, M. J., and Karn, J. (2002) Spt5 cooperates with human immunodeficiency virus type 1 Tat by preventing premature RNA release at terminator sequences. *Mol. Cell. Biol.* **22**, 1079–1093
25. Guo, S., Yamaguchi, Y., Schilbach, S., Wada, T., Lee, J., Goddard, A., French, D., Handa, H., and Rosenthal, A. (2000) A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**, 366–369
26. Mason, P. B., and Struhl, K. (2005) Distinction and relationship between elongation rate and processivity of RNA polymerase II *in vivo*. *Mol. Cell* **17**, 831–840
27. Swanson, M. S., Malone, E. A., and Winston, F. (1991) SPT5, an essential gene important for normal transcription in *Saccharomyces cerevisiae*, encodes an acidic nuclear protein with a carboxy-terminal repeat. *Mol. Cell. Biol.* **11**, 3009–3019
28. Missra, A., and Gilmour, D. S. (2010) Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the *Drosophila* RNA polymerase II transcription elongation complex. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11301–11306
29. Cheng, B., and Price, D. H. (2008) Analysis of factor interactions with RNA polymerase II elongation complexes using a new electrophoretic mobility shift assay. *Nucleic Acids Res.* **36**, e135
30. Lasko, P. (2010) Tudor domain. *Curr. Biol.* **20**, R666–R667
31. Charier, G., Couprie, J., Alpha-Bazin, B., Meyer, V., Quéméneur, E., Guérois, R., Callebaut, I., Gilquin, B., and Zinn-Justin, S. (2004) The tudor tandem of 53BP1: a new structural motif involved in DNA and RG-rich peptide binding. *Structure* **12**, 1551–1562
32. Meyer, P. A., Li, S., Zhang, M., Yamada, K., Takagi, Y., Hartzog, G. A., and Fu, J. (2015) Structures and functions of the multiple KOW domains of transcription elongation factor Spt5. *Mol. Cell. Biol.* **35**, 3354–3369
33. Wier, A. D., Mayekar, M. K., Héroux, A., Arndt, K. M., and VanDemark, A. P. (2013) Structural basis for Spt5-mediated recruitment of the Paf1 complex to chromatin. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17290–17295
34. Mayer, A., Schreieck, A., Lidschreiber, M., Leike, K., Martin, D. E., and Cramer, P. (2012) The Spt5 C-terminal region recruits yeast 3' RNA cleavage factor I. *Mol. Cell. Biol.* **32**, 1321–1331
35. Mayekar, M. K., Gardner, R. G., and Arndt, K. M. (2013) The recruitment of the *Saccharomyces cerevisiae* Paf1 complex to active genes requires a domain of Rtf1 that directly interacts with the Spt4–Spt5 complex. *Mol. Cell. Biol.* **33**, 3259–3273
36. Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W. H., Ranish, J., Shokat, K. M., and Hahn, S. (2009) Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Mol. Cell. Biol.* **29**, 4852–4863
37. Qiu, H., Hu, C., Gaur, N. A., and Hinnebusch, A. G. (2012) Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. *EMBO J.* **31**, 3494–3505
38. Viktorovskaya, O. V., Appling, F. D., and Schneider, D. A. (2011) Yeast transcription elongation factor Spt5 associates with RNA polymerase I and RNA polymerase II directly. *J. Biol. Chem.* **286**, 18825–18833
39. Tan, S., Kern, R. C., and Selleck, W. (2005) The pST44 polycistronic expression system for producing protein complexes in *Escherichia coli*. *Protein Expr. Purif.* **40**, 385–395
40. Longtine, M. S., McKenzie, A., 3rd, Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J. R. (1998) Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961
41. Cox, J. S., Chapman, R. E., and Walter, P. (1997) The unfolded protein response coordinates the production of endoplasmic reticulum protein and endoplasmic reticulum membrane. *Mol. Biol. Cell* **8**, 1805–1814
42. Suh, M.-H., Ye, P., Zhang, M., Hausmann, S., Shuman, S., Gnatt, A. L., and Fu, J. (2005) Fcp1 directly recognizes the C-terminal domain (CTD) and interacts with a site on RNA polymerase II distinct from the CTD. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17314–17319
43. Kruk, J. A., Dutta, A., Fu, J., Gilmour, D. S., and Reese, J. C. (2011) The multifunctional Ccr4–Not complex directly promotes transcription elongation. *Genes Dev.* **25**, 581–593
44. Zhang, Z., Fu, J., and Gilmour, D. S. (2005) CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev.* **19**, 1572–1580
45. Dutta, A., Babbarwal, V., Fu, J., Brunke-Reese, D., Libert, D. M., Willis, J., and Reese, J. C. (2015) Ccr4–Not and TFIIS function cooperatively to rescue arrested RNA polymerase II. *Mol. Cell. Biol.* **35**, 1915–1925
46. Babbarwal, V., Fu, J., and Reese, J. C. (2014) The Rpb4/7 of RNA polymerase II is required for carbon catabolite repressor protein 4-negative on TATA (Ccr4–Not) complex to promote elongation. *J. Biol. Chem.* **289**, 33125–33130
47. Dedrick, R. L., and Chamberlin, M. J. (1985) Studies on transcription of 3'-extended templates by mammalian RNA polymerase II. Parameters

Spt4/5 Interaction with DNA Prevents Arrest of RNAPII

- that affect the initiation and elongation reactions. *Biochemistry* **24**, 2245–2253
48. Kadesch, T. R., and Chamberlin, M. J. (1982) Studies of *in vitro* transcription by calf thymus RNA polymerase II using a novel duplex DNA template. *J. Biol. Chem.* **257**, 5286–5295
49. Sluder, A. E., Price, D. H., and Greenleaf, A. L. (1988) Elongation by *Drosophila* RNA polymerase II. Transcription of 3'-extended DNA templates. *J. Biol. Chem.* **263**, 9917–9925
50. Kireeva, M. L., Komissarova, N., Waugh, D. S., and Kashlev, M. (2000) The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J. Biol. Chem.* **275**, 6530–6536
51. Saeki, H., and Svejstrup, J. Q. (2009) Stability, flexibility, and dynamic interactions of colliding RNA polymerase II elongation complexes. *Mol. Cell* **35**, 191–205
52. Reines, D., Chamberlin, M. J., and Kane, C. M. (1989) Transcription elongation factor SII (TFIIS) enables RNA polymerase II to elongate through a block to transcription in a human gene *in vitro*. *J. Biol. Chem.* **264**, 10799–10809
53. Komissarova, N., and Kashlev, M. (1998) Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14699–14704
54. Gu, W., Wind, M., and Reines, D. (1996) Increased accommodation of nascent RNA in a product site on RNA polymerase II during arrest. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6935–6940
55. Zhu, W., Wada, T., Okabe, S., Taneda, T., Yamaguchi, Y., and Handa, H. (2007) DSIF contributes to transcriptional activation by DNA-binding activators by preventing pausing during transcription elongation. *Nucleic Acids Res.* **35**, 4064–4075
56. Kim, D.-K., Inukai, N., Yamada, T., Furuya, A., Sato, H., Yamaguchi, Y., Wada, T., and Handa, H. (2003) Structure-function analysis of human Spt4: evidence that hSpt4 and hSpt5 exert their roles in transcriptional elongation as parts of the DSIF complex. *Genes Cells* **8**, 371–378
57. Burova, E., Hung, S. C., Sagitov, V., Stitt, B. L., and Gottesman, M. E. (1995) *Escherichia coli* NusG protein stimulates transcription elongation rates *in vivo* and *in vitro*. *J. Bacteriol.* **177**, 1388–1392
58. Ujvári, A., and Luse, D. S. (2006) RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat. Struct. Mol. Biol.* **13**, 49–54
59. Yakhnin, A. V., Murakami, K. S., and Babitzke, P. (2016) NusG is a sequence-specific RNA polymerase pause factor that binds to the non-template DNA within the paused transcription bubble. *J. Biol. Chem.* **291**, 5299–5308
60. Artsimovitch, I., and Landick, R. (2002) The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* **109**, 193–203
61. Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. (1989) Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549–561
62. Haruki, H., Nishikawa, J., and Laemmli, U. K. (2008) The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Mol. Cell* **31**, 925–932
63. Zhou, K., Kuo, W. H., Fillingham, J., and Greenblatt, J. F. (2009) Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6956–6961
64. Xiao, T., Kao, C.-F., Krogan, N. J., Sun, Z.-W., Greenblatt, J. F., Osley, M. A., and Strahl, B. D. (2005) Histone H2B ubiquitylation is associated with elongating RNA polymerase II. *Mol. Cell Biol.* **25**, 637–651
65. Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobar, M. S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S., and Greenblatt, J. F. (2002) RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol. Cell Biol.* **22**, 6979–6992
66. Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* **17**, 1272–1278
67. Hartzog, G. A., and Kaplan, C. D. (2011) Competing for the clamp: promoting RNA polymerase processivity and managing the transition from initiation to elongation. *Mol. Cell* **43**, 161–163
68. Guo, G., Gao, Y., Zhu, Z., Zhao, D., Liu, Z., Zhou, H., Niu, L., and Teng, M. (2015) Structural and biochemical insights into the DNA-binding mode of MjSpt4p:Spt5 complex at the exit tunnel of RNAPII. *J. Struct. Biol.* **192**, 418–425
69. Belogurov, G. A., Sevostyanova, A., Svetlov, V., and Artsimovitch, I. (2010) Functional regions of the N-terminal domain of the antiterminator RfaH. *Mol. Microbiol.* **76**, 286–301
70. Pei, Y., and Shuman, S. (2002) Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J. Biol. Chem.* **277**, 19639–19648
71. Anderson, S. J., Sikes, M. L., Zhang, Y., French, S. L., Salgia, S., Beyer, A. L., Nomura, M., and Schneider, D. A. (2011) The transcription elongation factor Spt5 influences transcription by RNA polymerase I positively and negatively. *J. Biol. Chem.* **286**, 18816–18824
72. Sevostyanova, A., and Artsimovitch, I. (2010) Functional analysis of *Thermus thermophilus* transcription factor NusG. *Nucleic Acids Res.* **38**, 7432–7445
73. Andrecka, J., Lewis, R., Brückner, F., Lehmann, E., Cramer, P., and Michaelis, J. (2008) Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 135–140
74. Guo, M., Xu, F., Yamada, J., Egelhofer, T., Gao, Y., Hartzog, G. A., Teng, M., and Niu, L. (2008) Core structure of the yeast Spt4-Spt5 complex: a conserved module for regulation of transcription elongation. *Structure* **16**, 1649–1658

Preferential Phosphorylation on Old Histones during Early Mitosis in Human Cells^{*[S]}

Received for publication, March 7, 2016, and in revised form, May 10, 2016 Published, JBC Papers in Press, May 19, 2016, DOI 10.1074/jbc.M116.726067

Shu Lin^{†1}, Zuo-Fei Yuan[‡], Yumiao Han[‡], Dylan M. Marchione[§], and Benjamin A. Garcia^{‡2}

From the [†]Epigenetics Program, Department of Biochemistry and Biophysics, and the [§]Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104

How histone post-translational modifications (PTMs) are inherited through the cell cycle remains poorly understood. Canonical histones are made in the S phase of the cell cycle. Combining mass spectrometry-based technologies and stable isotope labeling by amino acids in cell culture, we question the distribution of multiple histone PTMs on old *versus* new histones in synchronized human cells. We show that histone PTMs can be grouped into three categories according to their distributions. Most lysine mono-methylation and acetylation PTMs are either symmetrically distributed on old and new histones or are enriched on new histones. In contrast, most di- and tri-methylation PTMs are enriched on old histones, suggesting that the inheritance of different PTMs is regulated distinctly. Intriguingly, old and new histones are distinct in their phosphorylation status during early mitosis in the following three human cell types: HeLa, 293T, and human foreskin fibroblast cells. The mitotic hallmark H3S10ph is predominantly associated with old H3 at early mitosis and becomes symmetric with the progression of mitosis. This same distribution was observed with other mitotic phosphorylation marks, including H3T3/T6ph, H3.1/2S28ph, and H1.4S26ph but not S28/S31ph on the H3 variant H3.3. Although H3S10ph often associates with the neighboring Lys-9 di- or tri-methylations, they are not required for the asymmetric distribution of Ser-10 phosphorylation on the same H3 tail. Inhibition of the kinase Aurora B does not change the distribution despite significant reduction of H3S10ph levels. However, K9me2 abundance on the new H3 is significantly reduced after Aurora B inhibition, suggesting a cross-talk between H3S10ph and H3K9me2.

In eukaryotes, histone proteins facilitate the packaging of DNA molecules. The DNA double helix wraps around histone octamers to form nucleosomes. A histone octamer contains two copies of each core histone H3, H4, H2A, and H2B. A 5th histone, histone H1, is associated with the linker DNA which lies between the nucleosomes. Canonical histone proteins are

cell cycle-dependent and are produced in S phase (1, 2), whereas cell cycle-independent histone variants (*e.g.* H3.3) are synthesized throughout the cell cycle (3). Histone proteins carry numerous post-translational modifications (PTMs)³ that are involved in multiple functions such as epigenetic regulation of transcription, DNA damage repair, and cell cycle progression (4, 5). To maintain lineage identity and to guide proper transcription, cells must replicate PTMs from old histones onto new histones at each cell division. Major efforts have been devoted to understanding how histones themselves are transmitted through the DNA replication fork in S phase (6). In principle, the newly deposited nucleosomes could contain entirely old or newly synthesized histone proteins, or a mixture of both. Accumulating evidence suggests that most H3/H4 tetramers remain intact, with the exception of some H3.3/H4 tetramers, indicating that nucleosomes should contain either new or old H3 and H4 rather than a mixture. Conversely, H2A/H2B dimers exchange freely during replication (6–8).

Determining the PTM profiles of newly deposited nucleosomes after replication, and how these profiles differ between old and new histone proteins, will help elucidate the mechanisms of histone PTM inheritance during the cell cycle. We and others have reported histone lysine methylation kinetics throughout the human cell cycle (9, 10). Although histone PTM inheritance is completed after one cell cycle, important repressive marks like H3K9me3 and H3K27me3 are not fully replenished until the next G₁ phase (9). Groth and co-workers (11) reported an overview of multiple histone PTMs at the replication fork and made very similar observations. However, much remains unclear about how different histone PTMs are transmitted through mitosis.

Interestingly, a number of histone PTMs regulate cell cycle stage-specific processes and therefore may not need to be inherited from the old histones to new histones. For example, histone H3K56ac was shown to be added onto new histones during S phase and rapidly erased in G₂ phase (12, 13). Mono-methylation of H4K20 is temporally added by G₂ and M phase-specific activities of the methyltransferase PR-Set7/SET8 and is linked to cell cycle progression (14). Furthermore, a handful of histone phosphorylation (ph) marks are highly abundant in mitosis and are present at very low levels in the interphase, including H3S10ph, H3S28ph, H3T3ph, H1.4S26ph, etc. (15–

^{*} This work was supported by a National Science Foundation CAREER award and National Institutes of Health Grants GM110174 and AI118891 (to B. A. G.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

^[S] This article contains [supplemental Tables S1–S9](#).

¹ Present address: Thayer School of Engineering, Dartmouth College, Hanover, NH 03755.

² To whom correspondence should be addressed. Tel.: 215-573-9423; Fax: 215-898-4217; E-mail: bgarci@mail.med.upenn.edu.

³ The abbreviations used are: PTM, post-translational modification; SILAC, stable isotope labeling by amino acids in cell culture; HFF, human foreskin fibroblast; ACN, acetonitrile; Hes, hesperadin; GSC, germ line stem cell; ac, acetylation; ph, phosphorylation; PI, propidium iodide; AdoMet, S-adenosylmethionine.

21). The major kinase for these histone phosphorylation marks is Aurora B, which is part of the chromosomal passenger complex and plays essential roles in chromosome condensation, segregation, and cytokinesis during mitotic progression (22). Aurora B phosphorylates histones directly (17, 21, 23–26) or indirectly through activation of another kinase Haspin (27). The levels of these phosphorylation marks peak after the new histones are synthesized in S phase; therefore, they are not likely being transmitted from old to new histones. However, it remains unclear whether these histone phosphorylation marks play a role in facilitating epigenetic inheritance of other PTMs.

We report here a systematic analysis of the distribution of histone PTMs in mitosis. We show that most histone Kme2/3s were biased toward old histones, consistent with previous studies (9–11). H3K4me2/3, however, was symmetrically distributed on old and new H3. We also show that most Kme1 and Kac events were either symmetric or enriched on new histones, with the exception of H4K5acK8acK12acK16ac (H4 4–17 4-ac). Surprisingly, although the mitotic histone phosphorylation marks do not need to be inherited, they were predominantly associated with the old histones in early mitosis and only became more symmetrically distributed in late mitosis. This phenomenon was observed for four histone phosphorylation marks, including H3S10ph on both canonical histone H3.1/2 and the variant H3.3, S28ph on H3.1/2, H3T3/T6ph, and S26ph on a linker histone H1.4. In contrast, S28/S31ph on H3.3 was distributed symmetrically on old and new histones. We additionally demonstrate that the H3K9 residue is not required for the asymmetric distribution of Ser-10 phosphorylation on the same H3 tail.

Results

Systematic Analysis of the Distribution of Histone PTMs in Mitosis—To investigate the distribution of histone PTMs on old *versus* new histones during mitosis, we used pulse-SILAC (stable isotope labeling by amino acids in cell culture) (28) followed by mass spectrometry techniques. HeLa cells were synchronized at the G₁/S boundary by double thymidine block (Fig. 1A). Typically, the majority of HeLa cells entered G₂/M by 7–8 h after release (Fig. 1, D and E). Upon release, we cultured the cells with SILAC media containing ¹³C- and ¹⁵N-labeled arginine. Because canonical histone proteins are only synthesized in S phase (1, 2), any heavy Arg-labeled histones were therefore newly synthesized histones and could be detected by the subsequent mass spectrometry analysis. In addition, we included L-[methyl-¹³C,¹⁵N]methionine in some experiments to enable characterization of new methylation events. Methionine can be converted into S-adenosylmethionine (AdoMet), the only methyl donor in the cell (29). Therefore, heavy-labeled methionine can mark both new methylation events as well as new proteins (9). Incorporation of the supplemented heavy-labeled amino acids may not happen immediately because they take time to enter the cell and get incorporated into proteins. To test whether the cells can utilize supplemented stable isotope-labeled amino acids rapidly enough, we performed metabolomics analysis in one pulse-SILAC experiment. As shown in Fig. 1C, the majority of cellular arginine and AdoMet was already heavy-labeled by 2 h. By 4.5 h, both of them had

reached a plateau of 98.9 and 90.8%, respectively. Minimum increase of extra labeling was achieved after 4.5 h. The maximum labeling efficiency was 99.1% for heavy arginine and 93.8% for heavy AdoMet. We also monitored heavy proline levels in addition to supplementing extra light proline in the medium because arginine can be converted into proline (30). Only a trace amount (<0.3%) of heavy proline was detected throughout the labeling time course, which would not likely affect our data analyses. Based on these data, we concluded that the pulse-SILAC was fast enough for labeling G₂/M histones.

As shown in Fig. 2A, all canonical histone proteins quantified, including H3.1/2, H4, and H1.4, showed very similar labeling between 8.5 and 11 h, suggesting the majority of histone protein synthesis had finished by 8.5 h in our experimental paradigm. Consistently, flow cytometry measurements of DNA contents showed the majority of cells had finished S phase and entered G₂/M phase by 8.5 h (Fig. 1D). However, the cell cycle-independent H3 variant H3.3 had a lower labeling rate compared with other histones, and the labeling continued to increase from 8.5 to 11 h (Fig. 2A), in agreement with the fact that H3.3 protein synthesis happens throughout the cell cycle (3). In general, old histones were over-represented in these pulse-SILAC experiments, likely due to recycling of cellular metabolites during histone synthesis (9). Taking the protein labeling efficiencies into consideration, we presented data in a normalized distribution manner as shown in Equation 1.

normalized PTM distribution

$$= \frac{\text{PTM on old histone/PTM on new histone}}{\text{old histone/new histone}} \quad (\text{Eq. 1})$$

The larger the number, the more asymmetric a particular PTM is biased toward the old histones. Fig. 2B shows a distribution map for a total of 57 histone PTM statuses (including unmodified) and their relative abundance in mitosis. The PTM statuses could be clustered into three categories according to their normalized distributions (Table 1) as follows: symmetrically distributed (29 statuses), enriched on new histones (8 statuses), and enriched on old histones (20 statuses). For PTMs requiring inheritance from the old histones to new histones, symmetric distributions implied that the inheritance had been completed before onset of mitosis. In contrast, enrichments on the old histones indicated that inheritance of such PTMs persisted into mitosis and even the next G₁.

Interestingly, the same type of histone PTMs tended to have a similar distribution. As listed in Table 1, most lysine mono-methylations and acetylations were either symmetrically distributed or enriched on the new histones. In contrast, most lysine di- and tri-methylations were enriched on the old histones, including H4K20me2/3, H3K79me2/3, H3K9me3, H3.1/2K27me2/3, H3.1/2K36me2/3, H3.3K27me3, and H3.3K36me2/3. These results extended previous findings from us and others demonstrating that new H3K9me3 and H3K27me3 are synthesized rather slowly and do not complete until next G₁ (9–11). However, a few histone PTMs did not follow the general rules. For instance, K4me2 and me3 were symmetrically distributed,

Asymmetric Histone Phosphorylation in the Human Cell Cycle

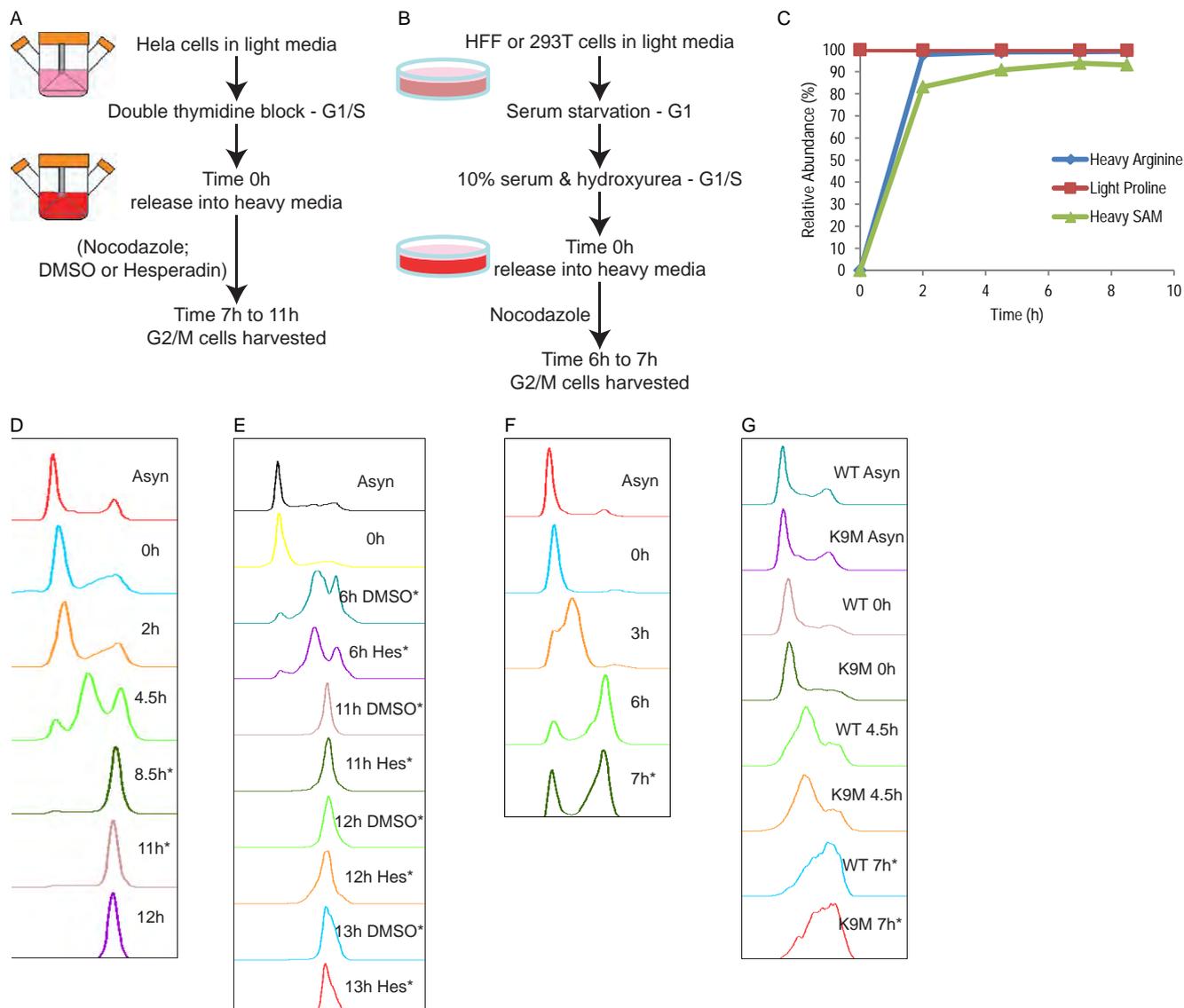


FIGURE 1. Cell synchronization and SILAC. *A*, pulse-SILAC experimental paradigm for HeLa cells. *B*, pulse-SILAC experimental paradigm for HFF and 293T cells. *C*, metabolite analysis of one pulse-SILAC experiment in HeLa cells. Relative abundance of heavy arginine, light proline, and heavy AdoMet (*SAM*) was shown in a time course manner. Sum of both light and heavy forms of the same metabolite was set to be 100%. *D*, example flow cytometry data of a HeLa cell synchronization experiment shown in Fig. 2. *E*, example flow cytometry data showing HeLa cell synchronization and DMSO and Hes treatments; no apparent defects were found in Hes-treated cells. *F*, example flow cytometry data showing HFF cell synchronization. *G*, flow cytometry data showing synchronization experiment conducted in 293T cells transfected with either wild type or K9M H3.3 transgenes. Both cell lines had very similar DNA content distribution across different time points. * shows time points when cells were harvested for mass spectrometry analyses.

whereas H3K4me1 was enriched on old H3 (Table 1). This result demonstrates for the first time that me2/3 marks could be restored before mitosis and suggests that the mechanisms underlying the inheritance of H3K4me and other Kme marks are distinct. Another example is that H3K79me1/2/3s were all asymmetrically distributed on the old H3 in mitosis (Table 1), suggesting that the sole methyltransferase for H3K79, Dot1l (31), was preferentially recruited to old histone H3. Interestingly, Dot1/Dot1l deficiency causes various cell cycle defects in human cells and in other organisms (32), which might be related to its enrichment on the old H3. In addition, although K27me2 was asymmetrically distributed on the old histone H3.1/2, its counterpart was symmetrically distributed on H3.3. Finally, we found that H4K5acK8acK12acK16ac was enriched

on old H4 at both time points, in contrast to most lysine acetylation marks (Table 1).

Enrichment of Histone Phosphorylation Marks on Old Histones in Mitosis—We also investigated the distribution of serine/threonine phosphorylations, which in principle do not require inheritance as they are mostly mitosis-specific (15–21). Surprisingly, four out of five Ser/Thr phosphorylations analyzed, namely H3S10ph, H3.1/2S28ph, H3T3/T6ph, and H1.4S26ph, were enriched on the old histones in early mitosis (Table 1 and Fig. 2*B*). The only exception was H3.3S28/S31ph, which was symmetrically distributed at both time points (Table 1 and Fig. 2*B*).

As H3S10ph was the most abundant form of histone Ser/Thr phosphorylation (Table 1 and Fig. 2*B*), we first confirmed the

Asymmetric Histone Phosphorylation in the Human Cell Cycle

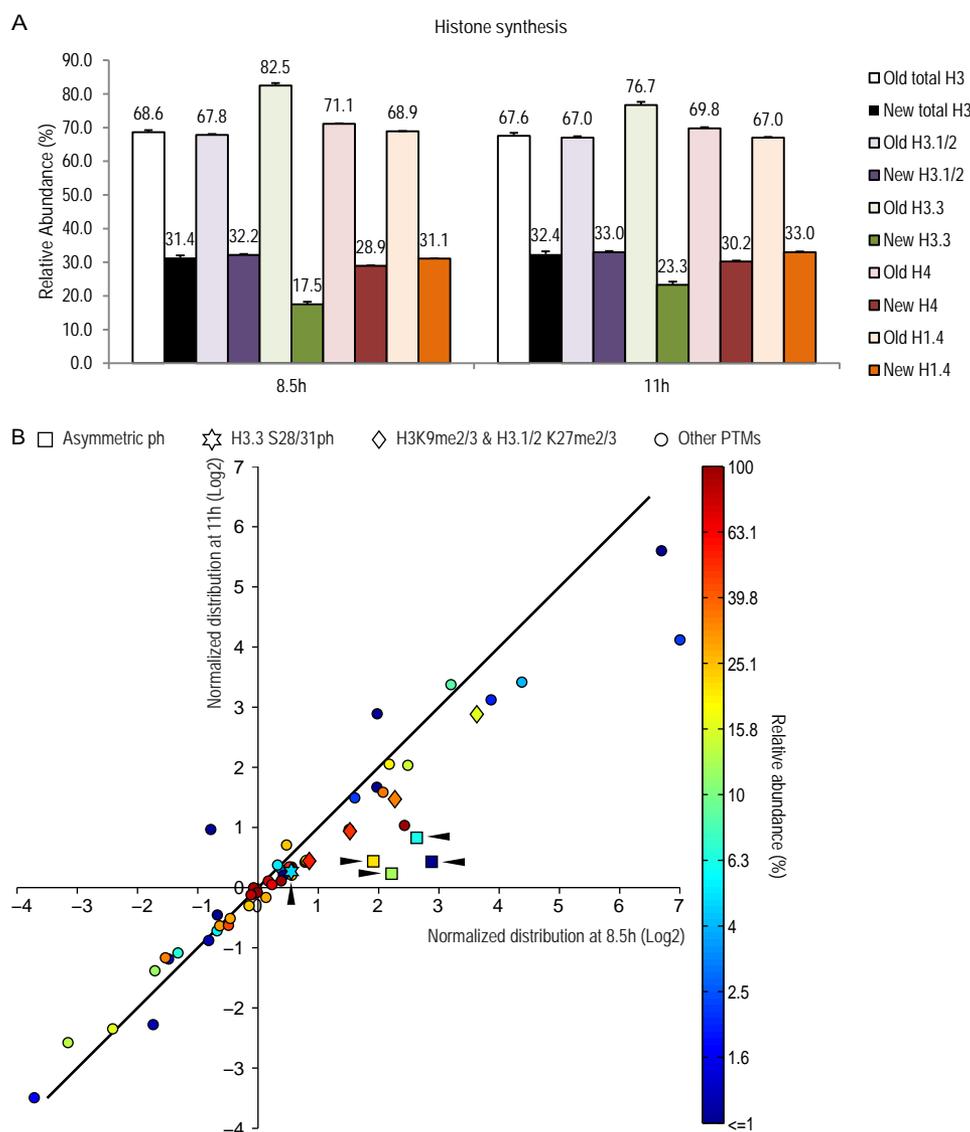


FIGURE 2. Distribution pattern of histone PTMs in mitosis. *A*, labeling of histone proteins by pulse-SILAC experiment. For the relative abundance of old versus new histone H3 (total H3, H3.1/2, and H3.3), H4 and H1.4 were generated from two time points, 8.5 and 11 h, respectively. Fig. 1*D* shows that cells had finished S phase and entered G₂/M phase at both time points. Numbers above each bar show average of three MS runs. Error bars show standard error. *B*, dynamic distribution of 57 histone PTM statuses in mitosis, including 16 unmodified statuses and 41 PTMs. All values used in this figure are listed in Table 1. The x and y axes show normalized distribution of histone PTM statuses at 8.5 and 11 h, respectively. The data were collected from the same HeLa cell pulse-SILAC experiment shown in *A*. The diagonal line indicates identical normalized distribution between the two time points. Each data dot was color-coded based on their averaged relative abundance between the two time points (Table 1). The color bar on the right was generated by log(2) value of the relative abundance (%). Square, asymmetric phosphorylation (ph) marks; star, H3.3S28/31ph; rhombus, H3K9me2/3 and H3.1/2 K27me2/3; circle, other PTMs. The arrowheads show the positions of five ph marks.

temporal pattern of its distribution in HeLa cells. Fig. 3*A* shows a clear inverse correlation of the asymmetry of H3S10ph against harvesting time, from six independent experiments with a total of 12 time points from the onset of mitosis. One of these experiments was conducted using heavy-to-light pulse labeling and produced a similar result as the other light-to-heavy experiments, ruling out any potential artifacts from SILAC labeling. In late G₂ phase and early mitosis (6–7 h) when the H3S10ph signal first appears (15), a much smaller proportion of the new H3 than the old H3 was phosphorylated (Fig. 3*B*). As mitosis progresses, the level of H3S10ph increased on both old and new H3 with a clear delay on the new H3 (Fig. 3, *B* and *C*). Eventually the level of Ser-10 phosphorylation

reached the same proportion on both old and new H3 (11 h and beyond, Fig. 3*B*).

In addition to HeLa cells, we examined the distribution of H3S10ph in two additional cell lines, 293T cells and the primary cell line human foreskin fibroblast (HFF) cells (33). Fig. 1, *F* and *G*, shows that the majority of 293T and HFF cells were synchronized, although to a less extent than the HeLa cells (Fig. 1*D*). Because of the incomplete synchronization of these two cell types, we only collected samples in early time points. As shown in Figs. 3 and 4, we performed two independent synchronization experiments for HFF cells and collected samples at 6 and 7 h, respectively. We performed one synchronization experiment for 293T cells and collected samples at 7 h. Importantly,

Asymmetric Histone Phosphorylation in the Human Cell Cycle

TABLE 1

Abundance and distribution pattern of histone PTMs in mitosis

Listed are data used to generate Fig. 2B. Data were collected from the same pulse-SILAC experiment shown in Fig. 2, A and B, and Fig. 3, A–F. 3rd and 4th columns show relative abundance of each PTM status at 8.5 and 11 h, respectively. The standard errors of multiple MS runs are also shown. 5th column shows average relative abundance between 8.5 and 11 h. 6th and 7th columns show normalized distributions of PTM statuses. All normalized distributions were calculated by the following formula: normalized distribution of a PTM status = (% on old histone/% on new histone)/(old histone/new histone. The categories (shown in 1st column) were generated using arbitrary cutoff; statuses falling between -1, -1 and 1, 1 were considered to be symmetric. Among the rest, the ones fell in the first quadrant were considered to be enriched on old histone, and the ones that fell in the fourth quadrant were considered to be enriched on new histone. The PTM statuses were listed in a descending order according to the value in the 4th column, normalized distribution at 8.5 h (log2).

Category	PTM status	Relative abundance (%)			Normalized distribution (log2)	
		8.5 h	11 h	Average	8.5 h	11 h
Enriched on old histone	H4K20me3	2.29 ± 0.06	1.90 ± 0.08	2.09	7.00 ^a	4.12
	H3K79me3	0.16 ± 0.01	0.67 ± 0.04	0.41	6.69	5.60
	H3.1/2K36me3	3.70 ± 0.06	3.94 ± 0.07	3.82	4.38	3.42
	H3K79me2	1.75 ± 0.15	1.98 ± 0.06	1.86	3.87	3.13
	H3.1/2K27me3	16.11 ± 0.22	15.47 ± 0.21	15.79	3.63	2.89
	H3.3K36me3	7.83 ± 0.25	8.29 ± 0.22	8.06	3.20	3.38
	H3T3/T6ph	0.39 ± 0.03	0.40 ± 0.09	0.39	2.88	0.44
	H3.1/2S28ph	4.51 ± 0.02	7.47 ± 0.22	5.99	2.63	0.83
	H3.3K27me3	16.12 ± 2.63	11.80 ± 1.75	13.96	2.48	2.04
	H4K20me2	94.69 ± 0.09	95.44 ± 0.26	95.06	2.43	1.04
	H3K9me3	33.29 ± 0.39	31.39 ± 0.24	32.34	2.27	1.48
	H1.4S26ph	5.43 ± 0.49	18.51 ± 0.87	11.97	2.22	0.24
	H3K79me1	20.49 ± 0.23	16.74 ± 0.71	18.61	2.18	2.06
	H3.3K36me2	37.27 ± 2.23	32.00 ± 1.98	34.63	2.07	1.59
	H3K18me1	0.66 ± 0.07	0.70 ± 0.11	0.68	1.98	2.89
	H4(4–17) four-ac	0.40 ± 0.04	0.14 ± 0.02	0.27	1.97	1.67
	H3S10ph	13.79 ± 1.17	26.96 ± 0.54	20.38	1.91	0.44
	H3K4me1	1.92 ± 0.04	2.28 ± 0.05	2.10	1.61	1.50
	H3.1/2K27me2	46.90 ± 1.11	45.03 ± 0.98	45.96	1.52	0.94
	H3.1/2 K36me2	37.68 ± 0.71	34.51 ± 0.91	36.10	1.51	0.97
	Symmetric	H3K9me2	50.80 ± 0.64	51.28 ± 0.58	51.04	0.85
H3.1/2K36me1		17.29 ± 0.14	19.34 ± 0.15	18.31	0.80	0.46
H3K14un		73.49 ± 1.11	81.97 ± 0.82	77.73	0.78	0.43
H3S10un		86.21 ± 1.17	73.04 ± 0.54	79.62	0.56	0.35
H3K14ac		26.50 ± 1.11	17.90 ± 0.82	22.20	0.56	0.21
H3.3S28/S31ph		2.46 ± 0.03	6.37 ± 0.29	4.41	0.55	0.27
H3.3K27me2		50.72 ± 3.45	47.76 ± 1.20	49.24	0.52	0.35
H3.3K27me1		21.68 ± 0.56	26.14 ± 0.64	23.91	0.47	0.71
H3K23me1		0.70 ± 0.05	0.36 ± 0.04	0.53	0.42	0.21
H3K4me2		4.70 ± 0.20	0.85 ± 0.19	2.77	0.39	0.30
H3.1/2S28un		95.49 ± 0.02	92.53 ± 0.22	94.01	0.38	0.12
H1.4K25me1		7.20 ± 0.37	2.72 ± 0.06	4.96	0.33	0.38
H4(4–17) un		61.30 ± 0.76	67.21 ± 2.34	64.25	0.23	0.05
H3(18–26) un		67.93 ± 1.63	72.02 ± 1.19	69.97	0.17	0.11
H3.3K36me1		24.59 ± 1.23	26.63 ± 1.40	25.61	0.13	-0.16
H3.3S28/S31un		97.54 ± 0.03	93.63 ± 0.29	95.59	-0.01	-0.08
H1. K25un		92.80 ± 0.37	97.28 ± 0.06	95.04	-0.02	-0.01
H3T3/T6un		99.61 ± 0.03	99.60 ± 0.09	99.61	-0.04	-0.06
H3K4un		92.23 ± 0.25	95.16 ± 0.52	93.70	-0.07	-0.10
H1.4S26un		94.57 ± 0.49	81.49 ± 0.87	88.03	-0.07	0.00
H3K79un		77.61 ± 0.35	80.62 ± 0.64	79.11	-0.12	-0.12
H3.1/2K27me1	21.65 ± 0.69	23.64 ± 0.54	22.65	-0.15	-0.30	
H4(4–17) one-ac	28.66 ± 0.87	26.11 ± 2.09	27.39	-0.46	-0.51	
H3.1/2K36un	41.33 ± 0.77	42.21 ± 0.90	41.77	-0.49	-0.62	
H3K18/23ac	29.49 ± 1.53	26.17 ± 1.13	27.83	-0.64	-0.63	
H3K9ac	0.53 ± 0.05	0.38 ± 0.09	0.45	-0.68	-0.45	
H3K9me1	4.94 ± 0.23	5.40 ± 0.05	5.17	-0.68	-0.71	
H3K4me3	0.77 ± 0.04	1.32 ± 0.22	1.04	-0.78	0.97	
H3K18ac23ac	1.54 ± 0.01	0.86 ± 0.08	1.20	-0.82	-0.87	
Enriched on new histone	H4(4–17) two-ac	7.45 ± 0.30	5.45 ± 0.33	6.45	-1.33	-1.08
	H4(4–17) three-ac	2.19 ± 0.04	1.08 ± 0.05	1.64	-1.49	-1.18
	H3.3K36un	30.32 ± 1.08	33.08 ± 0.41	31.70	-1.54	-1.16
	H3K9un	10.45 ± 0.80	11.54 ± 0.75	11.00	-1.71	-1.38
	H4K20me1	0.94 ± 0.04	1.45 ± 0.13	1.19	-1.74	-2.27
	H3.1/2K27un	15.35 ± 0.41	15.74 ± 0.76	15.55	-2.41	-2.34
	H3.3K27un	11.48 ± 1.04	14.29 ± 0.70	12.88	-3.15	-2.57
	H4K20un	2.08 ± 0.03	1.21 ± 0.10	1.65	-3.71	-3.49

^a The H4K20me3 was 100% on the old H4 at 8.5 h so the normalized distribution was ∞. It was arbitrarily assigned with a log2 value of 7 (larger than all the other values in the table) for plotting in Fig. 2B.

even with the less efficient cell cycle synchronization, H3S10ph was predominantly associated with the old H3 in both 293T cells and HFF cells (Fig. 3A). Taken together, we showed that Ser-10 phosphorylation was biased toward old histone H3 at early mitosis in all three cell lines tested and was likely common in other human cell types.

The other three asymmetrically distributed Ser/Thr phosphorylation marks shared the same temporal pattern as H3S10ph in HeLa cells. As shown in Fig. 3D, H3T3/T6ph, H3.1/2S28ph, and H1.4S26ph were enriched on old histones at earlier time points (6–8.5 h) and became more symmetric at 11 h. In contrast, Ser-28/Ser-31 phosphorylation on a cell cycle-inde-

Asymmetric Histone Phosphorylation in the Human Cell Cycle

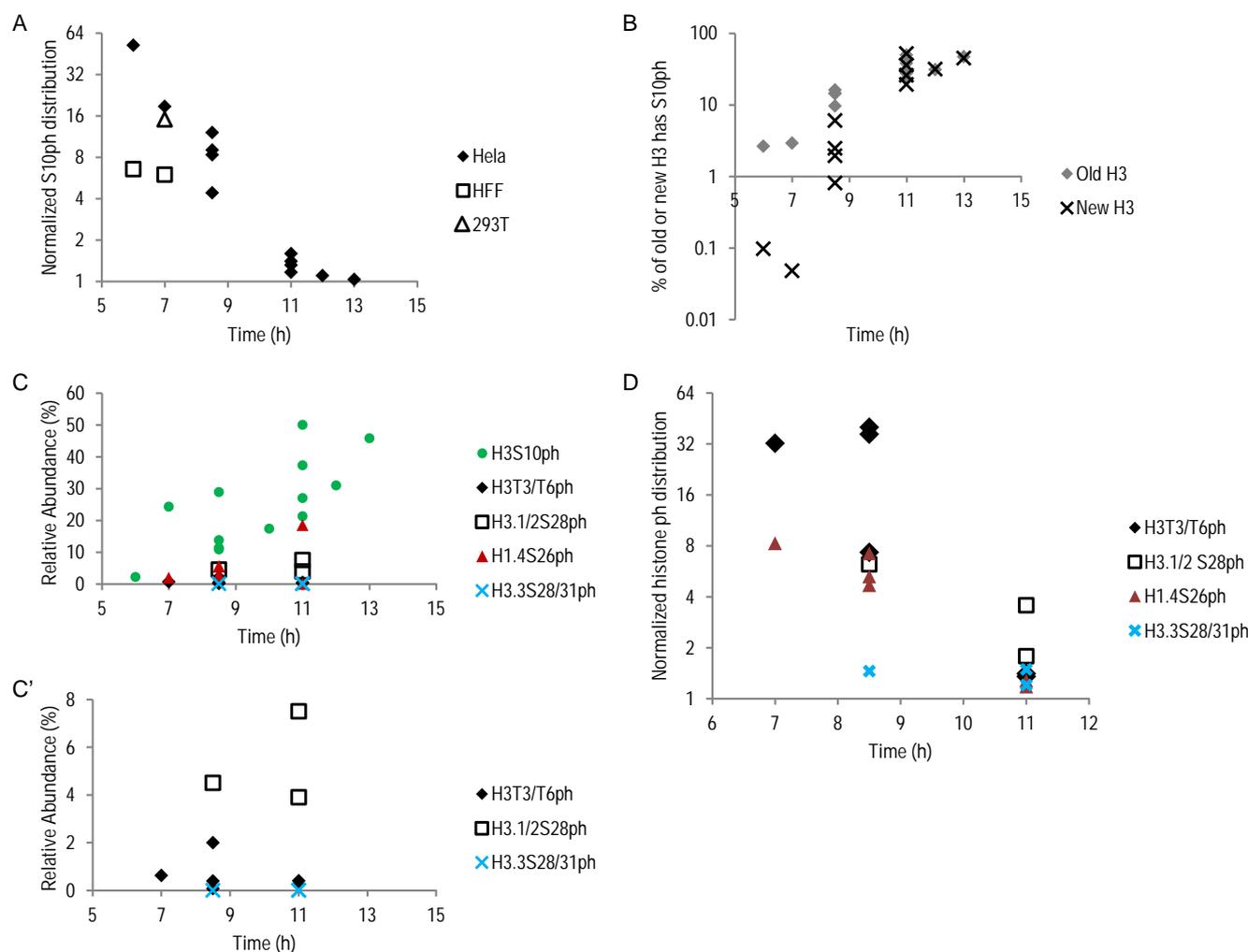


FIGURE 3. Histone phosphorylation marks were enriched on old histones in early mitosis. *A*, normalized H3S10ph distribution from three cell types: *rhombus*, six independent pulse-SILAC experiments with a total of 12 time points from HeLa cells; *square*, two independent pulse-SILAC experiments from HFF cells; *triangle*, one pulse-SILAC experiment from 293T cells. Normalized Ser-10 phosphorylation distribution = (S10ph on old H3/S10ph on new H3)/(old H3/new H3). y axis is in log₂ scale. *B*, Ser-10 phosphorylation on new H3 had a lagged phase compared with Ser-10 phosphorylation on old H3. y axis is in log₁₀ scale. *C*, relative abundance of histone phosphorylation marks in the pulse-SILAC experiments. *C*, all five phosphorylation marks are shown; *C'*, three less abundant phosphorylation marks are shown. *D*, normalized distribution of four additional histone phosphorylation marks. H3T3/T6ph, H3.1/2S28ph, and H1.4S26ph were enriched on old histones in early mitosis and became more symmetric in late mitosis, very similar to the distribution pattern of H3S10ph as shown in *A*. In contrast, distribution of H3.3S28/31ph was symmetric throughout mitosis. *B–D*, data for H3S10ph were from the same HeLa experiments shown in *A*. The data for H3T3/T6ph and H1.4S26ph were from four independent HeLa experiments with a total of six time points. The data for H3.1/2S28ph and H3.3S28/31ph were from two experiments with three time points. H3.3S28/31ph was only detected in TiO₂-enriched samples but not in input samples. Therefore the relative abundance was set to be 0% in *C*.

pendent histone H3 variant, H3.3, was symmetric throughout mitosis (Figs. 2*B* and 3*D*). We concluded that the time-dependent asymmetric distribution of histone phosphorylations was universal among canonical histones, including both core histone (H3) and linker histone (H1.4), although the abundance of these phosphorylation events was very different in mitosis (Table 1 and Figs. 2*B* and 3*C*). These findings suggested the existence of a mechanism that distinguished old *versus* new histones at early mitosis, which led to enrichment of Ser/Thr phosphorylation on old histones.

H3K9 Was Not Required for Asymmetric Distribution of the Neighboring Ser-10 Phosphorylation on the Same H3 Tail—The enrichment of serine phosphorylation on old histone resembled the pattern of di- and tri-methylations on lysine residues (Fig. 2*B* and Table 1). Interestingly, the me_{2/3} on Lys-9 often coexists with H3S10ph on the same histone tails. As shown in

Fig. 4*A*, the majority of the Lys-9 residue next to a phosphorylated Ser-10 phosphorylation was di- or tri-methylated in all three cell types we examined. We therefore sought to test whether the methylation status of the lysine residues could affect the phosphorylation status of the serine residues next to them.

To test whether H3K9 methylation affects the asymmetric Ser-10 phosphorylation distribution, we performed pulse-SILAC experiments in a 293T cell line carrying an H3.3K9M mutant transgene (34). As shown in Fig. 4, *B* and *C*, the distribution of Ser-10 phosphorylation on the mutant protein (MS10ph) was also enriched on the old mutant H3.3, strongly suggesting that the Lys-9 residue was not required for asymmetric distribution of Ser-10 phosphorylation on the same H3 tail. Furthermore, this result also suggested that Ser-10 phosphorylation on both H3.1/2 and H3.3 were enriched on old H3,

Asymmetric Histone Phosphorylation in the Human Cell Cycle

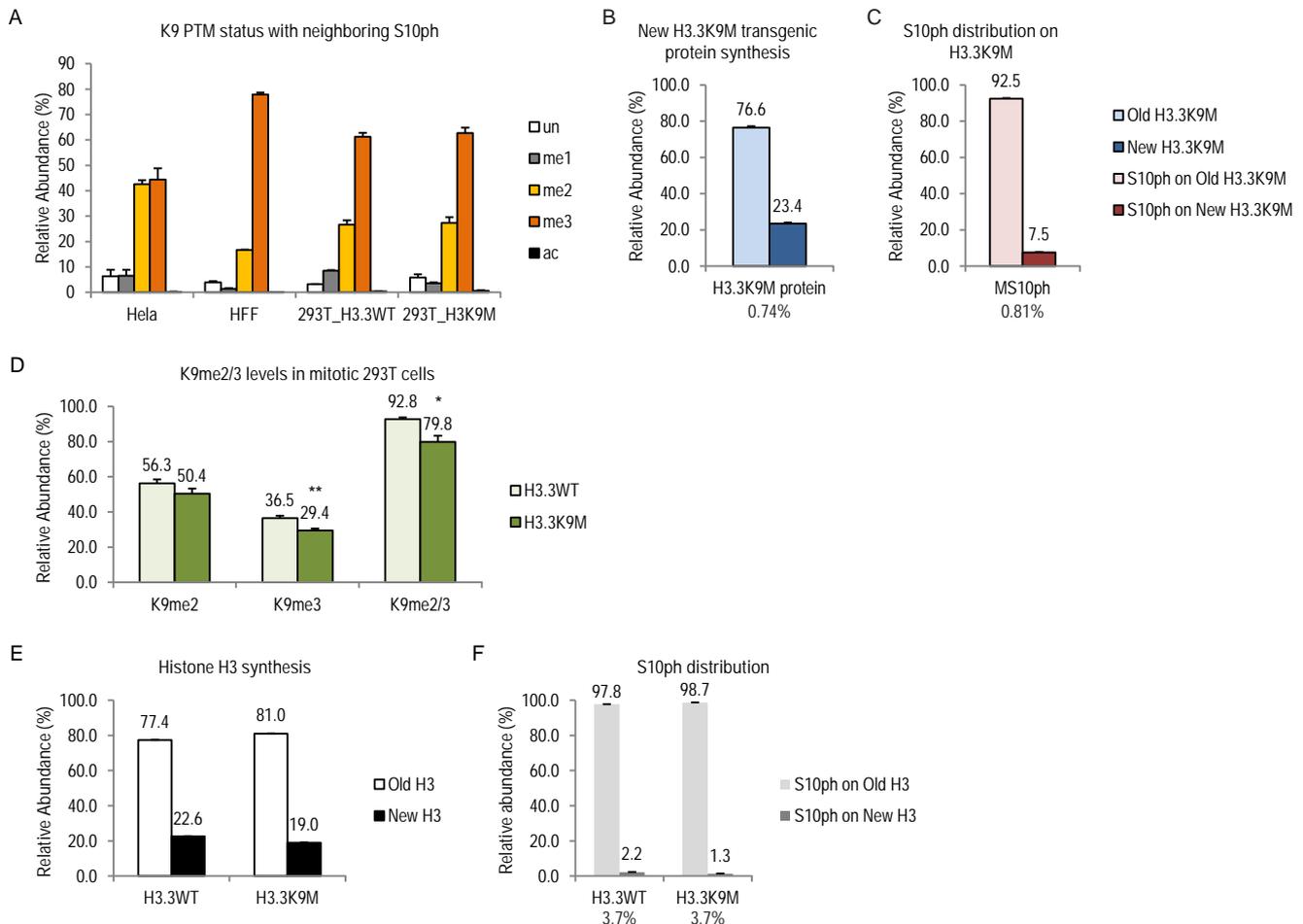


FIGURE 4. Neighboring H3K9 was not required for the asymmetric distribution of Ser-10 phosphorylation. *A*, H3S10ph frequently coexisted with K9me2/3 in mitotic human cells. *Bar graphs* show relative abundance of H3K9 PTM statuses with a neighboring Ser-10 phosphorylation. HeLa cell data were collected from three pulse-SILAC experiments with four time points. The *error bars* show standard error of biological replicates. HFF cell data were collected from two pulse-SILAC experiments. 293T cells were collected from one pulse-SILAC experiment. The *error bars* show standard error of multiple MS runs. *B* and *C*, H3K9 residue was not required for asymmetric distribution of its neighboring Ser-10 phosphorylation on a transgenic H3.3K9M protein in 293T cells. *Bar graphs* show H3.3K9M protein synthesis (*B*) and distribution of Ser-10 phosphorylation on the mutant protein (*C*). The *number below B* shows level of H3.3K9M protein is 0.74% of total H3. The *number below C* indicates 0.81% of the mutant protein had Ser-10 phosphorylation. The *numbers above each bar* show average of three MS runs. *Error bars* show standard error. *D–F*, H3S10ph level and distribution remained the same in 293T cells expressing H3.3K9M transgenic protein. *Bar graphs* showing relative abundance of the following items in 293T cells expressing either HA- and FLAG-tagged H3.3 wild type (WT) protein or H3.3K9M mutant protein. *D*, K9me2/3; *E*, new and old WT H3 protein, including endogenous protein; *F*, Ser-10 phosphorylation on old and new H3. The *numbers above each bar* show the average. *Error bars* show standard error. The *numbers below (F)* show the percentage of Ser-10 phosphorylation from the wild type H3 protein. *, $p < 0.05$; **, $p < 0.01$.

contrary to the symmetric distribution of H3.3S28/S31ph (Fig. 2*B* and 3*D*). Interestingly, there is only one amino acid difference (residue 31) in the N-terminal tails of canonical H3 and H3.3. The distribution of these phosphorylation marks was thereby regulated site-specifically even for the same histone variant.

The K9M mutant protein has been shown to have a dominant negative effect on global K9me2/3 levels (34). As shown in Fig. 4*D*, a total of 14% drop of K9me2/3 levels were observed in K9M transgenic cells, compared with a 293T cell line carrying a WT H3.3 transgene ($p < 0.05$). However, essentially no difference on both the level and the distribution of Ser-10 phosphorylation on the wild type protein was observed, further supporting our conclusion that K9me2/3 did not affect the distribution of the neighboring Ser-10 phosphorylation (Fig. 4, *E* and *F*). Our results were consistent with previous *in vitro* biochemical studies, which showed that Aurora B does not have substrate pref-

erence for a trimethylated form of H3K9 peptide, among other modified forms of the same peptide (35–37).

Crosstalk between H3S10ph and H3K9me2—As previously mentioned, the majority of the K9 residue next to a phosphorylated Ser-10 phosphorylation was di- or tri-methylated (Fig. 4*A*). Intriguingly, the appearance of H3S10ph coincides with the dissociation of HP1, a K9me2/3-binding protein, from mitotic chromosomes, both of which are Aurora B-dependent (36, 38). These observations raised the hypothesis that a functional cross-talk existed between Ser-10 phosphorylation and the neighboring K9me2/3 that is usually bound by HP1 during interphase (39, 40).

To investigate the roles of histone phosphorylation in mitosis and its interaction with other histone PTMs (e.g. H3K9me2/3), we inhibited the mitotic kinase Aurora B using hesperadin (Hes) (26, 41). As shown in Fig. 5*A*, treating cells with 100 nM Hes (Fig. 1*A*) drastically reduced H3S10ph level in mitotic HeLa

Asymmetric Histone Phosphorylation in the Human Cell Cycle

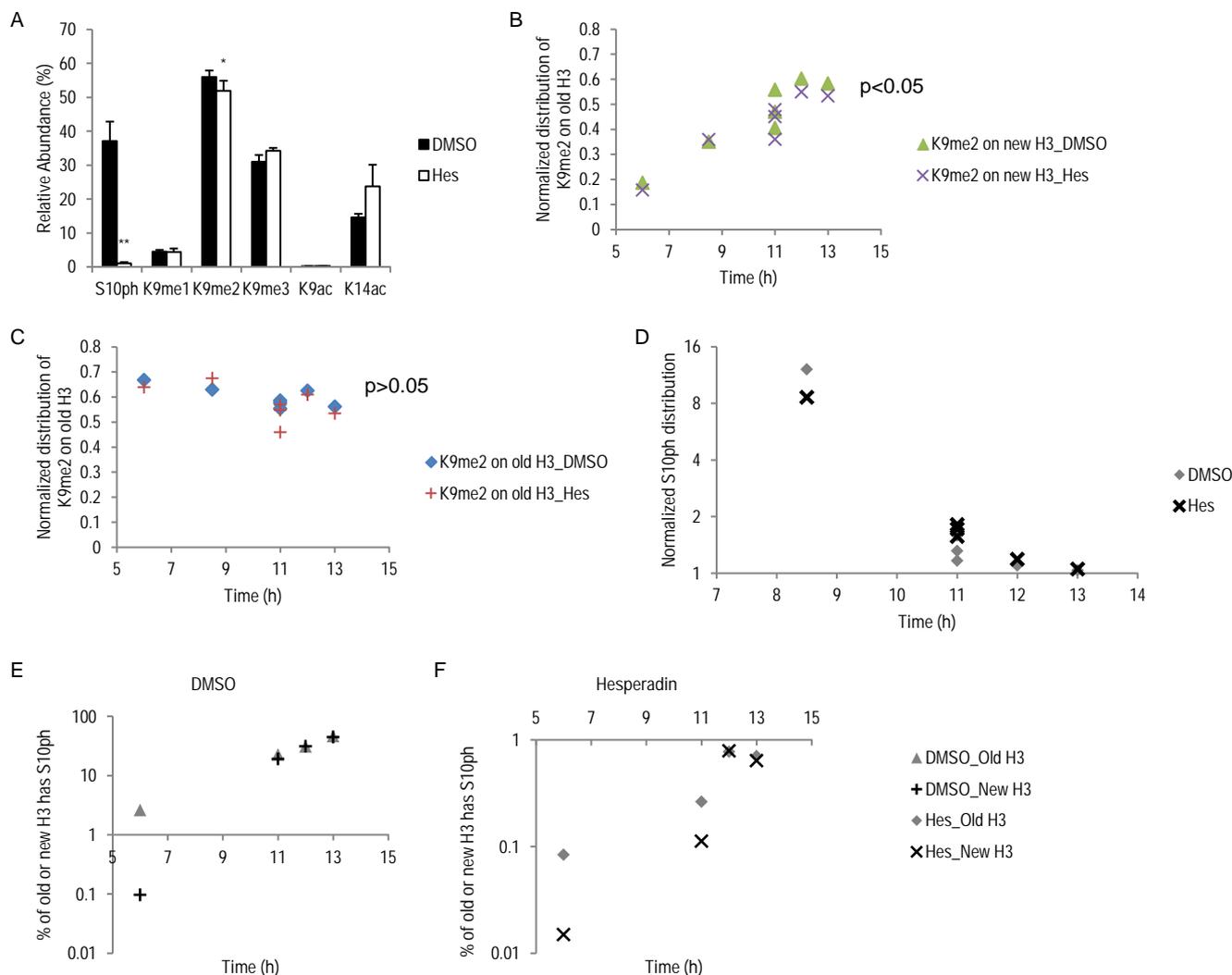


FIGURE 5. Inhibiting Aurora B kinase caused reduction of H3K9me2 level and did not impact H3S10ph distribution. Inhibition of Aurora B kinase caused a large drop in H3S10ph level and a small reduction in H3K9me2 level, but did not impact other histone PTMs. Three independent HeLa pulse-SILAC experiments were performed, and data were collected from the following time points: experiment 1, 8.5 and 11 h; experiment 2, 11 h; experiment 3, 6, 11, 12, and 13 h. *A*, bar graphs show averaged levels of PTMs on H3(9–17) peptide upon DMSO or Hes treatments. Data from late mitosis time points are shown (11 h and beyond). *, $p < 0.05$; **, $p < 0.01$. Error bars show standard error. *B* and *C*, Hes treatment caused less new K9me2 methylation on the new H3. Normalized K9me2 abundance on the new H3 (*B*) and old H3 (*C*) in DMSO- and Hes-treated cells. *p* values were generated using data from the same five time points as in *A*. *D–F*, inhibition of Aurora B did not change the distribution of H3S10ph across mitosis. *D*, normalized Ser-10 phosphorylation distribution was highly similar between DMSO and Hes treatments. Data were collected from early to late mitosis time points (8.5–13 h). *E* and *F*, new H3 lagged behind to get Ser-10 phosphorylation in both DMSO and Hes treatments, compared with the old H3. *E*, DMSO-treated cells; *F*, Hes-treated cells. Data were collected from the same pulse-SILAC experiment, namely experiment 3.

cells (harvested 11 h and beyond), without obviously delaying mitotic entrance (Fig. 1*E*). An average of 37.1% of H3 from DMSO-treated HeLa cells carried the Ser-10 phosphorylation mark, whereas only 1.0% of H3 from Hes-treated cells had Ser-10 phosphorylation (Fig. 5*A*). Interestingly, a small but significant reduction of H3K9me2 level was observed in Hes-treated cells (51.9%), compared with DMSO-treated cells (56.0%) (Fig. 5*A*). In contrast, no significant difference in the abundance of other lysine PTMs on the same peptide (Lys-9 or Lys-14) was seen. Surprisingly, the difference in K9me2 abundance between DMSO and Hes treatments was only seen on the new H3. As shown in Fig. 5*B*, K9me2 on the new H3 increased in a time-dependent manner in both DMSO- and Hes-treated cells. The increase in the Hes sample was less compared with DMSO ($p < 0.05$). In contrast, K9me2 on the old H3 stayed rather stable and did not differ between DMSO and Hes treat-

ments (Fig. 5*C*, $p > 0.05$). These results suggested that H3S10ph/Aurora B cross-talk specifically with K9me2 and were involved in either promoting new K9me2 events or the maintenance of the pre-existing K9me2 in mitosis, or both. We did not observe a significant difference of pre-existing K9me2 levels between DMSO- and Hes-treated cells (data not shown), arguing for the latter. Additionally, DMSO- and Hes-treated cells had similar amounts of new K9me2 on the old H3 (data not shown), suggesting that H3S10ph/Aurora B's role was only limited to the new H3. The effect of inhibiting Aurora B is specific to the H3K9me2 mark, as very little changes in abundance of overall histone PTMs were observed (Fig. 6). In addition, given the observation that the abundance of all asymmetric histone phosphorylation marks was drastically reduced when Aurora B was inhibited (Fig. 7, *A–C*), we conclude that the absolute abundance of histone phosphorylation did not

Asymmetric Histone Phosphorylation in the Human Cell Cycle

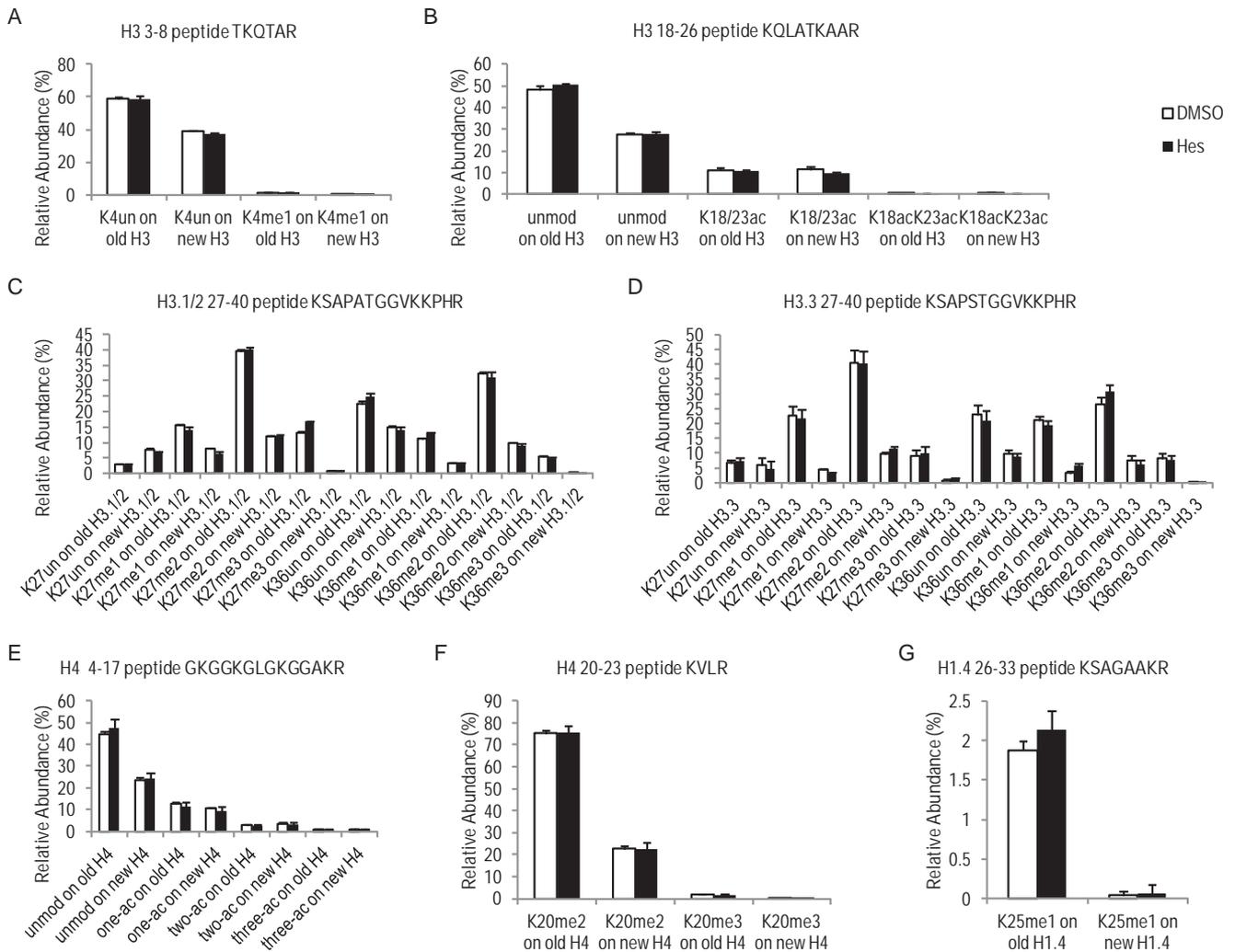


FIGURE 6. Hesperadin treatments did not affect relative levels of most histone PTMs. Bar graphs showing relative abundance of most commonly seen histone PTMs from three late mitosis time points (11 h and beyond). Only PTMs with relative abundance of over 0.5% are shown. Error bars show standard error. For all PTMs shown in this figure, there were no significant differences between DMSO- and Hes-treated cells ($p > 0.05$). Additional PTMs are shown in Figs. 5A and 7. *A*, PTMs on histone(3–8) peptide TKQTAR, very low levels of K4me2/3 were detected and not shown. *B*, PTMs on H3(18–26) peptide KQLATKAAR, very low levels of K18me1 and K23me1 were detected and not shown. *C*, PTMs on H3.1/2(27–40) peptide KSAPATGGVKKPHR, no K27ac was detected. *D*, PTMs on H3.3(27–40) peptide KSAPSTGGVKKPHR, no K27ac or Ser-28/Ser-31 phosphorylation was detected in the input samples. *E*, PTMs on H4(4–17) peptide GKGKGLGKGGAKR; shown are levels of acetylation events on one, two, or three lysine residues on this peptide, with very low levels of 4-ac detected and not shown. *F*, PTMs on H4(20–23) peptide KVLRL; Lys-20 methylation is cell cycle-regulated, and very low levels of K20un and K20me1 were detected in mitosis and not shown. *G*, K25me1 on H1.4(25–32) peptide KSAGAAKR.

affect epigenetic inheritance of other PTMs under the time points tested here.

Inhibiting Aurora B Does Not Affect the Asymmetry of Histone Phosphorylations—Next, we sought to investigate the impact of inhibiting Aurora B on the distribution of H3S10ph and other phosphorylation marks. Fig. 5, *D–F*, shows DMSO- and Hesperadin-treated HeLa cells, respectively, from the same pulse-SILAC experiment with four time points. As shown in Fig. 5*D*, essentially no difference was seen in the distribution of H3S10ph between DMSO- and Hesperadin-treated HeLa cells. Both populations showed enrichment of H3S10ph on the old H3 in an earlier time point (6 h) and symmetric distribution at later time points (11–13 h). In both samples, the percentage of Ser-10 phosphorylation on the new H3 lagged behind in the early 6-h time point, but ultimately it reached a similar level as the old H3 (11–13 h) (Fig. 5, *E* and *F*). In addition, the asymmetric distribution of other asymmetric histone phosphorylation marks was not affected by Hesperadin treatment

despite an overall decrease in their levels (Fig. 7, *D* and *E*). Thus, as the enzymatic activity of Aurora B was inhibited, the reduction in substrate phosphorylation manifested proportionally. These results suggested that Aurora B functioned downstream of the marker distinguishing old *versus* new histones.

Discussion

Mass Spectrometry Provides a Powerful Tool to Study Dynamics of Histone PTMs—Traditional antibody-based techniques are limited to recognize only one or a few PTMs at a time. They also suffer from potential off-target issues and epitope masking effects. Indeed, both of these issues have been reported with commercially available antibodies against H3S10ph (42, 43). Antibody-related issues may therefore contribute to some discrepancies in the literature regarding the function of Ser-10 phosphorylation (44–46). Over the last decade, mass spectrometry has become a powerful and unbiased tool for histone PTM analysis.

Asymmetric Histone Phosphorylation in the Human Cell Cycle

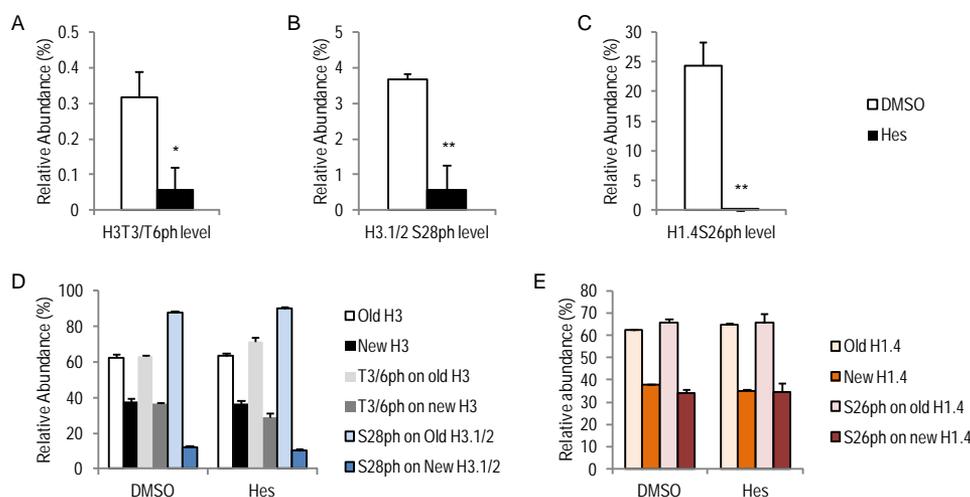


FIGURE 7. Hesperadin treatments reduced histone phosphorylation levels but did not impact their distribution patterns. Hes treatments reduced H3T3/T6ph (A), H3.1/2S28ph (B), and H1.4S26ph (C) levels, but it did not change the asymmetric distribution of these marks. *, $p < 0.05$; **, $p < 0.01$. D and E, bar graphs show relative abundance of histone protein SILAC labeling and phosphorylation mark distribution on H3 (D) and H1.4 (E). Error bars show standard error. Data were collected three late mitosis time points (11 h and beyond).

Using MS-based technologies, we not only can identify multiple PTMs simultaneously but also perform pulse-chase experiments on histone proteins. Combining MS and SILAC, we have shown that histone serine and threonine phosphorylations have asymmetric distributions in mitosis.

We used pulse-SILAC to label new histone synthesis. In theory, after one cell cycle division, the canonical histone protein abundance should double in the cells, and half of them should be labeled with SILAC media. However, we have never achieved greater than 40% labeling of canonical histones (Figs. 2A, 4, B and E, and 7, D and E). A number of factors could contribute to the incomplete labeling, including incomplete synchronization, impurities in the stable isotope-labeled compounds, and most importantly, the fact that cells recycle cellular metabolites to make new histones in S phase. Similar labeling efficiencies have been reported by us and others (9–11).

To rule out any potential artifacts that were linked to the heavy isotope labeling, one of our pulse-SILAC experiments was performed in a reverse fashion. HeLa cells were first cultured in heavy media for an extended period of time (more than 1 week with at least seven passages) to get complete labeling of the heavy-isotope amino acids (Arg and Met). Subsequently, these cells were subjected to a double thymidine block and were pulse-labeled in light media. These results, shown in Fig. 3, were highly consistent with light-to-heavy pulse-SILAC experiments.

H3.1/2 and H3.3 Show Different PTM Dynamics—To be noted, our analyses of histone H3 synthesis, H3S10ph, and Thr-3/Thr-6 phosphorylation include both the canonical histone H3.1/2 and H3.3 because we could not distinguish the variants in bottom-up mass spectrometry based on the peptides used (supplemental Tables S1–S3 and S5). In our experiments, the only peptide that can be used to distinguish between H3.1/2 and H3.3 levels was the H3(27–40) peptide as the 31st amino acid residue is different (H3.1/2, KSAPATGGVKKPHR; H3.3, KSAPSTGGVKKPHR) (Fig. 8B). However, this peptide has no less than 50 forms on each protein (supplemental Table S7) when labeled with heavy-Arg. The number of possible forms is

even larger when using both heavy-Arg and heavy-Met in the media, making it impractical to analyze from a technical standpoint. We therefore only showed data for this peptide from pulse-SILAC experiments that have solely heavy-Arg labeling (Figs. 2B, 3, C and D, and 7, B and D). Importantly, although H3.3 synthesis occurs throughout the cell cycle (3), its relative level remained to be a small portion of total H3. For example, H3.3 was 5.4 and 5.5% of total H3 in a HeLa cell pulse-SILAC experiment at two different time points (8.5 and 11 h), respectively. When we used the H3(27–40) peptide to estimate protein synthesis in our experiments (supplemental Table S7), we obtained very similar relative abundance of old and new H3 (difference $< 1\%$) between H3.1/2(27–40) peptide and four unmodified peptides that are shared by H3.1/2/3 (Fig. 2A).

Our examination of the H3(27–40) peptide in both H3.1/2 and H3.3 demonstrated that the distributions of PTMs on H3.1/2 and H3.3 were highly similar with only two exceptions, Ser-28 phosphorylation and K27me2. Although both were enriched on the old histone H3.1/2, their counterparts were symmetrically distributed on H3.3 (Table 1 and Fig. 8B).

Histone H3T80ph and H4S47ph Were Not Detected in Mitosis—We did not identify any H3T80 phosphorylation in our samples, despite the recent report about mitosis-specific Thr-80 phosphorylation (43). Additionally, we did not detect any histone H4S47ph in any of our samples. This result was not surprising because H4S47ph is catalyzed by the Pak2 kinase, which is not mitosis-specific (47).

Relative Abundance and Normalized Distribution—Because of technical limitations, including inconsistency in cell cycle synchronization and SILAC media incorporation, we observed a relatively large variation in the relative abundance measurements of mitosis-specific PTMs across different pulse-SILAC experiments. For example, the relative abundance of H3S10ph levels varied from ~ 20 to $\sim 50\%$ at the 11-h time point across four experiments (Fig. 3C). To address this issue, we introduced the normalized distribution of PTMs in this study. As shown in Fig. 3A, the normalized distributions of H3S10ph at the four

Asymmetric Histone Phosphorylation in the Human Cell Cycle

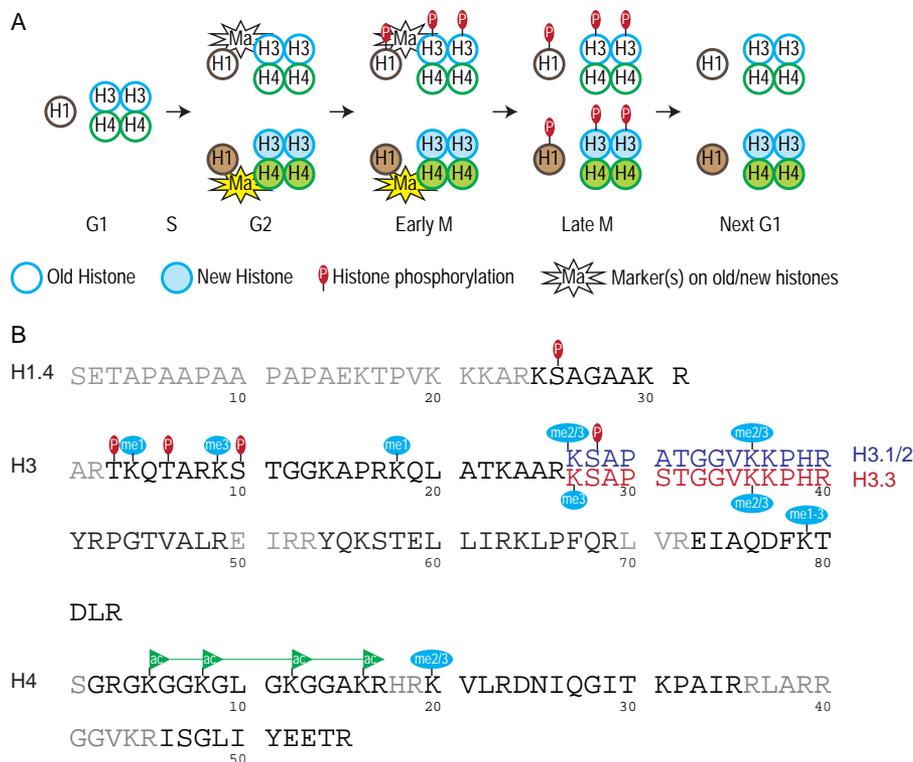


FIGURE 8. Preferential PTMs on old histones in mitosis. Schematic showing major findings of this study. *A*, preferential phosphorylations on old histones in early mitosis. Histone H1 and H3/H4 tetramers are shown. *Open circles* indicate old histones and *closed circles* indicate newly synthesized canonical histones. According to previous studies (6–8), histone H3/H4 tetramers have very low split rates and therefore old H3/H4 stay together. In early mitosis, most mitotic histone phosphorylation events preferentially happened on old histone H1 and H3. In late mitosis, both old and new histones were phosphorylated. The mitotic phosphorylations were erased before cells enter the next G₁ phase. Our data suggested that one or more markers (*Ma*) were involved in marking the old versus new histones in S, G₂, and early mitosis. *B*, shown are histone PTMs that were enriched on old histones in mitosis (Table 1). Only partial protein sequence is shown for each histone. Peptide sequences that were not analyzed in this study are in gray. The H3.1/2(27–40) peptide is in blue, and the H3.3(27–40) peptide is in red. The line connecting the four acetylation marks on H4(2–16) peptide indicates only the four-ac was enriched on old H4 but not two-ac or three-ac (Table 1).

11-h time points were more similar to each other than the relative abundances (Fig. 3C).

In addition, we provide a snapshot of the normalized distributions of 57 histone PTM statuses in Fig. 2B. We also showed the relative abundances of these PTM statuses in Fig. 2B and Table 1. The higher the abundance of a particular PTM was, the more confident we were. In general, we saw good reproducibility among experiments for any PTM status having a relative abundance above 1%.

Possible Mechanisms That Distinguish Old and New Histones—It was intriguing to discover the asymmetric distribution of histone phosphorylation events on old and new histones in mitosis (Fig. 8A). Based on our observations, we hypothesized that there was one or more marker(s) associated with either the old or new copies of histone proteins. The marker(s) should be added onto histones at S phase, either marking the old or newly synthesized histones. Subsequently, the marker(s) should remain associated with the old or new histones until late G₂ and/or M phase as the histone phosphorylation events start. The marker(s) should be recognized by mitotic kinases that are responsible for histone phosphorylation. Finally, the marker(s) should be highly abundant, at least as abundant as the phosphorylation marks. We found asymmetric phosphorylation marks on both core histone H3 and linker histone H1.4, suggesting a common mechanism for all histone phosphorylation

marks. Interestingly, Ser-10 phosphorylation but not Ser-28 phosphorylation on H3.3 is asymmetrically distributed (Figs. 3D and 4C), suggesting an alternative model that site-specific markers function locally. Our results suggested both mechanisms could exist and function coherently. Further investigation is needed to identify the molecular identity of the marker(s).

Our first candidate for the local marker of H3S10ph was H3K9me3 given its distribution pattern in mitosis (Fig. 2B), as well as the fact that H3S10ph is often associated with a neighboring K9me3 (Fig. 4A). Surprisingly, we showed in Fig. 4C that the Lys-9 residue was dispensable for the asymmetric distribution of Ser-10 phosphorylation. In addition, reduction in K9me2/3 levels did not change the global level of Ser-10 phosphorylation (Fig. 4, D–F). Therefore, K9me3 was unlikely to function as the marker for the asymmetric distribution of the neighboring Ser-10 phosphorylation on the same H3 tail.

Our data, however, did not rule out the possibility that the K9me3 on the other H3 molecule within the same nucleosome served as the marker for Ser-10 phosphorylation. Because the majority of H3/H4 tetramers (including most H3.3/H4 tetramers) do not split after replication (6–8), the H3/H4 tetramers are either old or new. Thus, H3K9me3 might still serve as the local marker for the distribution of the H3S10ph in the same nucleosome. Similarly, K27me2/3 might be the local marker for

H3.1/2S28ph because the K27me2/3-Ser(P)-28 pair resembled many features of K9me3-Ser(P)-10. In contrast, K9me2/3 and K27me2/3 were not likely to be the markers for H3T3/6ph or the linker histone H1.4S26ph, especially given that H1.4 did not reside in the nucleosomes. Further investigation is required to test these ideas.

Intriguingly, a similar phenomenon was recently reported in the *Drosophila* male germ line (48, 49), where the germ line stem cells (GSCs) undergo asymmetric cell division to generate one GSC and one differentiating daughter cell. Chen and co-workers (48) discovered that old/pre-existing canonical histone H3 but not H3.3 were selectively segregated to the daughter GSC. They further showed that H3T3ph was required for the asymmetric inheritance of H3, and they suggested it was required in a specific time window (prophase to metaphase) (49). The asymmetric cell division is only limited to *Drosophila* GSCs and is not seen in surrounding somatic tissues, so is the asymmetric distribution of H3T3ph and directional inheritance of histone H3 (48, 49). However, the observations are highly similar to ours in this study that mitotic histone phosphorylation mark(s) is(are) enriched on the old histones in early mitosis, suggesting the underlying mechanism is conserved between *Drosophila* and humans.

Epigenetic Inheritance of Histone PTMs—Two distinct models have been proposed regarding how the epigenetic information of histone PTMs is carried through DNA replication as follows: either the enzymes or the PTMs are associated with chromatin and presumably will guide the inheritance after replication. Mazo and co-workers (50, 51) have shown in *Drosophila* embryos that the H3K4me3 and H3K27me3 are lost during replication, whereas the methyltransferases responsible are retained on chromosome, suggesting the former mechanism. Contradicting results from Strome and co-workers (52) showed that the H3K27me3 remains on chromatin in *Caenorhabditis elegans* from germline to the next generation. We and others (9–11) have previously studied histone methylation kinetics in the human cell cycle, revealing that overall levels of H3K9me2/3 and H3K27me2/3 are not significantly reduced in S phase. The decreases in relative abundance can be explained by dilution effects from the newly synthesized histones. In this study we further showed that many other histone PTMs, including H3K36me2/3 and H3K79me1/2/3, were enriched on old histones during mitosis (Figs. 2A and 8B and Table 1), supporting the observations seen in *C. elegans* that histone PTMs themselves are maintained for epigenetic inheritance. In addition, our results also suggested that the corresponding enzymes responsible for these PTMs would propagate the information from old histones to new histones after histone synthesis in S phase. The differences in the spatial and temporal activities of these enzymes might be responsible for the distributions we observed in this study. For example, H3K9me2/3 and H3K27me2/3 do not recover to their original level until the end of the next G₁ phase, suggesting that the responsible enzymes are mainly active during G₁.

Histone acetylations are another major category of PTMs commonly seen on histones. It is long established that newly synthesized H3 and H4 carry acetylation (ac) marks and that deacetylation is essential for the assembly process (6), although

it remains unclear how the cell manages both deacetylation and epigenetic inheritance of acetylation at the same time. We showed in this study that most H3 and H4 ac marks, including H3K9ac, H3K14ac, H3K18/23ac, H3K18acK23ac, and H4(4–17) one-ac, were symmetrically distributed (Fig. 2B and Table 1). Two ac marks, namely H4(4–17) two-ac and H4(4–17) three-ac were enriched on new histones (Fig. 2B and Table 1). These two ac marks were of rather low abundance (<10%) and showed a decreasing trend from the 8.5-h time point to the 11-h time point, which suggested that they were from a small percentage of S phase cells (53, 54). Interestingly, H4(4–17) four-ac (K5acK8acK12acK16ac) was enriched on the old histones at both time points (Table 1 and Fig. 8B), suggesting that the epigenetic inheritance of this highly acetylated form (four-ac) of H4 was regulated differently from its less acetylated forms (one- to three-ac).

Biological Function of Histone Phosphorylations—Extensive work has been conducted to uncover the biological functions of histone phosphorylation marks (45, 46, 55–57). A common view is that they play distinct roles in interphase and mitosis. Several phosphorylation marks are implicated in transcriptional regulation of individual genes during interphase (58–63). However, their roles in mitosis remain elusive, despite the fact that the abundance of these marks is generally much higher in mitosis than in interphase (15–21). Our discovery about the asymmetric distribution of these marks might provide the first step toward understanding their function in mitosis. We originally hypothesized that histone phosphorylation marks served as indicators for distinguishing new *versus* old histones and were involved in guiding the inheritance of other histone PTMs. The reason that phosphorylation levels peak in mitosis may be because this is the only period during which the cell must differentiate old from new. However, H3K9me2 was the only other PTM affected by inhibiting Aurora B in mitosis, despite the fact that all four asymmetric histone phosphorylation marks were reduced drastically (Figs. 5A and 7, A–C). These results suggested that histone phosphorylation marks were not directly involved in epigenetic inheritance of most PTMs, with the caveat that the time frame of our experiments might have been too short to reveal the impact.

Although most Kme2/3s were enriched on the old histones in mitosis, H3K9me2 showed a symmetric distribution (Fig. 2B and Table 1). Therefore, it was surprising that H3K9me2, but not H3K9me3, was affected by Aurora B inhibition (Fig. 5A). As discussed previously, H3S10ph and/or Aurora B might be involved in either maintenance of old K9me2 or generation of new K9me2. Our results favored the latter model as we showed a significant difference in K9me2 levels on the new H3 but not the old H3 between DMSO and Hes treatments (Fig. 5, B and C). This observation was contradictory to previous work that showed Suv39h1 was incapable of methylating a Ser-10 phosphorylation peptide *in vitro* (35). The discrepancy may be due to either indirect interaction of H3K9me2-S10ph/Aurora B (*i.e.* mediated by other proteins) or technical limitations in the *in vitro* methyltransferase assay. In addition, the reduction in H3K9me2 levels was rather small, compared with the drop in H3S10ph and other phosphorylation mark levels. This could be due to an indirect impact of Hes treatments and therefore

Asymmetric Histone Phosphorylation in the Human Cell Cycle

did not represent a cross-talk between H3S10ph/Aurora B and H3K9me2. Further work is required to investigate this potential cross-talk.

Experimental Procedures

Tissue Culture and Synchronization—HeLa S3 cells were cultured and maintained in suspension as described previously (64). In brief, HeLa cells were cultured in Minimum Essential Medium Eagle's Joklik Modification with L-glutamine (Sigma) supplemented with 10% Hyclone® new calf serum (GE Healthcare) and 100 units/ml penicillin/streptomycin (Life Technologies, Inc.). HeLa cells were maintained in a density between 1×10^5 and 1×10^6 cells/ml. The cell synchronization procedure is illustrated in Fig. 1A. For the double thymidine block, 2.5 mM thymidine (Acros Organics) was added to the media. The first block was 18–19 h. Cells were subsequently released for 8.5–10 h in media without thymidine and subjected to a second block with 2.5 mM thymidine for 15–17 h. For experiments with nocodazole, 0.1 μ g/ml nocodazole (EMD Millipore) was supplemented with normal media after the double thymidine block. For experiments with the Aurora B inhibitor hesperadin, 100 nM hesperadin (EMD Millipore) or equivalent DMSO solvent was added to the media at 2 h after the double thymidine block. For different time points collected from the same experiment, a portion was removed from the suspension culture at each time point and analyzed by flow cytometry or mass spectrometry.

HFF cells and 293T cells were cultured and maintained as described previously (33, 34). Briefly, HFF and 293T cells were cultured in Dulbecco's modification of Eagle's medium (DMEM) with high glucose (Life Technologies, Inc.), supplemented with 10% fetal bovine serum (FBS) (Sigma). HFF cells were cultured below ~80% confluency to avoid a contact inhibition-induced quiescence state. Cells below 13 passages were used. 293T cells were supplemented with 3 μ g/ml puromycin (Santa Cruz Biotechnology) to maintain the HA- and FLAG-tagged H3.3 (WT or K9M) transgene. For synchronization of HFF and 293T cells, cells were plated at 20–25% confluency and subjected to serum starvation with 0.1% FBS for 24 h. Subsequently media with 10% FBS and 2 mM hydroxyurea were added for 18 h. Normal growth media with 0.1 μ g/ml nocodazole was subsequently introduced for releasing. G/2 M cells were collected 6–7 h after releasing (Fig. 1B). For different time points collected from the same experiment, cells were plated separately in different tissue culture wells/plates but at identical densities and treated the same.

SILAC Medium Preparation—DMEM lacking L-arginine and L-methionine was purchased (AthenaES or Thermo Fisher Scientific Inc.). Joklik medium lacking L-arginine and L-methionine was made from powdered Joklik mixture (Sigma) and supplemented with all the necessary vitamins and amino acids (excluding L-arginine and L-methionine) based on the standard recipe. The SILAC medium was supplemented with 10% dialyzed FBS (Thermo Fisher Scientific Inc.), as well as L-arginine- $^{13}\text{C}_6$, $^{15}\text{N}_4$ -HCl (Cambridge Isotope Labs.) and L-methionine-(methyl- ^{13}C , D_3) (Sigma), or regular light L-methionine (Sigma) according to the standard recipe. Light L-proline (Sigma) was also added to the medium at 200 mg/liter to prevent conversion from the heavy arginine (30).

Histone Extraction, Phosphorylated Peptide Enrichment—Histones were acid-extracted from cells and processed with two rounds of chemical derivatization, trypsin digestion, and desalting by Stage-Tips as described previously (65). These samples were either subjected directly to on-line nano-LC/MS/MS (input samples) or were enriched for phosphorylated peptides.

Phosphorylated peptide enrichment by TiO_2 was performed as described previously (65–68). Briefly, propionylated and desalted histone peptides were mixed with at least a 4-fold excess of TiO_2 beads (GL Sciences) in 200 μ l of loading buffer (2 M lactic acid (Sigma) in 50% acetonitrile (ACN)). TiO_2 micro-columns are made in the same way as Stage-Tips with a C8 mini-disk plug at the bottom of the tip. The histone peptide and TiO_2 /bead mixture were added onto the micro-column and centrifuged at $200 \times g$ for 10–20 min. The micro-column was then rinsed with 200 μ l of loading buffer and $2 \times 200 \mu$ l of wash buffer (50% ACN with 0.1% trifluoroacetic acid) at $1000 \times g$. Elution was performed twice with 200 μ l of elution buffer (50 mM KH_2PO_4 in 50% ACN, pH 10.0) at $200 \times g$. TiO_2 -enriched samples were acidified with glacial acetic acid. The samples were then desalted using Stage-Tips and subjected to on-line nano-LC/MS/MS.

Nano-Liquid Chromatography Electrospray Ionization Tandem Mass Spectrometry (Nano-LC/MS/MS) and Data Analysis—The samples were loaded onto one of the three instrument setups as shown in supplemental Table S1, all at 300 nl/min. Histone peptides were resolved on a two-step gradient from 2% ACN to 30% ACN in 0.1% formic acid over 40 min and then from 30 to 95% ACN in 0.1% formic acid over 20 min. The mass spectrometers were operated in the data-dependent mode with dynamic exclusion enabled (repeat count, 1; exclusion duration, 0.5 min). MS instrument methods were set up as reported previously (2). Settings for resolution, automatic gain control, and normalized collision energy are listed in supplemental Table S1. For every cycle, one full MS scan (m/z 290 to 1600) was collected and followed by 10 MS/MS scans using either high energy C-trap dissociation or collision-induced dissociation in the ion trap (supplemental Table S1). All isolation windows were set at 2.0 m/z . Ions with a charge state of 1 and a rejection list of common contaminant ions (including keratin, trypsin, and BSA) (exclusion width = 10 ppm) were excluded from MS/MS.

For each MS run, 1–2 μ g of input samples were loaded. For TiO_2 -enriched samples, 20–300 μ g of equivalent were loaded. A minimum of triplicates was performed for each sample, except for one time point from a HeLa cell pulse-SILAC that only two good quality runs were achieved. As illustrated previously (65), we targeted the m/z for the isobaric peptides and quantified the relative abundance of their unique b or y ions at the MS/MS level. Subsequently, we determined the relative abundance at the MS1 level based on the ratios we obtained, either manually or by a Matlab-based program, EpiProfile (69). Data were corrected for differences in detection efficiencies based on the correction factors generated by a synthetic peptide library as described previously (70). For phosphorylated peptides that were not included in the synthetic peptide library, an average correction factor generated from all the peptides with

the same histone backbone was used. For the histone H3.3(27–40) peptide, the H1.4(25–32) peptide, and the K9M(9–17) peptide, no correction factors were available, and thus no correction was performed.

MH⁺ masses of peptides quantified are listed in supplemental Tables S2 to S8. Most pulse-SILAC experiments were performed with the light-to-heavy media switch, with either both heavy-Arg and heavy-Met labeling or just heavy-Arg labeling. One pulse-SILAC experiment for HeLa cells was performed with heavy-to-light media switch, with both heavy-Arg and heavy-Met labeling in the original media. Histone H3(9–17) peptides quantified are listed in supplemental Table S2 with both heavy-Arg and heavy-Met labeling and supplemental Table S3 with just heavy-Arg labeling. For quantification of old versus new histone H3 and H4 proteins, four peptides from H3 and two peptides from H4 that are typically unmodified in mitosis were used to generate an average number for new protein synthesis (supplemental Table S4). For each peptide, the light and heavy peptides were quantified, and the sum was set to be 100%. We also monitored all peptides in supplemental Table S4 in this analysis for any potential PTMs but did not identify any in our experiments as expected. The supplemental Tables S5–S7 show additional peptides quantified. For differentially SILAC-labeled methylated peptides of the same modification (e.g. Lys-9me3:0, Lys-9me3:1, Lys-9me3:2, and Lys-9me3:3), the relative quantification of all these forms were combined and summarized. An example is given in supplemental Table S9.

PTM Distribution Normalization—The SILAC labeling of newly synthesized histone proteins was incomplete, *i.e.* we did not get 50% labeling of histones, presumably because of recycling of cellular amino acids. To account for this incomplete labeling, we normalized the distribution of PTMs based on the corresponding histone protein labeling efficiency. The normalized distribution was calculated for each pulse-SILAC time point as shown in Equation 2,

normalized PTM distribution

$$= \frac{\text{PTM on old histone/PTM on new histone}}{\text{old histone/new histone}} \quad (\text{Eq. 2})$$

For the K9me2 on old or new H3 shown in Fig. 5, B and C, the normalized distribution was calculated as shown in Equations 3 and 4,

normalized distribution of K9me2 on old H3

$$= \frac{\text{K9me2 on old H3}}{\text{old H3}} \quad (\text{Eq. 3})$$

normalized distribution of K9me2 on new H3

$$= \frac{\text{K9me2 on new H3}}{\text{new H3}} \quad (\text{Eq. 4})$$

Old and new H3 and H4 proteins were quantified based on multiple peptides listed in supplemental Table S4. The H1.4 protein was quantified based on the 25–32 peptide shown in supplemental Table S6. H3.1/2 and H3.3 proteins were quanti-

fied based on the 27–40 peptide shown in supplemental Table S7. H3.3K9M peptides were listed in supplemental Table S8.

Statistical Analysis—The *p* values were generated by Student's *t* test (two-tailed).

Mass Spectrometry Data—The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (71) partner repository with the dataset identifier PXD003709 and 10.6019/PXD003709.

Propidium Iodide (PI) Staining and Flow Cytometry Analysis—As described previously (9), cells were fixed in 70–80% ethanol and stained by PI for flow cytometry analysis. Briefly, one to three million cells were collected and rinsed with phosphate-buffered saline (PBS). Cells were subsequently resuspended in 1 ml of 4 °C PBS. 100% ethanol (–20 °C) was then added in a dropwise fashion to reach a total volume of >5 ml with moderate vortexing. After fixation overnight at 4 °C, cells were washed with PBS twice. Cells were then incubated with 0.08 mg/ml PI (EMD Millipore) and 0.02 mg/ml RNaseA (Roche Applied Science) in PBS for >1 h in the dark at room temperature. Flow cytometry was performed in either the Princeton University Flow Cytometry Resource Facility or the University of Pennsylvania Flow Cytometry and Cell Sorting Resource Laboratory. Data figures were generated using FlowJo.

Metabolomics Analyses—As described previously (9), metabolites were extracted from frozen cell pellets with 40% ACN, 40% methanol, and 0.1% formic acid on ice and centrifuged at 5000 × *g* to collect the supernatant. The supernatant was clarified by centrifugation at 16,000 × *g* prior to MS analyses. [¹²C₆,¹⁴N₄]- and [¹³C₆,¹⁵N₄]arginine, [¹²C₅,¹⁴N]- and [¹³C₅,¹⁵N]proline, and [¹²C,H₃]- and [¹³C,D₃]AdoMet were quantified as described previously (72).

Author Contributions—S. L. designed and performed the experiments and wrote the manuscript; Y. H. performed some 293T cell experiments; Z. Y. analyzed some data; D. M. M. performed some 293T cell experiments. B. A. G. designed the project.

Acknowledgments—We thank Adam G. Everts and Hilary A. Collier (UCLA) for HFF cells; Peter Lewis (Wisconsin Institute for Discovery) for H3.3K9M 293T cells; Wenyun Lu and Joshua Rabinowitz (Princeton University) for metabolomics analysis; Bing He, Yao Yao, Li Yang, and Kelly Karch for reading and revising the manuscript; Ben E. Black, Hua-Ying Fan, Roberto Bonasio, Johayra S. Williams, Simone Sidoli, Barry M. Zee, and Rosalynn C. Molden for useful discussions; Christina DeCoste (Princeton University Flow Cytometry Resource Facility) and Hank Pletcher (University of Pennsylvania Flow Cytometry and Cell Sorting Resource Laboratory) for flow cytometry analysis.

References

- Heintz, N. (1991) The regulation of histone gene expression during the cell cycle. *Biochim. Biophys. Acta* **1088**, 327–339
- Osley, M. A. (1991) The regulation of histone synthesis in the cell cycle. *Annu. Rev. Biochem.* **60**, 827–861
- Wu, R. S., Tsai, S., and Bonner, W. M. (1982) Patterns of histone variant synthesis can distinguish G0 from G1 cells. *Cell* **31**, 367–374
- Bird, A. (2007) Perceptions of epigenetics. *Nature* **447**, 396–398
- Goldberg, A. D., Allis, C. D., and Bernstein, E. (2007) Epigenetics: a land-

- scape takes shape. *Cell* **128**, 635–638
6. Annunziato, A. T. (2015) The fork in the road: histone partitioning during DNA replication. *Genes* **6**, 353–371
 7. Xu, M., Long, C., Chen, X., Huang, C., Chen, S., and Zhu, B. (2010) Partitioning of histone H3-H4 tetramers during DNA replication-dependent chromatin assembly. *Science* **328**, 94–98
 8. Huang, C., Zhang, Z., Xu, M., Li, Y., Li, Z., Ma, Y., Cai, T., and Zhu, B. (2013) H3.3-H4 tetramer splitting events feature cell-type specific enhancers. *PLoS Genet.* **9**, e1003558
 9. Zee, B. M., Britton, L.-M., Wolle, D., Haberman, D. M., and Garcia, B. A. (2012) Origins and formation of histone methylation across the human cell cycle. *Mol. Cell. Biol.* **32**, 2503–2514
 10. Xu, M., Wang, W., Chen, S., and Zhu, B. (2012) A model for mitotic inheritance of histone lysine methylation. *EMBO Rep.* **13**, 60–67
 11. Alabert, C., Barth, T. K., Reverón-Gómez, N., Sidoli, S., Schmidt, A., Jensen, O. N., Imhof, A., and Groth, A. (2015) Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes Dev.* **29**, 585–590
 12. Masumoto, H., Hawke, D., Kobayashi, R., and Verreault, A. (2005) A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature* **436**, 294–298
 13. Recht, J., Tsubota, T., Tanny, J. C., Diaz, R. L., Berger, J. M., Zhang, X., Garcia, B. A., Shabanowitz, J., Burlingame, A. L., Hunt, D. F., Kaufman, P. D., and Allis, C. D. (2006) Histone chaperone Asf1 is required for histone H3 lysine 56 acetylation, a modification associated with S phase in mitosis and meiosis. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6988–6993
 14. Jørgensen, S., Schotta, G., and Sørensen, C. S. (2013) Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. *Nucleic Acids Res.* **41**, 2797–2806
 15. Hendzel, M. J., Wei, Y., Mancini, M. A., Van Hooser, A., Ranalli, T., Brinkley, B. R., Bazett-Jones, D. P., and Allis, C. D. (1997) Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G₂ and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma* **106**, 348–360
 16. Goto, H., Tomono, Y., Ajiro, K., Kosako, H., Fujita, M., Sakurai, M., Okawa, K., Iwamatsu, A., Okigaki, T., Takahashi, T., and Inagaki, M. (1999) Identification of a novel phosphorylation site on histone H3 coupled with mitotic chromosome condensation. *J. Biol. Chem.* **274**, 25543–25549
 17. Goto, H., Yasui, Y., Nigg, E. A., and Inagaki, M. (2002) Aurora-B phosphorylates histone H3 at serine 28 with regard to the mitotic chromosome condensation. *Genes Cells* **7**, 11–17
 18. Preuss, U., Landsberg, G., and Scheidtmann, K. H. (2003) Novel mitosis-specific phosphorylation of histone H3 at Thr11 mediated by Dlk/ZIP kinase. *Nucleic Acids Res.* **31**, 878–885
 19. Polioudaki, H., Markaki, Y., Kourmouli, N., Dialynas, G., Theodoropoulos, P. A., Singh, P. B., and Georgatos, S. D. (2004) Mitotic phosphorylation of histone H3 at threonine 3. *FEBS Lett.* **560**, 39–44
 20. Hake, S. B., Garcia, B. A., Kauer, M., Baker, S. P., Shabanowitz, J., Hunt, D. F., and Allis, C. D. (2005) Serine 31 phosphorylation of histone variant H3.3 is specific to regions bordering centromeres in metaphase chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6344–6349
 21. Hergeth, S. P., Dunder, M., Tropberger, P., Zee, B. M., Garcia, B. A., Daujat, S., and Schneider, R. (2011) Isoform-specific phosphorylation of human linker histone H1.4 in mitosis by the kinase Aurora B. *J. Cell Sci.* **124**, 1623–1628
 22. Carmena, M., Wheelock, M., Funabiki, H., and Earnshaw, W. C. (2012) The chromosomal passenger complex (CPC): from easy rider to the godfather of mitosis. *Nat. Rev. Mol. Cell. Biol.* **13**, 789–803
 23. Hsu, J. Y., Sun, Z. W., Li, X., Reuben, M., Tatchell, K., Bishop, D. K., Grushcow, J. M., Brame, C. J., Caldwell, J. A., Hunt, D. F., Lin, R., Smith, M. M., and Allis, C. D. (2000) Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* **102**, 279–291
 24. Adams, R. R., Maiato, H., Earnshaw, W. C., and Carmena, M. (2001) Essential roles of Drosophila inner centromere protein (INCENP) and aurora B in histone H3 phosphorylation, metaphase chromosome alignment, kinetochore disjunction, and chromosome segregation. *J. Cell Biol.* **153**, 865–880
 25. Giet, R., and Glover, D. M. (2001) Drosophila aurora B kinase is required for histone H3 phosphorylation and condensin recruitment during chromosome condensation and to organize the central spindle during cytokinesis. *J. Cell Biol.* **152**, 669–682
 26. Hauf, S., Cole, R. W., LaTerra, S., Zimmer, C., Schnapp, G., Walter, R., Heckel, A., van Meel, J., Rieder, C. L., and Peters, J.-M. (2003) The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *J. Cell Biol.* **161**, 281–294
 27. Wang, F., Ulyanova, N. P., van der Waal, M. S., Patnaik, D., Lens, S. M., and Higgins, J. M. (2011) A positive feedback loop involving Haspin and Aurora B promotes CPC accumulation at centromeres in mitosis. *Curr. Biol.* **21**, 1061–1069
 28. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics MCP* **1**, 376–386
 29. Fontecave, M., Atta, M., and Mulliez, E. (2004) S-Adenosylmethionine: nothing goes to waste. *Trends Biochem. Sci.* **29**, 243–249
 30. Bendall, S. C., Hughes, C., Stewart, M. H., Doble, B., Bhatia, M., and Lajoie, G. A. (2008) Prevention of amino acid conversion in SILAC experiments with embryonic stem cells. *Mol. Cell. Proteomics* **7**, 1587–1597
 31. Feng, Q., Wang, H., Ng, H. H., Erdjument-Bromage, H., Tempst, P., Struhl, K., and Zhang, Y. (2002) Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* **12**, 1052–1058
 32. Kim, W., Choi, M., and Kim, J.-E. (2014) The histone methyltransferase Dot1/DOT1L as a critical regulator of the cell cycle. *Cell Cycle* **13**, 726–738
 33. Evertts, A. G., Zee, B. M., Dimaggio, P. A., Gonzales-Cope, M., Collier, H. A., and Garcia, B. A. (2013) Quantitative dynamics of the link between cellular metabolism and histone acetylation. *J. Biol. Chem.* **288**, 12142–12151
 34. Lewis, P. W., Müller, M. M., Koletsky, M. S., Cordero, F., Lin, S., Banaszynski, L. A., Garcia, B. A., Muir, T. W., Becher, O. J., and Allis, C. D. (2013) Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science* **340**, 857–861
 35. Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599
 36. Fischle, W., Tseng, B. S., Dormann, H. L., Ueberheide, B. M., Garcia, B. A., Shabanowitz, J., Hunt, D. F., Funabiki, H., and Allis, C. D. (2005) Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438**, 1116–1122
 37. Demidov, D., Hesse, S., Tewes, A., Rutten, T., Fuchs, J., Ashtiyani, R. K., Lein, S., Fischer, A., Reuter, G., and Houben, A. (2009) Aurora1 phosphorylation activity on histone H3 and its cross-talk with other post-translational histone modifications in *Arabidopsis*. *Plant J.* **59**, 221–230
 38. Hirota, T., Lipp, J. J., Toh, B.-H., and Peters, J.-M. (2005) Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature* **438**, 1176–1180
 39. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120
 40. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., and Grewal, S. I. (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110–113
 41. Sessa, F., Mapelli, M., Ciferri, C., Tarricone, C., Areces, L. B., Schneider, T. R., Stukenberg, P. T., and Musacchio, A. (2005) Mechanism of Aurora B activation by INCENP and inhibition by hesperadin. *Mol. Cell* **18**, 379–391
 42. Duan, Q., Chen, H., Costa, M., and Dai, W. (2008) Phosphorylation of H3S10 blocks the access of H3K9 by specific antibodies and histone methyltransferase. Implication in regulating chromatin dynamics and epigenetic inheritance during mitosis. *J. Biol. Chem.* **283**, 33585–33590
 43. Hammond, S. L., Byrum, S. D., Namjoshi, S., Graves, H. K., Dennehey, B. K., Tackett, A. J., and Tyler, J. K. (2014) Mitotic phosphorylation of histone H3 threonine 80. *Cell Cycle* **13**, 440–452

44. Prigent, C., and Dimitrov, S. (2003) Phosphorylation of serine 10 in histone H3, what for? *J. Cell Sci.* **116**, 3677–3685
45. Pérez-Cadahía, B., Drohic, B., and Davie, J. R. (2009) H3 phosphorylation: dual role in mitosis and interphase. *Biochem. Cell Biol.* **87**, 695–709
46. Sawicka, A., and Seiser, C. (2012) Histone H3 phosphorylation—a versatile chromatin modification for different occasions. *Biochimie* **94**, 2193–2201
47. Kang, B., Pu, M., Hu, G., Wen, W., Dong, Z., Zhao, K., Stillman, B., and Zhang, Z. (2011) Phosphorylation of H4 Ser 47 promotes HIRA-mediated nucleosome assembly. *Genes Dev.* **25**, 1359–1364
48. Tran, V., Lim, C., Xie, J., and Chen, X. (2012) Asymmetric division of *Drosophila* male germline stem cell shows asymmetric histone distribution. *Science* **338**, 679–682
49. Xie, J., Wooten, M., Tran, V., Chen, B.-C., Pozmanter, C., Simbolon, C., Betzig, E., and Chen, X. (2015) Histone H3 threonine phosphorylation regulates asymmetric histone inheritance in the *Drosophila* male germline. *Cell* **163**, 920–933
50. Petruk, S., Sedkov, Y., Johnston, D. M., Hodgson, J. W., Black, K. L., Kovermann, S. K., Beck, S., Canaan, E., Brock, H. W., and Mazo, A. (2012) TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. *Cell* **150**, 922–933
51. Petruk, S., Black, K. L., Kovermann, S. K., Brock, H. W., and Mazo, A. (2013) Stepwise histone modifications are mediated by multiple enzymes that rapidly associate with nascent DNA during replication. *Nat. Commun.* **4**, 2841
52. Gaydos, L. J., Wang, W., and Strome, S. (2014) Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development. *Science* **345**, 1515–1518
53. Sobel, R. E., Cook, R. G., Perry, C. A., Annunziato, A. T., and Allis, C. D. (1995) Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1237–1241
54. Jasencakova, Z., Scharf, A. N., Ask, K., Corpet, A., Imhof, A., Almouzni, G., and Groth, A. (2010) Replication stress interferes with histone recycling and predeposition marking of new histones. *Mol. Cell* **37**, 736–743
55. Oki, M., Aihara, H., and Ito, T. (2007) Role of histone phosphorylation in chromatin dynamics and its implications in diseases. *Subcell. Biochem.* **41**, 319–336
56. Cerutti, H., and Casas-Mollano, J. A. (2009) Histone H3 phosphorylation: universal code or lineage specific dialects? *Epigenetics* **4**, 71–75
57. Pérez-Cadahía, B., Drohic, B., Khan, P., Shivashankar, C. C., and Davie, J. R. (2010) Current understanding and importance of histone phosphorylation in regulating chromatin biology. *Curr. Opin. Drug Discov. Dev.* **13**, 613–622
58. Chadee, D. N., Hendzel, M. J., Tylipski, C. P., Allis, C. D., Bazett-Jones, D. P., Wright, J. A., and Davie, J. R. (1999) Increased Ser-10 phosphorylation of histone H3 in mitogen-stimulated and oncogene-transformed mouse fibroblasts. *J. Biol. Chem.* **274**, 24914–24920
59. Thomson, S., Mahadevan, L. C., and Clayton, A. L. (1999) MAP kinase-mediated signalling to nucleosomes and immediate-early gene induction. *Semin. Cell Dev. Biol.* **10**, 205–214
60. Dou, Y., Mizzen, C. A., Abrams, M., Allis, C. D., and Gorovsky, M. A. (1999) Phosphorylation of linker histone H1 regulates gene expression *in vivo* by mimicking H1 removal. *Mol. Cell* **4**, 641–647
61. Lo, W. S., Trievel, R. C., Rojas, J. R., Duggan, L., Hsu, J. Y., Allis, C. D., Marmorstein, R., and Berger, S. L. (2000) Phosphorylation of serine 10 in histone H3 is functionally linked *in vitro* and *in vivo* to Gcn5-mediated acetylation at lysine 14. *Mol. Cell* **5**, 917–926
62. Sarg, B., Helliger, W., Talasz, H., Förg, B., and Lindner, H. H. (2006) Histone H1 phosphorylation occurs site-specifically during interphase and mitosis: identification of a novel phosphorylation site on histone H1. *J. Biol. Chem.* **281**, 6573–6580
63. Zhang, Q., Zhong, Q., Evans, A. G., Levy, D., and Zhong, S. (2011) Phosphorylation of histone H3 serine 28 modulates RNA polymerase III-dependent transcription. *Oncogene* **30**, 3943–3952
64. Zee, B. M., Levin, R. S., Xu, B., LeRoy, G., Wingreen, N. S., and Garcia, B. A. (2010) *In vivo* residue-specific histone methylation dynamics. *J. Biol. Chem.* **285**, 3341–3350
65. Lin, S., and Garcia, B. A. (2012) Examining histone post-translational modification patterns by high-resolution mass spectrometry. *Methods Enzymol.* **512**, 3–28
66. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., and Jørgensen, T. J. (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* **4**, 873–886
67. Li, Q. R., Ning, Z. B., Tang, J. S., Nie, S., and Zeng, R. (2009) Effect of peptide-to-TiO₂ beads ratio on phosphopeptide enrichment select. *J. Proteome Res.* **8**, 5375–5381
68. Kettenbach, A. N., and Gerber, S. A. (2011) Rapid and reproducible single-stage phosphopeptide enrichment of complex peptide mixtures: application to general and phosphotyrosine-specific phosphoproteomics experiments. *Anal. Chem.* **83**, 7635–7644
69. Yuan, Z.-F., Lin, S., Molden, R. C., Cao, X.-J., Bhanu, N. V., Wang, X., Sidoli, S., Liu, S., and Garcia, B. A. (2015) EpiProfile quantifies histone peptides with modifications by extracting retention time and intensity in high-resolution mass spectra. *Mol. Cell. Proteomics* **14**, 1696–1707
70. Lin, S., Wein, S., Gonzales-Cope, M., Otte, G. L., Yuan, Z.-F., Afjehi-Sadat, L., Maile, T., Berger, S. L., Rush, J., Lill, J. R., Arnott, D., and Garcia, B. A. (2014) Stable isotope labeled histone peptide library for histone post-translational modification and variant quantification by mass spectrometry. *Mol. Cell. Proteomics* **13**, 2450–2466
71. Vizcaíno, J. A., Csordas, A., del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q.-W., Wang, R., and Hermjakob, H. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, D447–456
72. Bennett, B. D., Yuan, J., Kimball, E. H., and Rabinowitz, J. D. (2008) Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nat. Protoc.* **3**, 1299–1311

The E3 Ubiquitin Ligase Adaptor Protein Skp1 Is Glycosylated by an Evolutionarily Conserved Pathway That Regulates Protist Growth and Development^{*[S]♦}

Received for publication, November 11, 2015, and in revised form, December 19, 2015. Published, JBC Papers in Press, December 30, 2015, DOI 10.1074/jbc.M115.703751

Kazi Rahman^{‡§¶}, Peng Zhao^{§||}, Msano Mandalasi^{‡§}, Hanke van der Wel^{‡§}, Lance Wells^{§||}, Ira J. Blader^{¶**}, and Christopher M. West^{‡§¶1}

From the Departments of [‡]Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104, the [§]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602, the [¶]Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104, the ^{||}Complex Carbohydrate Research Center, Athens, Georgia 30602, and the ^{**}Department of Microbiology and Immunology, University at Buffalo, Buffalo, New York 14214

Toxoplasma gondii is a protist parasite of warm-blooded animals that causes disease by proliferating intracellularly in muscle and the central nervous system. Previous studies showed that a prolyl 4-hydroxylase related to animal HIF α prolyl hydroxylases is required for optimal parasite proliferation, especially at low O₂. We also observed that Pro-154 of Skp1, a subunit of the Skp1/Cullin-1/F-box protein (SCF)-class of E3-ubiquitin ligases, is a natural substrate of this enzyme. In an unrelated protist, *Dictyostelium discoideum*, Skp1 hydroxyproline is modified by five sugars via the action of three glycosyltransferases, Gnt1, PgtA, and AgtA, which are required for optimal O₂-dependent development. We show here that TgSkp1 hydroxyproline is modified by a similar pentasaccharide, based on mass spectrometry, and that assembly of the first three sugars is dependent on *Toxoplasma* homologs of Gnt1 and PgtA. Reconstitution of the glycosyltransferase reactions in extracts with radioactive sugar nucleotide substrates and appropriate Skp1 glycoforms, followed by chromatographic analysis of acid hydrolysates of the reaction products, confirmed the predicted sugar identities as GlcNAc, Gal, and Fuc. Disruptions of *gnt1* or *pgtA* resulted in decreased parasite growth. Off target effects were excluded based on restoration of the normal glycan chain and growth upon genetic complementation. By analogy to *Dictyostelium* Skp1, the mechanism may involve regulation of assembly of the SCF complex. Understanding the mechanism of *Toxoplasma* Skp1 glycosylation is expected to help develop it as a drug target for control of the pathogen, as the glycosyltransferases are absent from mammalian hosts.

Toxoplasma is a worldwide obligate intracellular apicomplexan parasite that infects most nucleated cells of warm-blooded animals (1). Toxoplasmosis, the disease caused by *Toxoplasma*, is an opportunistic infection in AIDS and other immune-suppressed patients (2). In addition, *in utero* infections can cause mental retardation, blindness, and death (3). *Toxoplasma* is transmitted by digesting parasites from feline feces (as oocysts) or undercooked meat (as tissue cysts). Once in the host, parasites convert to the tachyzoite form that disseminates to peripheral tissues (e.g. brain, retina, and muscle). The resulting immune response and/or drugs can control tachyzoite replication, but the parasite survives by converting into slow growing bradyzoites that encyst. Cysts sporadically burst, and the released parasites convert to tachyzoites whose unabated growth, as can occur in immune suppressed hosts, results in cell and tissue damage (4). Currently, no *Toxoplasma* vaccine exists; anti-toxoplasmosis drugs have severe side effects, and resistance to these drugs is occurring.

Recently, disruption of the gene for PhyA, the prolyl 4-hydroxylase that hydroxylates Pro-154 in Skp1, was observed to reduce tachyzoite proliferation in cell culture and fitness in a competition assay (5). Skp1 is an adaptor in the Skp1/Cullin-1/F-box protein (SCF)² class of E3 ubiquitin ligases, and its hydroxylation was hypothesized to contribute to O₂-dependent proliferation. That study noted that loss of hydroxylation resulted in increased migration in SDS-polyacrylamide gels suggesting a decrease in M_r of ~1000. Previous studies in an unrelated protist, the social soil amoeba *Dictyostelium*, had shown that the Skp1-hydroxyproline (Hyp) could be glycosylated by five glycosyltransferase activities encoded by three genes, resulting in assembly of a pentasaccharide at the equivalent Pro residue (6, 7). Because two of these genes, *gnt1* and *pgtA*, have apparent homologs in the *Toxoplasma* genome, we suspected that the gel shift might result from inability of the

* This work was supported in part by Grant 14-140 from the Mizutani Foundation for Glycoscience (to C. M. W. and I. J. B.), Oklahoma Center for the Advancement of Science and Technology Grant HR10-0181 (to C. M. W.), and National Institutes of Health Grants R01-GM084383 (to C. M. W. and I. J. B.) and R01-AI069986 (to I. J. B.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

♦ This article was selected as a Paper of the Week.

[S] This article contains supplemental Table S1 and Figs. S1 and S2.

¹ To whom correspondence should be addressed. Tel.: 706-542-4259; E-mail: westcm@uga.edu

² The abbreviations used are: SCF, Skp1/Cullin-1/FBP subcomplex of the Cullin-1/RING ligase class of E3 ubiquitin ligases; GlcNAcT, polypeptide N-acetyl- α -glucosaminyltransferase; GT, glycosyltransferase; HFF, human foreskin fibroblast; Hyp, 4R,2S-(trans)-hydroxyproline; S100, cytosolic extract prepared as the supernatant after 100,000 \times g centrifugation; SF-tag, a 51-amino acid peptide including 2 Strep-tag II epitopes and a FLAG epitope; CID, collision-induced dissociation; MPA, mycophenolic acid; dHex, deoxyhexose.

glycosyltransferases to modify Skp1 in the absence of formation of the Hyp anchor. In *Dictyostelium*, *Ddgent1* encodes a polypeptide α GlcNAc transferase that transfers GlcNAc from UDP-GlcNAc to form GlcNAc α 1-O-Skp1 (8). *Ddpgta* encodes a dual function diglycosyltransferase whose N-terminal domain then transfers Gal from UDP-Gal to form a Gal β 1–3GlcNAc linkage and whose C-terminal domain processively transfers Fuc from GDP-Fuc to form a Fuc α 1–2Gal linkage (9). However, the two domains are switched in the *Toxoplasma* version of the predicted protein (TGGT1_260650), and there is no evidence for *agtA*, the *Dictyostelium* gene that is responsible for addition of the final two sugars, both α Gal residues.

If the Skp1 Hyp of *Toxoplasma* can be glycosylated, the importance of hydroxylation for proliferation might be due to consequent loss of glycosylation rather than inability to hydroxylate *per se*. In *Dictyostelium*, hydroxylation alone partially rescues O₂-dependent development (10). Full recovery depends, however, on full glycosylation (11), and glycosylation is required to promote efficient assembly of the Skp1/F-box protein heterodimer, based on interactome studies (12). Therefore, we sought direct evidence for Hyp-dependent glycosylation of *Toxoplasma* Skp1 and the role of the predicted glycosyltransferase genes to test their contribution to parasite proliferation. The findings implicate Skp1 as the functional target of this novel post-translational modification pathway in *Toxoplasma* and indicate that the Skp1 modification pathway is evolutionarily conserved among protists.

Experimental Procedures

Parasites, Cell Culture, and Plaque Assays—*Toxoplasma* strain RH Δ ku80 Δ hxgprt (RH Δ) was cultured in association with human foreskin fibroblasts (HFFs) using Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) fetal bovine serum, 2 mM L-glutamine, and 100 units/ml penicillin/streptomycin (Complete medium) in a humidified CO₂ (5%) incubator at 37 °C. RH Δ ku80 Δ phyA (RHphyA Δ), RH Δ ku80 Δ gent1 (RHgent1 Δ), and RH Δ ku80 Δ pgtA (RHpgtA Δ) strains were cultured in the same medium supplemented with 25 μ g/ml mycophenolic acid (Sigma) and 25 μ g/ml xanthine (Sigma). RH Δ /SF and RHphyA Δ -1/SF strains, where TgSkp1 was tagged with the SF-tag and have the chloramphenicol acetyltransferase marker, were cultured in DMEM supplemented with 20 μ M chloramphenicol (Sigma). Strains were cloned by limiting dilution in 96-well plates.

To perform cell growth plaque assays, confluent HFF monolayers in 6-well tissue culture plates were infected with freshly lysed-out (see below) parasites at 250 parasites/well, equivalent to a multiplicity of infection of 0.002. After 3 h, unattached parasites were removed by two rinses with phosphate-buffered saline (15 mM sodium phosphate, pH 7.4, 135 mM NaCl). After undisturbed incubation in Complete medium for 5.5 days, monolayers were fixed with methanol and stained with crystal violet to detect plaques. Plaques ($n \geq 50$) from at least two wells were manually encircled, and areas were calculated by ImageJ software (National Institutes of Health). Data were presented and statistically analyzed using GraphPad Prism version 6.

TgphyA, Tggnt1, and TgpgtA Disruption Strains—DNAs for gene disruptions were generated from pminiGFP.ht (gift of Dr. Gustavo Arrizabalaga, University of Idaho), in which the *hxgprt* gene is flanked by multiple cloning sites. The approach was modeled after that used for the *TgphyA* disruption strain RHphyA Δ -1, in which exon 1 of *TgphyA* was replaced with *hxgprt* (5). To generate an independent *TgphyA* disruption strain, RHphyA Δ -2, the complete coding region was replaced with *hxgprt* by double crossover homologous recombination. First, the 5'- and 3'-flank targeting sequences of *TgphyA* from RH Δ were PCR-amplified with primer pairs a and a' and pairs b and b', respectively (supplemental Table S1). The 5'-fragment was digested with KpnI and HindIII and inserted into pminiGFP.ht between its KpnI and HindIII sites. The resulting plasmid was digested with XbaI and NotI and ligated to the XbaI- and NotI-digested 3'-flank. The resulting vector was linearized with KpnI and electroporated into RH Δ strain as described (5). Drug-resistant transformants were selected in the presence of 25 μ g/ml MPA and 25 μ g/ml xanthine and cloned by limiting dilution. Genomic DNA from three clones was screened by PCR to identify *TgphyA* disruption strain RHphyA Δ -2, as described under "Results." The PCRs were performed on extracts from 2×10^6 parasites, using *Taq* polymerase, and primers as listed in supplemental Table S1. Standard conditions included 1.5 mM MgCl₂, and reactions were run for 30 cycles of the following standard scheme: 94 °C, 30 s; 60 °C, 1 min; 68 °C, 3 min. Exact conditions were adjusted for specific reactions.

To disrupt *Tggnt1*, the 5'-flank and 3'-flank targeting sequences were PCR-amplified with primer pairs c and c' and d and d' (supplemental Table S1), respectively, and inserted into pminiGFP.ht as above. The vector was linearized with SapI and transfected into RH Δ , and drug-resistant clones were screened by PCR to generate the *Tggnt1* disruption strain (RHgent1 Δ). Similarly, the *TgpgtA* disruption construct was generated by PCR amplification and insertion into pminiGFP.ht of 5'-flank and 3'-flank targeting sequences using primer pairs e and e' and f and f', respectively. After digestion with PacI, the DNA was transfected into RH Δ , and RHpgtA Δ clones were screened by PCR.

Tggnt1 and TgpgtA Complemented Strains—pminiGFP.ht was used as the backbone for constructing the *Tggnt1* complementation construct after removing its HXGPRT cassette by KpnI and NotI digestion. A 7-kb DNA fragment containing the *Tggnt1* genomic region was PCR-amplified using primer pairs c and d' (supplemental Table S1), digested with KpnI and NotI, and ligated into the similarly digested pminiGFP.ht. The resultant vector was linearized with KpnI and electroporated into RHgent1 Δ . Transformants were selected under 300 μ g/ml 6-thioxanthine (Matrix Scientific), and clones were screened by PCR. To complement *TgpgtA* knock-out, the fosmid clone Rfos01M21 (13), containing a 36-kb fragment of RH strain chromosome VIII (2039542–2076165), which includes the *TgpgtA* gene (gift of Dr. Boris Striepen, University of Georgia), was linearized with ScaI and electroporated into RHpgtA Δ . Complemented clones were isolated as described for *Tggnt1*.

Epitope Tagging of Endogenous TgSkp1—To modify the C terminus of endogenous TgSkp1, the *skp1* genomic locus was

Complex Glycosylation of *Toxoplasma Skp1*

TABLE 1

Toxoplasma strains used in this study

Strain	Parental strain	Genotype	Gene targeted	Selection marker	Selection drug	Ref.
KU80ΔΔ	RH(1)	Δku80;Δhxpprt				26
RHΔ <i>phyA</i> -1	KU80ΔΔ	<i>phyA</i> Δ;Δku80	<i>phyA</i> -exon 1	Hxgprt	MPA, xanthine	5
RHΔ <i>phyA</i> -2	KU80ΔΔ	<i>phyA</i> Δ;Δku80	<i>phyA</i> -exons 1–9 (all)	Hxgprt	MPA, xanthine	TR ^a
RHΔ <i>gnt1</i>	KU80ΔΔ	<i>gnt1</i> Δ;Δku80	<i>gnt1</i> -exon 1 (all)	Hxgprt	MPA, xanthine	TR
RHΔ <i>gnt1</i> /complemented	RHΔ <i>gnt1</i>	Δku80;Δhxpprt		ΔHxgprt	6-Thioxanthine	TR
RHΔ <i>pgtA</i>	KU80ΔΔ	<i>pgtA</i> Δ;Δku80	<i>pgtA</i> -exons 1–14 (all)	Hxgprt	MPA, xanthine	TR
RHΔ <i>pgtA</i> /complemented	RHΔ <i>pgtA</i>	Δku80;Δhxpprt		ΔHxgprt	6-Thioxanthine	TR
RHΔΔ/Skp1-SF	KU80ΔΔ	<i>Skp1</i> ^{SF} ;Δku80; Δhxpprt;CAT ⁺	Skp1 C-terminus	CAT	Chloramphenicol	TR
RHΔ <i>phyA</i> -1/Skp1-SF	RH <i>phyA</i> Δ-1	<i>Skp1</i> ^{SF} ; <i>phyA</i> Δ; Δku80;CAT ⁺	<i>phyA</i> -exon 1; Skp1 C terminus	CAT	Chloramphenicol	TR

^a TR means this report.

modified by the insertion of SF-tag cDNA. A 1.5-kb region upstream of the *skp1* stop codon was PCR-amplified using primer pairs g and g'. Using a ligation-independent cloning strategy (14), the product was inserted into PacI-digested pSF-TAP-LIC-HXGPRT and pSF-TAP-LIC-CAT vectors (from Dr. Vern Carruthers, University of Michigan). 50 μg of the resulting constructs were linearized with EcoRV at position 527 of the insert, and the DNAs were electroporated into RHΔΔ and RH*phyA*Δ-1, respectively. RHΔ/SF transformants were selected under 25 μg/ml MPA/xanthine, and RH*phyA*Δ-1/SF was selected under 20 μM chloramphenicol. Site-specific integration was confirmed by PCR of clones using the primer pairs h and h'. DNA sequencing confirmed that the 3'-end of the Skp1 coding sequence encoded the native C terminus (... VREENKWCEDA) followed by a peptide containing two Strep-II tags and a FLAG tag (in boldface), **AKIGSGGR-EFWSHPQFEKGGGSGGGSGGGSSWSHPQFEKGASGEDYK-DDDDK**^{*}. Characteristics of the above strains are summarized in Table 1.

Purification of Endogenous TgSkp1—Tachyzoites from RHΔΔ, RH*phyA*Δ-1, RH*gnt1*Δ, and RH*pgtA*Δ strains were harvested from infected HFF monolayers by scraping and passage through a 27-gauge needle, centrifuged at 2000 × g for 8 min at room temperature, resuspended in sterile phosphate-buffered saline, and counted on a hemacytometer chamber as described (15). 6 × 10⁸ tachyzoites were pelleted, frozen at -80 °C, and subsequently thawed and solubilized in 8 M urea, 50 mM HEPES-NaOH, pH 7.4, supplemented with protease inhibitors (1 mM PMSE, 10 μg/ml aprotinin and 10 μg/ml leupeptin) at 4 °C for 30 min. The lysates were centrifuged at 16,000 × g for 15 min at 4 °C, and supernatants (S16) were collected and diluted 8-fold in IP buffer (0.2% Nonidet P-40 (v/v) in 50 mM HEPES-NaOH, pH 7.4, protease inhibitors) and incubated with 60 μl of rabbit polyclonal anti-TgSkp1 UOK75-Sepharose beads for 1 h at 4 °C. The UOK75 antiserum (5) was first affinity-purified against recombinant TgSkp1-Sepharose beads, performed as described for affinity purification of anti-DdSkp1 (12), and then coupled to CNBr-activated Sepharose CL-4B. After centrifuging and resuspending the beads three times with IP buffer and three times with wash buffer (10 mM Tris-HCl, pH 7.5, 154 mM NaCl), bound material was eluted twice with 150 μl of 133 mM triethylamine, pH 11.5, for 10 min and immediately neutralized with 150 μl of 200 mM acetic acid, pH 2.7. The pooled eluates (~400 μl) were divided into two equal parts, concentrated by vacuum centrifugation to ~10 μl, and snap-frozen in liquid nitrogen. To purify SF-tagged TgSkp1, soluble S16 fractions were prepared from RHΔ/SF and RH*phyA*Δ-1/SF

strains as described above and incubated with 100 μl of mouse anti-FLAG M2-agarose beads (Sigma) for 1 h at 4 °C. The beads were washed as above and eluted with 300 μl of 8 M urea in 25 mM NH₄HCO₃ (natural pH), supplemented with 40 mg of urea, and incubated for 15 min at room temperature, and the supernatants harboring TgSkp1-SF were collected at 2400 × g for 5 min in room temperature.

Mass Spectrometry of TgSkp1 Peptides—The untagged TgSkp1 samples were taken to dryness and solubilized in 100 μl of 8 M urea in 25 mM NH₄HCO₃. The untagged and SF-tagged TgSkp1 samples were reduced by addition of 0.5 M tris(2-carboxyethyl)phosphine to a final concentration of 5 mM for 20 min at 22 °C and alkylated by the addition of 0.5 M iodoacetamide to a final concentration of 10 mM for 15 min in the dark. Tris(2-carboxyethyl)phosphine was then added to a final concentration of 10 mM, and 300 μl of 50 mM NH₄HCO₃, pH 7.8, was added to dilute urea to 2 M. Samples were treated with 1 mg/ml mass spectrometry grade trypsin (Promega) at a final concentration 8.0 μg/ml and incubated overnight at 37 °C. Peptides were recovered by adsorption to a C18 Zip Tip (OMIX TIP C18 100 μl) and eluted with 0.1% trifluoroacetic acid in 50% (v/v) acetonitrile.

Dry peptides were reconstituted in 15.6 μl of solvent A (0.1% formic acid) and 0.4 μl of solvent B (0.1% formic acid in 80% acetonitrile) and loaded onto a 75-μm (inner diameter) × 115-mm C18 capillary column (YMC GEL ODS-AQ120ÅS-5, Waters) packed in-house with a nitrogen bomb. Peptides were eluted into the nanospray source of an LTQ OrbitrapTM mass spectrometer (Thermo Fisher Scientific) with a 160-min linear gradient consisting of 5–100% solvent B over 100 min at a flow rate of 250 nl/min. The spray voltage was set to 2.0 kV, and the temperature of the heated capillary was set to 210 °C. Full scan MS spectra were acquired from *m/z* 300 to 2000 at 30,000 resolution, and MS2 scans following collision-induced fragmentation were collected in the ion trap for the 12 most intense ions. The raw spectra were searched against a *Toxoplasma gondii* protein database (UniProt ATCC 50611/Me49, Sept. 2013) using SEQUEST (Proteome Discoverer 1.3, Thermo Fisher Scientific) with full MS peptide tolerance of 50 ppm and MS2 peptide fragment tolerance of 0.5 Da and filtered to generate a 1% target decoy peptide-spectrum-match false discovery rate for protein assignments. The spectra assigned as glycosylated TgSkp1 peptides were manually validated.

Cytosolic (S100) Extracts—Parasites were permeabilized as described (16) with slight modifications. Briefly, a pellet of 2.5 × 10⁹ frozen tachyzoites was resuspended in 500 μl of ice-cold water containing 10 μg/ml aprotinin, 10 μg/ml leupeptin,

1 mM PMSF, and 0.1 mM *N*- α -*p*-tosyl-L-lysine chloromethyl ketone and incubated for 10 min on ice. The suspension was transferred to a Dounce homogenizer and sheared by 10 strokes, diluted with an equal volume of 100 mM HEPES-NaOH, pH 7.4, 10 mM MgCl₂, 10 mM MnCl₂, 50 mM KCl, 10 μ g/ml aprotinin, 10 μ g/ml leupeptin, 1 mM PMSF, and sheared by 20 additional strokes. After confirmation of lysis using phase contrast microscopy, the lysate was centrifuged at 200,000 \times *g* at 5 °C for 35 min; the supernatant (S100) was immediately desalted over a PD10 column at 5 °C into 50 mM HEPES-NaOH, pH 7.4, 5 mM MgCl₂, 15% (v/v) glycerol, and 0.1 mM EDTA. Fractions with highest A₂₈₀ values (≥ 1 , 1-cm path length) were snap-frozen at -80 °C for enzyme assays.

Glycosyltransferase Assays—Skp1-dependent GlcNAcT activity was assayed in S100 fractions by the transfer of ³H from UDP-[³H]GlcNAc to exogenous *Dictyostelium* HO-DdSkp1 (17). Typically, a 50- μ l reaction volume containing 30 μ l of S100 fraction, 50 pmol of HO-DdSkp1 (18), and 0.5–2.5 μ M UDP-GlcNAc (including 1 μ Ci of UDP-[³H]GlcNAc at 37 Ci/mmol, PerkinElmer Life Sciences), in 50 mM HEPES-NaOH, pH 7.4, 10 mM MgCl₂, 2 mM DTT, 3 mM NaF, and protease inhibitors, was incubated at 30 °C for 0, 1, or 3 h. Reactions were stopped by addition of 4 \times Laemmli electrophoresis sample buffer, supplemented with 2 μ g of soybean trypsin inhibitor (Sigma) as a marker that comigrates with DdSkp1, boiled for 5 min, and resolved by SDS-PAGE (see below). The gel was stained for 1 h with 0.25% (w/v) Coomassie Blue in 45% (v/v) methanol, destained overnight in 5% (v/v) methanol, 7.5% (v/v) acetic acid, and rinsed in H₂O for 1 h. Five ~1-mm gel slices including and surrounding the soybean trypsin inhibitor band were excised and incubated in 7 ml of a scintillation mixture containing 100 ml of Soluene-350 (PerkinElmer Life Sciences) and 900 ml of 0.6 g/liter 2,5-diphenyloxazole and 0.15 g/liter dimethyl-1,4-bis(4-methyl-5-phenyl-2-oxazolyl)benzene in scintillation grade toluene. After 5 days, ³H was quantitated by scintillation counting in a Beckman LSC6500 instrument.

GalT activity was assayed similarly except that the donor was 1 μ M UDP-[³H]Gal and prepared from a mixture of 1 μ Ci of UDP-[³H]Gal (7) and unlabeled UDP-Gal, and the acceptor was recombinantly prepared GlcNAc-DdSkp1 (18). FucT activity was similarly assayed except that 2 μ M GDP-[³H]Fuc, prepared from a mixture of GDP-[³H]Fuc (1 μ Ci of 20 Ci/mmol, PerkinElmer Life Sciences) and unlabeled GDP-Fuc, was used in place of UDP-[³H]Gal, and 2 μ M unlabeled UDP-Gal was added to generate Gal-GlcNAc-DdSkp1 acceptor from the added GlcNAc-DdSkp1.

Radioactive Sugar Analyses—The chemical form of the radioactivity incorporated into DdSkp1 was determined by high pH anionic exchange chromatography analysis after acid hydrolysis as described (17). Briefly, Gnt1 reaction products were resolved by SDS-PAGE and electroblotted onto a 0.45- μ m PVDF membrane (EMD Millipore). The membrane was stained with 0.2% Ponceau S in 3% (w/v) TCA, and the Skp1 protein band was excised with a razor blade, submerged in 400 μ l of 6 M HCl, and incubated at 100 °C for 4 h. Hydrolysates were removed into microtubes, evaporated to dryness under vacuum centrifugation, dissolved in 500 μ l of H₂O, dried again twice, and reconstituted in 25 μ l of H₂O. A solution containing

1 nmol of GalNH₂ and GlcNH₂ in 9 μ l of H₂O was added to the hydrolysates and chromatographed on PA-1 column on a DX-600 Dionex high pH anionic exchange chromatography workstation in 16 mM NaOH at 1 ml/min with pulsed amperometric detection. Fractions were collected into EconoSafe (Research Products International) scintillation fluid and counted for ³H incorporation on a liquid scintillation counter. PgtA assay products were similarly hydrolyzed in 4 M TFA and mixed with a standard solution containing 1.5 nmol each of Glc, Gal, Man, and Fuc.

Western Blotting—Western blotting was performed as described (5). Briefly, tachyzoite pellets were solubilized in lysis buffer containing 8 M urea, 50 mM HEPES-NaOH, pH 7.4, supplemented with protease inhibitors at 4 °C for 30 min and centrifuged at 16,000 \times *g* for 15 min at 4 °C to generate a soluble S16 fraction. After combining with SDS-PAGE sample buffer, proteins were resolved on a 4–12% gradient SDS-polyacrylamide gel (NuPAGE Novex, Invitrogen) and transferred to a nitrocellulose membrane using an iBlot system (Invitrogen). After probing with a 1:500-fold dilution of the UOK75 anti-TgSkp1 antibody and a 1:10,000-fold dilution of Alexa-680-labeled goat anti-rabbit IgG secondary antibody (Invitrogen), blots were imaged on a Li-Cor Odyssey infrared scanner.

Results

***Toxoplasma Skp1* Is Modified by a Pentasaccharide**—Our previous study showed that disruption of exon 1 of the Skp1 prolyl 4-hydroxylase gene (*phyA*) in the parental type 1 strain RH Δ ku80 Δ hxgprt (RH $\Delta\Delta$) resulted in greater mobility of Skp1 during SDS-PAGE, corresponding to an *M_r* difference of about 1000 (5). To investigate the possibility that this was due to a loss of hydroxylation-dependent glycosylation as occurs in *Dictyostelium* (6), a previously described antiserum raised against recombinant TgSkp1 (5) was used to immunoprecipitate Skp1 from tachyzoite extracts, and its tryptic peptides were analyzed by conventional nano-LC/MS in an LTQ-XL Orbitrap mass spectrometer. Peptides covering 75% of the 170-amino acid sequence of Skp1, including ¹⁴⁵IFNIVNDFTP¹⁷⁰EEAAQVR containing unmodified Pro-154 (*m/z* 1011, [M + 2H]²⁺) eluting at 88.7 min, were observed. Potential hydroxylated glycopeptides were sought using a theoretical mass list of glycoforms of this peptide containing any combination of 1–8 residues of Hex, dHex, HexNAc, HexUA, and pentose. This search yielded, at 82.7 min elution time, a single glycopeptide ([M + 2H]²⁺, 1437.1464; [M + 3H]³⁺, 958.09) with an exact match (within 0.56 ppm) to a glycoform containing 1 HexNAc, 3 Hex, and 1 dHex residues (Fig. 1A). Similar results were obtained for a Skp1-SF preparation isolated by anti-FLAG immunoprecipitation from a strain in which Skp1 was C-terminally modified by an SF-tag (data not shown). The putative glycopeptide ion was absent from the RH Δ phyA Δ strain (summarized in Table 2), consistent with its identity as the predicted Skp1 glycopeptide.

The putative glycopeptide ion was subjected to MS/MS analysis to confirm its composition and characterize its organization. Fragmentation of the doubly charged parent ion by CID yielded a series of ions whose mass differences corresponded to loss of a combination of Hex, dHex, and/or HexNAc residues (Fig. 1B), resulting ultimately in the expected hydroxy peptide

Complex Glycosylation of *Toxoplasma Skp1*

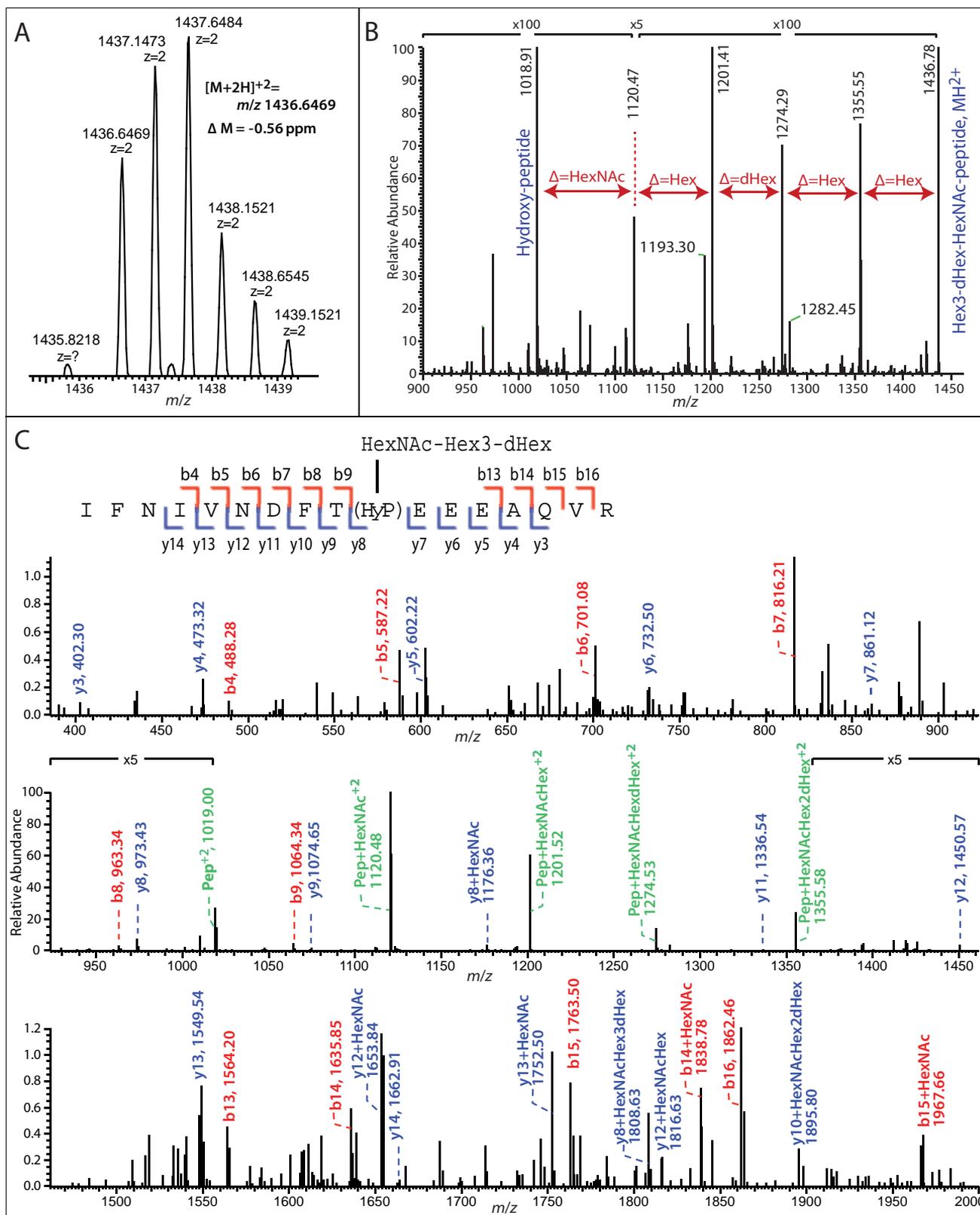


FIGURE 1. Orbitrap MS analysis of the TgSkp1 glycopeptide. RH Δ tachyzoites were lysed out of HFFs, urea-solubilized, and immunoprecipitated with bead-bound affinity-purified anti-TgSkp1 (pAb UOK75). The enriched preparation of TgSkp1 was eluted with triethylamine, reduced and alkylated, trypsinized, and analyzed by reverse phase-HPLC on an LTQ-XL Orbitrap MS. Extracted ion chromatograms showed coelution of a doubly charged (m/z 1436.6464) and a triply charged ion (m/z 958.0983) corresponding with a Δ mass of 0.56 ppm, to the predicted tryptic TgSkp1 peptide ¹⁴⁵IFNIVNDFT(HyP)EEEAQVR¹⁶¹ bearing a pentasaccharide with composition Hex3dHex1HexNAc1 (A). B, CID fragmentation of the doubly charged precursor ion yields a sequential loss of monosaccharide residues corresponding to Hex, Hex, dHex, Hex, and HexNAc, indicating the presence of a linear pentasaccharide. C, inspection of the full CID fragmentation spectrum shows b- (blue annotations) and y- (red annotations) ion series that match the predicted peptide sequence, as illustrated in the inset, and demonstrate that the glycan is linked via a hydroxylated derivative of Pro-154. Peptides with residual sugars are annotated in green.

TABLE 2

MS detection of Skp1 glycopeptides in strains

Isoforms of the Skp1 peptide ¹⁴⁵IFNIVNDFTPEEEAQR were detected as described in Fig. 1. The distribution of raw ion counts among the detected isoforms are shown for the strains analyzed.

Strain	Expt #	Unmodified peptide			Hyp peptide			Hyp-HexNAc peptide			Hyp-penta-saccharide peptide		
		Occu-pancy*	m/z**	ΔM (ppm)	Occu-pancy*	m/z**	ΔM (ppm)	Occu-pancy*	m/z**	ΔM (ppm)	Occu-pancy*	m/z**	ΔM (ppm)
RHΔΔ	1	73%	1011.0002	-2.18	0%			0%			27%	1436.6464	-0.92
RHΔΔ	2	64%	1011.0052	2.77	0%			0%			35%	1436.6538	4.23
RHΔΔ/SF	1	78%	1011.0012	-1.19	0%			0%			22%	1436.6456	-1.48
RHphyAΔ-1	1	100%	1011.0018	-0.59	0%			0%			0%		
RHphyAΔ-1/SF	1	100%	1011.0044	1.98	0%			0%			0%		
RHgnt1Δ	1	60%	1010.9994	-2.97	40%	1018.9995	-0.35	0%			0%		
RHgnt1Δ	2	55%	1011.0024	0.00	45%	1019.0004	0.53	0%			0%		
RHpgtAΔ	1	72%	1011.0031	0.69	0%			28%	1120.5431	3.19	0%		

* apparent occupancy based on raw spectral counts.

** values refer to [M+2H]²⁺ ions.

ion. The predominant ions were consistent with the presence of a linear pentasaccharide whose composition from the non-reducing end is Hex-Hex-dHex-Hex-HexNAc-. Fragmentation also yielded a series of a singly charged peptide and glycopeptide fragment b- and y-ions that confirmed the predicted amino acid sequence of the parent ion and demonstrated the position of the additional O atom as occurring on Pro (to yield Hyp) and the attachment of all sugars via Pro-154 (Fig. 1C). These data indicate that a fraction of Skp1 in *Toxoplasma* is modified by a linear pentasaccharide reminiscent of the linear pentasaccharide on DdSkp1.

Predicted TgSkp1-modifying Glycosyltransferases—BLASTP and TBLASTN searches for sequences corresponding to the three GT genes that catalyze formation of the pentasaccharide on DdSkp1 in ToxoDB (Version 7.3) yielded high scoring hits for DdGnt1 and DdPgtA. No candidates for a homolog of DdAgtA were detected using these algorithms or PSI- or PHI-BLAST toward either its catalytic or WD40 repeat domains. The *Toxoplasma* Ddgnt1-like sequence is represented by a one-exon gene model (Fig. 2A) in three sequenced strains of *Toxoplasma* (GT1, ME49, and VEG). The GT1 (type 1 strain) sequence (TGGT1_315885) exhibited 42% identity and 67% similarity to DdGnt1 over 214 amino acids of the ~250 amino acid catalytic domain (supplemental Fig. S1). Like DdGnt1, TGGT1_315885 is predicted to be a cytoplasmic protein because of the absence of detectable membrane or nuclear targeting motifs. However, at 1510 amino acids, TGGT1_315885 is substantially longer than DdGnt1 by 423 amino acids. As revealed by the amino acid sequence alignment (supplemental Fig. S1), and as illustrated in Fig. 2C, this results from multiple insertions throughout the length of the protein, a common occurrence in *Toxoplasma* genes as observed, e.g. in TgIF2Kb

(19). In addition, a C-terminal sequence referred to as domain A that lies outside of the predicted catalytic domain (10) but is required for DdGnt1 activity is weakly conserved across these predicted proteins (Fig. 2). The remaining intervening sequences are poorly conserved even among coccidian apicomplexans that have PhyA-, Gnt1-, and PgtA-like sequences (supplemental Fig. S1). The coccidian *Sarcosystis neurona* is predicted to contain even longer insert sequences. In comparison, sequences from *Chromera velia* and *Vitrella brassicaformis*, representatives of the closest known photosynthetic relatives of apicomplexans in the alveolate superphylum (20), largely lack these inserts and resemble the length of the *Dictyostelium* sequence. Although the TgGnt1-coding region remains to be confirmed experimentally, two of the inserts are present in expressed sequence tags derived from mRNA, and a third was detected in a shotgun proteomics screen (Fig. 2C; supplemental Fig. S1). Although the significance of these additional sequences is enigmatic, their low conservation suggests that they fulfill *Toxoplasma*-specific functions that are unlikely to be related to the proposed enzymatic activity.

The *Toxoplasma* DdpgtA-like sequence is represented by a 14-exon gene model (Fig. 2B). The predicted amino acid sequence of TGGT1_260650, from the type I GT1 strain, is 98 and 97% identical to that of the type II ME49 and type III VEG strain sequences, and the protein is predicted to be cytoplasmic. The *Toxoplasma* candidate is 1801 amino acids long, compared with the 768 amino acid length of DdPgtA, and the order of the two putative glycosyltransferase domains is reversed (Fig. 2D). The N-terminal CAZy GT2 family sequence of DdPgtA, which encodes a β3-GalT activity, has its sequence homolog in the C-terminal half of the *Toxoplasma* protein, whereas the C-terminal CAZy GT74 family sequence of DdPgtA, which encodes

Complex Glycosylation of *Toxoplasma* Skp1

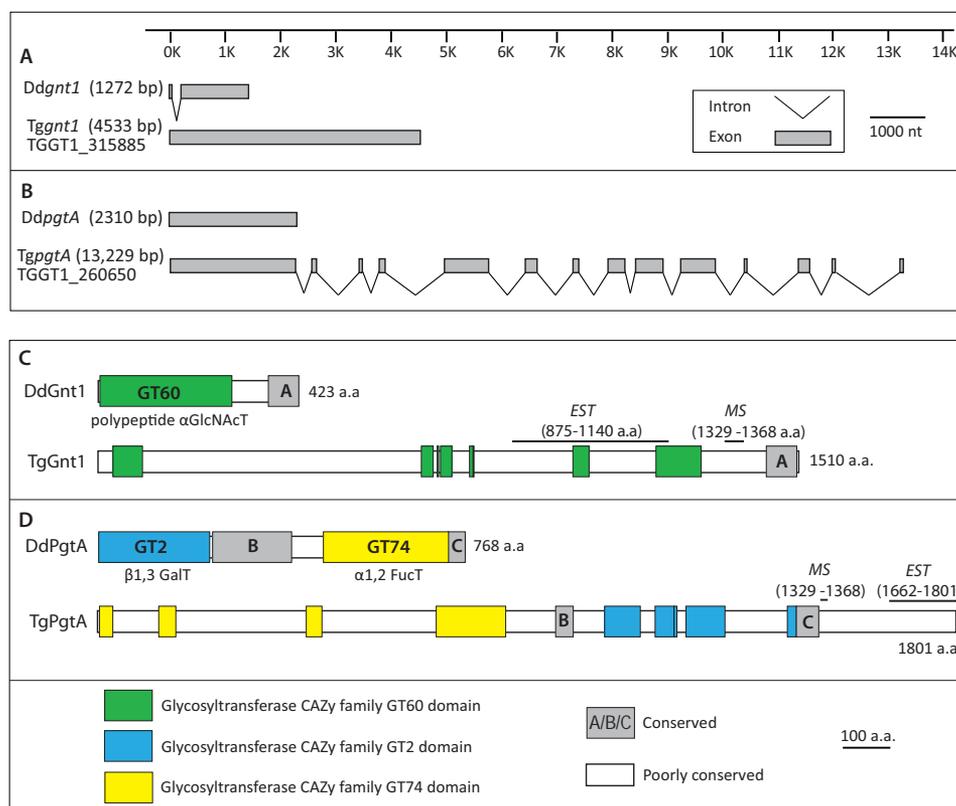


FIGURE 2. Comparative genomic and domain organization of Gnt1 and PgtA from *Dictyostelium* and *Toxoplasma*. A and B, exon-intron organization of the *gnt1* and *pgtA* genes. *gnt1* (A) and *pgtA* (B) gene models from *Dictyostelium discoideum* are from Refs. 8, 9 and available at dictybase.org (27). The gene models from *T. gondii* (Gt1 (type 1) strain) are from Ref. 28. The length from the start codon to the stop codon in nucleotides is in parentheses. C and D, domain organization of Gnt1 (C) and PgtA (D) proteins. Gnt1- and PgtA-like sequences from *Toxoplasma* and three other coccidian apicomplexans were aligned with corresponding sequences from *D. discoideum* and three other amoebozoans, and sequences from two chromerid alveolates, as shown in supplemental Figs. S1 and S2. Regions of high conservation among all 10 sequences are shown in color for the glycosyltransferase-like sequences and in gray for non-GT-like sequences. *Toxoplasma* sequences whose expression has been confirmed at the transcriptional (expressed sequence tags) or proteomic (MS) level are indicated in Ref. 28. Diagrams are shown to scale.

an α 2-FucT activity, has its homolog in the N-terminal half of the *Toxoplasma* protein. The GT2-like domain of the predicted TgPgtA is 22% identical and 60% similar to the DdPgtA- β GalT domain over the 191 most conserved amino acids, and the GT74-like domain is 33% identical and 66% similar to the DdPgtA- α FucT domain over 190 amino acids. The amino acid sequence alignment (supplemental Fig. S2) of the PgtA-like sequences reveal that, as for DdGnt1, the predicted apicomplexan PgtA-like sequences have numerous multiple inserts relative to the *Dictyostelium* prototype, as illustrated in Fig. 2D. These insert sequences, several of which occur in the protein based on expressed sequence tags and MS data (Fig. 2D), tend to diverge even within the coccidian apicomplexans and are minimal in the chromerids. Thus, they are unlikely to be critical for enzymatic activity.

Tggnt1 and TgpgtA Are Required for TgSkp1 Glycosylation—To determine whether Tggnt1 and TgpgtA are involved in TgSkp1 glycosylation, their genes were disrupted by double-crossover homologous recombination in the RH $\Delta\Delta$ strain, as described under “Experimental Procedures” and illustrated in Fig. 3A for Gnt1. Deletion of *gnt1* in recovered clones was demonstrated by loss of a PCR product for *gnt1*-coding DNA, and positive PCRs for the insertion of the selection marker *hxpprt* between *gnt1*-flanking sequences, as described in Fig. 3B. To control for off-target genetic modifications, a *gnt1* disruption

clone was complemented with a version of the original disruption DNA in which *hxpprt* was replaced by the deleted coding region and counter-selected for loss of *hxpprt* (Fig. 3A). The same set of PCRs was used to confirm the desired gene restoration in clonal isolates (Fig. 3C).

The effect of *gnt1* deletion on Skp1 glycosylation was evaluated initially by SDS-PAGE and Western blotting. As shown in Fig. 4A, Skp1 from parasites lacking *gnt1* (lane 3) migrated more rapidly than wild-type Skp1 (lane 1), and similarly to Skp1 from parasites whose *phyA* had been disrupted (lane 2). MS searches for the Skp1 glycopeptide were negative, but a novel hydroxy peptide corresponding to the hydroxylated but non-glycosylated Skp1 was obtained, in addition to the unmodified peptide (Table 2). Analysis of the complemented strain revealed that normal mobility of Skp1 in SDS-PAGE was at least partially restored. Thus, Skp1 HexNAcylation depends on Gnt1, and by analogy with the *Dictyostelium* example, Gnt1 is expected to directly catalyze addition of the first sugar in α -linkage onto Hyp154.

A similar analysis was performed on PgtA, whose genetic locus was manipulated as described in Fig. 3, D–F. In this case, a recently prepared genomic fosmid (13) was utilized to restore the genomic locus in the *pgtA*-disruption clone. Genomic DNA was utilized because of the large number of predicted introns and our difficulty in isolating a full-length cDNA using RT-PCR

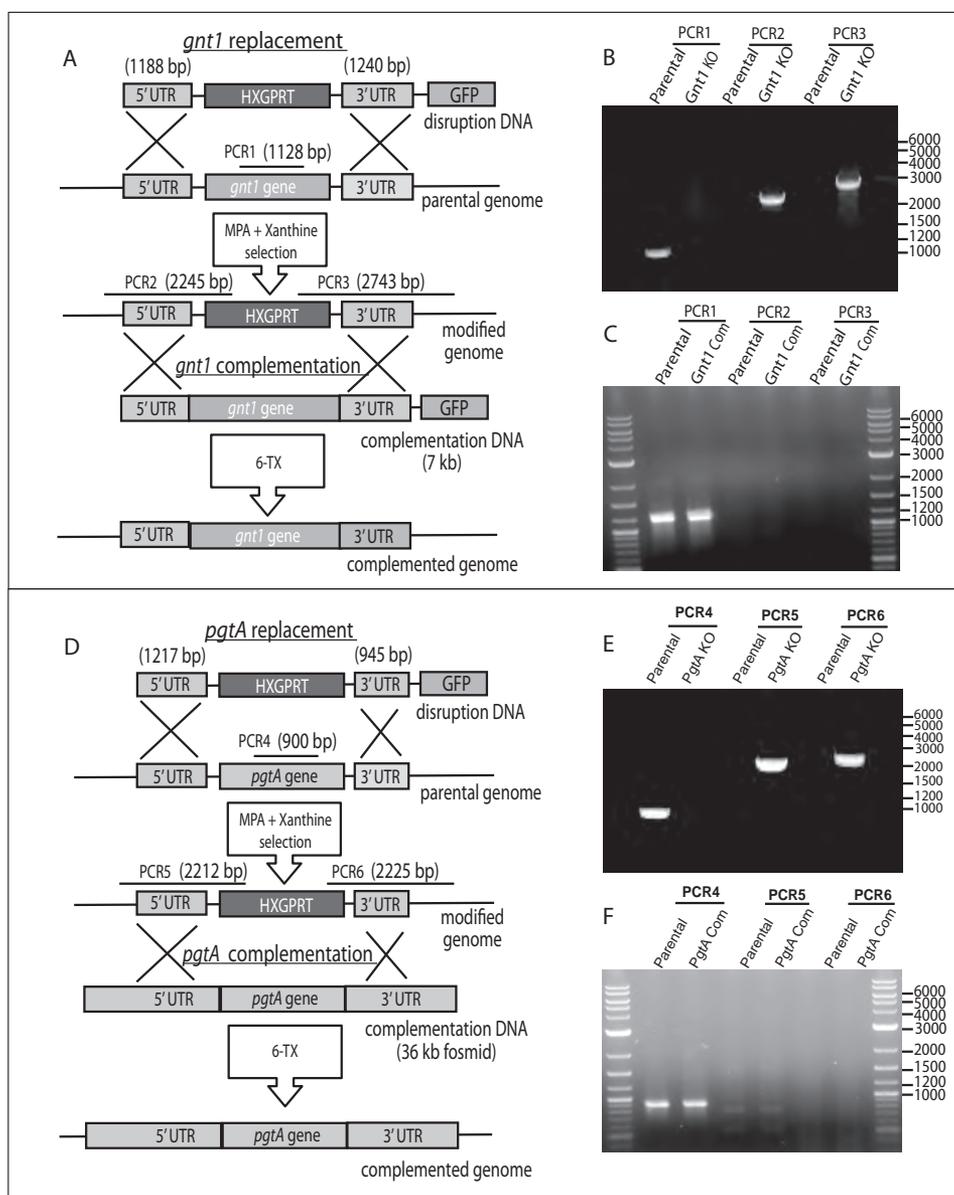


FIGURE 3. Disruption and complementation of *Tggnt1* and *TgpgtA*. A, strategy for deletion of *Tggnt1* and its subsequent complementation. The plasmid-derived disruption of DNA with homologous targeting sequences was electroporated into parasites. Recovery of *hxgpRT*-positive clones that were resistant to MPA and xanthine and were GFP-negative were candidates for double crossover gene replacement. B, gene replacement was confirmed by PCR-1, which demonstrated loss of *gnt1* coding DNA, and PCR-2 and -3, which demonstrated that the inserted *hxgpRT* DNA was flanked by neighboring *gnt1* DNA. To complement *Tggnt1* in the disruption strain, a plasmid containing an ~7-kb genomic locus, including *Tggnt1* coding region and 5'- and 3'-untranslated regions (A), was transfected. Complemented strains where the *hxgpRT* is replaced by *Tggnt1* locus were counter-selected under 6-thioxanthine. C, *gnt1* replacement was confirmed by the positive PCR-1 and negative reactions for PCR-2 and PCR-3, which depended on the presence of *hxgpRT*. D–F, *TgpgtA* was similarly targeted for disruption and complementation. Characteristics of the above strains are summarized in Table 1.

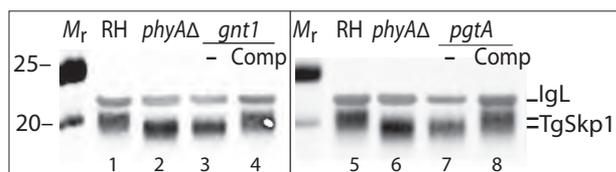


FIGURE 4. Disruption of *Tggnt1* or *TgpgtA* affects *TgSkp1* glycosylation. Soluble S16 fractions from equivalent numbers (3×10^6 cells) of parental RH $\Delta\Delta$ (RH) and RH $\text{phyA}\Delta$ -1 (*phyA*⁻), RH $\text{gnt1}\Delta$, RH $\text{pgtA}\Delta$, and their complemented cells were resolved by 4–12% SDS-PAGE, electroblotted, and probed using anti-TgSkp1 (UOK75) antiserum. Changes in glycosylation inferred from altered gel mobility were confirmed by mass spectrometry (Table 2). Similar results were obtained for independently derived clones of RH $\text{gnt1}\Delta$ and RH $\text{pgtA}\Delta$.

(data not shown), and the fosmid clone was used because of the predicted large size (13.2 kb) of its genomic locus. As shown in Fig. 4, lane 7, Skp1 glycosylation also appeared to be affected by the loss of *pgtA* based on SDS-PAGE/Western blotting, and this was confirmed by the accumulation of the HexNAc form of Skp1, the expected acceptor substrate of PgtA, based on MS analysis of tryptic peptides (Table 2). As expected, restoration of the *pgtA* locus resulted in at least partial recovery of Skp1 glycosylation (Fig. 4, lane 8). At a minimum, PgtA is thus required for addition of the second sugar to the Skp1 glycan.

TgGnt1 Has Properties of a Skp1 Polypeptide UDP-GlcNAc:HO-Skp1 GlcNAcT—A previous study detected a Gnt1-like activity in tachyzoite cytosolic extracts, based on transfer of ³H

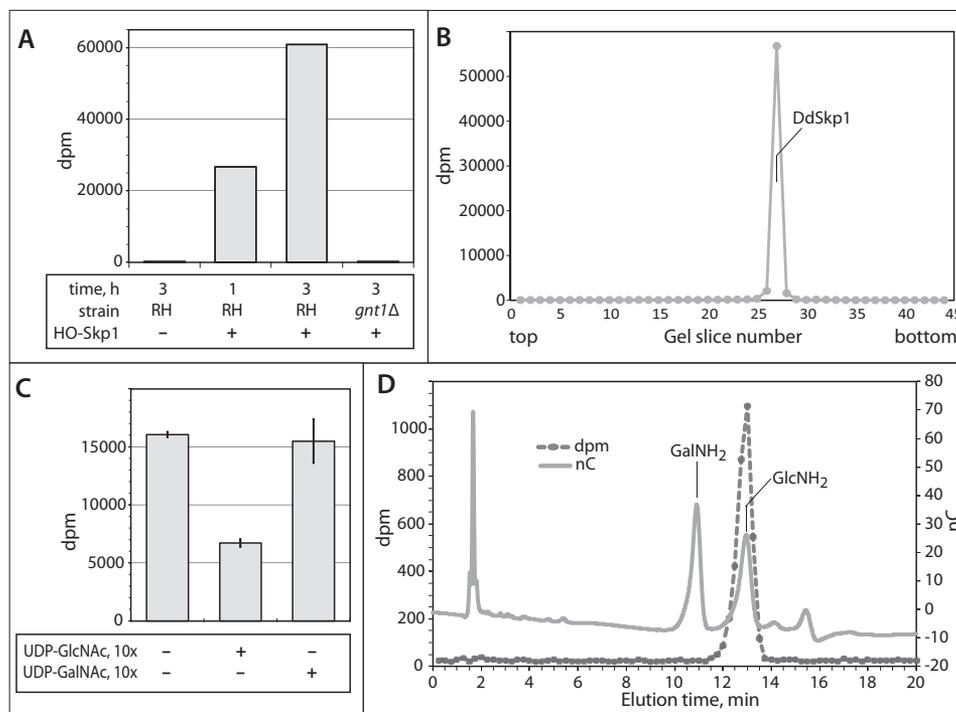


FIGURE 5. TgGnt1 is a Skp1 GlcNAcT. *A*, GlcNAcT activity in S100 cytosolic parasite was assayed based on transfer of ^3H from $0.5\ \mu\text{M}$ UDP- ^3H GlcNAc to exogenous HO-DdSkp1 for 1–3 h as described under “Experimental Procedures.” Reactions were loaded onto and separated on an SDS-polyacrylamide gel, and the Coomassie Blue-stained DdSkp1 bands were excised and subjected to liquid scintillation spectroscopy. The reaction time, presence of HO-Skp1, and source of the extract (RHΔΔ or RH, or RH*gnt1*Δ or *gnt1*Δ) were varied as indicated. *B*, entire lane from a parallel 3-h reaction (RH, +HO-Skp1) from *A* was analyzed for incorporation of ^3H . Incorporation was only detected at the migration position of DdSkp1. *C*, donor substrate specificity of the GlcNAcT activity was examined by including a 9-fold excess of unlabeled UDP-GlcNAc or UDP-GalNAc to reactions containing $10\ \mu\text{M}$ UDP- ^3H GlcNAc. Incorporation was measured as in *A*. Error bars show standard deviations of the mean of two replicates from each of two independent reactions. *D*, analysis of incorporated ^3H . The reacted Skp1 band was excised from a PVDF membrane electroblot of the SDS-polyacrylamide gel, subjected to acid hydrolysis in $6\ \text{N}$ HCl, and analyzed by high pH anion exchange chromatography. The hydrolysate was supplemented with GlcNH₂ and GalNH₂ and chromatographed on a Dionex PA-1 column. Elution of the sugar standards was monitored by a pulsed amperometric detector (*nC*), and fractions were collected to monitor the elution of ^3H by scintillation counting (*dpm*).

from UDP- ^3H GlcNAc to DdHis₁₀-Skp1 that was recovered from an SDS-polyacrylamide gel (21). This activity depended on addition of DdPhyA indicating dependence on Hyp. Replication of an optimized form of this assay (see “Experimental Procedures”) revealed time-dependent transfer of ^3H to HO-DdSkp1 that was absent from *gnt1*Δ extracts (Fig. 5*A*). As shown in Fig. 5*B*, no incorporation into endogenous proteins was detected based on analysis of an entire SDS-polyacrylamide gel lane, indicating absence of activity of endogenous GTs from other sources, such as the Golgi, that modify other targets in these cytosolic preparations. The lack of other radiolabeled proteins is inconsistent with the existence of an intermediate TgGnt1 substrate that itself mediates modification of Skp1. Incorporation was reduced as expected after addition of a 10-fold excess of unlabeled UDP-GlcNAc, but not of unlabeled UDP-GalNAc (Fig. 5*C*), indicating that the enzyme is selective for the GlcNAc isomer. Because a homolog of Gnt1 that resides in the Golgi transfers GalNAc to proteins (22), and *Toxoplasma* possesses an epimerase that can interconvert UDP-GlcNAc with UDP-GalNAc, the nature of the transferred ^3H was confirmed by another method. ^3H was found to be incorporated as GlcNAc, based on co-chromatography of ^3H released by HCl hydrolysis, which de-*N*-acetylates GlcNAc to GlcNH₂, with a GlcNH₂ standard (Fig. 5*D*). Although the evidence that TgGnt1 modifies Skp1 in this assay is indirect, its ho-

mology with DdGnt1, whose purified recombinant version can α GlcNAcylate HO-DdSkp1 *in vitro* (8), suggests that TgGnt1 also directly α GlcNAcylates TgSkp1.

TgPgtA Is a Bifunctional Glycosyltransferase with GalT and FucT Activities—To characterize the role of *pgtA* in extending the Skp1 glycan, the above assay was first modified by substituting UDP- ^3H Gal for UDP- ^3H GlcNAc and GlcNAc-Skp1 for HO-Skp1. Time-dependent incorporation of ^3H into GlcNAc-Skp1 was observed (Fig. 6*A*). Incorporation into the Skp1 band on the SDS-polyacrylamide gel required the inclusion of Skp1, and Skp1 was the only protein that incorporated detectable radioactivity (Fig. 6*B*). No incorporation was detected in extracts of *pgtA*Δ cells. Recovery and analysis of the ^3H after acid hydrolysis confirmed incorporation as Gal rather than a derivative (Fig. 6*C*). Similar findings were observed in a corresponding FucT assay, in which UDP- ^3H Gal was replaced by GDP- ^3H Fuc (Fig. 6, *D–F*). However, incorporation of ^3H depended on the inclusion of UDP-Gal (unlabeled), in contrast to the GalT reaction that did not require GDP-Fuc (Fig. 6*A*). This indicated that TgPgtA is, like DdPgtA (9), a processive diglycosyltransferase that catalyzes the sequential addition of Gal and then Fuc, an order consistent with the MS-MS data. Although >95% of incorporation of ^3H Fuc into the Skp1 band depended on Skp1 and *pgtA* (Fig. 6, *D* and *E*), residual incorporation was observed at this

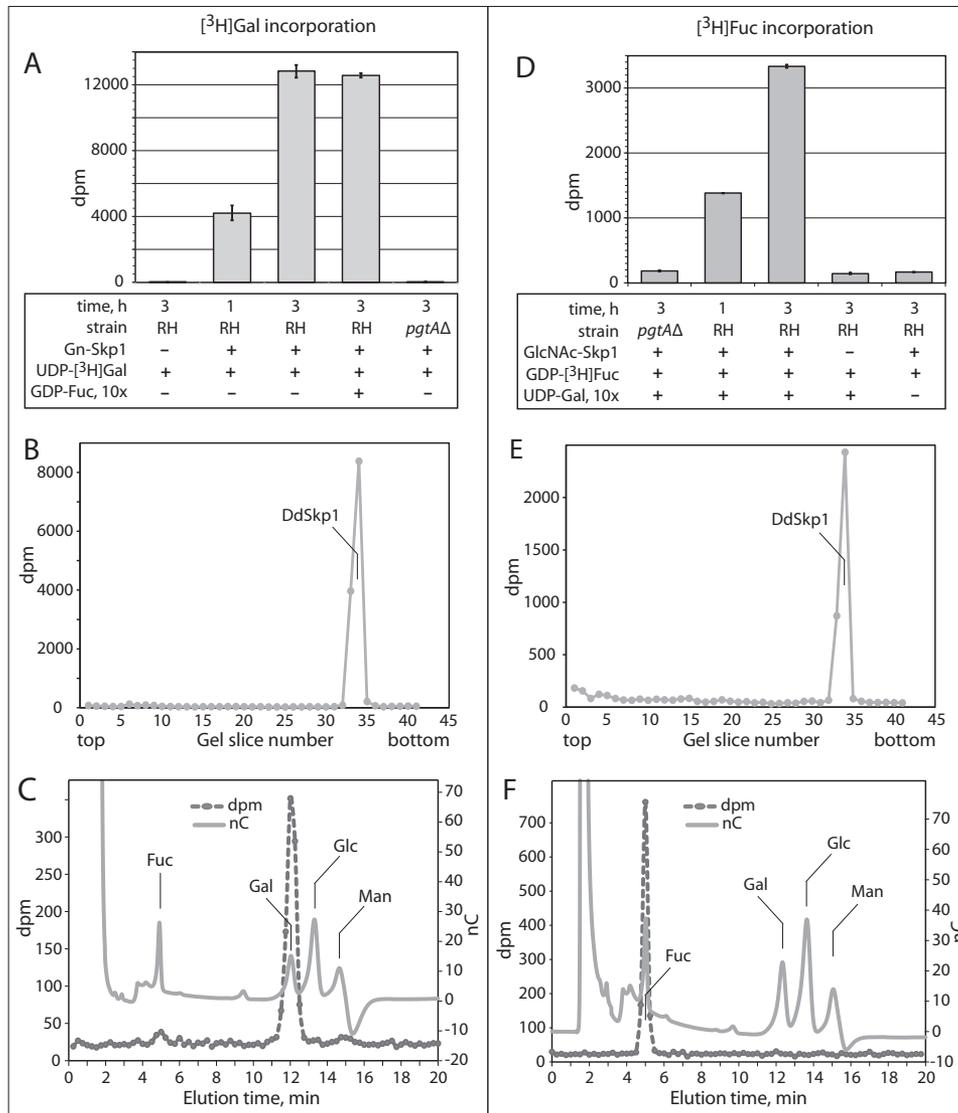


FIGURE 6. PgtA is a Skp1 GalT and FucT. *A*, GalT activity directed toward GlcNAc-Skp1 was assayed as described for GlcNAcT activity in Fig. 5A, except that GlcNAc-Skp1 and UDP-[³H]Gal were used in place of HO-Skp1 and UDP-[³H]GlcNAc. The reaction time, inclusion of GlcNAc-Skp1 and GDP-Fuc, and source of the extract (RHΔΔ or RH, or RH*pgtA*Δ or *pgtA*Δ) were varied as indicated. *B*, entire lane from a parallel 3-h reaction (RH, +GlcNAc-Skp1) from *A* was analyzed for incorporation of ³H. Incorporation was only detected at the migration position of DdSkp1. *C*, [³H]DdSkp1 from the 3-h GalT reaction in *A* was isolated as in Fig. 5D and hydrolyzed in 4 M TFA. The hydrolysate was chromatographed on a Dionex PA-1 column with internal standards of Gal, Glc, Man, and Fuc, and the elution of ³H was monitored by scintillation counting of collected fractions. *D–F*, FucT activity assays. Reactions were conducted as above except that GDP-[³H]Fuc replaced UDP-[³H]Gal, and the dependence of incorporation on a 10-fold concentration excess of UDP-Gal, GlcNAc-Skp1, and PgtA in the extract and time was examined.

position and elsewhere in the gel indicative of additional *pgtA*- and Skp1-independent FucT activity in the extract. These results are consistent with accumulation of GlcNAc-Skp1 and the absence of higher glycosylation states in *pgtA*Δ cells (Fig. 4 and Table 2).

TgGnt1 and TgPgtA Are Important for Toxoplasma Growth in Cell Culture—Previous studies revealed that disruption of exon 1 of *phyA* results in a parasite growth defect, which could be detected as reduced plaque areas after 5 days of replication on a fibroblast monolayer (5). To check that no residual *phyA* activity was present, all nine exons were deleted (RH*phyA*Δ-2). A similarly reduced ability to grow on monolayers was observed, as illustrated in Fig. 7A and quantified in *B*. Furthermore, deletion of exon 1 in a strain in which Skp1 was C-terminally modified with an SF-epitope tag, which itself did not affect

growth, also resulted in slowed growth (Fig. 7C). To examine the roles *gnt1* and *pgtA*, the plaque-forming abilities of the disruption strains described above were analyzed. As shown in Fig. 7B, *gnt1*Δ cells exhibited slow growth that was statistically indistinguishable from that of *phyA*Δ cells. *pgtA*Δ cells also exhibited a slow growth phenotype, which was intermediate between that of *gnt1*Δ and parental (RHΔΔ) cells. Complementation of *gnt1* and *pgtA* by gene replacement at their original loci restored normal growth (Fig. 7D), showing that the growth differences in the original disruption strains could be attributed to the GT targets. Although these findings do not demonstrate directly that the altered modifications of Skp1 are involved in reduced growth, the finding that three independent enzymes that share Skp1 as a target substrate exhibit a similar deficiency is consistent

Complex Glycosylation of *Toxoplasma Skp1*

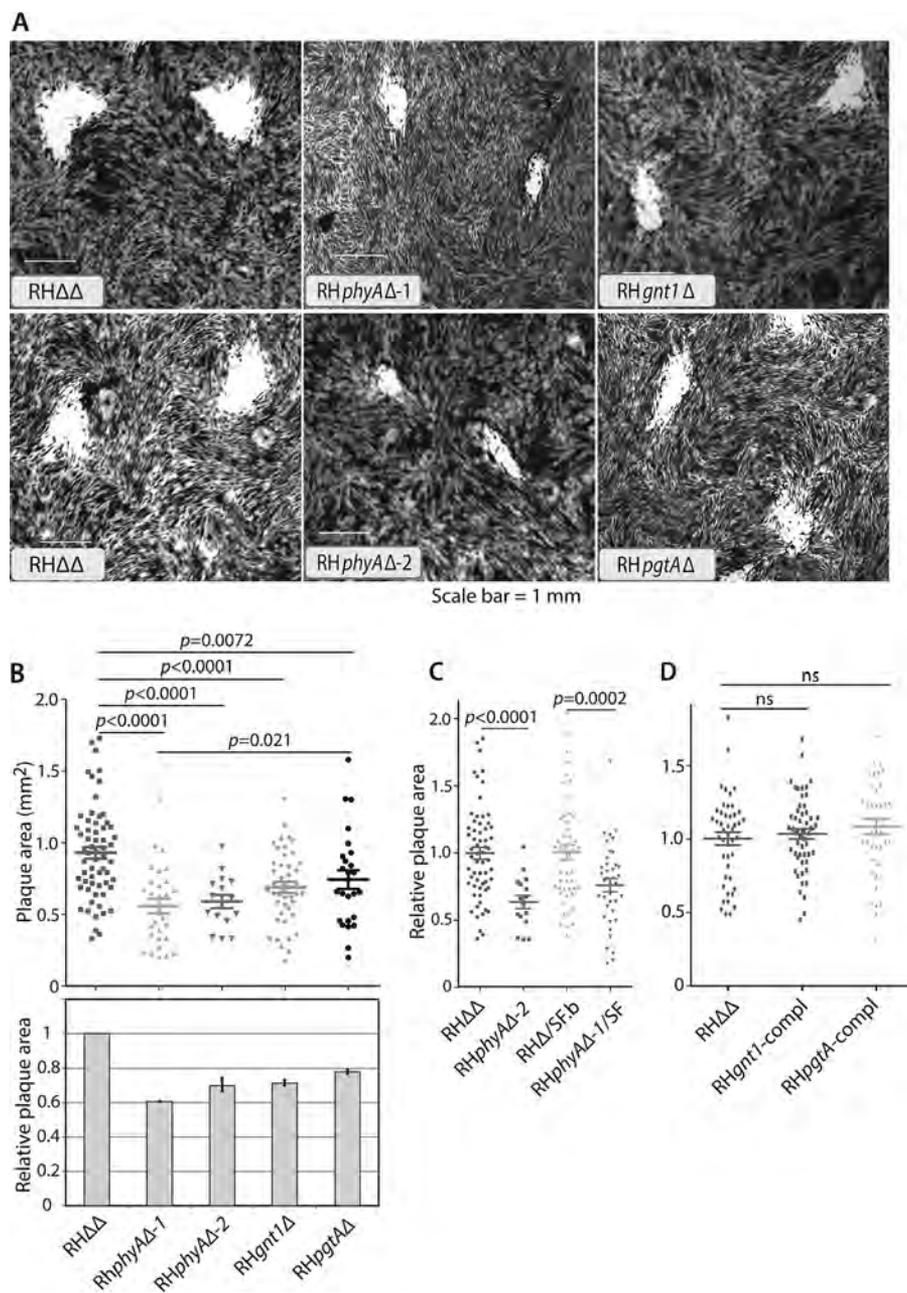


FIGURE 7. Role of *Tggt1* and *TgpgtA* in parasite proliferation. HFF monolayers were inoculated with freshly isolated tachyzoite stage parasites at a multiplicity of infection of 0.002. After 5.5 days, monolayers were stained with crystal violet. **A**, representative images of cleared areas of the host monolayers. **B–D**, images digitized and plaque areas were calculated. The dot plots show the area distributions and average values \pm S.E. from a representative of two independent experiments. Average parental strain areas ranged from 0.5 to 1.0 mm². *p* values for statistical significance of the differences, based on a one-way analysis of variance test, are shown above. *ns* = not significant. **B**, data from RH $\Delta\Delta$, RH $phyA\Delta-1$, and RH $phyA\Delta-2$, generated by different strategies, RH $gnt1\Delta$, and RH $pgtA\Delta$ strains. **Bar graph** shows average (\pm S.D.) from two independent experiments. **C**, data from strains in which *Skp1* was SF-tagged. **D**, data from *Tggt1* or *TgpgtA* complemented (*compl*) strains.

with a role for *Skp1*, a model that has both biochemical and genetic support in *Dictyostelium*.

Discussion

Skp1 isolated from the tachyzoite stage of *Toxoplasma* is partially modified by a glycan chain that consists of five sugars and is linked to the hydroxylated form of Pro-154. The monosaccharides are organized as a linear pentasaccharide with a sequence, reading from the peripheral non-reducing end, of Hex-Hex- α Fuc- β Gal- α GlcNAc-, based on mass spectrometry

and characterization of the core GTs. Remarkably, these properties match those of the glycan that was previously characterized on *Skp1* from an unrelated protist, the social amoeba *Dictyostelium* (6). Genetic disruption of glycosylation rendered a growth defect in host cell monolayers, implicating a role for *Toxoplasma* E3^{SCF} ubiquitin ligases in cell proliferation.

The predicted *Skp1* glycan was analyzed at the glycopeptide level because of the resistance of the glycan-Hyp linkage to known methods of cleavage. Native *Skp1* from parasites was analyzed to avoid overexpression artifacts, but this necessitated

application of highly sensitive methods due to the limited amount of material available because *Toxoplasma* can only be grown intracellularly. As is typical for the analysis of glycopeptides by mass spectrometry, their detection required manual inspection of spectra for candidate ions, because of suppression and abundance issues. The pentasaccharide peptide was detected in an exhaustive search of the primary spectra for tryptic peptides bearing Pro-154 (Fig. 1), the previously documented hydroxylation site (5), and any combination of up to eight monosaccharides. Decomposition analysis in the gas phase confirmed the glycan's location at Hyp154 and suggested that it is organized as a linear pentasaccharide with a sequence, from the non-reducing end, of Hex-Hex-dHex-Hex-HexNAc-. The only other peptide isoform detected was non-hydroxylated and therefore non-glycosylated (Table 2), suggesting limited glycosylation microheterogeneity.

The sugar identities were investigated by highly sensitive incorporation of radioactive sugars mediated by parasite extracts that harbor the biosynthetic enzymes. Cytosolic extracts were observed to specifically incorporate radioactivity from three different radioactive sugar nucleotides into recombinantly generated isoforms of *Dictyostelium* Skp1, which was previously shown to be an excellent substrate for DdPhyA (5). Based on the confirmed identities of the ^3H -sugars after incorporation, the core trisaccharide is concluded to consist of Fuc-Gal-GlcNAc, which matches the dHex-Hex-HexNAc found by MS.

Incorporation of [^3H]GlcNAc into HO-Skp1 depended on *Tggnt1* based on the absence of activity in *gnt1* Δ extracts (Fig. 5). By analogy with its *Dictyostelium* ortholog, TgGnt1 is inferred to transfer GlcNAc in an α -linkage to Hyp154 of TgSkp1. Incorporation of [^3H]Gal depended on *TgpgtA* (Fig. 6), and the addition of [^3H]Fuc, which also depended upon *TgpgtA*, required prior addition of Gal (Fig. 6). These linkages require further study to determine whether they are conserved with the *Dictyostelium* example. Thus, although the physical order of the two GT domains is reversed relative to DdPgtA (Fig. 2), the order of addition of the two sugars is conserved. The existence of the two terminal Hex residues was unexpected because of the absence of a homolog for AgtA in the *Toxoplasma* genome, and their identities are under current investigation.

Both *Tggnt1* and *TgpgtA* were important for efficient plaque-forming ability in tissue culture monolayers cultivated under standard conditions (Fig. 7). Loss of *gnt1* was as severe as loss of *phyA*, the prolyl 4-hydroxylase that is required for Gnt1 action on Skp1. Loss of *pgtA* resulted in an intermediate effect, i.e. the plaque-forming ability was improved relative to cells lacking *phyA*. These effects were specific for the GT genes as normal plaque-forming ability was restored upon genetic complementation to the original genotype. Thus, Gnt1 appears to make a major contribution to *phyA*-dependent growth, but complete realization appears to depend on Gnt1-dependent glycosylation contributed by, at least on Skp1, PgtA and the additional unknown GTs. In *Dictyostelium*, biochemical studies indicate that Skp1 is the only substrate for the orthologs of these GTs, and disruption of either *gnt1* or *pgtA* inhibits development in a way that is related to, but less severe than, disruption of *phyA*

(10). Gene dosage manipulations on Skp1 expression show inverse effects on development consistent with the modification genes acting via Skp1 in this organism, a model that is supported by point mutations that remove the target Pro residue, and from a double mutant between a GT gene and one of two Skp1 genes (23). By analogy, and based on similar phenotypes of *TgphyA* Δ and *Tggnt1* Δ , we propose that the modification genes render their effects in a common pathway that affects Skp1. Further studies on Skp1 itself are needed to evaluate this possibility.

The reduced plaque areas are consistent with a role for Skp1 modification in cell cycle progression as demonstrated in yeast and mammalian cells, where SCF-type E3 ubiquitin ligases are important for signaling proteasome-dependent turnover of cell cycle kinase inhibitors (24). However, SCF ligases represent a family of enzymes that includes many F-box proteins with distinct substrate receptor activities (25). Given the large variety of substrates known to be recognized by yeast, plant, and human F-box proteins, further studies are needed to evaluate whether other processes required for plaque formation, such as binding of the parasite to host cells, ingress and/or egress are affected. In future studies, it will be interesting to evaluate whether any of these potential mechanisms are more severely affected in low O_2 or altered metabolic states that are anticipated to influence Skp1 modification enzyme activities in cells (11).

PhyA-, Gnt1-, and PgtA-like sequences are selectively conserved across a broad spectrum of unicellular eukaryotes, including representatives of all major protist clades (6). Validation of their shared functions in both *Toxoplasma* and *Dictyostelium*, which are highly diverged, suggests that Skp1 hydroxylation and glycosylation occurred in ancestral eukaryotes prior to their loss in fungi, higher plants, animals, and select protists. The relatively rapid evolution of F-box proteins, with which Skp1 partners for many if not all of its functions, suggests that Skp1 modifications serve an outsized role in environmental regulation of unique lineage and species-specific functions in many unicellular organisms.

Author Contributions—K. R. conducted most of the experiments, analyzed the results, assembled the figures, and wrote most of the first draft. P. Z. conducted the mass spectrometry and interpreted the data together with L. W. M. M. and H. vdW. contributed to the enzymatic assays and HPAEC analyses. I. J. B. and C. M. W. conceived the ideas and the experimental strategies for the project. C. M. W. wrote the final draft.

Acknowledgments—We thank Drs. Boris Striepen, Gustavo Arrizabalaga, and Vern Carruthers for the gifts of plasmids.

References

- Kim, K., and Weiss, L. M. (2004) *Toxoplasma gondii*: the model apicomplexan. *Int. J. Parasitol.* **34**, 423–432
- Pereira-Chioccola, V. L., Vidal, J. E., and Su, C. (2009) *Toxoplasma gondii* infection and cerebral toxoplasmosis in HIV-infected patients. *Future Microbiol.* **4**, 1363–1379
- Luft, B. J., and Remington, J. S. (1992) Toxoplasmic encephalitis in AIDS. *Clin. Infect. Dis.* **15**, 211–222
- Lambert, H., and Barragan, A. (2010) Modelling parasite dissemination: host cell subversion and immune evasion by *Toxoplasma gondii*. *Cell*

Complex Glycosylation of *Toxoplasma Skp1*

- Microbiol.* **12**, 292–300
- Xu, Y., Brown, K. M., Wang, Z. A., van der Wel, H., Teygong, C., Zhang, D., Blader, I. J., and West, C. M. (2012) The Skp1 protein from *Toxoplasma* is modified by a cytoplasmic prolyl 4-hydroxylase associated with oxygen sensing in the social amoeba *Dictyostelium*. *J. Biol. Chem.* **287**, 25098–25110
 - West, C. M., Wang, Z. A., and van der Wel, H. (2010) A cytoplasmic prolyl hydroxylation and glycosylation pathway modifies Skp1 and regulates O₂-dependent development in *Dictyostelium*. *Biochim. Biophys. Acta* **1800**, 160–171
 - Schafer, C. M., Sheikh, M. O., Zhang, D., and West, C. M. (2014) Novel regulation of Skp1 by the *Dictyostelium* AgtA α -galactosyltransferase involves the Skp1-binding activity of its WD40 repeat domain. *J. Biol. Chem.* **289**, 9076–9088
 - van der Wel, H., Morris, H. R., Panico, M., Paxton, T., Dell, A., Kaplan, L., and West, C. M. (2002) Molecular cloning and expression of a UDP-*N*-acetylglucosamine (GlcNAc):hydroxyproline polypeptide GlcNAc-transferase that modifies Skp1 in the cytoplasm of *Dictyostelium*. *J. Biol. Chem.* **277**, 46328–46337
 - van der Wel, H., Fisher, S. Z., and West, C. M. (2002) A bifunctional diglycosyltransferase forms the Fuca1,2Gal β 1,3-disaccharide on Skp1 in the cytoplasm of *Dictyostelium*. *J. Biol. Chem.* **277**, 46527–46534
 - Zhang, D., van der Wel, H., Johnson, J. M., and West, C. M. (2012) The Skp1 prolyl 4-hydroxylase of *Dictyostelium* contributes glycosylation-independent and -dependent effects on O₂-dependent development without affecting Skp1 stability. *J. Biol. Chem.* **287**, 2006–2016
 - West, C. M., and Blader, I. J. (2015) Oxygen sensing by protozoans: How they catch their breath. *Curr. Opin. Microbiol.* **26**, 41–47
 - Sheikh, M. O., Xu, Y., van der Wel, H., Walden, P., Hartson, S. D., and West, C. M. (2015) Glycosylation of Skp1 promotes formation of Skp1-Cullin-1-F-box protein complexes in *Dictyostelium*. *Mol. Cell. Proteomics* **14**, 66–80
 - Vinayak, S., Brooks, C. F., Naumov, A., Suvorova, E. S., White, M. W., and Striepen, B. (2014) Genetic manipulation of the *Toxoplasma gondii* genome by fosmid recombineering. *MBio* **5**, e02021
 - Huynh, M. H., and Carruthers, V. B. (2009) Tagging of endogenous genes in a *Toxoplasma gondii* strain lacking Ku80. *Eukaryot. Cell* **8**, 530–539
 - Figueras, M. J., Martin, O. A., Echeverria, P. C., de Miguel, N., Naguleswaran, A., Sullivan, W. J., Jr., Corvi, M. M., and Angel, S. O. (2012) *Toxoplasma gondii* Sis1-like J-domain protein is a cytosolic chaperone associated to HSP90/HSP70 complex. *Int. J. Biol. Macromol.* **50**, 725–733
 - Kimmel, J., Smith, T. K., Azzouz, N., Gerold, P., Seeber, F., Lingelbach, K., Dubremetz, J. F., and Schwarz, R. T. (2006) Membrane topology and transient acylation of *Toxoplasma gondii* glycosylphosphatidylinositols. *Eukaryot. Cell* **5**, 1420–1429
 - Teng-Umuay, P., van der Wel, H., and West, C. M. (1999) Identification of a UDP-GlcNAc:Skp1-hydroxyproline GlcNAc-transferase in the cytoplasm of *Dictyostelium*. *J. Biol. Chem.* **274**, 36392–36402
 - Sheikh, M. O., Schafer, C. M., Powell, J. T., Rodgers, K. K., Mooers, B. H., and West, C. M. (2014) Glycosylation of Skp1 affects its conformation and promotes binding to a model F-box protein. *Biochemistry* **53**, 1657–1669
 - Narasimhan, J., Joyce, B. R., Naguleswaran, A., Smith, A. T., Livingston, M. R., Dixon, S. E., Coppens, I., Wek, R. C., and Sullivan, W. J., Jr. (2008) Translation regulation by eukaryotic initiation factor-2 kinases in the development of latent cysts in *Toxoplasma gondii*. *J. Biol. Chem.* **283**, 16591–16601
 - Moore, R. B., Oborník, M., Janouskovec, J., Chrudimský, T., Vancová, M., Green, D. H., Wright, S. W., Davies, N. W., Bolch, C. J., Heimann, K., Slapeta, J., Hoegh-Guldberg, O., Logsdon, J. M., Carter, D. A. (2008) A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451**, 959–963
 - West, C. M., van der Wel, H., and Blader, I. J. (2006) Detection of cytoplasmic glycosylation associated with hydroxyproline. *Methods Enzymol.* **417**, 389–404
 - Wojczyk, B. S., Stwora-Wojczyk, M. M., Hagen, F. K., Striepen, B., Hang, H. C., Bertozzi, C. R., Roos, D. S., and Spitalnik, S. L. (2003) cDNA cloning and expression of UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetyl-galactosaminyltransferase T1 from *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **131**, 93–107
 - Wang, Z. A., Singh, D., van der Wel, H., and West, C. M. (2011) Prolyl hydroxylation- and glycosylation-dependent functions of Skp1 in O₂-regulated development of *Dictyostelium*. *Dev. Biol.* **349**, 283–295
 - Willems, A. R., Schwab, M., and Tyers, M. (2004) A hitchhiker's guide to the cullin ubiquitin ligases: SCF and its kin. *Biochim. Biophys. Acta* **1695**, 133–170
 - Skaar, J. R., Pagan, J. K., and Pagano, M. (2013) Mechanisms and function of substrate recruitment by F-box proteins. *Nat. Rev. Mol. Cell Biol.* **14**, 369–381
 - Fox, B. A., Ristuccia, J. G., Gigley, J. P., and Bzik, D. J. (2009) Efficient gene replacements in *Toxoplasma gondii* strains deficient for nonhomologous end joining. *Eukaryot. Cell* **8**, 520–529
 - Basu, S., Fey, P., Jimenez-Morales, D., Dodson, R. J., and Chisholm, R. L. (2015) DictyBase 2013: integrating multiple Dictyostelid species. *Genesis* **53**, 523–534
 - Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X., Heiges, M., Iodice, J., Kissinger, J. C., Mackey, A. J., Pinney, D. F., Roos, D. S., Stoeckert, C. J., Jr., Wang, H., and Brunk, B. P. (2008) ToxoDB: an integrated *Toxoplasma gondii* database resource. *Nucleic Acids Res.* **36**, D553–D556

A Genome-wide CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) Screen Identifies NEK7 as an Essential Component of NLRP3 Inflammasome Activation^{*[5]}

Received for publication, October 26, 2015, and in revised form, November 5, 2015
Published, JBC Papers in Press, November 9, 2015, DOI 10.1074/jbc.C115.700492

Jonathan L. Schmid-Burgk[‡], Dhruv Chauhan[‡], Tobias Schmidt[‡],
Thomas S. Ebert[‡], Julia Reinhardt[§], Elmar Endl[‡],
and Veit Hornung^{‡¶1}

From the [‡]Institute of Molecular Medicine and [§]Institute of Clinical Chemistry and Clinical Pharmacology, University Hospital, University of Bonn, 53127 Bonn and the [¶]Gene Center and Department of Biochemistry, Ludwig-Maximilians-University Munich, 81377 Munich, Germany

Inflammasomes are high molecular weight protein complexes that assemble in the cytosol upon pathogen encounter. This results in caspase-1-dependent pro-inflammatory cytokine maturation, as well as a special type of cell death, known as pyroptosis. The Nlrp3 inflammasome plays a pivotal role in pathogen defense, but at the same time, its activity has also been implicated in many common sterile inflammatory conditions. To this effect, several studies have identified Nlrp3 inflammasome engagement in a number of common human diseases such as atherosclerosis, type 2 diabetes, Alzheimer disease, or gout. Although it has been shown that known Nlrp3 stimuli converge on potassium ion efflux upstream of Nlrp3 activation, the exact molecular mechanism of Nlrp3 activation remains elusive. Here, we describe a genome-wide CRISPR/Cas9 screen in immortalized mouse macrophages aiming at the unbiased identification of gene products involved in Nlrp3 inflammasome activation. We employed a FACS-based screen for Nlrp3-dependent cell death, using the ionophoric compound nigericin as a potassium efflux-inducing stimulus. Using a genome-wide guide RNA (gRNA) library, we found that targeting *Nek7* rescued macrophages from nigericin-induced lethality. Subsequent studies revealed that murine macrophages deficient in *Nek7* displayed a largely blunted Nlrp3 inflammasome response, whereas Aim2-mediated inflammasome activation proved to be fully intact. Although the mechanism of *Nek7* functioning upstream of Nlrp3 yet remains elusive, these studies provide a first genetic handle of a component that specifically functions upstream of Nlrp3.

* This work was supported European Research Council Grant 647858–GENESIS and Deutsche Forschungsgemeinschaft Grant EXC1023. The authors declare that they have no conflicts of interest with the contents of this article.

[5] This article contains supplemental Figs. 1 and 2 and supplemental Tables 1–3.
¹ To whom correspondence should be addressed: Gene Center and Department of Biochemistry, Ludwig-Maximilians-University Munich, 81377 Munich, Germany. Tel.: 49-89-218076951; Fax: 49-89-218076999; E-mail: hornung@genzentrum.lmu.de.

Employing an evolutionary conserved set of pattern recognition receptors (PRRs),² the innate immune system senses the presence of microbial pathogens (1). Although PRRs can directly detect microbe-associated molecular patterns, some PRRs also respond to endogenous, host-derived signals that are formed or released upon perturbation or damage that is caused by microbial infections. These signals, which are commonly referred to as damage-associated molecular patterns, can also trigger PRR activation in the context of sterile inflammatory conditions (2).

At the cell-autonomous level, PRR engagement and associated signaling cascades can result in a diverse set of responses, such as the induction of pro-inflammatory gene expression, the control of cytoskeletal rearrangement (e.g. in the context of phagocytosis), as well as the activation of proteolytic cascades, such as the inflammasome pathway. The inflammasome is a cytosolic multiprotein complex that regulates the activation and processing of caspase-1 (3). Inflammasome sensor proteins employ the adapter protein ASC to recruit caspase-1, which in turn results in the proximity-induced autoprocessing and activation of caspase-1. Active caspase-1 cleaves and thereby matures pro-inflammatory cytokines such as IL-1 β and IL-18, and at the same time, it triggers an inflammatory type of cell death known as pyroptosis (3). The exact mechanisms of this caspase-1-dependent cell death are currently unknown, but recent evidence suggests that cleavage of the cytosolic protein Gasdermin-D by caspase-1 might in part be responsible for cell death induction (4, 5).

Of all known inflammasome-forming sensors, the Nlrp3 inflammasome plays a central role in antimicrobial defense (6). To this end, a large array of microbial pathogens and microbe-derived molecules have been described to trigger Nlrp3 inflammasome activation. As such, it appears most plausible that Nlrp3 functions as an indirect sensor of cellular perturbation rather than a receptor that would be dedicated to a certain class of exogenous ligands. This notion of Nlrp3 being a general sensor of cellular damage is well in line with its dominant role in sterile inflammatory diseases (7). In fact many sterile inflammatory conditions, in which the cellular integrity of myeloid cells is compromised, display an involvement of Nlrp3-dependent inflammation. Unlike other inflammasome sensors, Nlrp3 activity is subject to a number of regulatory steps that function as additional safeguard mechanisms to prevent unwanted Nlrp3 activation (6). These inputs, commonly known as priming steps (or signal 1), are required to facilitate subsequent Nlrp3 inflammasome activation (or signal 2). On the one hand, given its limited expression under steady state conditions in most cell types, a pro-inflammatory stimulus is required to up-regulate Nlrp3 expression (8). Of note, this requirement for “transcriptional priming” can be overcome by heterologous expression of Nlrp3 (8, 9). On the other hand, pro-inflammatory stimuli can

² The abbreviations used are: PRR, pattern recognition receptor; gRNA, guide RNA; CRISPR, clustered regularly interspaced short palindromic repeats; LDH, lactate dehydrogenase; PI, propidium iodide; MOI, multiplicity of infection.

also convey a rapid, “post-translational priming” signal that appears to be required to de-ubiquitinate and thereby to pre-activate Nlrp3 (10, 11). In most studies, LPS is used as a signal 1 stimulus, as it not only functions to prime Nlrp3, but also robustly induces pro-IL-1 β expression.

The exact mechanism of Nlrp3 inflammasome activation (signal 2) yet remains to be determined. Nevertheless, a common denominator of its activation seems to be the cytosolic efflux of potassium ions, which appears to be necessary and sufficient to trigger Nlrp3 inflammasome activation (12, 13). Despite enormous research efforts, no molecule has been identified that would specifically and non-redundantly function downstream of potassium efflux, as well as upstream of Nlrp3 inflammasome activation. Driven by the assumption that a genetically encoded factor would indeed serve such a purpose, we carried out a genome-wide, CRISPR/Cas9 loss-of-function screen to identify proteins that would specifically regulate Nlrp3 activation.

Experimental Procedures

gRNA Library Design—A library design tool was written in C++ which for every annotated protein coding gene isoform picks four independent 18-mer gRNA target sites in the first half of the coding region. If more than four target sites were found, target sites fulfilling the scoring criteria detailed in Ref. 14 were prioritized. Duplicate target sites resulting from redundant or similar gene annotations were removed. The resulting library contained 73,177 constructs, which are provided in [supplemental Table 1](#).

gRNA Library Synthesis—Oligonucleotide pools were obtained from CustomArray, and gRNA plasmid libraries were assembled into pLenti-gRNA-GFP using ligation-independent cloning as described (pL-U6-gRNA (14)).

Lentiviral Packaging—For each 15-cm dish of HEK 293T cells grown to a confluence of 70%, 12 μ g of pVSV-G, 30 μ g of pCMV Δ -8.91, and 40 μ g of library plasmid preparation were diluted in 2 ml of 1 \times HEPES-buffered saline, pH 7.0, and mixed with 100 μ l of 2.5 mM CaCl₂, vortexed briefly, and incubated for 20 min at room temperature before dropwise addition to the dish. 8 h later, the medium was replaced with DMEM containing 30% FCS. After 48 h, the supernatant was harvested, centrifuged briefly, and filtered through a 0.45- μ m filter (GE Healthcare). Virus supernatants were stored at -80°C in small aliquots.

Cell Culture—Cells were cultured in DMEM supplemented with 10% FCS, 1 mM sodium pyruvate, and 10 μ g/ml Ciprofloxacin. Cells were grown at 37 $^{\circ}\text{C}$ and 5% CO₂. Wild-type mouse macrophages are immortalized macrophages as described in Ref. 8.

Screening Cell Line—Immortalized murine macrophages stably expressing mNlrp3-FLAG and hASC-CFP (9) were transduced with a lentivirus stably encoding for mCherry-T2A-spCas9 under the CMV promoter. After limiting dilution cloning, a clone with strong and homogenous red fluorescence was selected and expanded (from now on referred to as Nlrp3-Cas9 macrophages throughout this study).

Genome-wide CRISPR/Cas9 Screen—The screen was performed in 12 replicates, and for each replicate, 3 \times 10⁶ Nlrp3-

Cas9 macrophages were plated in six wells and infected with the library pool at an MOI of 0.05. Starting the next day, cells were trypsinized and expanded every second day. After 1 week, half of the cells were stimulated by adding nigericin to a final concentration of 10 μ M, whereas the other half was left unstimulated. Cells were incubated in 15-ml tubes with loose lids at 37 $^{\circ}\text{C}$ 5% CO₂ for 6 h. Propidium iodide was added to a final concentration of 5 μ g/ml, and cells were incubated on ice for 5 min. Cells were sorted on a BD INFLUX device at 4 $^{\circ}\text{C}$ at a rate of 10,000 cells/s. 2–3 \times 10⁷ cells were sorted per stimulated sample, or 5 \times 10⁶ cells were sorted per unstimulated sample. Cells were sorted directly into small polypropylene plastic wells pre-filled with 50 μ l of a proteinase K-containing lysis buffer (0.2 mg/ml proteinase K, 1 mM CaCl₂, 3 mM MgCl₂, 1 mM EDTA, 1% Triton X-100, 10 mM Tris, pH 7.5). After sorting, lysates were immediately heated to 65 $^{\circ}\text{C}$ for 10 min and to 95 $^{\circ}\text{C}$ for 15 min and kept at 4 $^{\circ}\text{C}$ until analysis.

Deep Sequencing-based Digital Cell Counting—Each lysate was split into 16 (stimulated samples) or 48 (unstimulated samples) PCR reactions using Phusion HF polymerase (Thermo Fisher) and the primers screen_fwd_2 (stimulated samples, 5'-ACACTCTTTCCTACACGACGCTCTTCCGATCTCCACTTTTTCAAGTTGATAACGGAC-3'), screen_fwd_1 (unstimulated samples, 5'-ACACTCTTTCCTACACGACGCTCTTCCGATCTNNNNNNNNCTCGGTGCCACTTTTTCAAGTTG-3'), and screen_rev_1 (5'-TGACTGGAGTTCA-GACGTGTGCTCTTCCGATCTTACGATACAAGGCTGT-TAGAGAG-3'). PCR amplification was performed according to the manufacturer's instructions using an annealing temperature of 65 $^{\circ}\text{C}$, an elongation time of 30 s, and 19 cycles. From each PCR reaction, 1 μ l was transferred to a second PCR reaction using the same cycling conditions, but individual combinations of barcode primers described in Ref. 15. PCR products were pooled, gel-purified, precipitated as described (15) and sequenced using the MiSeq deep sequencing platform. Raw data were evaluated using custom-written software that counts the number of primer barcode combinations with which each individual library gRNA sequence was sequenced, which corresponds to the absolute number of initially sorted cells.

Screening Hit Re-validation—gRNA sequences enriched in nigericin-surviving cells were ordered as single oligonucleotides ([supplemental Table 2](#)) and cloned into pLenti-gRNA-GFP as described (14). After lentiviral packaging and infection of Nlrp3-Cas9 macrophages, cells were stimulated in the same way as during the screening and were analyzed on a BD LSR-II flow cytometer. Relative numbers of cells determined to be GFP⁺ PI⁻ after nigericin stimulation were normalized to the transduction rate in the corresponding unstimulated control sample using the same virus.

KO Clone Generation—Nlrp3-Cas9 macrophages transduced with a gRNA were cloned using limiting dilution. Subsequently, growing cell clones were picked and duplicated. One of the duplicates was lysed for deep sequencing-based genotyping (15, 16). Two all-allelic *Nek7* knock-out cell clones as well as two clones generated with a control gRNA transduction were expanded for functional analysis.

Functional Testing of KO Clones—Clones were plated at a density of 50,000 cells per flat-bottom 96-well. On the next day,

cells were primed for 3 h using 200 ng/ml LPS (InvivoGen). Cells were stimulated using 6.5 μ M nigericin (InvivoGen) for 2–6 h or 200 ng/well poly(dA:dT) (InvivoGen) or plasmid DNA transfected with Lipofectamine 2000 (Thermo Fisher) for 6 h. Unprimed cells were stimulated with 200 ng/well IVT4 5'-triphosphate-dsRNA transfected with Lipofectamine 2000, 1 μ g/ml LPS, or 2 μ g/ml Pam3CSK4 (InvivoGen) for 16 h. Cytotoxicity was measured by LDH release assay according to the manufacturer's instructions (Pierce), and cytokine release was assessed by ELISA (BD Biosciences).

Immunoblotting—Cells were lysed in 1 \times Laemmli buffer and heated to 95 °C for 10 min, and supernatants were precipitated as described (17). Blots were incubated in primary antibodies (Nlrp3, Cryo-2, Adipogen International; Nek7, EPR4900, Abcam; α -tubulin, DM1A, New England Biolabs; acetylated α -tubulin, 6-11B-1, Santa Cruz Biotechnology; IL-1 β , AF-401-NA, R&D Systems; caspase-1, AG-20B-0042-C100, Adipogen International) overnight in PBS with 0.5% milk. Secondary IgG-HRP conjugates (Santa Cruz Biotechnology) were bound at room temperature for 2 h.

Results and Discussion

To devise a loss-of-function screening strategy aimed at identifying genetic factors functioning upstream of Nlrp3, we made use of a murine macrophage cell line that stably expresses murine Nlrp3. This cell line does not require a priming signal (signal 1) to activate Nlrp3 in response to a potassium efflux-inducing stimulus (signal 2). This poses an important advantage as it focuses a perturbation screen on genuine Nlrp3 activation, rendering the cellular model under study insensitive to the loss of components that are involved in Nlrp3 priming (supplemental Fig. 1A). Employing the ionophoric compound nigericin as a stimulus, pyroptosis as measured by LDH release was readily induced in these cells without additional priming (data not shown). To render these cells amenable to CRISPR-mediated screening, Cas9 was stably transduced into the cells (from now on referred to as Nlrp3-Cas9 macrophages).

Studying different readouts of Nlrp3 activation at the single cell level, we observed that measuring loss of cytosolic GFP expression with gain in propidium iodide (PI) positivity served as the best signal for Nlrp3-induced cell death. To identify cells resistant to Nlrp3-induced cell death, we defined a stringent gate that contained GFP-positive, yet PI-negative cells (supplemental Fig. 1B). Using these settings, we transduced Nlrp3-Cas9 cells with a gRNA targeting Nlrp3 or a negative control gRNA at an MOI of 0.01. Subsequently, we stimulated cells with nigericin or left them untreated (Fig. 1A). Analyzing these cells via flow cytometry revealed that 27.9% of all cells that had been successfully transduced with a gRNA targeting *Nlrp3* still remained in the stringent live gate, whereas only 0.53% of cells transduced with a control gRNA were found in this gate (Fig. 1A). Altogether, these results indicated that targeting *Nlrp3* in these cells provided a 52.2-fold survival benefit (27.9%:0.53%) when analyzing cells within this live gate.

Based on this technical setup allowing us to efficiently discriminate between Nlrp3-proficient and Nlrp3-defective cells at a single cell level, we conducted a polyclonal genome-wide CRISPR screen by transducing Nlrp3-Cas9 macrophages with a

pool of lentiviral particles, each encoding for both GFP and a gene-targeting gRNA (Fig. 1B). After 1 week of expansion, we stimulated half of the cells with nigericin and subsequently enriched the GFP⁺ PI⁻ population of cells using FACS sorting. As the absolute cell numbers that can be sorted in a reasonable time frame are technically limited, we used a digital PCR barcoding strategy to allow counting single cells that bear an individual gRNA library construct instead of bulk PCR amplification and end point quantification of gRNA sequences as employed in other CRISPR screens (18, 19). From the unstimulated control cells, gRNA sequences targeting 15,600 annotated genes could be retrieved, covering 79.9% of the genes targeted by the library with an average of 1.71 gRNAs per gene. When plotting the numbers of sorted cells bearing each library gRNA with or without nigericin stimulation, five gRNA sequences were readily identified as specifically enriched in the population of cells being resistant to nigericin-induced cell death (Fig. 1C and supplemental Table 3). Of these, two gRNAs targeted *Nlrp3* itself, demonstrating the specificity of the screening approach, whereas the three other gRNAs targeted genes that had not been associated with Nlrp3 signaling before (*Nek7*, *Fam83c*, and *Cdkn2a*) (Fig. 1D). We next tested these gRNAs individually for their survival benefit in Nlrp3-Cas9 macrophages using the same FACS readout as during the screening. Here, we found that both the gRNAs targeting *Nek7* and *Fam83c* induced a similarly strong survival benefit as the *Nlrp3* control gRNA, whereas a non-targeting control gRNA and a gRNA targeting the unrelated gene *Emc3* failed to do so (Fig. 1E). At the same time, the survival benefit of the *Cdkn2a*-targeting gRNA was only partial, whereas this gRNA induced rapid proliferation in positively transduced cells (data not shown). Therefore, we only validated the target genes *Nek7* and *Fam83c* for their functional involvement in Nlrp3 activation using two alternative gRNA sequences that were not found in the screen (Fig. 1F). Both gRNAs targeting *Nek7* could reproduce the survival benefit of the *Nek7* gRNA found in the initial screen. However, both alternative gRNAs targeting *Fam83c* failed to rescue nigericin-induced cell death, despite their target sites being located close to the original hit gRNA within the genetic locus. Altogether, this indicated that the *Fam83c* gRNA enriched in the screen might be active due to an off-target effect within an unknown second gene, whereas *Nek7* was a *bona fide* hit.

To validate the role of *Nek7* in inflammasome signaling and to explore its epistatic role in Nlrp3 activation, we generated single cell clones of Nlrp3-Cas9 macrophages that had been transduced with a gRNA targeting *Nek7* (15). Following this targeting approach, we obtained several cell clones bearing allelic frameshift mutations within the *Nek7* target region. We picked two *Nek7* knock-out clones and two clones with a non-targeting control gRNA for further analysis (supplemental Fig. 2). First, we wanted to explore whether *Nek7* deficiency had a general impact on pro-inflammatory gene expression. To study this, we stimulated *Nek7*-deficient and wild-type cells with Pam3CSK4, LPS, and 5'-triphosphate dsRNA to trigger TLR2, TLR4, and RIG-I, respectively. Doing so revealed that *Nek7*-competent and -deficient cells responded equally well to these stimuli with regard to pro-inflammatory gene expression (Fig. 2A). On the other hand, *Nek7*-competent macrophages dis-

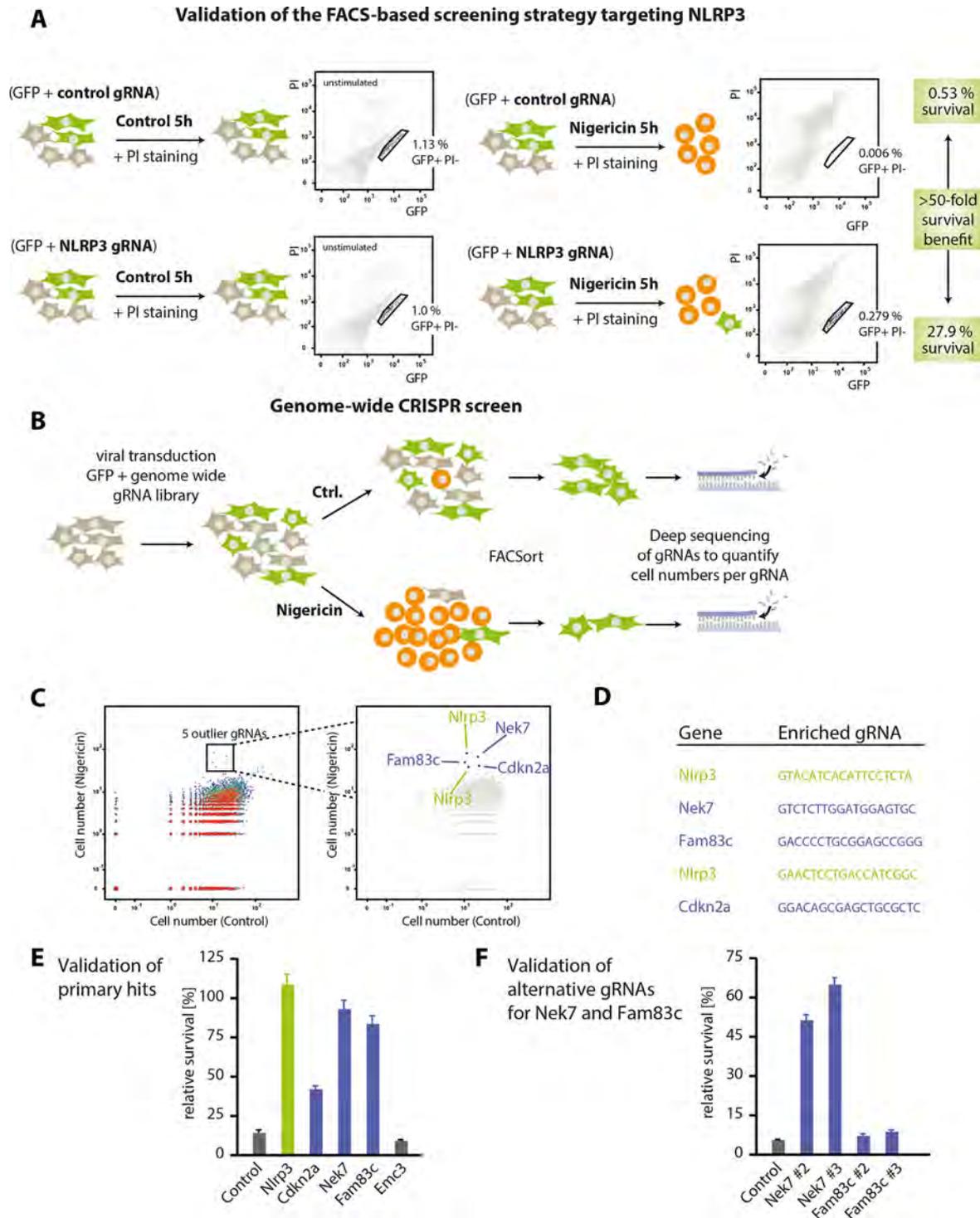
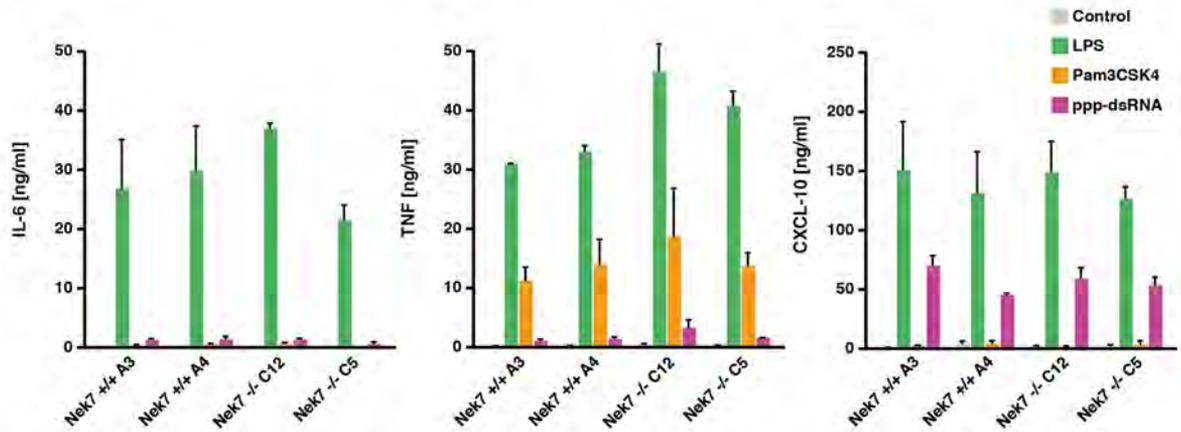


FIGURE 1. A CRISPR loss-of-function screen identifies *Nek7* as a component involved in *Nlrp3* signaling. *A*, *Nlrp3*-Cas9 macrophages were transduced with a control gRNA or a gRNA targeting *Nlrp3*, which additionally encoded for GFP at an MOI of 0.01. Following stable transduction, macrophages were either stimulated with nigericin for 5 h or left untreated. Subsequently, cells were labeled with PI and subjected to FACS analysis. Depicted are schematic views of the transduction and stimulation modalities as well as FACS plots of a representative result. Highlighted are the frequencies of cells in the stringent live gate that contains GFP-positive and PI-negative cells. Data are representative of three independent experiments with comparable results. *B*, schematic view of the genome-wide screening approach that was undertaken to identify factors that confer resistance to nigericin-induced cell death. Please see “Experimental Procedures” for details. *Ctrl.*, control. *C*, dot plot representation of the screening results. Each dot represents the frequency of cells carrying a specific gRNA that were found in the live gate of mock-treated cells (*x* axis) or of nigericin-treated cells (*y* axis). Highlighted are five outlier gRNAs targeting *Nlrp3* (*x*2), *Nek7*, *Fam83c*, or *Cdkn2a*. *D*, depicted are the gRNA sequences found in the outlier population. *E*, *Nlrp3*-Cas9 macrophages were transduced with the same gRNAs identified in the screen and a control gRNA targeting *Emc3*. Subsequently, cells were subjected to nigericin stimulation, and cell survival was analyzed by FACS. The relative survival benefit was calculated by dividing the gated cell fraction from the stimulated condition by that from the unstimulated condition and is depicted as a mean value \pm S.E. of three independent experiments. *F*, two alternative gRNAs targeting *Nek7* and *Fam83c* were used to transduce *Nlrp3*-Cas9 macrophages. Nigericin stimulation and subsequent cell survival were analyzed as in *E*.

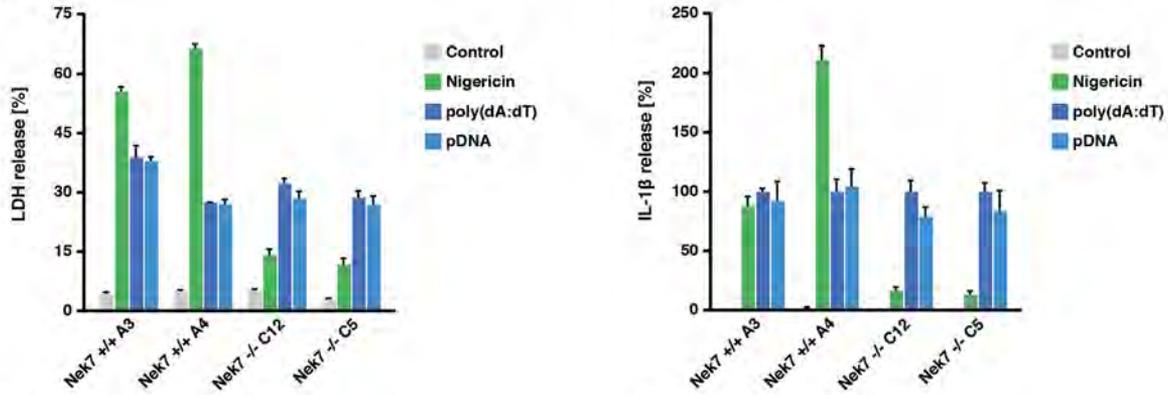
A



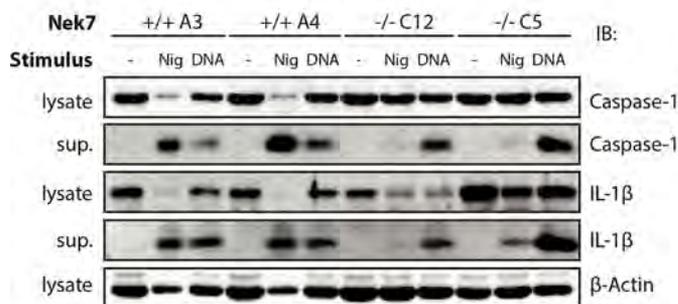
B



C



D



E

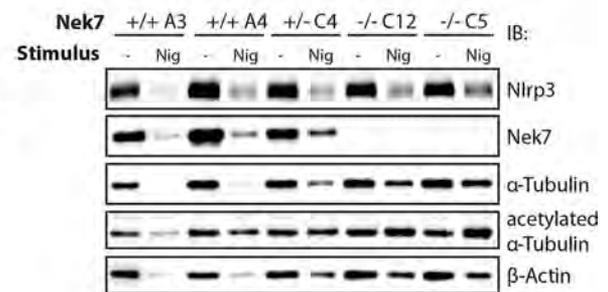


FIGURE 2. Nek7 functions specifically upstream of the Nlrp3 inflammasome. *A*, Nek7-deficient Nlrp3-Cas9 macrophages or wild-type cells were stimulated with Pam3CSK4, LPS, or 5'-triphosphate dsRNA. 16 h later, supernatants were collected and analyzed for TNF, IL-6, and CXCL10 production. Data are presented as mean values \pm S.E. of two independent experiments. *B*, wild-type macrophages or Nlrp3-Cas9 macrophages of the indicated genotype were stimulated with 200 ng/ml LPS for the durations indicated. Nek7 protein expression was analyzed by immunoblot (IB), whereas β -actin served as a loading control. Representative data from one experiment out of two independent experiments are depicted. *C*, macrophages of the indicated genotypes were primed with LPS and subsequently stimulated with nigericin, poly(dA:dT), or plasmid DNA (pDNA). 6 h after stimulation, supernatants were analyzed for LDH release (left panel) or IL-1 β production (right panel). IL-1 β data were normalized to the poly(dA:dT) condition. *D*, lysates or supernatants (sup.) of cells stimulated with nigericin (Nig) for 2 h or poly(dA:dT) for 6 h were analyzed for IL-1 β , caspase-1, or β -actin by immunoblot. Representative data from one experiment out of two independent experiments are depicted. *E*, cells of the indicated genotypes were stimulated with nigericin or left untreated (note that one heterozygote Nek7 cell clone was included in the analysis). 6 h after stimulation, cells were analyzed for Nlrp3, Nek7, α -tubulin, acetylated α -tubulin, and β -actin expression by immunoblot. Representative data from one experiment out of two independent experiments are depicted.

played no change in Nek7 protein expression upon TLR4 ligation (Fig. 2B), which is in line with mRNA expression data of primary mouse macrophages stimulated with LPS (20). Studying nigericin- and DNA-mediated inflammasome activation in these cells revealed a different picture. Although Nek7-competent cells readily displayed cell death as well as IL-1 β release upon nigericin stimulation, *Nek7* knock-out clones showed a greatly blunted cell death as well as IL-1 β response following Nlrp3 activation. Of note, Aim2-dependent inflammasome activation was operational in both wild-type cells as well as Nek7-deficient cells. (Fig. 2C). Assessing caspase-1 cleavage as well as IL-1 β maturation by immunoblot confirmed the specific role of Nek7 in Nlrp3 inflammasome activation (Fig. 2D). Both caspase-1 cleavage as well as IL-1 β cleavage were readily induced by nigericin treatment in Nek7-competent cells, but severely blunted in the absence of Nek7. Again, DNA-mediated inflammasome activation was equally potent in cells of either genotype. Given the fact that Nlrp3-dependent pyroptosome formation was also blunted in the absence of Nek7 (data not shown), these results indicated that Nek7 functioned specifically upstream of Nlrp3 in inflammasome activation.

In keeping with the fact that the cells under study were stably expressing Nlrp3, we observed considerable Nlrp3 expression in unprimed macrophages (Fig. 2E). Moreover, there was no alteration in Nlrp3 expression in relation to Nek7 gene targeting.

It has been described that microtubule acetylation triggered by mitochondrial damage is a prerequisite for Nlrp3 inflammasome activation, allowing Nlrp3 and Asc to get into close proximity via active transport in a dynein-dependent mechanism (21). Although Nek7 has been primarily characterized as a factor regulating microtubule network nucleation and spindle formation during mitosis, additional functions within the microtubule network during interphase have also been described. To this effect, it has been shown that Nek7 accelerates microtubule dynamic instability during interphase (22) and that it is able to phosphorylate α - and β -tubulin *in vitro* (23). Moreover, it has been demonstrated that the related family member Nek3 is able to control microtubule acetylation (24). Consequently, in light of the aforementioned study linking microtubule acetylation and Nlrp3 inflammasome signaling, we wondered whether Nek7 was involved in microtubule acetylation during Nlrp3 inflammasome activation. To test this hypothesis, we blotted α -tubulin and acetylated α -tubulin in wild-type and Nek7-deficient macrophages with or without nigericin stimulation (Fig. 2E). After doing so, we did not detect reduced α -tubulin acetylation in Nek7 knock-out cells; on the contrary, acetylated α -tubulin levels were even increased following nigericin stimulation in the absence of Nek7. Although the latter phenomenon could be due to reduced cell death in response to nigericin, these data clearly indicated that Nek7 was not acting upstream of α -tubulin acetylation in this setting.

Although the mechanism of Nek7 functioning upstream of Nlrp3 currently remains elusive, we consider the unbiased discovery of Nek7 a significant advance in our understanding of Nlrp3 inflammasome biology, as it provides the first “genetic handle” upstream of Nlrp3, yet downstream of potassium efflux. Our screening approach used a stringent setup in which

Nlrp3 is steadily expressed to focus only on genetic components relaying signal 2. In light of the fact that Nek7 deficiency did not impact on pro-inflammatory gene expression and that Nek7 expression itself was not regulated as such, we consider it unlikely that Nek7 is involved in providing signal 1. Nevertheless, to fully rule out this possibility, it should be informative to study the role of Nek7 in unmodified macrophages, in which both signal 1 and signal 2 are required. At the same time, it would be interesting to explore the role of Nek7 *in vivo*, which would, however, require a strategy that circumvents the lethality of Nek7 deficiency in mice (25).

Given its evolutionary conservation (26), its broad expression (20), and its pivotal role in mitosis (27), we consider it unlikely that Nek7 functions as a dedicated and sufficient activator of the Nlrp3 inflammasome. Instead, we favor the hypothesis that Nek7 is involved in the co-regulation of a pathway that impacts on the ability of Nlrp3 to respond to an upstream signal or to relay this signal toward its adapter protein Asc. To this end, Nek7 could be involved in the formation or provision of a common signal that functions upstream of Nlrp3. Although this could be an indirect mechanism, *e.g.* a phosphorylation event mediated by Nek7, this could also directly involve Nek7 as a protein itself. On the other hand, through its function as a regulator of microtubule dynamics, Nek7 could also be required to facilitate the interaction of Nlrp3 and Asc (21). Without doubt, additional studies are required to obtain insight into the molecular mechanisms of Nek7 facilitating Nlrp3 activation. In this context, it should also be interesting to address the role of its closely related homologue Nek6, as well as its common upstream activator Nek9. Nevertheless, apart from its currently unclear mode of action, the fact that Nek7 is a kinase already makes it a potentially interesting drug target for the treatment of sterile inflammatory conditions known to involve Nlrp3.

Author Contributions—J. L. S. B. and V. H. conceived the project idea and methodology. J. L. S. B. conducted most of the experiments and also analyzed the results. T. S. E., T. S., D. C., and J. R. helped with cell culture experiments, immunoblotting studies and ELISA measurements. E. E. supervised the FACS sorting experiments. J. L. S. B. and V. H. wrote the paper. V. H. supervised the study.

Acknowledgments—We thank Drs. Bernardo Franklin and Eicke Latz (Institute of Innate Immunity, University Hospital, University of Bonn) for providing us with immortalized macrophages expressing Nlrp3.

References

1. Medzhitov, R. (2007) Recognition of microorganisms and activation of the immune response. *Nature* **449**, 819–826
2. Chen, G. Y., and Nuñez, G. (2010) Sterile inflammation: sensing and reacting to damage. *Nat. Rev. Immunol.* **10**, 826–837
3. Gross, O., Thomas, C. J., Guarda, G., and Tschopp, J. (2011) The inflammasome: an integrated view. *Immunol. Rev.* **243**, 136–151
4. Kayagaki, N., Stowe, I. B., Lee, B. L., O'Rourke, K., Anderson, K., Warming, S., Cuellar, T., Haley, B., Roose-Girma, M., Phung, Q. T., Liu, P. S., Lill, J. R., Li, H., Wu, J., Kummerfeld, S., Zhang, J., Lee, W. P., Snipas, S. J., Salvesen, G. S., Morris, L. X., Fitzgerald, L., Zhang, Y., Bertram, E. M., Goodnow, C. C., and Dixit, V. M. (2015) Caspase-11 cleaves gasdermin D for non-canonical inflammasome signaling. *Nature* **526**, 666–671

5. Shi, J., Zhao, Y., Wang, K., Shi, X., Wang, Y., Huang, H., Zhuang, Y., Cai, T., Wang, F., and Shao, F. (2015) Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. *Nature* **526**, 660–665
6. Latz, E., Xiao, T. S., and Stutz, A. (2013) Activation and regulation of the inflammasomes. *Nat. Rev. Immunol.* **13**, 397–411
7. Masters, S. L., Latz, E., and O'Neill, L. A. (2011) The inflammasome in atherosclerosis and type 2 diabetes. *Sci. Transl. Med.* **3**, 81ps17
8. Bauernfeind, F. G., Horvath, G., Stutz, A., Alnemri, E. S., MacDonald, K., Speert, D., Fernandes-Alnemri, T., Wu, J., Monks, B. G., Fitzgerald, K. A., Hornung, V., and Latz, E. (2009) Cutting edge: NF- κ B activating pattern recognition and cytokine receptors license NLRP3 inflammasome activation by regulating NLRP3 expression. *J. Immunol.* **183**, 787–791
9. Franklin, B. S., Bossaller, L., De Nardo, D., Ratter, J. M., Stutz, A., Engels, G., Brenker, C., Nordhoff, M., Mirandola, S. R., Al-Amoudi, A., Mangan, M. S., Zimmer, S., Monks, B. G., Fricke, M., Schmidt, R. E., Espevik, T., Jones, B., Jarnicki, A. G., Hansbro, P. M., Busto, P., Marshak-Rothstein, A., Hornemann, S., Aguzzi, A., Kastenmüller, W., and Latz, E. (2014) The adaptor ASC has extracellular and 'prionoid' activities that propagate inflammation. *Nat. Immunol.* **15**, 727–737
10. Juliana, C., Fernandes-Alnemri, T., Kang, S., Farias, A., Qin, F., and Alnemri, E. S. (2012) Non-transcriptional priming and deubiquitination regulate NLRP3 inflammasome activation. *J. Biol. Chem.* **287**, 36617–36622
11. Schroder, K., Sagulenko, V., Zamoshnikova, A., Richards, A. A., Cridland, J. A., Irvine, K. M., Stacey, K. J., and Sweet, M. J. (2012) Acute lipopolysaccharide priming boosts inflammasome activation independently of inflammasome sensor induction. *Immunobiology* **217**, 1325–1329
12. Pétrilli, V., Papin, S., Dostert, C., Mayor, A., Martinon, F., and Tschopp, J. (2007) Activation of the NALP3 inflammasome is triggered by low intracellular potassium concentration. *Cell Death Differ.* **14**, 1583–1589
13. Muñoz-Planillo, R., Kuffa, P., Martínez-Colón, G., Smith, B. L., Rajendiran, T. M., and Núñez, G. (2013) K⁺ efflux is the common trigger of NLRP3 inflammasome activation by bacterial toxins and particulate matter. *Immunity* **38**, 1142–1153
14. Schmidt, T., Schmid-Burgk, J. L., and Hornung, V. (2015) Synthesis of an arrayed sgRNA library targeting the human genome. *Sci. Rep.* **5**, 14987
15. Schmid-Burgk, J. L., Schmidt, T., Gaidt, M. M., Pelka, K., Latz, E., Ebert, T. S., and Hornung, V. (2014) OutKnocker: a web tool for rapid and simple genotyping of designer nuclease edited cell lines. *Genome Res.* **24**, 1719–1723
16. Schmidt, T., Schmid-Burgk, J. L., Ebert, T. S., Gaidt, M. M., and Hornung, V. (2016) Designer nuclease-mediated generation of knockout THP1 cells. *Methods Mol. Biol.* **1338**, 261–272
17. Jakobs, C., Bartok, E., Kubarenko, A., Bauernfeind, F., and Hornung, V. (2013) Immunoblotting for active caspase-1. *Methods Mol. Biol.* **1040**, 103–115
18. Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G., and Zhang, F. (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87
19. Wang, T., Wei, J. J., Sabatini, D. M., and Lander, E. S. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84
20. Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C. L., Haase, J., Janes, J., Huss, J. W., 3rd, and Su, A. I. (2009) BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* **10**, R130
21. Misawa, T., Takahama, M., Kozaki, T., Lee, H., Zou, J., Saitoh, T., and Akira, S. (2013) Microtubule-driven spatial arrangement of mitochondria promotes activation of the NLRP3 inflammasome. *Nat. Immunol.* **14**, 454–460
22. Cohen, S., Aizer, A., Shav-Tal, Y., Yanai, A., and Motro, B. (2013) Nek7 kinase accelerates microtubule dynamic instability. *Biochim. Biophys. Acta* **1833**, 1104–1113
23. O'Regan, L., and Fry, A. M. (2009) The Nek6 and Nek7 protein kinases are required for robust mitotic spindle formation and cytokinesis. *Mol. Cell Biol.* **29**, 3975–3990
24. Chang, J., Baloh, R. H., and Milbrandt, J. (2009) The NIMA-family kinase Nek3 regulates microtubule acetylation in neurons. *J. Cell Sci.* **122**, 2274–2282
25. Salem, H., Rachmin, I., Yissachar, N., Cohen, S., Amiel, A., Haffner, R., Lavi, L., and Motro, B. (2010) Nek7 kinase targeting leads to early mortality, cytokinesis disturbance and polyploidy. *Oncogene* **29**, 4046–4057
26. Quarumby, L. M., and Mahjoub, M. R. (2005) Caught Nek-ing: cilia and centrioles. *J. Cell Sci.* **118**, 5161–5169
27. Fry, A. M., O'Regan, L., Sabir, S. R., and Bayliss, R. (2012) Cell cycle regulation by the NEK family of protein kinases. *J. Cell Sci.* **125**, 4423–4433

Differential Roles of Cell Death-inducing DNA Fragmentation Factor- α -like Effector (CIDE) Proteins in Promoting Lipid Droplet Fusion and Growth in Subpopulations of Hepatocytes*[♦]

Received for publication, November 6, 2015, and in revised form, December 27, 2015. Published, JBC Papers in Press, January 5, 2016, DOI 10.1074/jbc.M115.701094

Wenyi Xu⁺¹, Lizhen Wu⁺¹, Miao Yu[‡], Feng-Jung Chen[‡], Muhammad Arshad[§], Xiayu Xia[¶], Hao Ren[‡], Jinhai Yu[‡], Li Xu^{||}, Dijin Xu[‡], John Zhong Li^{**}, Peng Li^{‡2}, and Linkang Zhou^{‡3}

From the [‡]MOE Key Laboratory of Bioinformatics and Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China, the [§]Department of Bioinformatics and Biotechnology, International Islamic University, Islamabad 44000, Pakistan, the [¶]Institute of Laboratory Animal Sciences, Chinese Academy of Medical Sciences, Beijing 100021, China, the ^{||}Key Laboratory for Feed Biotechnology of the Ministry of Agriculture, Feed Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China, and the ^{**}Jiangsu Province Key Laboratory of Human Functional Genomics, Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Nanjing Medical University, Nanjing 210029, China

Lipid droplets (LDs) are dynamic subcellular organelles whose growth is closely linked to obesity and hepatic steatosis. Cell death-inducing DNA fragmentation factor- α -like effector (CIDE) proteins, including Cidea, Cideb, and Cidec (also called Fsp27), play important roles in lipid metabolism. Cidea and Cidec are LD-associated proteins that promote atypical LD fusion in adipocytes. Here, we find that CIDE proteins are all localized to LD-LD contact sites (LDCSs) and promote lipid transfer, LD fusion, and growth in hepatocytes. We have identified two types of hepatocytes, one with small LDs (small LD-containing hepatocytes, SLHs) and one with large LDs (large LD-containing hepatocytes, LLHs) in the liver. Cideb is localized to LDCSs and promotes lipid exchange and LD fusion in both SLHs and LLHs, whereas Cidea and Cidec are specifically localized to the LDCSs and promote lipid exchange and LD fusion in LLHs. *Cideb*-deficient SLHs have reduced LD sizes and lower lipid exchange activities. Fasting dramatically induces the expression of Cidea/Cidec and increases the percentage of LLHs in the liver. The majority of the hepatocytes from the liver of obese mice are Cidea/Cidec-positive LLHs. Knocking down Cidea or Cidec significantly reduced lipid storage in the livers of obese animals. Our data reveal that CIDE proteins play differential roles in promoting LD fusion and lipid storage; Cideb promotes lipid storage under normal diet conditions, whereas Cidea and Cidec are responsible for liver steatosis under fasting and obese conditions.

Obesity and its associated diseases, including type II diabetes, cardiovascular disease, and hepatic steatosis, have become alarmingly common diseases (1). Liver plays a major regulatory role in whole-body lipid metabolism. Disruption of the hepatic lipid metabolism could lead to the initiation and progression of several metabolic disorders (2, 3). The accumulation of fat in the form of lipid droplets (LDs)⁴ is an early pathophysiological feature of altered liver metabolism that is linked to insulin resistance and the potential progression of severe liver diseases, such as liver steatosis, liver cirrhosis, and hepatocellular carcinoma (4). LDs are subcellular organelles composed of a neutral lipid core surrounded by a phospholipid monolayer that is coated with various types of proteins (5). LDs are involved in several biologically significant processes, including neutral lipid storage, protein storage and degradation, as well as viral packaging (6, 7). LDs can grow in size via a tightly regulated mechanism (8, 9). The sizes of LDs reflect different biological processes. Several models projecting the growth of LDs have been proposed, including targeted lipid delivery from the endoplasmic reticulum to LDs mediated by fat storage-inducing transmembrane proteins 1 and 2 (FITM1/2) (10), local lipid synthesis on LDs mediated by CTP:phosphocholine cytidylyltransferase and diacylglycerol acyltransferase 2 (DGAT2) (11, 12), and the fusion of smaller LDs into larger LDs mediated by CIDEs in adipocytes (13).

Many proteins are localized on the surface of LDs. These proteins play important roles in regulating the size and function of LDs (14). Perilipin1/2/3/4/5 (Plin1/2/3/4/5) are LD-associated proteins in mammalian cells (15). The expression levels of Plin2, Plin3, and Plin5 are up-regulated in fatty livers (16). CIDE proteins, including Cidea, Cideb, and Cidec (also called Fsp27), are novel LD-associated proteins (17, 18). CIDEs play important roles in LD morphology and function. Cidea and Cidec are

* This work was supported by National Basic Research Program Grants 2013CB530602 (to P. L.), National Natural Science Foundation of China Grants 31430040 and 31321003 (to P. L.) and Grant 31501089 (to J. Y.), and Grants 31271268 and 81471079 (to J. Z. L.), and China Postdoctoral Science Foundation Grants 2012M520249 and 2013T60103 (to L. Z.). The authors declare that they have no conflicts of interest regarding the contents of this article.

[♦] This article was selected as a Paper of the Week.

¹ Both authors contributed equally to this work.

² To whom correspondence may be addressed. Tel.: 86-10-62797121; E-mail: li-peng@mail.tsinghua.edu.cn.

³ To whom correspondence may be addressed. Tel.: 86-10-62797133; E-mail: zhoulinkang@mail.tsinghua.edu.cn.

⁴ The abbreviations used are: LD, lipid droplet; CIDE, cell death-inducing DNA fragmentation factor- α -like effector; LDCS, LD-LD contact site; LLHs, large lipid droplet-containing hepatocytes; and SLHs, small lipid droplet-containing hepatocytes; TAG, triacylglycerol.

predominantly expressed in adipocytes, although Cideb is specifically expressed in the liver of wild-type mice. Cidea and Cidec are up-regulated in the steatotic liver (19, 20). Fasting can induce the expression of Cidec in the liver of wild-type mice (21–24). Both Cidea and Cidec localize on the surface of LDs and are particularly enriched at LD-LD contact sites (LDCSs) to promote atypical LD fusion and growth by lipid exchange and transfer in adipocytes (25–28). Several factors are involved in Cidea/Cidec-mediated LD fusion, including Plin1, Rab8a, MSS4, and AS160 (29–31). However, the role of Cidea/Cidec in LD fusion in the liver has not been investigated. Cideb is localized on the endoplasmic reticulum and LDs (32), and its deficiency results in the accumulation of smaller LDs in the liver (33). However, whether Cideb has the ability to localize on LDCSs and promote lipid exchange and LD fusion is still unknown.

Here, we systematically analyzed the role of CIDEs in LD fusion and lipid storage in hepatocytes under normal, fasting, and obese conditions. We found that Cideb is localized to LDCSs and promotes LD fusion and growth in hepatocytes. In the livers of normal diet-fed mice, the majority of hepatocytes express Cideb alone and contain small LDs. A small proportion of hepatocytes express Cidea and Cidec and contain large LDs. Under fasting and obese conditions, the percentage of hepatocytes expressing Cidea and Cidec increases dramatically. Finally, using *Cideb*^{-/-}, *ob/ob* hepatocytes, we found that CIDEs play important roles in LD size and lipid storage in hepatocytes.

Materials and Methods

Mice—Wild-type, *Cideb*^{-/-}, *ob/ob* mice were maintained as described previously (33–35). All mice used were on a C57BL/6j background. Four-month-old male mice were used for hepatocyte isolation. *In vivo* delivery of siRNAs was performed using InvivoFectamine 2.0 (Invitrogen, 1377501). One week after the injection of siRNAs, the mice were fasted for 16 h. Liver tissues were harvested for further analysis. Liver TAG was measured as described previously (20). All animals were maintained in the animal facility of the Center of Biomedical Analysis, Tsinghua University (Beijing, China). The laboratory animal facility has been accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC). The Institutional Animal Care and Use Committee (IACUC) of Tsinghua University approved all animal protocols.

Cell Culture and Transfection—293T cells and HepG2 cells were cultured in DMEM (Invitrogen) containing 10% FBS (Invitrogen). Mouse primary hepatocytes were isolated as described previously (35). The isolated hepatocytes were seeded at a density of 10⁶ cells per dish in glass bottom microwell dishes (P35G-1.5-14-C, MatTek Corp.) in DMEM (Invitrogen) containing 10% FBS (Invitrogen). Plasmid DNAs were transfected into 293T cells, HepG2 cells, and primary hepatocytes using Lipofectamine 2000 according to the manufacturer's instruction (Invitrogen, 11668019). For the siRNA experiment, 24 h after seeding, hepatocytes were transfected with siRNA using Lipofectamine 2000 (Invitrogen). The sequences used to target the CIDEs and Plin2 are as follows: Cidea, ACA-

CGCATTTTCATGATCTT; Cideb, CCTCTGCATGGAGTACCTT; Cidec, AATCGTGGAGACAGAAGAATA; Plin2, GAATATGCACAGTGCCAAC. Cells were visualized using an Axiovert 200 M microscope (Carl Zeiss) or an LSM710 confocal microscope (Carl Zeiss) 48 h after transfection.

Fluorescent Microscopic Imaging—Twelve hours after seeding, hepatocytes were fixed with 4% paraformaldehyde for 1 h at room temperature. Cells were then treated with 0.4% Triton X-100 for 20 min and then blocked with 10% goat serum for 1 h at room temperature. The primary antibodies for Cidea, Cideb and Cidec were added and the reaction was incubated for 1 h at room temperature. Anti-rabbit IgG antibodies conjugated with Alexa Fluor 568 (Molecular Probes, A11011) were used as secondary antibodies. Bodipy 493/503 (Molecular Probes, D3922) was used for neutral lipid staining. For the Cideb and Cidea co-staining and the Cideb and Cidec co-staining (Figs. 2E and 6, C and D), a Cideb antibody (from goat) obtained from Santa Cruz Biotechnology (sc-8733, 1:50) was used. Cidea and Cidec antibodies were generated from rabbit (20). Donkey anti-rabbit 488 (Molecular Probes, A21206) and donkey anti-goat 568 (Molecular Probes, A11057) were used as secondary antibodies. LDs were stained with LipidTox (Molecular Probes, H34477). For the staining in Fig. 3C, a Cideb antibody (generated from rabbit) was used. Anti-rabbit IgG antibodies conjugated with Alexa Fluor 405 (Molecular Probes, A31556) were used as secondary antibodies. The sections were observed using a Zeiss 200 M inverted microscope, and the images were collected using an AxioCam MRm camera and Axio Vision software.

Antibodies and Western Blot Analysis—Methods for LD isolation, tissue homogenization, immunoprecipitation, two-step co-immunoprecipitation, and Western blot sample preparation were previously described (30, 35). The proteins were subjected to Western blot analysis with the desired antibodies. The antibodies against Cidea, Cideb, and Cidec were used as described previously (35). Antibodies against β -actin (Sigma, A5441, 1:2000), FLAG (Sigma, F1804, 1:1000), HA (Santa Cruz Biotechnology, sc-7392, 1:1000), and Plin2 (Fitzgerald Industries, 20R-Ap002, 1:8000) were used for Western blot analysis. The blots were detected using HRP-conjugated secondary antibodies (GE Healthcare, UK) and the ECL-Plus system.

Measurement of LD Sizes—Quantitative analysis of LD size in hepatocytes has been previously described (26, 29, 30). The diameter of the largest LD in each hepatocyte was measured. At least 50 hepatocytes were analyzed for each condition.

Calculation of Neutral Lipid Exchange Rate—The calculation of the lipid exchange rate was essentially the same as described previously (30). In brief, hepatocytes transfected with siRNAs were incubated with 200 μ M BSA-bound oleic acid (Sigma) and 1 μ g/ml Bodipy 558/568 C12 fatty acids (Molecular Probes, D3835) for 15 h and then transferred to fresh medium 1 h before the experiment. Live cells were viewed under a confocal microscope (LSM710) using a \times 63 oil immersion objective. LD pairs were selected for bleaching. Selected regions were bleached by 500 interactions at 100% laser power (543 diode laser), followed by time-lapse scanning with 12.5-s intervals for 2 min for Cidea/Cidec-mediated exchange and a total recovery of 6 min for Cideb-mediated exchange.

CIDE Proteins and Liver Lipid Storage

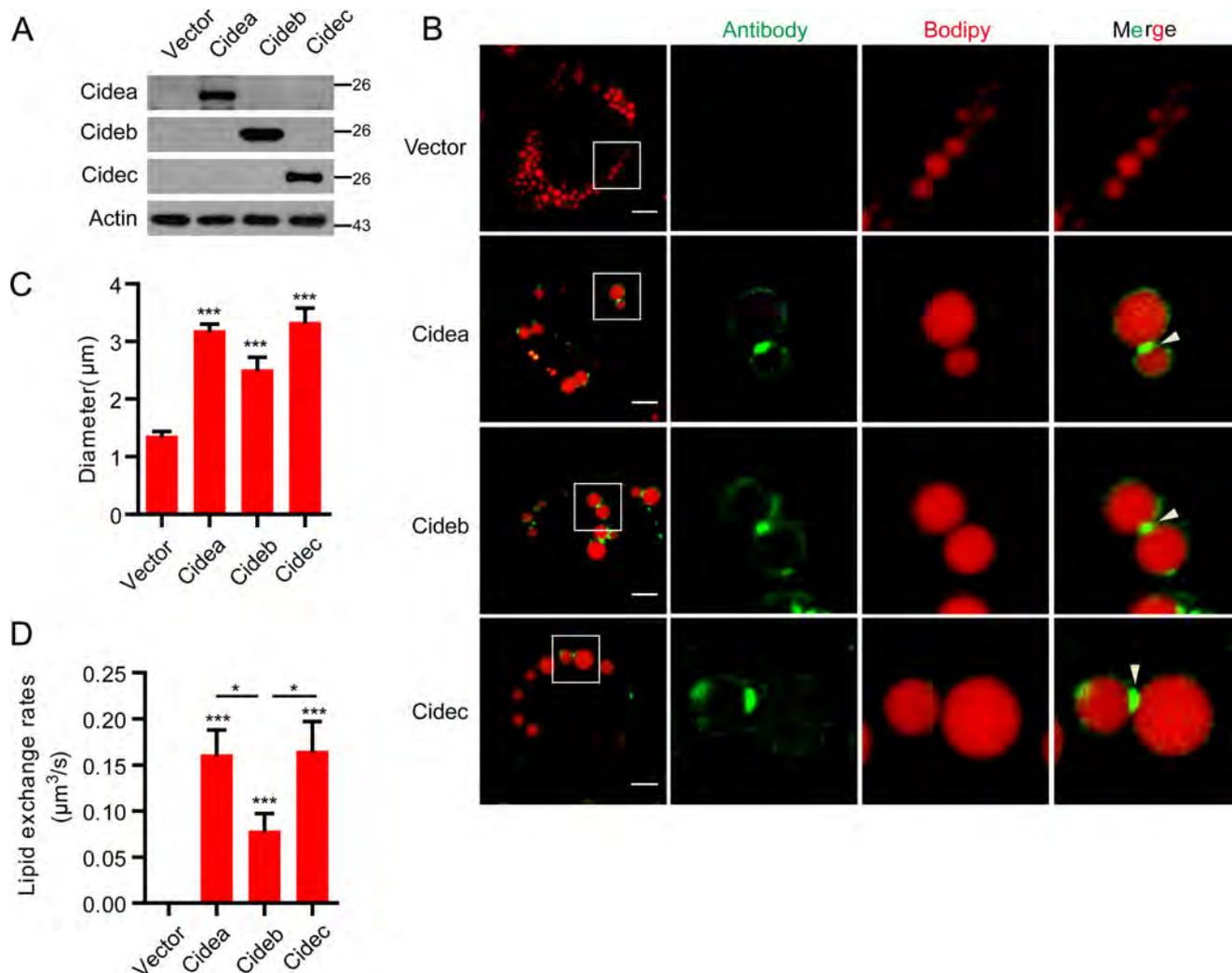


FIGURE 1. CIDE proteins promote LD fusion and large LD formation in HepG2 cells. *A*, Vector, Cidea, Cideb, and Cidec plasmids were transfected into HepG2 cells. Expression levels of CIDEs were detected by Western blot. *B*, protein localization of CIDEs in HepG2 cells as in *A*. Oleic acid was added to promote the formation of LDs for 15 h. LDs were labeled with Bodipy 556/568 (C12, red). Scale bars, 10 μm. Arrowheads point to LDCSs. *C*, largest lipid droplet size per cell was measured in *A*. Ten cells were analyzed in each group. *D*, lipid exchange rates were measured in *A*. Five pairs of LDs were measured. Quantitative data are presented as the mean ± S.E. Differences were considered significant at $p < 0.05$. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

Statistics—The statistical data reported include results from at least three biological replicates. All results are expressed as the mean ± S.E. Quantitation of the Western blot bands was performed using Quantity One software (Bio-Rad) and was expressed as the fold change after correction for relative control levels. All statistical analyses were performed in GraphPad Prism Version 5 (GraphPad Software). Significance was established using a two-tailed Student's *t* test. Differences were considered significant at $p < 0.05$. *p* values are indicated in each figure as follows: *, $p < 0.05$; **, $p < 0.01$, or ***, $p < 0.001$.

Results

CIDE Proteins Promote LD Fusion and Growth When Overexpressed in HepG2 Cells—To investigate the role of CIDE proteins in LD fusion and growth in hepatocytes, we transfected HepG2 cells with no-tagged CIDE cDNAs. The expression levels of these proteins were detected by their corresponding antibodies (Fig. 1*A*). Similar to our observation in adipocytes (26), Cidea and Cidec were enriched at the LDCS in HepG2 cells (Fig. 1*B*). Interestingly, Cideb proteins were also enriched at LDCSs

(Fig. 1*B*). We then measured the LD sizes and observed that cells expressing Cidea and Cidec accumulated larger LDs compared with control cells (Fig. 1*C*). Cideb also showed high activity in promoting large LD accumulation, albeit lower than that of Cidea and Cidec (Fig. 1*C*). Next, we measured lipid exchange activity, a hallmark of Cidea/Cidec-mediated LD fusion, between LD pairs that were positive for CIDE proteins. Lipid exchange activity for LD pairs positive for Cidea or Cidec was high (all are 0.16 μm³/s) and was similar to that in adipocytes (0.13 μm³/s) (26). Lipid exchange was also observed in Cideb-positive LD pairs and appeared to be lower (0.078 μm³/s) than that of Cidea and Cidec (Fig. 1*D*), consistent with the accumulation of smaller LDs in Cideb-expressing HepG2 cells. Overall, these data indicate that all CIDE proteins have the ability to promote LD fusion and growth in hepatocytes.

Identification of Two Types of Hepatocytes with Differential Expression of CIDE Proteins and Lipid Storage Capacity—We then investigated the precise subcellular distribution and function of endogenous CIDEs in isolated wild-type hepatocytes. We found that ~95% of hepatocytes had small LDs, and the size

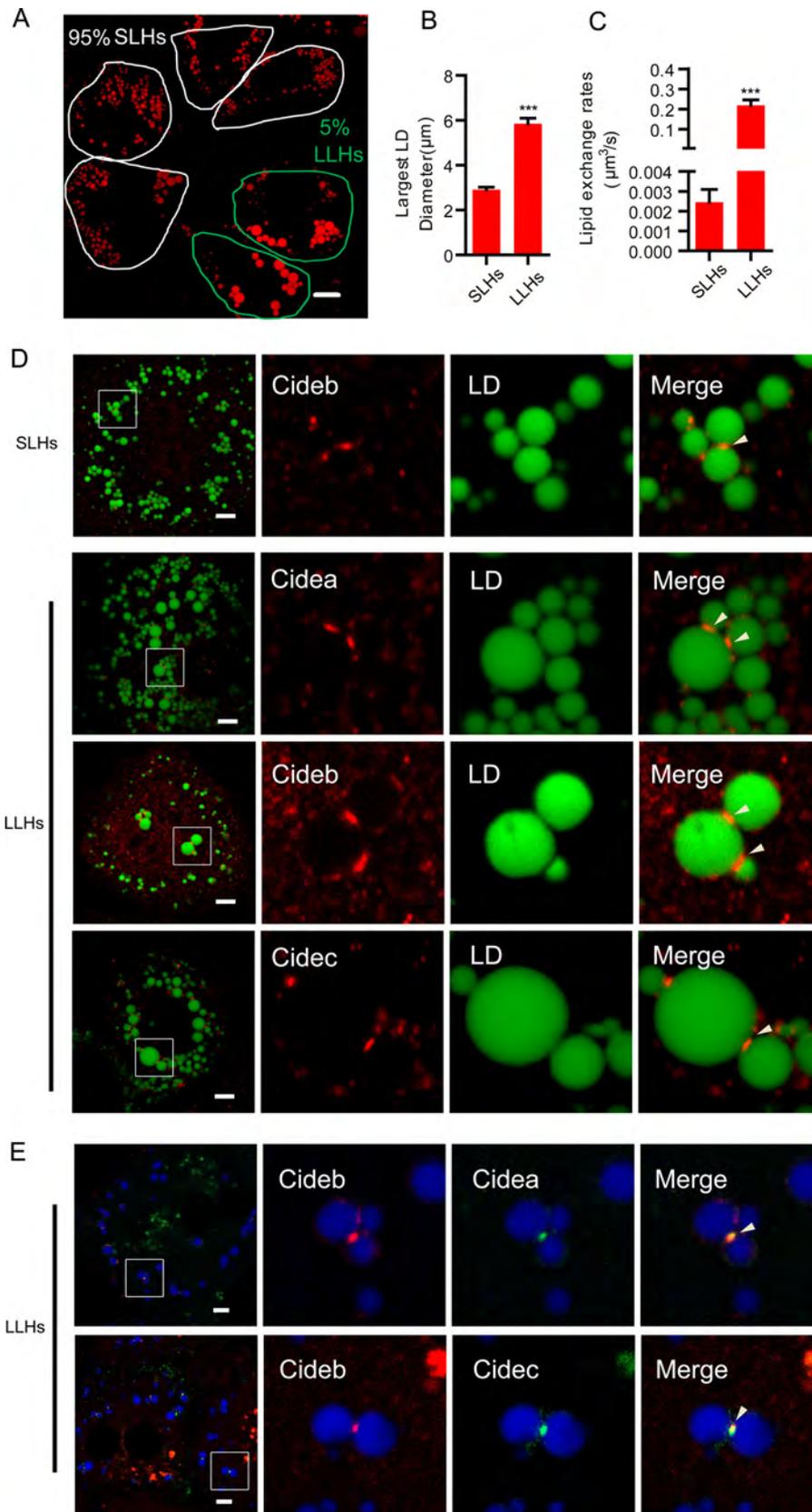


FIGURE 2. **Endogenous CIDE proteins are localized at LDCs in wild-type hepatocytes.** *A*, representative image showing the LD morphology in wild-type hepatocytes. A total of 95% of the hepatocytes have small LDs with a diameter below $4 \mu\text{m}$ (SLHs). A total of 5% of the hepatocytes have large LDs with a diameter above $4 \mu\text{m}$ (LLHs). LDs were labeled with Bodipy 556/568 (C12, red). Scale bars, $10 \mu\text{m}$. *B*, largest LD diameter in SLHs and LLHs. $n = 20$ for each group. *C*, lipid exchange rates in SLHs and LLHs. $n = 8$ for each group. *D*, wild-type hepatocytes were stained with antibodies against Cidea, Cideb, and Cidec. LDs were labeled with Bodipy 493/503 (green). Scale bars, $5 \mu\text{m}$. *E*, LLHs were stained with antibodies against Cideb (red) and Cidea/Cidec (green). Scale bars, $5 \mu\text{m}$. *D* and *E*, arrowheads point to LDCs. Quantitative data are presented as the mean \pm S.E. Differences were considered significant at $p < 0.05$. ***, $p < 0.001$.

CIDE Proteins and Liver Lipid Storage

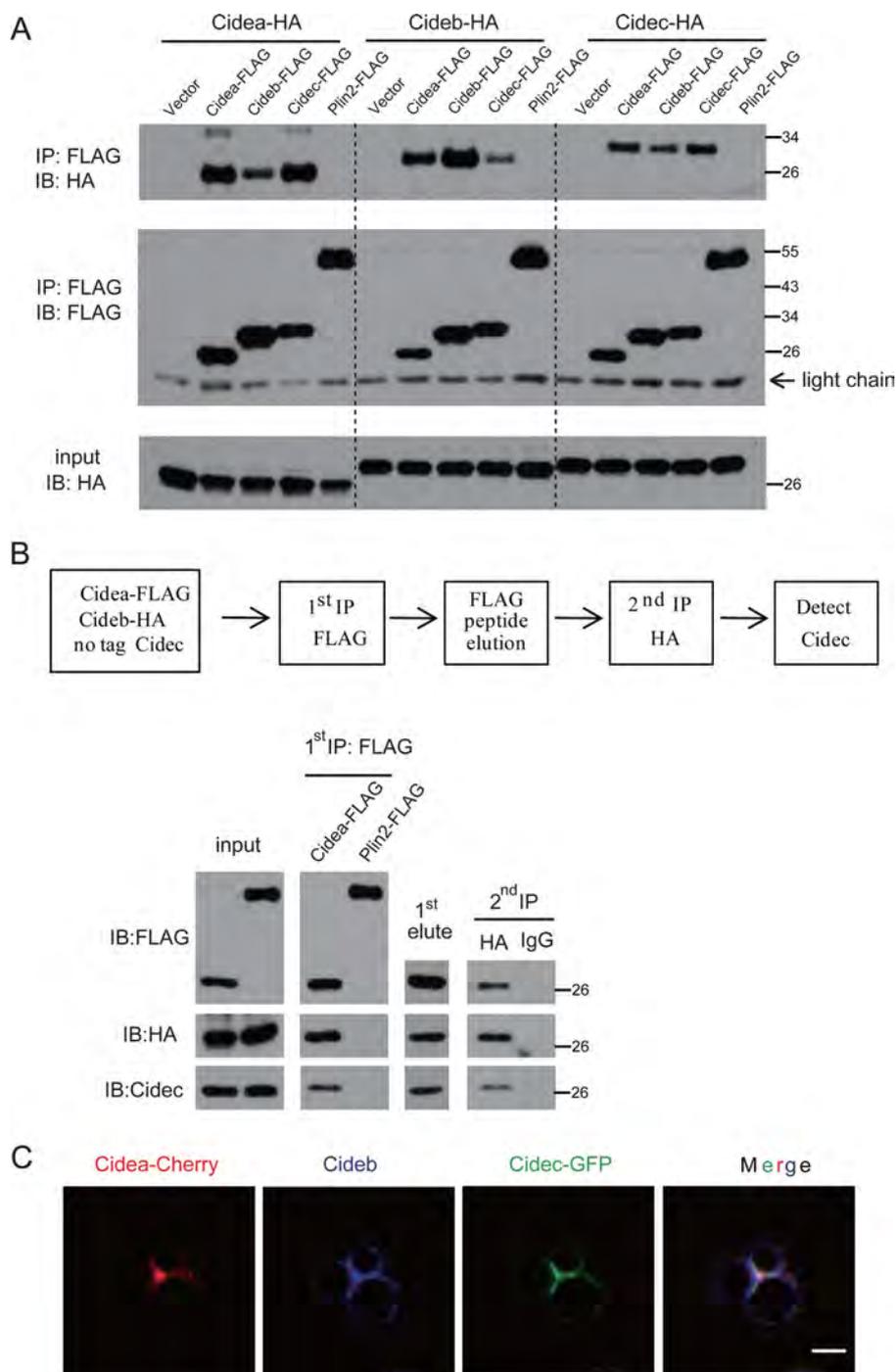


FIGURE 3. CIDE proteins interact with each other and form a complex at LDCSs. *A*, FLAG-tagged CIDEs and Plin2 together with HA-tagged CIDEs were coexpressed in 293T cells. Anti-FLAG M2 beads were used for immunoprecipitation. The immunoprecipitated products were detected by antibodies against FLAG or HA. *IP*, immunoprecipitation; *IB*, immunoblot. *B*, two-step coimmunoprecipitation of complex containing Cidea, Cideb, and Cidec. *Top*, schematic showing procedures for two-step coimmunoprecipitation assay. Cidea-FLAG or Plin2-FLAG (as a control) was transfected into 293T cells with Cideb-HA and untagged Cidec. *C*, Cidea, Cideb, and Cidec were colocalized at the LDCSs. HepG2 cells were cotransfected with Cidea-Cherry (red), Cidec-GFP (green), and Cideb. Cells were stained with antibodies against Cideb (blue). Scale bar, 2 μm .

(in diameter) of the largest LD in these cells was below 4 μm . We defined these hepatocytes as small LD-containing hepatocytes (SLHs). The remaining $\sim 5\%$ of hepatocytes had large LDs, and the size (diameter) of the largest LD in these cells was above 4 μm . We defined those hepatocytes as large LD-containing hepatocytes (LLHs) (Fig. 2, *A* and *B*). In LLHs, the average size of the largest LD in each cell was $\sim 6 \mu\text{m}$ in diameter, and the average size of the largest LD in SLHs was 3 μm in diameter

(Fig. 2*B*). Lipid exchange activity between contacted LD pairs in SLHs was 0.0025 $\mu\text{m}^3/\text{s}$. Lipid exchange activity was 0.22 $\mu\text{m}^3/\text{s}$ in LLHs, nearly 90-fold higher than that in the SLHs (Fig. 2*C*). The heterogeneity in LD size and lipid exchange activity in the hepatocytes prompted us to check the levels of CIDE proteins in SLHs and LLHs using their corresponding antibodies. Cideb proteins were detected in both SLHs and LLHs and were found to be enriched at the LDCSs (Fig. 2*D*). However, Cidea and

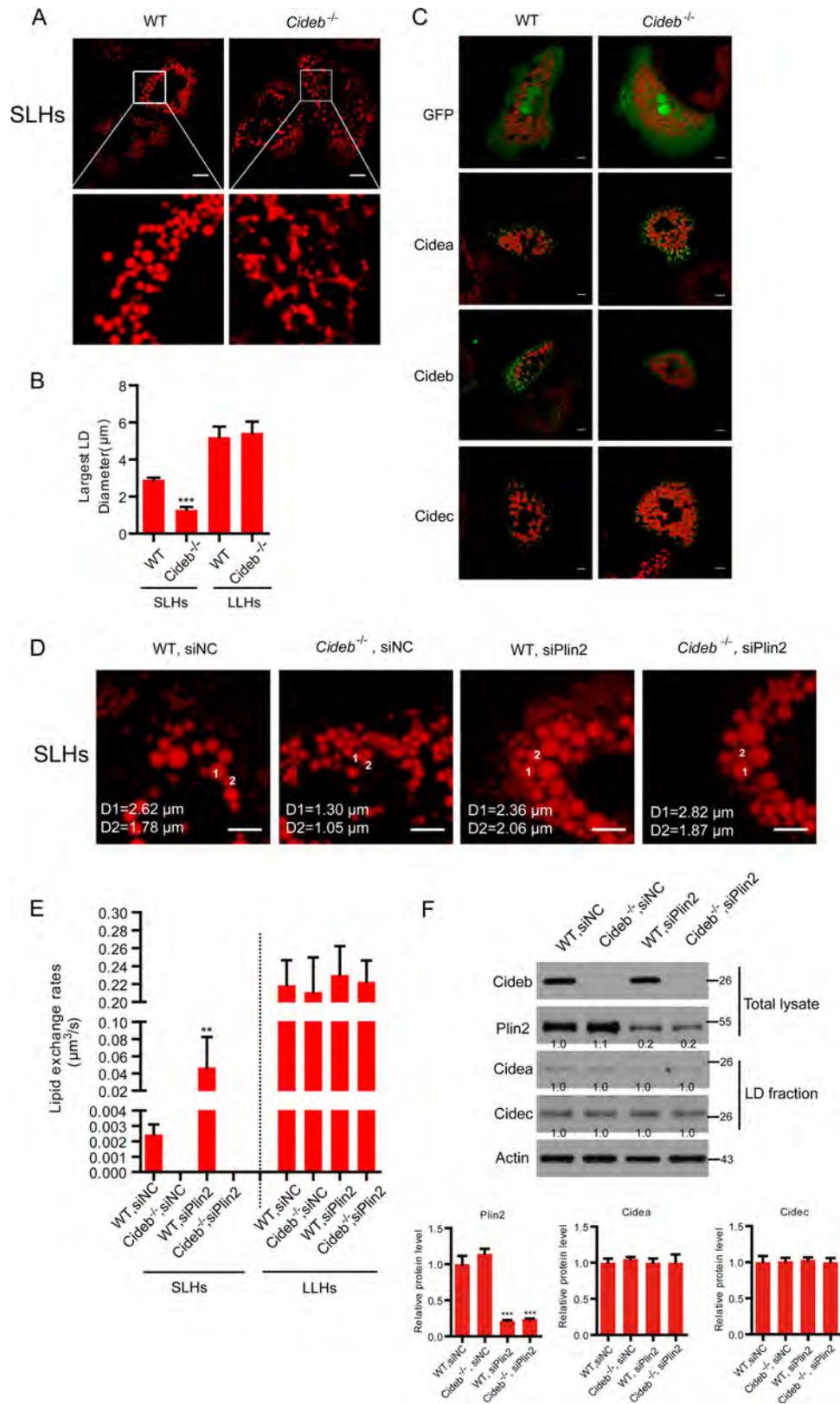
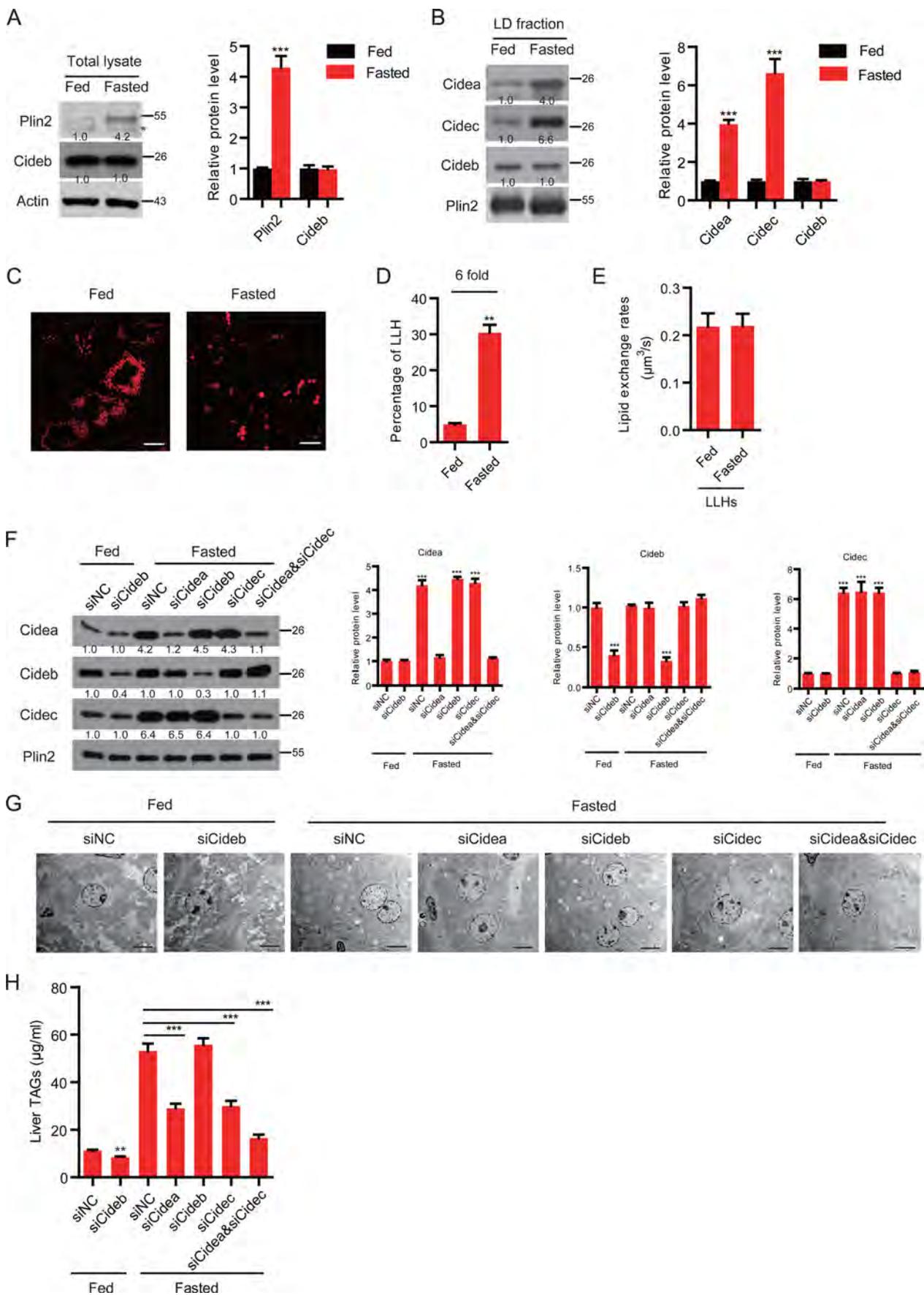


FIGURE 4. Lower LD fusion activity and reduced LD size in *Cideb*^{-/-} hepatocytes. *A*, reduced LD sizes in *Cideb*^{-/-} SLHs. LDs were labeled with Bodipy 556/568 (C12, red). Scale bar, 5 μm. *B*, largest LD sizes in SLHs and LLHs were measured ($n = 20$). *C*, representative image showing the LD size of wild-type (WT) and *Cideb*^{-/-} hepatocytes transfected with CIDEs. Non-tagged CIDE proteins were transfected to hepatocytes. These proteins were stained by their corresponding antibodies. LDs were labeled with Bodipy 556/568 (C12, red). Scale bar, 5 μm. *D*, representative image showing the lipid droplet size of SLHs from WT, *Cideb*^{-/-}, WT with siPlin2, and *Cideb*^{-/-} with siPlin2. Scale bar, 5 μm. *E*, lipid exchange rates in SLHs as in *D* and LLHs ($n = 6$). *F*, protein expression levels in the total lysate and LD fraction of the indicated liver tissues. Quantitation of the bands was performed using Quantity One software and are expressed as the fold change, after correction for actin levels. Values are averages obtained from three independent experiments. Quantitative data are presented as the mean \pm S.E. Differences were considered significant at $p < 0.05$. **, $p < 0.01$; ***, $p < 0.001$.

CIDE Proteins and Liver Lipid Storage



Cidec were only detected in LLHs and were enriched at LDCSs of contacted LDs (Fig. 2D). Co-staining of Cidea- or Cidec-positive hepatocytes with the Cideb antibody showed that Cideb was also present in LLHs and was enriched at the LDCSs (Fig. 2E). These data indicate that hepatocytes can be separated into two types according to their lipid storage capacity. Hepatocytes with high lipid storage capacity express all three CIDE proteins, whereas hepatocytes with low lipid storage capacity express only Cideb.

CIDE Proteins Form a Complex at LDCSs—We then investigated whether CIDEs interact with each other at LDCSs. We co-expressed Cidea-FLAG, Cideb-FLAG, Cidec-FLAG, or Plin2-FLAG with Cidea-HA in 293T cells. Then we used anti-FLAG M2 beads to immunoprecipitate FLAG-tagged proteins. The immunoprecipitated products were detected by using antibody against HA. Cidea was pulled down by Cidea, Cideb, and Cidec but not Plin2 (Fig. 3A). The interaction between Cidea and Cideb was slightly weaker than that of the Cidea-Cidea or Cidea-Cidec interaction. We also found that Cidea-FLAG, Cideb-FLAG, and Cidec-FLAG, but not Plin2-FLAG, were able to immunoprecipitate Cideb-HA or Cidec-HA (Fig. 3A). These data clearly showed that CIDE proteins can interact with each other. We then did a two-step coimmunoprecipitation analysis and to check whether Cidea, Cideb, and Cidec could form a ternary complex. Cidea-FLAG, Cideb-HA, and non-tagged Cidec were coexpressed in 293T cells. 1st step immunoprecipitation was carried out using anti-FLAG antibody. The coimmunoprecipitate components were then eluted with the FLAG peptide, and immunoprecipitation was carried out for the second round using HA antibody. Cideb and Cidec were observed in the final products (Fig. 3B). Contrary to that, Plin2 was not pulled down by either Cideb or Cidec. Consistent with their complex formation, Cidea, Cideb, and Cidec were colocalized and enriched at the LDCSs (Fig. 3C).

Reduced LD Size and Lipid Exchange Activity in Cideb^{-/-} SLHs—To investigate the physiological role of Cideb in controlling LD fusion and lipid storage in hepatocytes, we isolated wild-type and Cideb^{-/-} hepatocytes. The ratio of SLHs and LLHs was similar between the wild-type and Cideb^{-/-} hepatocytes (data not shown). However, the sizes of the LDs were dramatically reduced in the Cideb^{-/-} SLHs (Fig. 4, A and B), although the sizes of LDs in the LLHs were not affected by Cideb depletion (Fig. 4B). Introduction of Cidea, Cideb, or Cidec into Cideb^{-/-} hepatocytes was able to increase the LD sizes (Fig. 4C). Therefore, Cidea and Cidec could functionally substitute for Cideb in enlargement of LDs. No lipid exchange activity was detected in Cideb^{-/-} SLHs (Fig. 4, D and E), consistent with their reduced LD sizes and lower lipid storage capacity. Our previous studies have shown that knockdown of Plin2 enhanced LD size in WT and Cideb^{-/-} hepatocytes (36).

We then tested the lipid exchange rate in hepatocytes with Plin2 knockdown (with a knockdown efficiency of ~80%, Fig. 4F). Plin2 knockdown did not affect the expression level of Cidea and Cidec (Fig. 4F). The lipid exchange rate was dramatically increased in wild-type SLHs with the depletion of Plin2 (18-fold higher, Fig. 4, D–F). However, the knockdown of Plin2 did not affect the lipid exchange activity of Cideb^{-/-} SLHs and LLHs (Fig. 4E). Therefore, Cideb is responsible for LD fusion and lipid storage in SLHs.

Fasting-induced LD Growth and Lipid Storage in the Liver Is Controlled by Cidea and Cidec—Many studies have shown that fasting induces the accumulation of large LDs in hepatocytes and induces higher hepatic lipid storage in mice (21–23). To investigate the role of CIDE proteins in fasting-induced liver steatosis, we analyzed expression levels of CIDE proteins, LD sizes, and lipid exchange and fusion activity in the livers and in isolated hepatocytes of animals fasted up to 16 h. The expression levels of Cideb in the liver were similar before and after fasting (Fig. 5A). Levels of Plin2 in the fasted liver were increased compared with that in the fed liver (4-fold, Fig. 5A). Interestingly, the levels of Cidea and Cidec were both increased in the LD fraction of liver tissue under the fasting condition (Fig. 5B). The percentage of LLHs isolated from the livers of fasted mice was dramatically increased compared with that of fed animals (30% versus 5%, 6-fold higher in fasted livers, Fig. 5, C and D). The lipid exchange rate in LLHs was similar between the fed and fasted hepatocytes (Fig. 5E). To better understand the relevance of Cidea, Cideb, and Cidec on the metabolic adaptation of the liver to fasting, we knocked down Cidea, Cideb, or Cidec using an InvivoFectamine-mediated siRNA delivery system. As shown in Fig. 5F, the expression levels of Cidea, Cideb, and Cidec in the liver were significantly reduced by targeted delivery of siRNAs against Cidea, Cideb, or Cidec. Cideb knockdown under fed conditions indeed led to 25% reduction in TAG storage in liver (Fig. 5, G and H). However, in fasting wild-type mice, Cideb knockdown alone did not affect hepatic TAG levels (Fig. 5H). Lipid accumulation was dramatically reduced in the livers of Cidea- or Cidec-depleted animals that were fasted for 16 h (Fig. 5, G and H). The knockdown of Cidea and Cidec led to a further reduction in hepatic TAG levels and to the accumulation of smaller LDs under fasting conditions (Fig. 5, G and H). These data indicate that Cidea and Cidec mediate the elevated hepatic lipid storage under the fasting condition by increasing the LD fusion in LLHs and switching SLHs to LLHs.

Cidea and Cidec Promote Hepatic Lipid Storage in ob/ob Mice—Cidea and Cidec are both shown to be dramatically unregulated in the livers of ob/ob mice (Fig. 6A). Hepatocytes isolated from ob/ob mice contained very large LDs (Fig. 6B). Expression levels of Cidea and Cidec were detected in nearly all hepatocytes and were enriched at LDCSs (Fig. 6, C and D).

FIGURE 5. Fasting induced the expression of Cidea and Cidec. *A, left*, representative Western blot showing the protein expression profiles of the liver total lysate. The asterisk designates a nonspecific band. *Right*, quantitative analysis of the relative level of Plin2 and Cideb. $n = 3$. Actin was used as a loading control. *B, left*, representative Western blot showing the protein expression profiles of the LD fraction. *Right*, quantitative analysis of the relative level of CIDE proteins. $n = 3$. Plin2 was used as a loading control. *C*, representative image showing the lipid droplet morphology in fed and fasted wild-type (WT) hepatocytes. Scale bar, 10 μm . *D*, ratio of LLHs was increased in fasted conditions ($n = 4$). *E*, lipid exchange rates in LLHs ($n = 6$). *F*, protein expression levels in the LD fraction of the liver. Wild-type mice were injected with different siRNAs. $n = 3$. Quantitation of the bands was performed using Quantity One software and are expressed as the fold change, after correction for Plin2 levels. Values are averages obtained from three independent experiments. *G*, electron microscopy (EM) showing the morphology of the liver. *H*, liver TAG level. $n = 3$. Quantitative data are presented as the mean \pm S.E. Differences were considered significant at $p < 0.05$. **, $p < 0.01$; ***, $p < 0.001$.

CIDE Proteins and Liver Lipid Storage

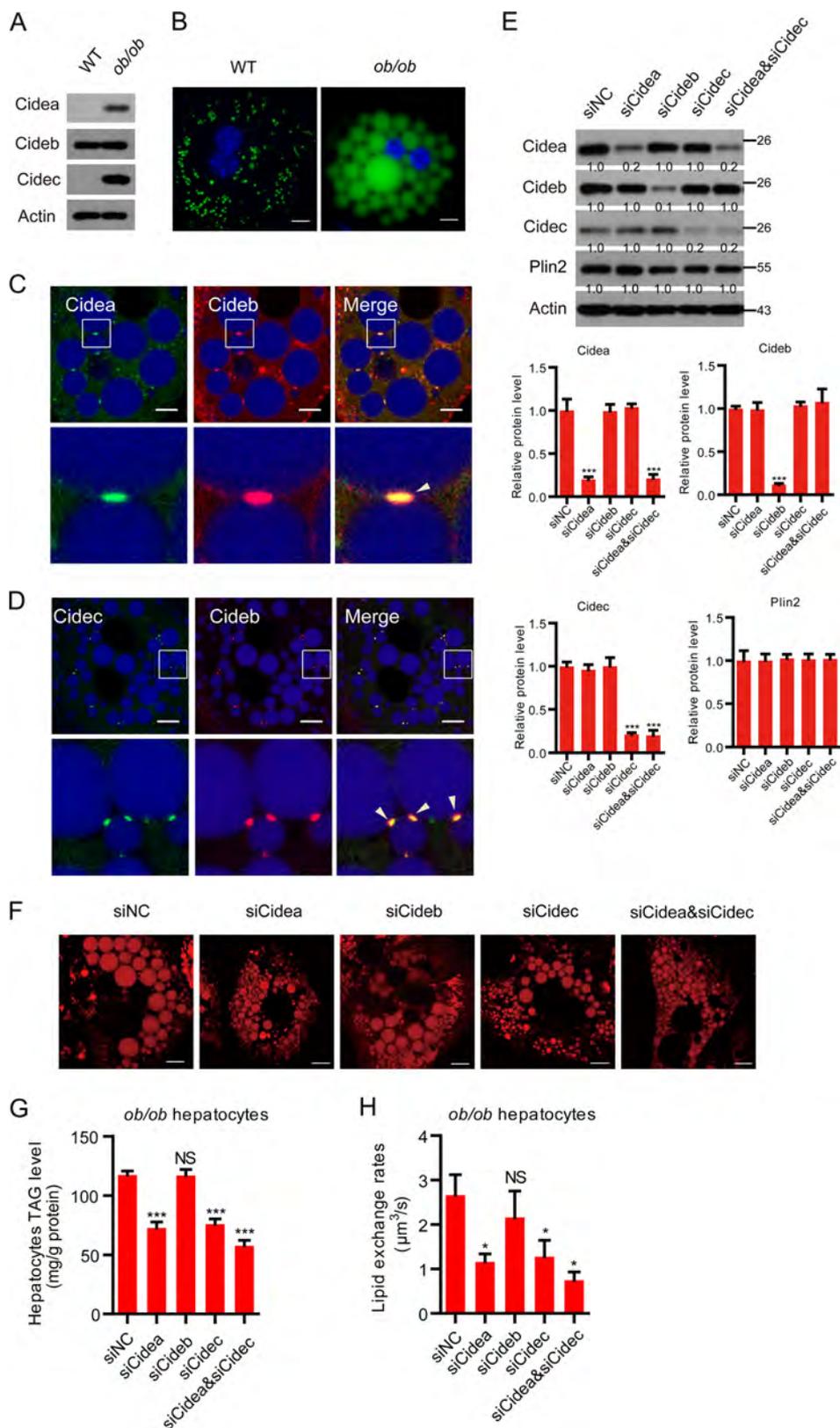


FIGURE 6. CIDE proteins promote lipid storage in *ob/ob* hepatocytes. *A*, expression levels of CIDEs in the livers of wild-type (*WT*) and *ob/ob* mice. *B*, representative image showing the lipid droplet morphology in wild-type and *ob/ob* hepatocytes. Scale bar, 10 μm . *C*, Cidea and Cideb localized on LDCs. *ob/ob* hepatocytes were stained with antibodies against Cidea (green) and Cideb (red). Scale bar, 10 μm . *D*, Cidec and Cideb localized on LDCs. *ob/ob* hepatocytes were stained with antibodies against Cidec (green) and Cideb (red). Scale bar, 10 μm . Arrowheads point to LDCs. *E*, expression levels of the indicated proteins. Quantitation of the bands was performed using Quantity One software and are expressed as the fold change, after correction for actin levels. Values are averages obtained from three experiments. *F*, representative image showing the LD morphology after the depletion of different proteins. Scale bar, 10 μm . *G*, hepatocytes TAG level ($n = 4$). *H*, lipid exchange rates ($n = 4$). Quantitative data are presented as the mean \pm S.E. Differences were considered significant at $p < 0.05$. *, $p < 0.05$; ***, $p < 0.001$; NS, no significant difference.

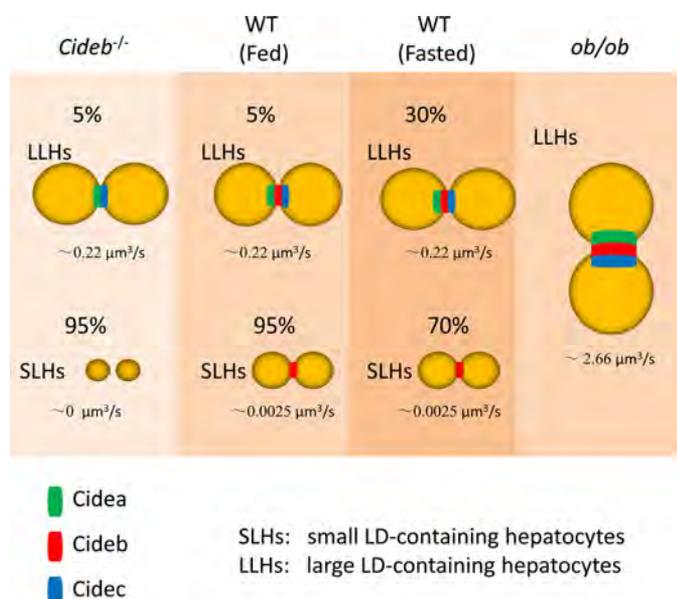


FIGURE 7. Proposed model of CIDE-mediated LD fusion and growth in hepatocytes. In wild-type (*WT*) hepatocytes, two types (SLHs and LLHs) are observed. Cideb is expressed in SLHs and LLHs. Cidea and Cidec are expressed in LLHs. Fasting enhances the percentage of LLHs. *Cideb* deficiency leads to reduced LD size in SLHs. *ob/ob* hepatocytes have higher Cidea and Cidec expression levels and higher lipid exchange rates.

Cideb was also detected and enriched at LDCSs of all *ob/ob* hepatocytes (Fig. 6, C and D). We knocked down individual CIDE protein alone or in combination in *ob/ob* hepatocytes using a Lipofectamine-mediated siRNA delivery system. This strategy resulted in a substantial knockdown of Cidea (80% knockdown efficiency), Cideb (90% knockdown efficiency), and Cidec (80% knockdown efficiency), respectively (Fig. 6E). The knockdown of Cidea and Cidec alone or in combination in the *ob/ob* hepatocytes resulted in reduced lipid storage and accumulation of smaller LDs (Fig. 6, F and G). However, the knockdown of Cideb did not affect the LD size of *ob/ob* hepatocytes (Fig. 6, F and G). Consistently, the depletion of Cidea or Cidec in *ob/ob* hepatocytes led to reduced lipid exchange activity (Fig. 6H). The knockdown of Cideb did not affect the lipid exchange activity of contacted LD pairs (Fig. 6H). The knockdown of both Cidea and Cidec reduced the lipid exchange rate further compared with Cidea or Cidec single depletion (Fig. 6H). These data indicate that Cidea and Cidec play important roles in hepatic lipid storage in *ob/ob* animals.

Discussion

CIDE family proteins (Cidea and Cidec) act as important regulators of lipid storage and lipid metabolism in adipocytes by promoting LD fusion and growth (17). Here, we examined the precise function of individual CIDEs in controlling LD fusion and growth in the liver. We found that overexpression of CIDE proteins dramatically induced the formation of large LDs and promoted the fusion and growth of LDs in hepatocytes. In addition, all CIDE proteins are enriched at LDCSs in hepatocytes or when they were overexpressed in liver cell lines.

Interestingly, we identified two populations of hepatocytes based on the size of LDs (SLHs and LLHs). We further observed that $\sim 5\%$ of wild-type hepatocytes are LLHs that express all

three CIDE proteins and have higher LD fusion activity. SLHs express Cideb alone and have lower LD fusion activity (Fig. 7). *Cideb*^{-/-} SLHs have smaller LDs and lower lipid exchange activity. In contrast, depletion of Cideb did not affect the LD sizes and LD fusion activity of LLHs. The lipid storage capacity in SLHs is primarily determined by Cideb, whereas Cidea and Cidec regulate lipid storage in LLHs. Under fasting conditions, the percentage of LLHs and the expression levels of both Cidea and Cidec increased dramatically. However, the lipid exchange activity in individual LLHs was similar to that of control hepatocytes. Therefore, fasting induced the expression of Cidea/Cidec in SLHs, promoting the conversion of SLHs to LLHs, but did not affect the Cidea/Cidec expression in LLHs. Moreover, the percentage of Cidea/Cidec-expressing cells reached nearly 100% in *ob/ob* hepatocytes. All hepatocytes from *ob/ob* livers exhibited high expression levels of Cidea/Cidec and had higher LD fusion activity (Fig. 7). The increased expression of Cidea and Cidec in LLHs could be due to their transcriptional regulation by various factors (22–24). Alternatively, increased Cidea and Cidec protein levels in fasted or obese hepatocytes are most likely due to the enhanced stability resulting from exposure to high fatty acid levels and TAG synthesis (20–23, 37). We think the expansion of LLHs that express Cidea and Cidec contributes primarily to the development of hepatic steatosis. In addition, the percentage of LLHs may reflect the nutrition supply to the liver and may play an important role in the development of liver steatosis. Further characterization of gene expression profiles and physiological function (insulin sensitivity) of SLHs and LLHs will be useful in the investigation of the origin and regulation of SLHs and LLHs.

Previous data showed that the expressions levels of Cidea and Cidec were significantly increased in human steatotic liver (20, 38, 39). A single nucleotide polymorphism of a G to T transversion in CIDEA exon 4, which is equivalent to a V115F substitution, is associated with body mass index in Swedish male and female obese patients (40). In Japanese and Chinese patients, the CIDEA V115F polymorphism is associated with obesity and metabolic syndrome (41, 42). A single nucleotide polymorphism involving a G to T transversion in CIDEA exon 6, which causes a E186X nonsense mutation was also observed in a female patient who showed partial lipodystrophy (43). Controlling the correct amount of CIDE proteins was important in maintaining lipid homeostasis in the liver. If there was insufficient CIDE proteins in the liver, this would lead to reduced LD sizes and lower lipid storage capacity. Free fatty acid levels may increase in hepatocytes, resulting in lipotoxicity and increased oxidative stress and inflammatory response (44). Previous data showed that overexpression of CIDE proteins in many cell types resulted in a caspase-independent cell death (45). However, the rate of cell death was significantly reduced in CIDE-overexpressing cells when exogenous oleic acid was introduced (46). In the presence of lipid-rich medium, CIDE proteins localized to LDs promote LD fusion and lipid storage, resulting in reduced lipotoxicity and oxidative stress.

In conclusion, we have shown that CIDE proteins are localized to LDCSs and promote LD fusion in hepatocytes. According to the differential expression of CIDE proteins and the sizes of LDs, hepatocytes can be characterized into two populations

(SLHs and LLHs). Cideb plays an important role in promoting lipid storage and maintaining lipid homeostasis in the liver under normal conditions. Cidea and Cidec are crucial regulators of hepatic lipid storage under fasting and obese conditions.

Author Contributions—L. Z., W. X., L. W., and P. L. designed the research; W. X., L. Z., L. W., M. Y., H. R., F. C., and J. Y. performed the experiments; L. Z., L. W., L. X., D. X., J. Z. L., X. X., and M. A. analyzed the data; and P. L. and L. Z. wrote the paper. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments—We thank the members of the P. Li Laboratory at Tsinghua University for their helpful discussions.

References

- Cornier, M. A., Dabelea, D., Hernandez, T. L., Lindstrom, R. C., Steig, A. J., Stob, N. R., Van Pelt, R. E., Wang, H., and Eckel, R. H. (2008) The metabolic syndrome. *Endocr. Rev.* **29**, 777–822
- Tessari, P., Coracina, A., Cosma, A., and Tiengo, A. (2009) Hepatic lipid metabolism and non-alcoholic fatty liver disease. *Nutr. Metab. Cardiovasc. Dis.* **19**, 291–302
- Musso, G., Gambino, R., and Cassader, M. (2009) Recent insights into hepatic lipid metabolism in non-alcoholic fatty liver disease (NAFLD). *Prog. Lipid Res.* **48**, 1–26
- Dietrich, P., and Hellerbrand, C. (2014) Non-alcoholic fatty liver disease, obesity and the metabolic syndrome. *Best Pract. Res. Clin. Gastroenterol.* **28**, 637–653
- Wilfling, F., Haas, J. T., Walther, T. C., and Farese, R. V., Jr. (2014) Lipid droplet biogenesis. *Curr. Opin. Cell Biol.* **29**, 39–45
- Farese, R. V., Jr., and Walther, T. C. (2009) Lipid droplets finally get a little R-E-S-P-E-C-T. *Cell* **139**, 855–860
- Thiam, A. R., Farese, R. V., Jr., and Walther, T. C. (2013) The biophysics and cell biology of lipid droplets. *Nat. Rev. Mol. Cell Biol.* **14**, 775–786
- Yang, H., Galea, A., Sytnyk, V., and Crossley, M. (2012) Controlling the size of lipid droplets: lipid and protein factors. *Curr. Opin. Cell Biol.* **24**, 509–516
- Wilfling, F., Thiam, A. R., Olarte, M. J., Wang, J., Beck, R., Gould, T. J., Allgeyer, E. S., Pincet, F., Bewersdorf, J., Farese, R. V., Jr., and Walther, T. C. (2014) Arf1/COPI machinery acts directly on lipid droplets and enables their connection to the ER for protein targeting. *Elife* **3**, e01607
- Gross, D. A., Zhan, C., and Silver, D. L. (2011) Direct binding of triglyceride to fat storage-inducing transmembrane proteins 1 and 2 is important for lipid droplet formation. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19581–19586
- Krahmer, N., Guo, Y., Wilfling, F., Hilger, M., Lingrell, S., Heger, K., Newman, H. W., Schmidt-Suppran, M., Vance, D. E., Mann, M., Farese, R. V., Jr., and Walther, T. C. (2011) Phosphatidylcholine synthesis for lipid droplet expansion is mediated by localized activation of CTP:phosphocholine cytidyltransferase. *Cell Metab.* **14**, 504–515
- McFie, P. J., Banman, S. L., Kary, S., and Stone, S. J. (2011) Murine diacylglycerol acyltransferase-2 (DGAT2) can catalyze triacylglycerol synthesis and promote lipid droplet formation independent of its localization to the endoplasmic reticulum. *J. Biol. Chem.* **286**, 28235–28246
- Xu, L., Zhou, L., and Li, P. (2012) CIDE proteins and lipid metabolism. *Arterioscler. Thromb. Vasc. Biol.* **32**, 1094–1098
- Crunk, A. E., Monks, J., Murakami, A., Jackman, M., Maclean, P. S., Ladinsky, M., Bales, E. S., Cain, S., Orlicky, D. J., and McManaman, J. L. (2013) Dynamic regulation of hepatic lipid droplet properties by diet. *PLoS One* **8**, e67631
- Bickel, P. E., Tansey, J. T., and Welte, M. A. (2009) PAT proteins, an ancient family of lipid droplet proteins that regulate cellular lipid stores. *Biochim. Biophys. Acta* **1791**, 419–440
- Okumura, T. (2011) Role of lipid droplet proteins in liver steatosis. *J. Physiol. Biochem.* **67**, 629–636
- Gong, J., Sun, Z., and Li, P. (2009) CIDE proteins and metabolic disorders. *Curr. Opin. Lipidol.* **20**, 121–126
- Yonezawa, T., Kurata, R., Kimura, M., and Inoko, H. (2011) Which CIDE are you on? Apoptosis and energy metabolism. *Mol. Biosyst.* **7**, 91–100
- Matsusue, K., Kusakabe, T., Noguchi, T., Takiguchi, S., Suzuki, T., Yamano, S., and Gonzalez, F. J. (2008) Hepatic steatosis in leptin-deficient mice is promoted by the PPAR γ target gene Fsp27. *Cell Metab.* **7**, 302–311
- Zhou, L., Xu, L., Ye, J., Li, D., Wang, W., Li, X., Wu, L., Wang, H., Guan, F., and Li, P. (2012) Cidea promotes hepatic steatosis by sensing dietary fatty acids. *Hepatology* **56**, 95–107
- Vilà-Brau, A., De Sousa-Coelho, A. L., Gonçalves, J. F., Haro, D., and Marrero, P. F. (2013) Fsp27/CIDEc is a CREB target gene induced during early fasting in liver and regulated by FA oxidation rate. *J. Lipid Res.* **54**, 592–601
- Langhi, C., and Baldán, Á. (2015) CIDEc/FSP27 is regulated by peroxisome proliferator-activated receptor α and plays a critical role in fasting- and diet-induced hepatosteatosis. *Hepatology* **61**, 1227–1238
- Xu, X., Park, J. G., So, J. S., and Lee, A. H. (2015) Transcriptional activation of Fsp27 by the liver-enriched transcription factor CREBH promotes lipid droplet growth and hepatic steatosis. *Hepatology* **61**, 857–869
- Puri, V. (2013) Fasting regulates FSP27 expression in the liver. *J. Lipid Res.* **54**, 569–570
- Jambunathan, S., Yin, J., Khan, W., Tamori, Y., and Puri, V. (2011) FSP27 promotes lipid droplet clustering and then fusion to regulate triglyceride accumulation. *PLoS One* **6**, e28614
- Gong, J., Sun, Z., Wu, L., Xu, W., Schieber, N., Xu, D., Shui, G., Yang, H., Parton, R. G., and Li, P. (2011) Fsp27 promotes lipid droplet growth by lipid exchange and transfer at lipid droplet contact sites. *J. Cell Biol.* **195**, 953–963
- Wu, L., Zhou, L., Chen, C., Gong, J., Xu, L., Ye, J., Li, D., and Li, P. (2014) Cidea controls lipid droplet fusion and lipid storage in brown and white adipose tissue. *Sci. China Life Sci.* **57**, 107–116
- Barneda, D., Planas-Iglesias, J., Gaspar, M. L., Mohammadyani, D., Prasan-nan, S., Dormann, D., Han, G. S., Jesch, S. A., Carman, G. M., Kagan, V., Parker, M. G., Ktistakis, N. T., Dixon, A. M., Klein-Seetharaman, J., Henry, S., and Christian, M. (2015) The brown adipocyte protein CIDEA promotes lipid droplet fusion via a phosphatidic acid-binding amphipathic helix. *Elife* **4**, e07485
- Sun, Z., Gong, J., Wu, H., Xu, W., Wu, L., Xu, D., Gao, J., Wu, J. W., Yang, H., Yang, M., and Li, P. (2013) Perilipin1 promotes unilocular lipid droplet formation through the activation of Fsp27 in adipocytes. *Nat. Commun.* **4**, 1594
- Wu, L., Xu, D., Zhou, L., Xie, B., Yu, L., Yang, H., Huang, L., Ye, J., Deng, H., Yuan, Y. A., Chen, S., and Li, P. (2014) Rab8a-AS160-MSS4 regulatory circuit controls lipid droplet fusion and growth. *Dev. Cell* **30**, 378–393
- Grahn, T. H., Zhang, Y., Lee, M. J., Sommer, A. G., Mostoslavsky, G., Fried, S. K., Greenberg, A. S., and Puri, V. (2013) FSP27 and PLIN1 interaction promotes the formation of large lipid droplets in human adipocytes. *Biochem. Biophys. Res. Commun.* **432**, 296–301
- Ye, J., Li, J. Z., Liu, Y., Li, X., Yang, T., Ma, X., Li, Q., Yao, Z., and Li, P. (2009) Cideb, an ER- and lipid droplet-associated protein, mediates VLDL lipidation and maturation by interacting with apolipoprotein B. *Cell Metab.* **9**, 177–190
- Li, J. Z., Ye, J., Xue, B., Qi, J., Zhang, J., Zhou, Z., Li, Q., Wen, Z., and Li, P. (2007) Cideb regulates diet-induced obesity, liver steatosis, and insulin sensitivity by controlling lipogenesis and fatty acid oxidation. *Diabetes* **56**, 2523–2532
- Toh, S. Y., Gong, J., Du, G., Li, J. Z., Yang, S., Ye, J., Yao, H., Zhang, Y., Xue, B., Li, Q., Yang, H., Wen, Z., and Li, P. (2008) Up-regulation of mitochondrial activity and acquirement of brown adipose tissue-like property in the white adipose tissue of fsp27 deficient mice. *PLoS One* **3**, e2890
- Zhou, L., Park, S. Y., Xu, L., Xia, X., Ye, J., Su, L., Jeong, K. H., Hur, J. H., Oh, H., Tamori, Y., Zingaretti, C. M., Cinti, S., Argente, J., Yu, M., Wu, L., et al. (2015) Insulin resistance and white adipose tissue inflammation are uncoupled in energetically challenged Fsp27-deficient mice. *Nat. Commun.* **6**, 5949
- Li, X., Ye, J., Zhou, L., Gu, W., Fisher, E. A., and Li, P. (2012) Opposing roles of cell death-inducing DFF45-like effector B and perilipin 2 in controlling hepatic VLDL lipidation. *J. Lipid Res.* **53**, 1877–1889

37. Nian, Z., Sun, Z., Yu, L., Toh, S. Y., Sang, J., and Li, P. (2010) Fat-specific protein 27 undergoes ubiquitin-dependent degradation regulated by triacylglycerol synthesis and lipid droplet formation. *J. Biol. Chem.* **285**, 9604–9615
38. Hall, A. M., Brunt, E. M., Klein, S., and Finck, B. N. (2010) Hepatic expression of cell death-inducing DFFA-like effector C in obese subjects is reduced by marked weight loss. *Obesity* **18**, 417–419
39. Xu, M. J., Cai, Y., Wang, H., Altamirano, J., Chang, B., Bertola, A., Odena, G., Lu, J., Tanaka, N., Matsusue, K., Matsubara, T., Mukhopadhyay, P., Kimura, S., Pacher, P., Gonzalez, F. J., Bataller, R., and Gao, B. (2015) Fat-specific protein 27/CIDEA promotes development of alcoholic steatohepatitis in mice and humans. *Gastroenterology* **149**, 1030–1041. e1036
40. Dahlman, I., Kaaman, M., Jiao, H., Kere, J., Laakso, M., and Arner, P. (2005) The CIDEA gene V115F polymorphism is associated with obesity in Swedish subjects. *Diabetes* **54**, 3032–3034
41. Zhang, L., Miyaki, K., Nakayama, T., and Muramatsu, M. (2008) Cell death-inducing DNA fragmentation factor α -like effector A (CIDEA) gene V115F (G→T) polymorphism is associated with phenotypes of metabolic syndrome in Japanese men. *Metabolism* **57**, 502–505
42. Wu, J., Zhang, L., Zhang, J., Dai, Y., Bian, L., Song, M., Russell, A., and Wang, W. (2013) The genetic contribution of CIDEA polymorphisms, haplotypes and loci interaction to obesity in a Han Chinese population. *Mol. Biol. Rep.* **40**, 5691–5699
43. Rubio-Cabezas, O., Puri, V., Murano, I., Saudek, V., Semple, R. K., Dash, S., Hyden, C. S., Bottomley, W., Vigouroux, C., Magré, J., Raymond-Barker, P., Murgatroyd, P. R., Chawla, A., Skepper, J. N., Chatterjee, V. K., *et al.* (2009) Partial lipodystrophy and insulin resistant diabetes in a patient with a homozygous nonsense mutation in CIDEA. *EMBO Mol. Med.* **1**, 280–287
44. Anstee, Q. M., and Goldin, R. D. (2006) Mouse models in non-alcoholic fatty liver disease and steatohepatitis research. *Int. J. Exp. Pathol.* **87**, 1–16
45. Chen, Z., Guo, K., Toh, S. Y., Zhou, Z., and Li, P. (2000) Mitochondria localization and dimerization are required for CIDE-B to induce apoptosis. *J. Biol. Chem.* **275**, 22619–22622
46. Liu, K., Zhou, S., Kim, J. Y., Tillison, K., Majors, D., Rearick, D., Lee, J. H., Fernandez-Boyanapalli, R. F., Barricklow, K., Houston, M. S., and Smas, C. M. (2009) Functional analysis of FSP27 protein regions for lipid droplet localization, caspase-dependent apoptosis, and dimerization with CIDEA. *Am. J. Physiol. Endocrinol. Metab.* **297**, E1395–E1413

Sar1 GTPase Activity Is Regulated by Membrane Curvature^{*□}

Received for publication, June 12, 2015, and in revised form, October 27, 2015. Published, JBC Papers in Press, November 6, 2015, DOI 10.1074/jbc.M115.672287

Michael G. Hanna IV[‡], Ioanna Mela^{§1}, Lei Wang[‡], Robert M. Henderson^{§1}, Edwin R. Chapman^{¶12},
 J. Michael Edwardson^{§1}, and Anjon Audhya^{‡3}

From the [‡]Department of Biomolecular Chemistry, University of Wisconsin-Madison School of Medicine and Public Health, Madison, Wisconsin 53706, the [§]Department of Pharmacology, University of Cambridge, Tennis Court Road, CB2 1PD Cambridge, United Kingdom, and the [¶]Department of Neuroscience, University of Wisconsin-Madison School of Medicine and Public Health, Madison, Wisconsin 53705

The majority of biosynthetic secretory proteins initiate their journey through the endomembrane system from specific subdomains of the endoplasmic reticulum. At these locations, coated transport carriers are generated, with the Sar1 GTPase playing a critical role in membrane bending, recruitment of coat components, and nascent vesicle formation. How these events are appropriately coordinated remains poorly understood. Here, we demonstrate that Sar1 acts as the curvature-sensing component of the COPII coat complex and highlight the ability of Sar1 to bind more avidly to membranes of high curvature. Additionally, using an atomic force microscopy-based approach, we further show that the intrinsic GTPase activity of Sar1 is necessary for remodeling lipid bilayers. Consistent with this idea, Sar1-mediated membrane remodeling is dramatically accelerated in the presence of its guanine nucleotide-activating protein (GAP), Sec23–Sec24, and blocked upon addition of guanosine-5'-[(β,γ -imido)triphosphate, a poorly hydrolysable analog of GTP. Our results also indicate that Sar1 GTPase activity is stimulated by membranes that exhibit elevated curvature, potentially enabling Sar1 membrane scission activity to be spatially restricted to highly bent membranes that are characteristic of a bud neck. Taken together, our data support a stepwise model in which the amino-terminal amphipathic helix of GTP-bound Sar1 stably penetrates the endoplasmic reticulum membrane, promoting local membrane deformation. As membrane bending increases, Sar1 membrane binding is elevated, ultimately culminating in GTP hydrolysis, which may destabilize the bilayer sufficiently to facilitate membrane fission.

ative environment of the endoplasmic reticulum (ER)⁴ lumen and subsequently packaged into coated vesicles or tubules for export. Components of the cytosolic coat protein II (COPII) complex play key roles in ER membrane remodeling and scission to bud transport carriers (1–4). In particular, the conserved Sar1 GTPase has been implicated in multiple steps of this process. Current models suggest that in the presence of its guanine nucleotide exchange factor (GEF) Sec12, which is enriched at subdomains of the ER that synthesize COPII-coated carriers, GTP-bound Sar1 exposes an amphipathic helix that penetrates the outer leaflet of the membrane (5–9). This type of membrane penetration induces positive curvature that is further stabilized when Sar1-GTP is bound by its major effector, the Sec23–24 heterodimer, which exhibits a concave surface (10, 11). Additionally, recruited Sec23–24 complexes generate an adaptor layer for Sec13–31 lattice assembly, completing the COPII vesicle coat (12). Recent studies have highlighted the flexibility exhibited by the inner and outer COPII layers, enabling the formation of small vesicles (~50–100 nm in diameter) and elongated tubules (~300–400 nm in length), which can capture and traffic both small and large cargoes, respectively, away from the ER (13, 14). However, mechanisms governing the location and timing of carrier scission have not been clearly defined. In particular, how GTP hydrolysis on Sar1 contributes to this process continues to be debated.

In cell-free assays conducted more than 2 decades ago, poorly hydrolysable analogs of GTP were shown to be sufficient to promote COPII-mediated vesicle formation and release, suggesting that GTP hydrolysis was dispensable for the budding reaction (1). Similar studies were conducted more recently and confirmed this finding (14). However, in both cases, Sar1-dependent vesicle release was measured following differential centrifugation and/or prolonged periods of incubation with

Approximately one-third of all translated proteins in mammalian cells are predicted to enter the secretory pathway (1). These diverse cargoes must be folded properly within the oxi-

* This work was supported in part by National Institutes of Health Grants GM110567 and GM088151 (to A. A.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

□ Author's Choice—Final version free via Creative Commons CC-BY license.

◆ This article was selected as a Paper of the Week.

□ This article contains supplemental Videos S1–S10.

¹ Supported by Grant BB/J018236/1 from the Biotechnology and Biological Sciences Research Council.

² Investigator of the Howard Hughes Medical Institute.

³ To whom correspondence should be addressed: Dept. of Biomolecular Chemistry, University of Wisconsin-Madison School of Medicine and Public Health, 440 Henry Mall, Madison, WI 53706. Tel.: 608-262-3761; Fax: 608-262-5253; E-mail: audhya@wisc.edu.

⁴ The abbreviations used are: ER, endoplasmic reticulum; DOPC, 1,2-dioleoyl-*sn*-glycero-3-phosphocholine; DOPE, 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine; DOPS, 1,2-dioleoyl-*sn*-glycero-3-phosphoserine; DOPA, 1,2-dioleoyl-*sn*-glycero-3-phosphate; PI, phosphatidylinositol; PI4P, phosphatidylinositol 4-phosphate; PI4,5P₂, phosphatidylinositol 4,5-bisphosphate; DAG, diacylglycerol; NBD-PE, 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-*N*-(7-nitro-2-*l*,3-benzoxadiazol-4-yl); GEF, guanine nucleotide exchange factor; GUV, giant unilamellar vesicle; AFM, atomic force microscopy; SLB, supported lipid bilayer; COPII, coat protein II; MALS, multi-angle light scattering; GMP-PNP, guanosine-5'-[(β,γ -imido)triphosphate; dansyl, 5-dimethylaminonaphthalene-1-sulfonyl; GAP, guanine nucleotide-activating protein; PE, phosphatidylethanolamine; 6,7-Br₂-PC, 1-palmitoyl-2-stearoyl-(6,7)-dibromo-*sn*-glycero-3-phosphocholine; 11,12-Br₂-PC, 1-palmitoyl-2-stearoyl-(11,12)-dibromo-*sn*-glycero-3-phosphocholine.

COPII components and nucleotide analogs, which may have led to the unintentional detachment of nascent transport carriers from ER membranes. Using synthetic giant unilamellar vesicles (GUVs), another report showed that Sar1 alone was capable of deforming membranes and generating small vesicles independently of GTP hydrolysis (15). However, a high concentration of Sar1 ($>7 \mu\text{M}$) was necessary for this effect, which may not be physiologically relevant in the context of COPII-mediated budding *in vivo*. Moreover, the electron microscopy (EM)-based methods employed in this study required mechanical perturbations, including fixation and dehydration in negative staining preparations and liquid blotting in cryogenic approaches, which can perturb samples sufficiently to induce artifacts (16). In contrast to these studies, others have shown that inhibition of GTP hydrolysis on Sar1 strongly inhibits vesicle budding from ER microsomes or vesiculation of synthetic GUVs (7, 8, 13, 16–18), raising the possibility that nucleotide hydrolysis contributes to the membrane fission reaction necessary to detach transport carriers. Although difficult to reconcile, these conflicting reports highlight the need for minimally invasive techniques to study the role of Sar1 in membrane remodeling, while still maintaining sufficiently high spatial and temporal resolution to interrogate the process.

Another consistent observation from these studies indicates that Sar1 is capable of associating with lipid bilayers that exhibit a range of curvatures, in a manner dependent on the amino-terminal amphipathic helix of Sar1. Specifically, Sar1 has been shown to bind to both highly curved liposomes (nanometers in diameter) and flat GUVs (microns in diameter) (7, 15). However, Sar1 remodels GUVs over relatively long time scales (minutes to hours), generating heterogeneous tubules that range in diameter from ~ 40 to 250 nm (15). Depending on the size of the tubule, Sar1 can form structurally unique scaffolds, varying from ordered lattices on membranes possessing shallow curvature to fragmented arrays on bilayers of higher curvature (15). The difference in its membrane binding properties may be related to its dual role in the formation and budding of COPII-coated transport carriers. However, apart from structural studies and largely qualitative measurements of bilayer association, the biophysical characteristics of Sar1 membrane binding have not been clearly defined. Additionally, it remains unknown how membrane curvature impacts GTP hydrolysis on Sar1. Here, we investigate these properties using Sar1 isoforms isolated from three distinct species. Our data demonstrate that membrane curvature strongly influences both the affinity of Sar1 for membranes and its rate of GTP hydrolysis in an evolutionarily conserved manner. Collectively, our data are most consistent with a model in which Sar1 promotes membrane scission at bud necks in a curvature-dependent process that involves GTP hydrolysis.

Experimental Procedures

Caenorhabditis elegans Growth, Maintenance, RNA Interference, and Live Imaging—All *C. elegans* strains used in this study were derived from the Bristol strain N2, which was described previously (19, 20). Double-stranded RNA (dsRNA) was synthesized from templates prepared by PCR to amplify *C. elegans* genomic DNA. For RNAi experiments, early L4 stage hermaph-

rodites were soaked in dsRNA for 24 h at 20 °C within a humidified chamber. Animals were then allowed to recover for 24–48 h before analyzing them for embryo production or mounting onto a 10% agarose pad in a 4- μl suspension of polystyrene beads to immobilize them for imaging (21). Images were acquired on a swept-field confocal microscope (Nikon Ti-E), using a Nikon $\times 60$, 1.4 numerical aperture Planapo oil objective lens and a Roper CoolSnap HQ2 CCD camera. Acquisition parameters were controlled by Nikon Elements software, and image analysis was conducted using Metamorph software.

Recombinant Protein Expression, Purification, and Mass Determination—All *C. elegans* COPII components were amplified from a *C. elegans* cDNA library, and Sanger sequencing was used to confirm their identity. Proteins were expressed as His₆-SUMO fusions and purified using nickel-nitrilotriacetic acid-agarose resin in Sar1 buffer (25 mM HEPES, pH 7.2, 100 mM NaCl, and 1 mM MgCl₂). Sumo protease was used to remove the His₆-Sumo tag, and the cleaved proteins were subjected to size-exclusion chromatography, which was coupled to a Wyatt mini-DAWN TREOS three-angle light scattering detector and a Wyatt Optilab T-rEX refractive index detector. Data were collected at a flow rate of 0.5 ml/min and analyzed using ASTRA software to determine molecular mass (22). For *C. elegans* SAR-1, protein purification was carried out in the presence of either GDP or GTP (500 μM each). Specifically, a 30-fold molar excess of nucleotide was present during Sumo protease cleavage (16 h at 4 °C) to ensure complete incorporation onto SAR-1, as described previously (11). Furthermore, nucleotide-bound SAR-1 was additionally gel-filtered in the presence of the appropriate nucleotide (500 μM) prior to use in all assays. *C. elegans* SEC-23/SEC-24.2 and *Saccharomyces cerevisiae* Sec23p-Sec24p complexes were purified similarly, with the exception of the buffer used (25 mM HEPES, pH 7.2, 160 mM KOAc, and 1 mM MgCl₂). Human and yeast forms of Sar1 were purified as described previously (1).

Synthetic Liposome Generation and Size Determination—Liposomes were generated as described previously (13). Phospholipids in chloroform were mixed to generate lipid mixtures, including the “major/minor mix” (lipid, mol %) as follows: 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC), 50; 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE), 21; 1,2-dioleoyl-*sn*-glycero-3-phosphoserine (DOPS), 8; 1,2-dioleoyl-*sn*-glycero-3-phosphate (DOPA), 5; phosphatidylinositol (PI), 9; phosphatidylinositol 4-phosphate (PI4P), 2.2; phosphatidylinositol 4,5-bisphosphate (PI(4,5)P₂), 0.8; diacylglycerol (DAG), 2; 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-*N*-(7-nitro-2- β -1,3-benzoxadiazol-4-yl), 2, and supplemented with 20% cholesterol, which was shown previously to facilitate maximal binding of Sar1 isoforms (13). When *S. cerevisiae* proteins were utilized, 20% ergosterol was used in place of cholesterol. Lipids were dried, resuspended in Sar1 buffer, and subjected to extrusion through a nitrocellulose filter (Whatman) of a desired pore size. Liposomes were analyzed by dynamic light scattering using a Wyatt DynaPro NanoStar to determine their average diameters.

Stopped-flow Rapid Kinetic Analysis and Co-sedimentation Analysis—Stopped-flow rapid kinetic analysis and co-sedimentation experiments were carried out as described previously

Sar1 Senses Membrane Curvature

(23). Briefly, SAR-1 (8 μM) was incubated with GTP (100 μM) in Sar1 buffer before being injected into the stopped-flow chamber. An equal volume of Sar1 buffer containing a desired liposome dilution was simultaneously injected to initiate SAR-1 binding. Liposomes contained the following phospholipid composition (lipid, mol %): DOPC, 71; DOPE, 15; DOPS, 8; DOPA, 5; dansyl-PE, 1, and were supplemented with 20% cholesterol (major/minor mix lacking phosphoinositides and DAG). For co-sedimentation analysis, SAR-1 (4 μM) was incubated with 1 mM GMP-PNP in Sar1 buffer and supplemented with a desired liposome concentration. Liposomes were composed of 66% DOPC, 21% DOPE, 8% DOPS, 5% DOPA, and supplemented with 20% cholesterol. Samples were mixed in a TLA100 tube for 15 min at room temperature and then centrifuged at $100,000 \times g$ for 45 min. The supernatant was collected, mixed with sample buffer, and resolved on an SDS-polyacrylamide gel for densitometry to determine SAR-1 depletion.

Sar1 GTPase Activity Assays—The GTPase/GAP/GEF-Glo system (Promega) was used to measure the concentration of GTP remaining after incubation with Sar1 isoforms (24). Sar1 (3.1 μM , unless otherwise noted) was added to the reaction before the addition of GTP (5 μM) in a 25- μl reaction (Corning Costar 3912 solid white 96-well plate) in the supplied GEF buffer containing MgCl_2 . The addition of GTP initiated the reaction, which was conducted at 25 °C. After 2 h, an equal volume of the GTPase-Glo reagent was added to stop the reaction, incubated for 30 min, and followed by the addition of an equal volume of the detection reagent for 10 min. Luminescence was measured on a 96-well plate reader. All reactions were performed in triplicate and repeated independently on three separate occasions. Reactions with *S. cerevisiae* proteins were performed similarly, except that the initial reactions were carried out at 30 °C. Malachite Green GTPase assays were conducted by first incubating SAR-1^{GDP} with liposomes (1.27 mM) of a specific diameter for 10 min at room temperature. The GTPase reaction was initiated with the addition of GTP (300 μM) to a final reaction volume of 20 μl . In some cases, recombinant Sec23-Sec24 (10 μM , unless otherwise noted) was also present in the reaction prior to GTP addition. The reaction was allowed to continue for 1 h at room temperature. Malachite Green reagent (w/v %: Malachite Green, 0.027; poly(vinyl alcohol), 0.773; ammonium molybdate, 1.01) was added to the reaction to a final volume of 440 μl and allowed to incubate at room temperature for 2 h for the reaction with free inorganic phosphate to proceed to completion (indicated by a colorimetric change). After incubation with the Malachite Green reagent, the reaction was quenched with a sodium citrate solution (w/v %: 34; 50 μl) and incubated at room temperature for 5 min before absorbance measurements were taken (636 nm).

Fluorescence Quenching Assays—SAR1B^{W86F,W192F} (SAR1B Trp reporter) was engineered using site-directed mutagenesis and purified similarly to the wild type protein. Acrylamide quenching measurements were carried out as described previously (25). Briefly, a 2.0 M stock of acrylamide was dissolved in Sar1 buffer (with the addition of 1 mM DTT). The aqueous acrylamide quencher was added to the SAR1B-Trp reporter (1 μM) and supplemented with either 1 mM GDP or 1 mM GTP in the presence or absence of 2 mM lipids (mol %: DOPC, 75;

DOPE, 12; DOPS, 8; DOPA, 5) as indicated. Emission spectra were subtracted for blank, lipids, nucleotide, and acrylamide and integrated from 300 to 400 nm to calculate fluorescence intensity. The fluorescence intensity was plotted *versus* quencher concentration (0–140 mM), and the degree of quenching was analyzed using the Stern-Volmer equation (Equation 1),

$$F_0/F = 1 + K_{SV}[Q] \quad (\text{Eq. 1})$$

where F_0 and F are the fluorescence intensities in the absence and presence of acrylamide, respectively; K_{SV} is the Stern-Volmer constant for collisional quenching, and $[Q]$ is the concentration of aqueous quencher.

To measure the depth of penetration of the SAR1B Trp reporter into the membrane bilayer, parallax analysis was performed (25). Briefly, the fluorescence of the SAR1B-Trp reporter was individually measured in the presence of two separate membrane-embedded quenchers of distinct depths. The distance of the single Trp reporter from the bilayer center (Z_{CF}) was determined using parallax analysis as given by Equation 2,

$$Z_{CF} = L_{C1} + [-\ln(F_1/F_2)/\pi C - L^2]/2L \quad (\text{Eq. 2})$$

where L_{C1} represents the distance from the bilayer center to the shallow quencher; C is the mole fraction of the quencher divided by the lipid area; F_1 and F_2 are the relative fluorescence intensities of the shallow (6,7-Br₂-PC) and deep quenchers (11,12-Br₂-PC), respectively, and L is the difference in depth between the two quenchers. Brominated lipid concentrations were varied by substituting DOPC.

Atomic Force Microscopy—AFM imaging was performed using a Bruker Dimension FastScan instrument. All imaging was conducted under fluid using FastScan D cantilevers (Bruker). Their resonant frequencies under fluid were 110–140 kHz, and the actual scanning frequencies were ~5% below the maximal resonance peak. Lipid mixtures containing either a “basic mix” of lipids (70% DOPC, 15% DOPE, and 15% DOPS) or the major/minor mix (52% DOPC, 21% DOPE, 8% DOPS, 5% DOPA, 9% PI, 2.2% PI4P, 0.8% PI(4,5)P₂, 2% DAG, and supplemented with 20% cholesterol) were dried under nitrogen and hydrated in Biotechnology Performance Certified water overnight. Suspensions were probe-sonicated at an amplitude of 10 μA until the mixture became transparent. Liposomes were incubated in the presence or absence of proteins for 30 min and then placed on freshly cleaved mica. In each case, 40 μl of the liposome mixture and an equal volume of buffer (50 mM HEPES, pH 7.6, 100 mM NaCl, and 1 mM MgCl_2) were applied to the mica surface. The mica was washed three times with the same buffer and placed in the fluid cell of the atomic force microscope. The assembled lipid bilayer was immersed in 150 μl of buffer, and all imaging was performed at room temperature. For the buffer-exchange experiments, buffer containing 1 mM GTP, GMP-PNP, or GDP was introduced into the fluid cell of the microscope. AFM images were acquired at a rate of four frames/min and plane-fitted to remove tilt. Each scan line was fitted to a first-order equation.

Results

C. elegans Sar1 Regulates the Secretion of ER-derived Transport Carriers to Maintain Germ Line Tissue Architecture—The Sar1 GTPase plays a conserved role in protein and lipid secretion in several organisms, including yeast, *Drosophila*, and mammals (1, 26, 27). In humans, several distinct mutations affecting one of the Sar1 isoforms (SAR1B) have been implicated in chylomicron retention disease (Anderson disease), an early onset, inherited lipid malabsorption disorder characterized by hypocholesterolemia (28, 29). However, the contribution of Sar1 to other developmental processes, including tissue morphogenesis, remains largely uncharacterized. To investigate a potential role for Sar1 GTPases in this process and to verify that *C. elegans* SAR-1 regulates protein secretion from the ER, we conducted a series of depletion experiments using the *C. elegans* germ line as a model system. An important feature of the *C. elegans* reproductive system is its amenability to RNA interference (RNAi)-mediated depletion (30). Consistent with previous work, inhibition of the single Sar1 isoform expressed in worms (ZK180.4) results in potent sterility (Fig. 1A) (31). Further examination of germ line tissue depleted of SAR-1 revealed a dramatic defect in the organization of membranes that normally partition individual nuclei into compartments within the syncytial stem cell niche (Fig. 1B). A nearly identical phenotype was observed upon depletion of the COPII subunit SEC-23, suggesting that the defect was a consequence of inhibiting COPII-mediated secretion (Fig. 1B). Consistent with this idea, trafficking of integral membrane secretory cargoes, including the secretory vesicle-associated SNARE synaptobrevin (SNB-1), from the ER was potently blocked following depletion of SAR-1 or SEC-23 (Fig. 1C). These data highlight a conserved function for *C. elegans* SAR-1 and COPII in protein secretion and further demonstrate a key role for COPII-mediated vesicle transport during tissue morphogenesis.

Stable Membrane Penetration of Sar1 Is Dependent upon an Association with GTP—Because our findings indicated that *C. elegans* SAR-1 is functionally conserved, we decided to purify a recombinant form of the GTPase and analyze its membrane binding capabilities *in vitro*. For these experiments, nucleotide was loaded onto SAR-1 by purifying the protein in the presence of a large molar excess of GTP or GDP (500 μM), as described previously (11). Because previous work indicates that Sar1 GTPases exhibit minimal intrinsic activity in the absence of a guanine nucleotide-activating protein (GAP), we decided to use GTP in these experiments as opposed to a poorly hydrolysable analog (32). Size-exclusion chromatography and multiangle light scattering conducted in the presence of nucleotide (500 μM GTP or GDP) indicated that recombinant SAR-1 loaded with either GTP or GDP formed mainly monodisperse monomers in solution, with small populations of dimers and larger oligomers (Fig. 2, A–C, and Table 1). Similar results were obtained using recombinant forms of yeast Sar1p and human SAR1B (Fig. 2, A–C, and Table 1). These data indicate that Sar1 isoforms exist largely as monomers in solution, but they are also capable of self-association.

We next investigated the manner by which Sar1 binds to lipid bilayers. Although the amphipathic amino terminus of Sar1 is

known to play an essential role in membrane association and tubulation, relatively little is known about its dynamic properties. To address this issue, we used a technique that measures tryptophan fluorescence quenching by aqueous acrylamide, both in the presence and absence of membranes (33). For these experiments, we used a form of human SAR1B, which harbors only a single naturally occurring tryptophan reporter at position 7 within the amino-terminal amphipathic helix (SAR1B^{W86F,W192F}), enabling us to examine its accessibility. The protein was purified in the presence of GDP and subsequently incubated with a 100-fold molar excess of GTP for 30 min at room temperature, which was shown previously to be sufficient for nucleotide exchange (7). This form of SAR1B (1 μM) exhibited efficient quenching of fluorescence upon the addition of acrylamide, suggesting that the amphipathic helix was exposed in solution (Fig. 2, D and E, and Table 2). We also measured changes in tryptophan fluorescence when GDP-bound SAR1B^{W86F,W192F} was incubated with increasing concentrations of acrylamide. In contrast to the idea that the amphipathic helix is buried within the core of Sar1 when bound to GDP, we found that the addition of acrylamide resulted in equally efficient quenching with nearly identical kinetics to the GTP-bound protein (Fig. 2, D and E, and Table 2). These data argue that the amphipathic helix of Sar1 is solvent-exposed irrespective of its nucleotide-bound state.

When examined in the presence of membranes, we found that the GTP-bound form of SAR1B was highly resistant to tryptophan fluorescence quenching, suggesting that the amphipathic helix becomes buried within the hydrophobic core of the lipid bilayer under these conditions and is protected from acrylamide in solution (Fig. 2, D and E, and Table 2). Surprisingly, GDP-bound SAR1B also exhibited reduced fluorescence quenching when membranes were present, although the effect was not as pronounced as compared with GTP-bound SAR1B (Fig. 2, D and E, and Table 2). These data are consistent with the idea that SAR1B can associate with membranes, irrespective of the bound nucleotide. However, GTP-binding appears to promote more stable membrane penetration of the SAR1B amphipathic helix, which likely supports membrane bending during the formation of COPII-coated transport carriers.

To determine the extent to which the SAR1B amino-terminal amphipathic helix inserts into lipid bilayers, we measured tryptophan fluorescence quenching mediated by brominated phospholipids that are modified at specific positions along their acyl chains (34). Parallax analysis demonstrated that the single tryptophan reporter in the helix penetrates ~ 5 Å into the hydrophobic core of the outer leaflet, just below the hydrophilic headgroups, in the presence of GTP (Fig. 2, F and G; see “Experimental Procedures” for calculations). These data provide a mechanistic basis for the ability of Sar1 isoforms to bind and bend membranes.

Sar1 Is the Curvature-sensing Component of the COPII Complex—The formation of all transport carriers requires stepwise changes in local membrane curvature that ultimately result in membrane scission. We hypothesized that a curva-

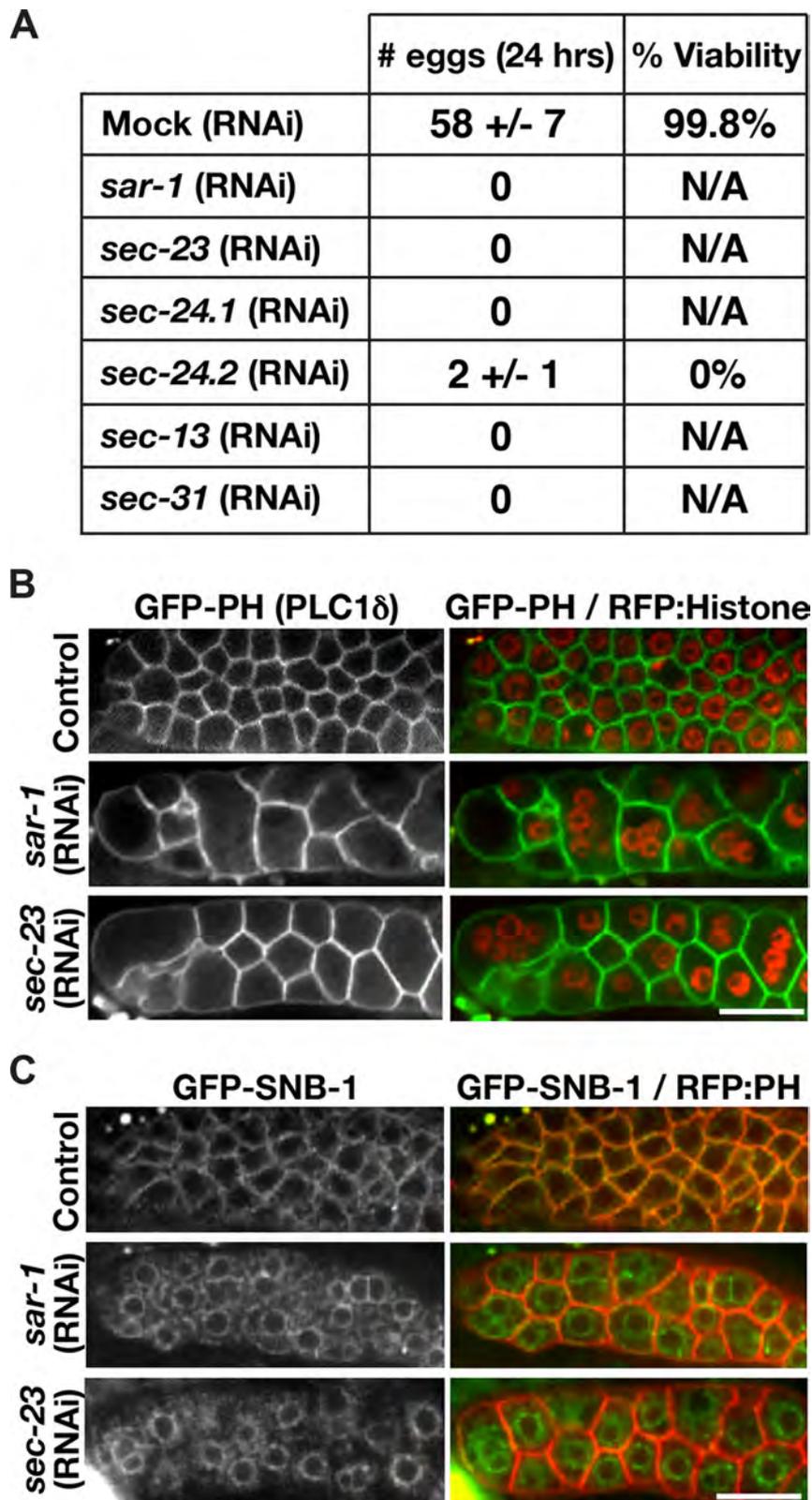


FIGURE 1. ***C. elegans* SAR-1 is functionally conserved.** *A*, effect of depleting each of the *C. elegans* COPII subunits was assessed by measuring the ability of treated animals to produce embryos during a 24-h period following exposure to various dsRNAs. Embryos produced following depletion of SEC-24.2 were osmotically sensitive and did not remain viable beyond the one-cell stage. *N/A*, not applicable. *B*, transgenic animals co-expressing a GFP fusion to the pleckstrin homology (PH) domain of PLC1 δ , which binds to phosphatidylinositol 4,5-bisphosphate present on the surface of membrane compartments throughout the germ line, and a red fluorescent protein (RFP) fusion to histone H2B (HIS-58) were treated with dsRNAs targeting COPII subunits for 24 h, and the distal portions of their germ lines were imaged 24 or 48 h later (for SEC-23 and SAR-1, respectively) using swept field confocal optics. Scale bar, 5 μ m. *C*, transgenic animals co-expressing a GFP fusion to SNB-1, an integral membrane protein produced in the ER and trafficked to the surface of compartments throughout the germ line, and a red fluorescent protein fusion to the PLC1 δ pleckstrin homology domain, were treated as described in *B*. Scale bar, 5 μ m.

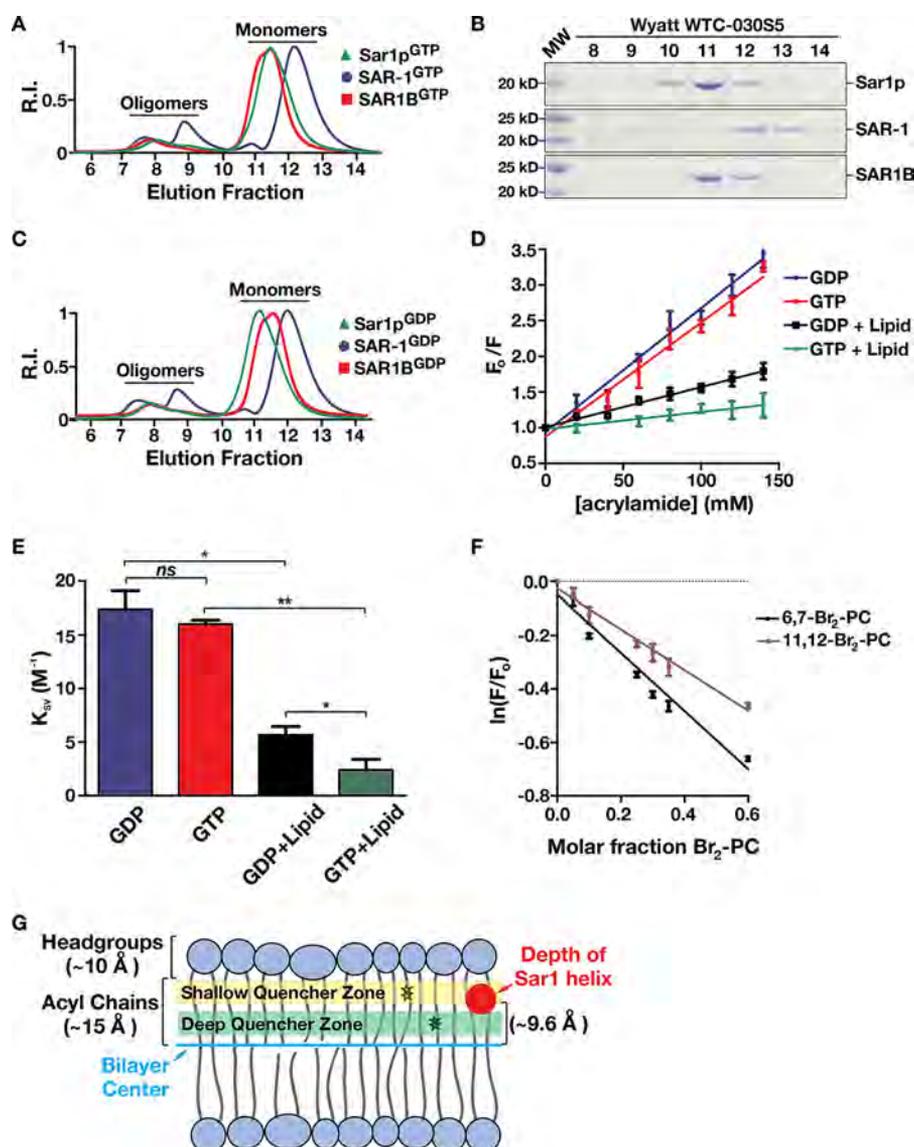


FIGURE 2. GTP facilitates stable Sar1 membrane penetration. *A*, refractometer traces from multiangle light scattering measurements of Sar1^{GTP} isoforms from various species are shown. Samples were applied onto a size-exclusion chromatography column equilibrated in 500 μM GTP before light-scattering measurements were taken. The relative intensities (*R.I.*) of fractions eluted from a Wyatt WTC-030S5 size-exclusion column subsequent to multiangle light scattering analysis were separated by SDS-PAGE analysis and stained with Coomassie. *B*, samples (Sar1^{GTP} isoforms) eluted from a Wyatt WTC-030S5 size-exclusion column subsequent to multiangle light scattering analysis were separated by SDS-PAGE analysis and stained with Coomassie. *C*, refractometer traces from multiangle light scattering measurements of Sar1^{GDP} isoforms from various species are shown. Samples were applied onto a size-exclusion chromatography column equilibrated in 500 μM GDP before light-scattering measurements were taken. The relative intensity values of fractions eluted are shown following normalization. *D*, various concentrations of aqueous acrylamide were added to SAR1B^{W86F,W192F} (1 μM) in the presence of either GDP alone, GTP alone, GDP with liposomes, or GTP with liposomes. Liposomes were composed of 66% DOPC, 21% DOPE, 8% DOPS, 5% DOPA and supplemented with 20% cholesterol. Tryptophan fluorescence emission spectra were recorded from 300 to 400 nm and integrated to determine fluorescence intensity. *E*, Stern-Volmer constant (K_{sv}) calculations reveal that the amphipathic tail of SAR1B^{GTP} is solvent-exposed when the GTPase is bound to either GDP or GTP. In the presence of liposomes, the amphipathic tail of SAR1B^{GTP} is embedded more stably in the membrane bilayer and protected from the aqueous quencher, as compared with SAR1B^{GDP}. The statistical significance of pairwise differences was calculated using a *t* test. *ns*, not significant; *, $p < 0.05$; **, $p < 0.01$. *F*, SAR1B^{GTP} was added to liposomes harboring an embedded membrane quencher (shallow: 6,7-Br₂-PC; deep: 11,12-Br₂-PC) at the indicated mol %. Tryptophan fluorescence emission spectra were recorded from 300 to 400 nm and integrated to determine fluorescence intensity. Only the 60 mol % liposome condition was used for parallax calculation analysis, which revealed that the SAR1B amphipathic tail penetrates to ~ 9.6 Å from the bilayer center. *G*, schematic illustrating the penetration depth of the SAR1B amphipathic helix relative to the positions of a shallow membrane quencher (6,7-Br₂-PC), a deep membrane quencher (11,12-Br₂-PC), and the center of the bilayer. Approximate sizes of the lipid headgroups and acyl chains are indicated.

TABLE 1

Molecular masses of monomeric Sar1 isoforms determined experimentally using size-exclusion chromatography-multiangle light scattering

Sar1 isoform	Predicted molecular mass (kDa) based on amino acid content	Experimentally determined mass (kDa) when bound to GDP	Experimentally determined mass (kDa) when bound to GTP
<i>S. cerevisiae</i> Sar1p	21.5	20.3 \pm 0.5	20.7 \pm 0.4
<i>C. elegans</i> SAR-1	21.7	21.6 \pm 0.4	21.5 \pm 0.4
<i>Homo sapiens</i> SAR1B	22.4	16.8 \pm 0.5	19.3 \pm 0.4

Sar1 Senses Membrane Curvature

TABLE 2

Stern-Volmer constants determined for the SAR1B amphipathic helix under various conditions

Condition	K_{SV} M^{-1}
GDP (in solution)	17.2 ± 1.0
GTP (in solution)	16.0 ± 0.9
GDP + liposomes	6.5 ± 0.5
GTP + liposomes	2.4 ± 0.5

ture-sensing mechanism may exist to facilitate budding of COPII-coated carriers from the ER. Based on its ability to bind and bend membranes, Sar1 is a logical candidate for acting as a curvature-sensitive module within the COPII complex. To test this idea, we used two independent approaches to explore the ability of Sar1 isoforms to discriminate between membranes of varying curvatures. First, we compared the binding affinities of SAR-1 for liposomes ranging in size from 105 to 225 nm in diameter, using a co-sedimentation assay. Chemically defined liposomes composed of 66% DOPC, 21% DOPE, 8% DOPS, 5% DOPA, and supplemented with 20% cholesterol, were utilized for these experiments as described previously (13). Because this assay was carried out under equilibrium conditions, we loaded SAR-1 (4 μ M) with the poorly hydrolysable GTP analog GMP-PNP to ensure that it remained in an active state throughout the experiment. To monitor binding, we mixed SAR-1 with different concentrations of liposomes and tracked depletion of the protein from the supernatants after centrifugation. The resulting titration curves indicated a 2-fold difference in the apparent dissociation constants between the smallest and largest liposomes, suggesting that SAR-1 prefers membranes of higher curvature (Fig. 3, A and B).

Because Sar1 is known to remodel membranes, which may influence the binding affinities determined following the prolonged incubation periods needed for sedimentation assays, we decided to use a second approach that measures the association between proteins and membranes instantaneously. Specifically, a rapid mixing fluorescence resonance energy transfer (FRET)-based assay was used to monitor the association of Sar1 with liposomes ranging in average size from 105 to 225 nm in diameter. Importantly, we took advantage of two native tryptophan residues within the *C. elegans* SAR-1 amino terminus, which could act as FRET donors for headgroup-dansylated phosphatidylethanolamine (PE) that was incorporated into liposomes. Using stopped-flow spectroscopy, GTP-loaded SAR-1 (8 μ M) was rapidly mixed with liposomes of varying diameters, enabling us to calculate relative binding affinities. Our data demonstrated again that SAR-1 binds preferentially to more highly curved membranes, exhibiting an \sim 4.6-fold higher affinity for 105-nm liposomes as compared with 225-nm liposomes (Fig. 3, C–E). The difference in apparent binding affinities determined using the two methods likely reflects the ability of SAR-1 to tubulate the larger liposomes over time, thereby increasing its affinity for the 225-nm liposomes during the 1-h co-sedimentation process. Taken together, these data strongly suggest that Sar1 binds to membranes in a curvature-sensitive manner.

Sar1 Remodels Supported Lipid Bilayers in a Manner Dependent on GTP Hydrolysis—Current approaches to study the role of Sar1 in membrane remodeling include the use of COPII budding assays in semi-intact cells followed by differential centrifugation or imaging of synthetic liposomes and GUVs by light and electron microscopy (14–16, 26, 35). In cell-free assays, contradictory findings argue both for and against the necessity of Sar1 GTP hydrolysis during cargo secretion from ER membranes (14, 26, 35). In some reports, poorly hydrolysable analogs of GTP potentially inhibit the formation of COPII transport carriers, whereas other studies have found that these analogs work equivalently to GTP (14, 26, 35). An explanation for these discrepancies remains unclear but may depend on sample handling. Nevertheless, these data suggest that cell-free assays for COPII budding exhibit substantial variability and are therefore difficult to interpret. EM-based studies using recombinant Sar1 and chemically defined liposomes or GUVs have yielded similarly conflicting results regarding the importance of GTP hydrolysis in membrane scission (7, 15, 18). Again, EM methods employed may be highly sensitive to sample manipulation. Although live imaging assays cause the least perturbation, the resolution afforded by this approach is typically limited and has only demonstrated that Sar1 tubulates membranes in the presence of GTP (16).

To study the action of Sar1 on lipid bilayers, we took an alternative approach that uses AFM imaging under fluid. In this method, Sar1 assembly was monitored at nanometer resolution on supported lipid bilayers (SLBs) assembled on mica, with very limited mechanical perturbation. We used SLBs composed of either a basic mix of lipids (70% phosphatidylcholine, 15% PE, and 15% phosphatidylserine) or the major/minor mix (52% DOPC, 21% DOPE, 8% DOPS, 5% DOPA, 9% PI, 2.2% PI4P, 0.8% PI(4,5)P₂, 2% DAG), which was shown previously to be important for maximal Sar1 membrane binding and facilitating Sec23–Sec24-mediated GAP activity on Sar1 (13). Gaps in the bilayer that formed spontaneously, irrespective of protein addition or lipid composition, produced edges with high curvature (Fig. 4A), enabling us to determine the distribution of Sar1 in the presence of both flat and highly curved membrane surfaces. We previously used a similar AFM-based assay to study the formation and activity of the ESCRT-III complex, another protein machinery implicated in membrane bending and scission, and demonstrated that it assembled in a curvature-dependent manner and promoted remodeling of the bent bilayer edges (36). Importantly, by imaging in a fast-scan mode, we were able to visualize the effects of various nucleotides, nucleotide analogs, and other proteins on Sar1 membrane association and remodeling.

Supported lipid bilayers composed of either the basic mix or the major/minor mix formed in the absence of proteins exhibited mild movements on the mica surface, indicating that they remained mobile throughout the period of AFM imaging (Fig. 4B and supplemental Video 1). Cross-sections showed that the height of the bilayer was largely uniform across flat regions but exhibited a steep decline at its edges where the mica was exposed (Fig. 4C). When supplemented with SAR-1 (400 nM) that had been purified in the presence of GDP, we observed the appearance of particles, which were distributed across the

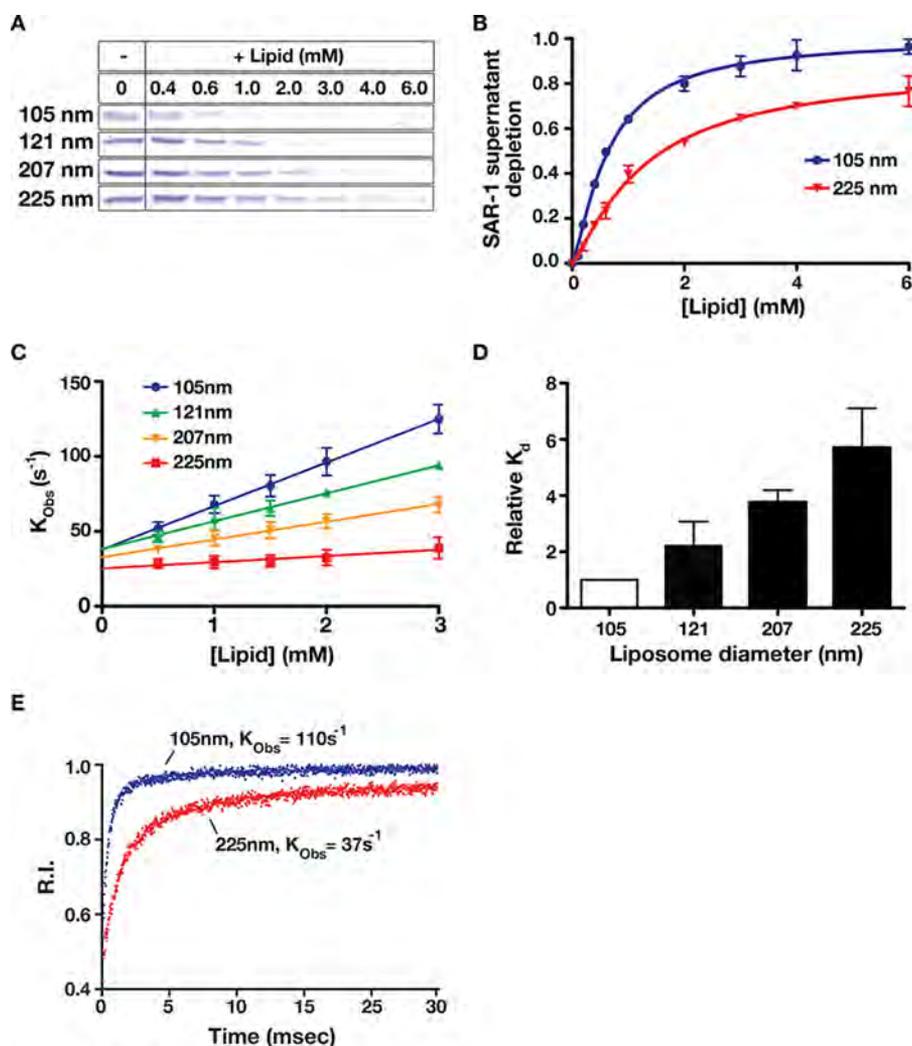


FIGURE 3. Sar1 preferentially binds to membranes that exhibit elevated curvature. *A*, SAR-1^{GMP-PNP} (4 μ M) was incubated in the presence of a range of liposome concentrations of varying diameter (105–225 nm) before being subjected to ultracentrifugation. Supernatants were collected after centrifugation, resolved by SDS-PAGE, and stained using Coomassie. *B*, densitometry measurements were used to determine relative depletion of SAR-1 from the supernatant, and these values were plotted against liposome concentration. *C*, SAR-1^{GTP} (8 μ M) was rapidly mixed with an equal volume of liposomes (composed of 71% DOPC, 15% DOPE, 8% DOPS, 5% DOPA, 1% dansyl-PE, and supplemented with 20% cholesterol) of varying diameters at the indicated concentrations using stopped-flow rapid kinetic analysis. Liposomes harboring the membrane-embedded dansyl moiety were used to measure SAR-1 fluorescence resonance energy transfer after solutions were mixed. K_{obs} for each condition was plotted against liposome concentration. *D*, dissociation constants (K_d) were measured for each liposome diameter described in *C* and divided by the calculated K_d of the 105 nm condition to determine the relative changes in apparent affinity between liposome conditions. As liposome diameter increases, K_d increases indicating lower binding affinities. *E*, representative stopped-flow data recorded upon mixing SAR-1 (8 μ M) with liposomes (3 mM) of two different sizes. Fluorescence resonance energy transfer between the tryptophan residues in the SAR-1 amphipathic helix and membrane-embedded dansyl was measured over 30 ms.

bilayer and at the periphery of the gaps that formed during bilayer assembly (Fig. 4*D*). Decoration of the bilayer edges was particularly evident when excessive force was applied onto the bilayer with the AFM tip (Fig. 4*E*). To verify that these particles corresponded to SAR-1, we varied the concentration of the protein used (100 nM to 1.6 μ M) and measured the number of particles observed by AFM. Our findings showed that increased levels of SAR-1 led to the appearance of more particles on the bilayer, which accumulated predominantly at the highly bent edges of the membranes (Fig. 4, *F–H*). Together, these data highlight our ability to monitor SAR-1 accumulation on supported lipid bilayers using AFM.

To facilitate nucleotide exchange onto SAR-1 in the absence of its GEF, a buffer containing 1 mM GTP (equivalent to \sim 2-fold the concentration found in most mammalian cell types) was flowed onto a bilayer composed of the basic mix of lipids.

Within moments, the bilayer began to undergo a remodeling process that resulted in its systematic removal from the mica surface, which we quantified by measuring the percentage of the mica surface covered with membrane over time (Fig. 4, *I* and *J*, and [supplemental Video 2](#)). Based on time-lapse AFM imaging, membrane removal appeared to occur from the highly curved edges of the bilayer (Fig. 4*I*), consistent with our data indicating that SAR-1 can sense membrane curvature. In contrast to the effect of SAR-1 at bent membranes, its presence on the flat regions of the bilayer failed to affect membrane topology under these conditions. The addition of GTP to protein-free bilayers had no substantial effect on their architecture as compared with buffer alone, and the addition of GDP to bilayers harboring SAR-1 similarly had no effect ([supplemental Videos 3 and 4](#)). Notably, bilayer removal mediated by SAR-1 was distinct from that of a detergent, which caused the membrane to

Sar1 Senses Membrane Curvature

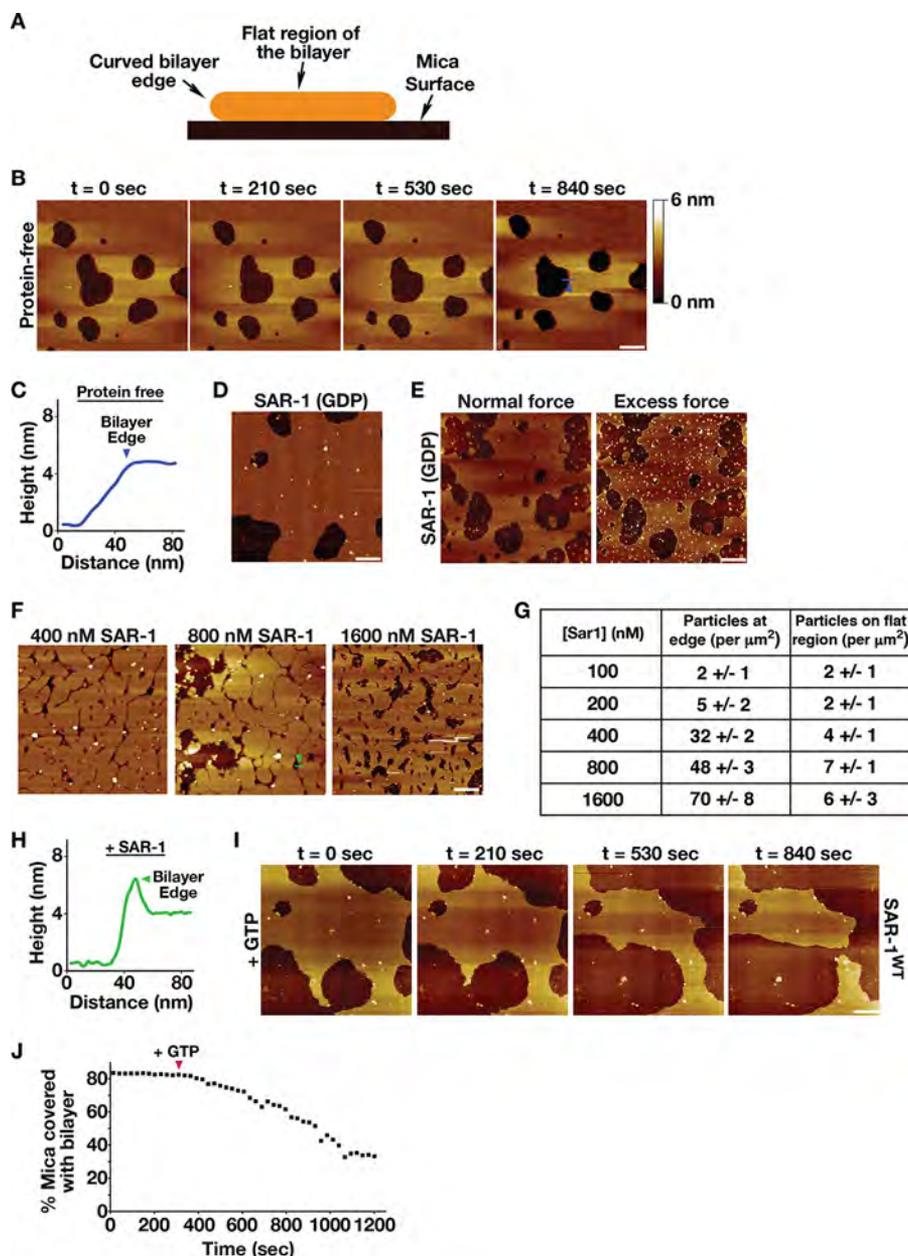


FIGURE 4. GTP hydrolysis is necessary for Sar1-mediated membrane remodeling. *A*, schematic illustrating a lipid bilayer (orange) assembled on a mica surface (black). The central portion of the bilayer is flat (~ 4 – 5 nm in thickness), and the edges of the bilayer are highly curved. *B*, representative AFM images of supported lipid bilayers (composed of the basic mix of lipids) acquired over time, following assembly in the absence of protein. A shade-height bar is shown on the right. Bar, 250 nm. *C*, height relative to the mica surface is plotted along the line shown in *B* (right panel, highlighted in blue). The position of the arrowhead (at the bilayer edge) along the line is also shown in *B*. *D*, representative AFM image of supported lipid bilayer (composed of the basic mix of lipids) assembled in the presence of SAR-1^{GDP} (400 nM). Bar, 250 nm. *E*, representative AFM images of the same supported lipid bilayer (composed of the basic mix of lipids), following assembly in the presence of SAR-1 (400 nM), imaged with normal (left) or excessive (right) force by the AFM tip. Bar, 250 nm. *F*, representative AFM images of supported lipid bilayers (composed of the basic mix of lipids) generated in the presence of different concentrations of SAR-1. Bar, 250 nm. *G*, table showing the number of particles (per μm^2) present at the bilayer edge or the flat portion of the bilayer generated in the presence of varying concentrations of SAR-1. *H*, height relative to the mica surface is plotted along the line shown in *F* (right panel, highlighted in green). The position of the arrowhead (at the bilayer edge) along the line is also shown in *F*, which is raised above the height of the bilayer surface suggesting the presence of protein. *I*, representative AFM images of supported lipid bilayers (composed of the basic mix of lipids) imaged over time, following assembly in the presence of SAR-1^{GDP} (400 nM) and supplementation with a buffer containing 1 mM GTP. The times shown are relative to the timing of GTP addition. Bar, 250 nm. *J*, representative plot of the area of the mica surface coated with membrane (assembled in the presence of SAR-1^{GDP}) over time. The timing of GTP addition is shown with an arrowhead.

dissociate from the mica surface in a manner that did not depend on membrane curvature (supplemental Video 5). These data support the idea that GTP-bound SAR-1 is capable of remodeling curved lipid bilayers.

To address the question of whether GTP hydrolysis is necessary for membrane remodeling and removal mediated by Sar1, SLBs composed of the basic mix and harboring SAR-1^{GDP} (400

nM) were exposed to an excess of GMP-PNP (1 mM). In contrast to the effect of GTP, we observed only a transient change in bilayer appearance, which was arrested within moments after addition of the nucleotide analog (Fig. 5, *A* and *B*, and supplemental Video 6). We also examined the ability of a mutant form of Sar1 (H75G), which harbors a point mutation within its catalytic switch II region and renders the protein

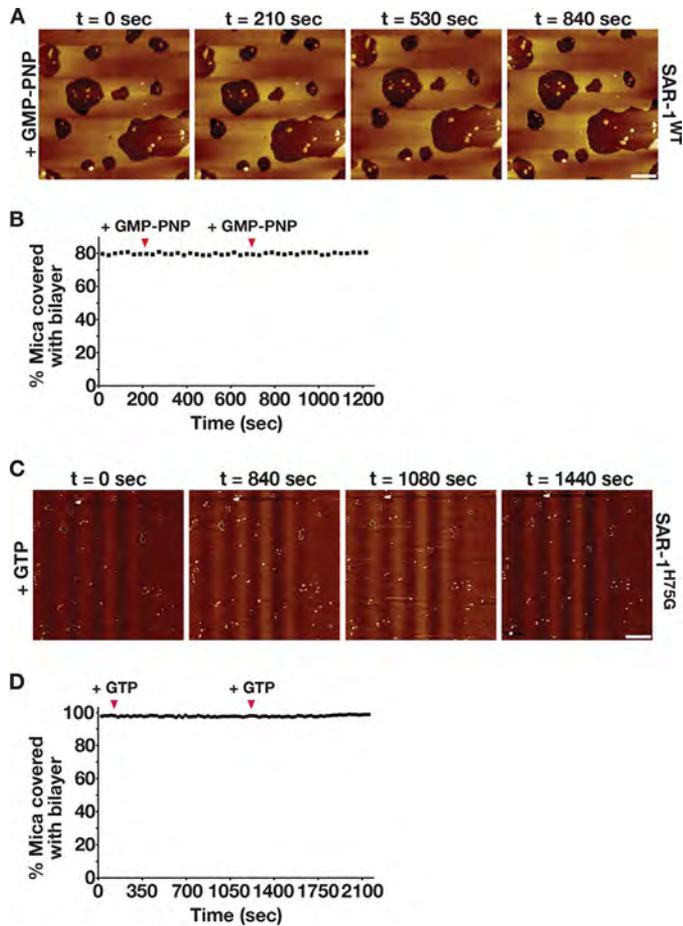


FIGURE 5. In the absence of GTP hydrolysis, SAR-1 is unable to remodel membranes. *A*, representative AFM images of supported lipid bilayers (composed of the basic mix of lipids) imaged over time, following assembly in the presence of SAR-1^{GDP} (1 μ M) and supplementation with a buffer containing 1 mM GMP-PNP. *Bar*, 250 nm. *B*, representative plot of the area of the mica surface coated with membrane (assembled in the presence of 1 μ M SAR-1^{GDP}) over time. The time points of GMP-PNP addition are highlighted with *arrowheads*. *C*, representative AFM images of supported lipid bilayers (composed of the basic mix of lipids) imaged over time, following assembly in the presence of SAR-1^{H75G} (1 μ M) and repeated supplementation with a buffer containing 1 mM GTP. *Bar*, 250 nm. *D*, representative plot of the area of the mica surface coated with membrane (assembled in the presence of 1 μ M SAR-1^{H75G}) over time. The timepoints of GTP addition are highlighted with *arrowheads*.

incapable of GTP hydrolysis at any concentration (37), to act on supported lipid bilayers. Although we found SAR-1^{H75G} was highly enriched along membrane edges, the addition of GTP (1 mM) did not stimulate membrane removal, even upon repeated exposure (Fig. 5, *C* and *D*, and [supplemental Video 7](#)). Together, these data argue that GTP hydrolysis is necessary for the continuous function of SAR-1 to remodel and sever lipid bilayers.

To further confirm the importance of GTP hydrolysis for SAR-1-mediated membrane remodeling, we also examined the effect of adding *C. elegans* Sec23-Sec24 (500 nM), the SAR-1 GAP, together with GTP (1 mM) to bilayers. For these studies, we used membranes composed of the major/minor mix, which was shown previously to be necessary for efficient Sec23-Sec24-mediated GAP activity on Sar1 (13). Similar to our findings using bilayers composed of the basic mix, the addition of GTP alone to the major/minor mix membranes containing SAR-1 initiated its steady removal (Fig. 6, *A* and *B*, and [supplemental Video 8](#)), although the kinetics of the process was slower and multiple addi-

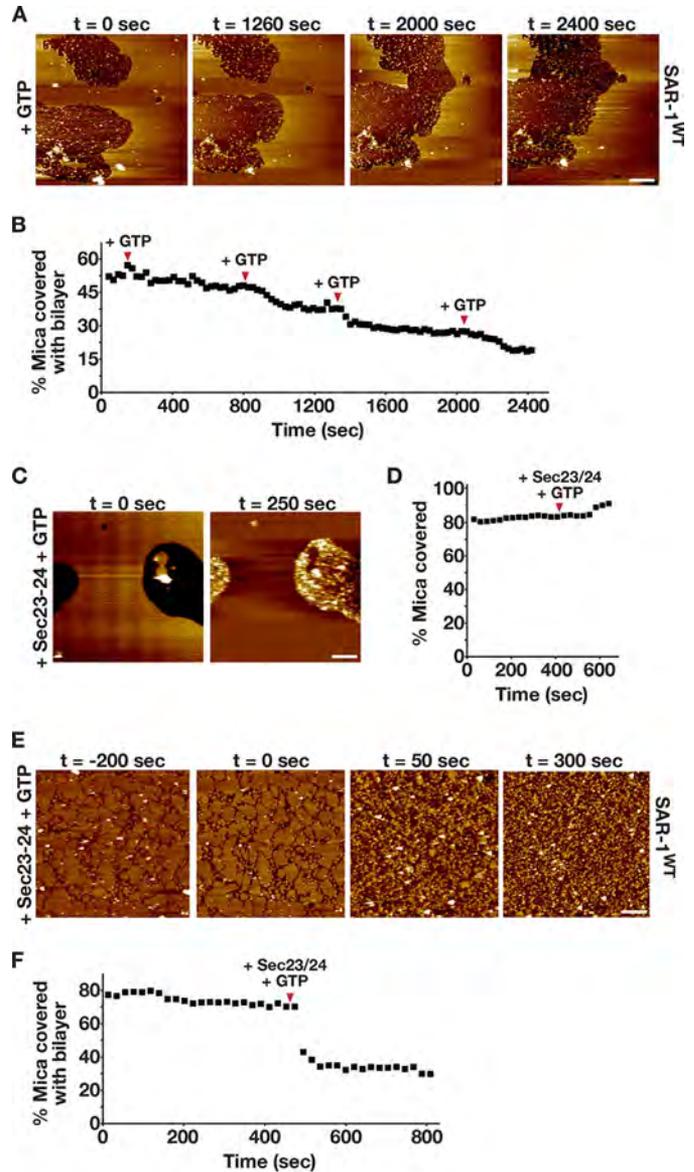


FIGURE 6. GAP activity on SAR-1 dramatically accelerates the kinetics of lipid bilayer remodeling. *A*, representative AFM images of supported lipid bilayers (composed of the minor/minor mix of lipids) imaged over time, following assembly in the presence of SAR-1^{GDP} (400 nM) and supplementation with a buffer containing 1 mM GTP. *Bar*, 250 nm. *B*, representative plot of the area of the mica surface coated with membrane (assembled in the presence of 400 nM SAR-1^{GDP}) over time. The time points of GTP addition are highlighted with *arrowheads*. *C*, representative AFM images of supported lipid bilayers (composed of the major/minor mix of lipids) imaged before and after the addition of purified Sec23-Sec24 (500 nM) and GTP (1 mM). *Bar*, 250 nm. *D*, representative plot of the area of the mica surface coated with membrane and protein (assembled in the presence of 400 nM SAR-1^{GDP}) over time. The time point of Sec23-Sec24 and GTP addition is highlighted with an *arrowhead*. *E*, representative AFM images of supported lipid bilayers (composed of the major/minor mix of lipids) imaged over time, following assembly in the presence of 400 nM SAR-1^{GDP} and supplementation with a buffer containing 1 mM GTP and 500 nM Sec23-Sec24 (at the 0-s time point). *Bar*, 250 nm. *F*, representative plot of the area of the mica surface coated with membrane and protein (assembled in the presence of 400 nM SAR-1^{GDP}) over time. The time point of GTP and Sec23-Sec24 addition is highlighted with an *arrowhead*.

tions of GTP were required. The addition of Sec23-Sec24 and GTP to bilayers lacking SAR-1 had no effect on the membrane (Fig. 6, *C* and *D*, and [supplemental Video 9](#)), consistent with the inability of Sec23-Sec24 to bind lipid bilayers without SAR-1 (13). However, we observed substantial, nonspecific accumulation of

Sar1 Senses Membrane Curvature

Sec23-Sec24 heterodimers on the mica surface (Fig. 6, C and D, and supplemental Video 9). We next added Sec23-Sec24 and GTP to major/minor mix bilayers containing SAR-1 and observed an extremely rapid removal of the membrane (Fig. 6, E and F, and supplemental Video 10). These data are again consistent with the idea that GTP hydrolysis on SAR-1, which is stimulated by Sec23-Sec24, is critical for promoting membrane remodeling by SAR-1.

Membrane Curvature Regulates Sar1 GTP Hydrolysis—Given the ability of SAR-1 to bind preferentially to membranes of high curvature and remodel them in a manner that depends on GTP hydrolysis, we questioned whether curvature may also influence the rate of GTP hydrolysis on Sar1. To examine this possibility, we employed a new, facile, and highly reproducible bioluminescence assay for monitoring the enzymatic activity of Sar1 in the presence of liposomes that ranged in average size from 105 to 225 nm in diameter. To first validate the use of this assay, we measured Sar1-mediated GTP hydrolysis over a range of protein concentrations. These data highlight the ability of SAR-1 to hydrolyze GTP in solution, but only at relatively high concentrations (above 4 μM) (Fig. 7A). The mutant form of SAR-1 (H75G) was largely incapable of GTP hydrolysis at any concentration, as predicted by previous work (Fig. 7A). Additionally, the supplementation of the reaction with recombinant Sec23-Sec24 dramatically increased the rate of GTP hydrolysis by wild type SAR-1 but not SAR-1^{H75G} (Fig. 7A). These findings collectively indicate that the bioluminescent assay provides a faithful readout of Sar1 GTPase activity.

To test whether membrane curvature affects Sar1 GTPase activity, we incubated nucleotide-free SAR-1 (3.1 μM) with Sec23-Sec24 heterodimers (255 nm) and liposomes (1 μM) of various curvatures and initiated each reaction by the addition of 5 μM GTP. Based on our sedimentation assays, equivalent amounts of SAR-1 are bound to each liposome population under these conditions, irrespective of their diameter (Fig. 3, A and B). Our studies indicated that GTPase activity was inversely related to liposome size, with the 105-nm liposomes promoting SAR-1 activity significantly better than liposomes greater than 200 nm in diameter (Fig. 7B). Additionally, by comparing these data to GTPase activity measured in the absence of membranes, we found that the largest liposomes used (225 nm in diameter) failed to significantly stimulate GTP hydrolysis on SAR-1, in contrast to the smaller liposomes (Fig. 7B). We also confirmed that SAR-1 exhibits curvature sensitivity in the presence of Sec23-Sec24 using a standard colorimetric assay that measures inorganic phosphate release as a function of GTPase activity (Fig. 7C). To further verify that SAR-1, and not its GAP, exhibits curvature sensitivity, we conducted similar experiments using SAR-1 alone. To detect inorganic phosphate release, a higher concentration of SAR-1 was necessary (14 μM). Under these conditions, we again found that GTP hydrolysis on SAR-1 was increased in the presence of small liposomes (105 nm) as compared with large liposomes (225 nm) (Fig. 7D).

To determine whether this curvature-sensitive property of SAR-1 is conserved, we also measured the ability of yeast Sar1p to hydrolyze GTP in the presence of differently sized liposomes. Strikingly, the impact of curvature was even more pronounced, with a significant difference in Sar1p GTPase activity in the

presence of 105-nm liposomes as compared with 225-nm liposomes (Fig. 7E). These data highlight a new regulatory mechanism that controls Sar1 function and demonstrates the role of membrane curvature in influencing GTP hydrolysis on Sar1. Together, our findings suggest that the various curvatures adopted during individual steps of COPII-mediated transport carrier budding directly influence Sar1 distribution and activity.

Discussion

The amino-terminal region of Sar1 has been shown to play an essential role in membrane deformation and generating COPII-coated transport carriers, both *in vivo* and *in vitro* (7, 8, 38). However, the mechanisms by which this region binds and manipulates membranes remain unclear. To date, the intact domain has not been crystallographically resolved, but it is predicted to form an amphipathic α -helix and is capable of destabilizing lipid bilayers on its own (39). Models suggest that GTP binding by Sar1 relieves it from an autoinhibited state and enables exposure of its amino terminus to facilitate membrane association (8). In contrast, our data suggest that the amphipathic helix transitions constantly through solvent-exposed states, irrespective of GTP binding. However, GTP loading promotes stable membrane penetration, which is necessary to generate curvature. Thus, our results are consistent with the idea that Sar1 continually samples membranes, but only in the presence of its GEF Sec12 can it stably insert to facilitate the formation of COPII transport carriers.

In addition to its role during the initial stages of COPII carrier formation, Sar1 has also been implicated in membrane sculpting and scission steps (7, 8). Recent work highlighted the propensity of Sar1 to dimerize and form ordered lattices on membrane tubules exhibiting diameters larger than 60 nm, equivalent to the size of the smallest COPII-coated transport carriers observed *in vitro* or *in vivo* (3, 15). These data are consistent with our results demonstrating that Sar1 can form dimers in solution. In cases where membrane tubules fall below 60 nm in diameter, ordered arrays of Sar1 are no longer observed, implying that curvature plays an important role in Sar1 membrane association (15). We explored this idea further by examining the relative binding affinities of Sar1 for a range of membrane curvatures, which COPII transport carriers likely adopt during their formation. Our findings indicate that Sar1 associates more tightly with membranes of elevated curvature. Importantly, discrepancies in the relative affinities obtained using equilibrium binding and stopped-flow rapid kinetics further suggest that Sar1 actively bends the membrane, diminishing lipid density in the outer leaflet as compared with the inner leaflet and facilitating membrane penetration of additional molecules of Sar1. A localized increase in Sar1 concentration is likely instrumental in achieving sufficient bending, potentially through molecular crowding, to ultimately drive membrane scission (39).

In many ways, Sar1 appears to act similarly to epsin, a regulator of clathrin-coated vesicle formation, which harbors an ENTH domain that penetrates the inner leaflet of the plasma membrane to a depth similar to what we have measured for the Sar1 amphipathic helix (40–43). This type of shallow penetra-

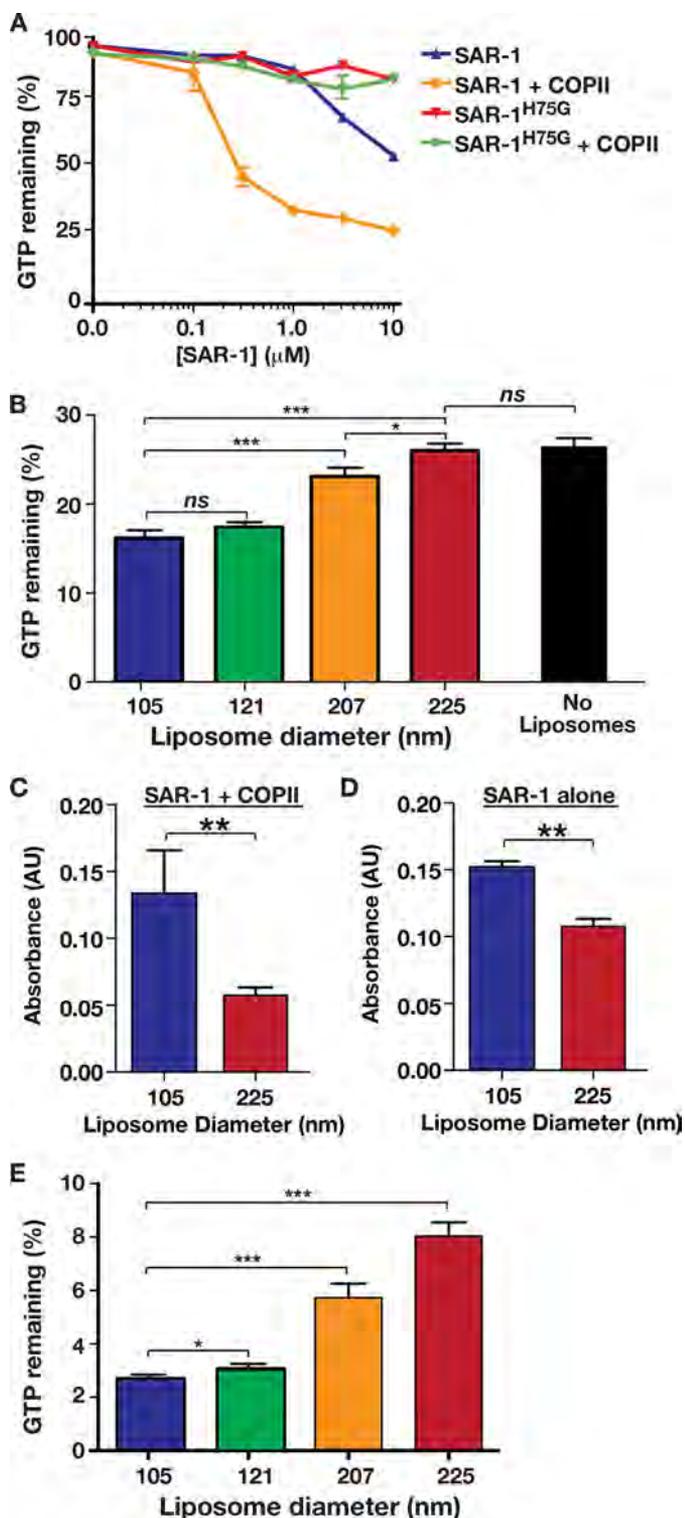


FIGURE 7. Membrane curvature regulates GTP hydrolysis on Sar1. A, GTP hydrolysis on SAR-1 was measured using a bioluminescence assay, in the presence or absence of additional COPII subunits and GTP. B, GTP hydrolysis on *C. elegans* SAR-1 (3.1 μM) was measured using a bioluminescence assay in the presence of the inner COPII coat components (255 nm) and liposomes composed of the major/minor mix lipids (1 μM) of varying diameter. For comparison, GTP hydrolysis on SAR-1 (3.1 μM) was also measured in the presence of inner COPII coat subunits (255 nm) but in the absence of liposomes (right bar). The statistical significance of pairwise differences was calculated using a *t* test. *ns*, not significant; *, $p < 0.05$; ***, $p < 0.001$. C, GTP hydrolysis on SAR-1 (14 μM) was measured in the presence of two differently sized liposomes composed of the major/minor mix lipids (1.27 μM) and Sec23-Sec24 (10 μM) using a Malachite Green phosphate assay. The statistical significance of the

tion leads to a major asymmetry in the structure of a membrane monolayer, enabling positive curvature generation (44). Although controversial, the epsin ENTH domain has been proposed to have an active role in the vesicle scission process during endocytosis (45, 46). Analogously, by acting as the curvature sensor during COPII transport carrier formation, Sar1 may possess a similar function. Consistent with this idea, at sufficiently high concentrations, Sar1 is capable of transforming membrane tubules into vesicles independently of GTP hydrolysis (15). However, it remains unclear whether the concentration of Sar1 necessary for this effect *in vitro* is physiologically relevant.

Although the current situation is confusing, abundant evidence now exists to support both active and passive roles for Sar1 GTP hydrolysis in generating transport carriers from the ER (1, 7, 8, 13–18). However, fission reactions that occur under conditions where the GTPase cycle is halted typically require high levels of Sar1 to be present, raising concerns. To address this issue, we took advantage of an *in vitro* assay to study the impact of various nucleotides and nucleotide analogs on Sar1-mediated membrane remodeling using a limited concentration of the protein (400 nM). Irrespective of the artificial nature of these experiments, our combined use of supported lipid bilayers and AFM imaging enabled us to visualize the effect of Sar1 under various conditions in real time. Our findings demonstrated that GTP hydrolysis is necessary for membrane remodeling mediated by Sar1, suggesting a requirement to expend energy during this process. These data were particularly surprising in light of our results showing that the intrinsic rate of Sar1 GTP hydrolysis in the absence of its GAP is extremely slow. Nevertheless, this minimal amount of GTP consumption appears to be required for Sar1 to perform work on supported lipid bilayers, as neither GTP-locked Sar1 (H75G) nor wild type Sar1 in the presence of poorly hydrolysable GTP analogs were able to mimic the effect of adding GTP to wild type Sar1 assembled on the edges of membranes.

One possible explanation for these findings stems from our work showing that GTP hydrolysis on Sar1 is enhanced in the presence of membranes with elevated curvature. Although the differences in GTPase activity that we measured may appear modest, we were only able to examine effects over a narrow range of curvatures (~ 0.004 – 0.009 nm^{-1} , where curvature is defined as $1/\text{radius}$). The edges of supported lipid bilayers are predicted to exhibit the highest degree of curvature feasible ($\sim 0.4 \text{ nm}^{-1}$), which appears to promote substantial GTP hydrolysis on Sar1, sufficient to remodel the bilayer even in the absence of its GAP. An analogous situation may arise during the fission of COPII-coated transport carriers from the ER. Over the course of this final step, the ER membrane must achieve a

pairwise difference was calculated using a *t* test. **, $p < 0.01$. D, GTP hydrolysis on SAR-1 (14 μM) was measured in the presence of two differently sized liposomes composed of the major/minor mix lipids (1.27 μM) using a Malachite Green phosphate assay. The statistical significance of the pairwise difference was calculated using a *t* test. **, $p < 0.01$. E, GTP hydrolysis on *S. cerevisiae* Sar1p (3.1 μM) was measured using a bioluminescent assay in the presence of the inner COPII coat components (255 nm) and liposomes composed of the major/minor mix lipids (1 μM) of varying diameter. The statistical significance of pairwise differences was calculated using a *t* test. *ns*, not significant; *, $p < 0.05$; ***, $p < 0.001$.

Sar1 Senses Membrane Curvature

high degree of curvature, significantly higher than that of vesicles we can generate by extrusion, which may be sufficient to promote GTP hydrolysis on Sar1 and thereby disrupt the bilayer to enable fission (3). The role of Sar1 GAP activity must also be considered, especially in light of its dramatic enhancement of the kinetics of SAR-1-mediated membrane remodeling. It is feasible that Sec23 behaves similarly to ArfGAP1, the GAP that stimulates GTP hydrolysis on the Arf1 GTPase in a curvature-dependent manner and directs COPI-mediated trafficking (47, 48). However, unlike ArfGAP1, which harbors an ArfGAP1 lipid packing sensor motif that inserts into one leaflet of highly bent bilayers, none of the COPII subunits bind directly to membranes (13). Thus, our data are more consistent with the idea that Sar1, and not its GAP, acts as the curvature sensor in COPII-mediated membrane transport.

Taken together, our findings support a model in which Sec12-mediated GTP loading onto Sar1 allows for stable membrane penetration of its amphipathic tail to initiate membrane curvature on subdomains of the ER and recruitment of additional coat components. A number of regulatory factors, including Sec16, likely contribute to the control of Sec23-dependent GAP activity to limit release of Sar1 from the membrane (49, 50). Ultimately, we speculate that the progressive ability of Sar1 to bend membranes, likely in coordination with COPII coat assembly, leads to an increased propensity for GTP hydrolysis, which is able to support membrane fission and the release of coated transport carriers.

Author Contributions—M. G. H., I. M., R. M. H., E. R. C., J. M. E., and A. A. conceived and designed the experiments. M. G. H., I. M., L. W., E. R. C., J. M. E., and A. A. performed the experiments and analyzed the data. M. G. H. and A. A. wrote the paper.

Acknowledgments—We thank Elizabeth Miller for providing purified yeast COPII components, Subhanjan Mondal and Said Goueli at Promega Corp., for providing us access to the GTPase-Glo system ahead of release, and members of the Audhya laboratory for critically reading this manuscript.

References

1. Barlowe, C., Orci, L., Yeung, T., Hosobuchi, M., Hamamoto, S., Salama, N., Rexach, M. F., Ravazzola, M., Amherdt, M., and Schekman, R. (1994) COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. *Cell* **77**, 895–907
2. Brandizzi, F., and Barlowe, C. (2013) Organization of the ER-Golgi interface for membrane traffic control. *Nat. Rev. Mol. Cell Biol.* **14**, 382–392
3. D'Arcangelo, J. G., Stahmer, K. R., and Miller, E. A. (2013) Vesicle-mediated export from the ER: COPII coat function and regulation. *Biochim. Biophys. Acta* **1833**, 2464–2472
4. Venditti, R., Wilson, C., and De Matteis, M.A. (2014) Exiting the ER: what we know and what we don't. *Trends Cell Biol.* **24**, 9–18
5. d'Enfert, C., Wuestehube, L. J., Lila, T., and Schekman, R. (1991) Sec12p-dependent membrane binding of the small GTP-binding protein Sar1p promotes formation of transport vesicles from the ER. *J. Cell Biol.* **114**, 663–670
6. Futai, E., Hamamoto, S., Orci, L., and Schekman, R. (2004) GTP/GDP exchange by Sec12p enables COPII vesicle bud formation on synthetic liposomes. *EMBO J.* **23**, 4146–4155
7. Lee, M. C., Orci, L., Hamamoto, S., Futai, E., Ravazzola, M., and Schekman, R. (2005) Sar1p N-terminal helix initiates membrane curvature and completes the fission of a COPII vesicle. *Cell* **122**, 605–617
8. Bielli, A., Haney, C. J., Gabreski, G., Watkins, S. C., Bannykh, S. I., and Aridor, M. (2005) Regulation of Sar1 NH₂ terminus by GTP binding and hydrolysis promotes membrane deformation to control COPII vesicle fission. *J. Cell Biol.* **171**, 919–924
9. Montegna, E. A., Bhave, M., Liu, Y., Bhattacharyya, D., and Glick, B. S. (2012) Sec12 binds to Sec16 at transitional ER sites. *PLoS ONE* **7**, e31156
10. Bi, X., Corpina, R. A., and Goldberg, J. (2002) Structure of the Sec23/24-Sar1 pre-budding complex of the COPII vesicle coat. *Nature* **419**, 271–277
11. Bi, X., Mancias, J. D., and Goldberg, J. (2007) Insights into COPII coat nucleation from the structure of Sec23.Sar1 complexed with the active fragment of Sec31. *Dev. Cell* **13**, 635–645
12. Stagg, S. M., LaPointe, P., Razvi, A., Gürkan, C., Potter, C. S., Carragher, B., and Balch, W. E. (2008) Structural basis for cargo regulation of COPII coat assembly. *Cell* **134**, 474–484
13. Matsuoka, K., Orci, L., Amherdt, M., Bednarek, S. Y., Hamamoto, S., Schekman, R., and Yeung, T. (1998) COPII-coated vesicle formation reconstituted with purified coat proteins and chemically defined liposomes. *Cell* **93**, 263–275
14. Adolf, F., Herrmann, A., Hellwig, A., Beck, R., Brügger, B., and Wieland, F. T. (2013) Scission of COPI and COPII vesicles is independent of GTP hydrolysis. *Traffic* **14**, 922–932
15. Harii, H., Bhattacharya, N., Johnson, K., Noble, A. J., and Stagg, S. M. (2014) Insights into the mechanisms of membrane curvature and vesicle scission by the small GTPase Sar1 in the early secretory pathway. *J. Mol. Biol.* **426**, 3811–3826
16. Daum, S., Krüger, D., Meister, A., Auerswald, J., Prinz, S., Briggs, J. A., and Bacia, K. (2014) Insights from reconstitution reactions of COPII vesicle formation using pure components and low mechanical perturbation. *Biol. Chem.* **395**, 801–812
17. Bannykh, S. I., Rowe, T., and Balch, W. E. (1996) The organization of endoplasmic reticulum export complexes. *J. Cell Biol.* **135**, 19–35
18. Bacia, K., Futai, E., Prinz, S., Meister, A., Daum, S., Glatte, D., Briggs, J. A., and Schekman, R. (2011) Multibudded tubules formed by COPII on artificial liposomes. *Sci. Rep.* **1**, 17
19. Shi, A., Chen, C. C., Banerjee, R., Glodowski, D., Audhya, A., Rongo, C., and Grant, B. D. (2010) EHB1 functions with RAB-10 during endocytic recycling in *Caenorhabditis elegans*. *Mol. Biol. Cell* **21**, 2930–2943
20. Green, R. A., Kao, H. L., Audhya, A., Arur, S., Mayers, J. R., Fridolfsson, H. N., Schulman, M., Schloissnig, S., Niessen, S., Laband, K., Wang, S., Starr, D. A., Hyman, A. A., Schedl, T., Desai, A., et al. (2011) A high-resolution *C. elegans* essential gene network based on phenotypic profiling of a complex tissue. *Cell* **145**, 470–482
21. Wang, L., and Audhya, A. (2014) *In vivo* imaging of *C. elegans* endocytosis. *Methods* **68**, 518–528
22. Schuh, A. L., Hanna, M., Quinney, K., Wang, L., Sarkeshik, A., Yates, J. R., 3rd, and Audhya, A. (2015) The VPS-20 subunit of the endosomal sorting complex ESCRT-III exhibits an open conformation in the absence of upstream activation. *Biochem. J.* **466**, 625–637
23. Hui, E., Gaffaney, J. D., Wang, Z., Johnson, C. P., Evans, C. S., and Chapman, E. R. (2011) Mechanism and function of synaptotagmin-mediated membrane apposition. *Nat. Struct. Mol. Biol.* **18**, 813–821
24. Mondal, S., Hsiao, K., and Goueli, S. A. (2015) A homogenous bioluminescent system for measuring GTPase, GTPase activating protein, and guanine nucleotide exchange factor activities. *Assay Drug Dev. Technol.* **13**, 444–455
25. Bai, J., Earles, C. A., Lewis, J. L., and Chapman, E. R. (2000) Membrane-embedded synaptotagmin penetrates cis or trans target membranes and clusters via a novel mechanism. *J. Biol. Chem.* **275**, 25427–25435
26. Ivan, V., de Voer, G., Xanthakis, D., Spoorendonk, K. M., Kondylis, V., and Rabouille, C. (2008) *Drosophila* Sec16 mediates the biogenesis of tER sites upstream of Sar1 through an arginine-rich motif. *Mol. Biol. Cell* **19**, 4352–4365
27. Kuge, O., Dascher, C., Orci, L., Rowe, T., Amherdt, M., Plutner, H., Ravazzola, M., Tanigawa, G., Rothman, J. E., and Balch, W. E. (1994) Sar1 promotes vesicle budding from the endoplasmic reticulum but not Golgi compartments. *J. Cell Biol.* **125**, 51–65
28. Jones, B., Jones, E. L., Bonney, S. A., Patel, H. N., Mensenkamp, A. R.,

- Eichenbaum-Voline, S., Rudling, M., Myrdal, U., Annesi, G., Naik, S., Meadows, N., Quattrone, A., Islam, S. A., Naoumova, R. P., Angelin, B., *et al.* (2003) Mutations in a Sar1 GTPase of COPII vesicles are associated with lipid absorption disorders. *Nat. Genet.* **34**, 29–31
29. Siddiqi, S., Saleem, U., Abumrad, N. A., Davidson, N. O., Storch, J., Siddiqi, S. A., and Mansbach, C. M., 2nd. (2010) A novel multiprotein complex is required to generate the prechylomicron transport vesicle from intestinal ER. *J. Lipid Res.* **51**, 1918–1928
 30. Oegema, K., and Hyman, A. A. (2006) Cell division. *WormBook* **19**, 1–40
 31. Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D. P., Zipperlen, P., and Ahringer, J. (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237
 32. Yoshihisa, T., Barlowe, C., and Schekman, R. (1993) Requirement for a GTPase-activating protein in vesicle budding from the endoplasmic reticulum. *Science* **259**, 1466–1468
 33. Tallmadge, D. H., Huebner, J. S., and Borkman, R. F. (1989) Acrylamide quenching of tryptophan photochemistry and photophysics. *Photochem. Photobiol.* **49**, 381–386
 34. Bolen, E. J., and Holloway, P. W. (1990) Quenching of tryptophan fluorescence by brominated phospholipid. *Biochemistry* **29**, 9638–9643
 35. Schindler, A. J., and Schekman, R. (2009) *In vitro* reconstitution of ER-stress induced ATF6 transport in COPII vesicles. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17775–17780
 36. Fyfe, I., Schuh, A. L., Edwardson, J. M., and Audhya, A. (2011) Association of the endosomal sorting complex ESCRT-II with the Vps20 subunit of ESCRT-III generates a curvature-sensitive complex capable of nucleating ESCRT-III filaments. *J. Biol. Chem.* **286**, 34262–34270
 37. Oka, T., Nishikawa, S., and Nakano, A. (1991) Reconstitution of GTP-binding Sar1 protein function in ER to Golgi transport. *J. Cell Biol.* **114**, 671–679
 38. Long, K. R., Yamamoto, Y., Baker, A. L., Watkins, S. C., Coyne, C. B., Conway, J. F., and Aridor, M. (2010) Sar1 assembly regulates membrane constriction and ER export. *J. Cell Biol.* **190**, 115–128
 39. Stachowiak, J. C., Schmid, E. M., Ryan, C. J., Ann, H. S., Sasaki, D. Y., Sherman, M. B., Geissler, P. L., Fletcher, D. A., and Hayden, C. C. (2012) Membrane bending by protein-protein crowding. *Nat. Cell Biol.* **14**, 944–949
 40. Ford, M. G., Mills, I. G., Peter, B. J., Vallis, Y., Praefcke, G. J., Evans, P. R., and McMahon, H. T. (2002) Curvature of clathrin-coated pits driven by epsin. *Nature* **419**, 361–366
 41. Itoh, T., Koshihara, S., Kigawa, T., Kikuchi, A., Yokoyama, S., and Takenawa, T. (2001) Role of the ENTH domain in phosphatidylinositol-4,5-bisphosphate binding and endocytosis. *Science* **291**, 1047–1051
 42. Stahelin, R. V., Long, F., Peter, B. J., Murray, D., De Camilli, P., McMahon, H. T., and Cho, W. (2003) Contrasting membrane interaction mechanisms of AP180 N-terminal homology (ANTH) and epsin N-terminal homology (ENTH) domains. *J. Biol. Chem.* **278**, 28993–28999
 43. Kweon, D. H., Shin, Y. K., Shin, J. Y., Lee, J. H., Lee, J. B., Seo, J. H., and Kim, Y. S. (2006) Membrane topology of helix 0 of the Epsin N-terminal homology domain. *Mol. Cells* **21**, 428–435
 44. Graham, T. R., and Kozlov, M. M. (2010) Interplay of proteins and lipids in generating membrane curvature. *Curr. Opin. Cell Biol.* **22**, 430–436
 45. Boucrot, E., Pick, A., Çamdere, G., Liska, N., Evergren, E., McMahon, H. T., and Kozlov, M. M. (2012) Membrane fission is promoted by insertion of amphipathic helices and is restricted by crescent BAR domains. *Cell* **149**, 124–136
 46. Messa, M., Fernández-Busnadiego, R., Sun, E. W., Chen, H., Czapla, H., Wrasman, K., Wu, Y., Ko, G., Ross, T., Wendland, B., and De Camilli, P. (2014) Epsin deficiency impairs endocytosis by stalling the actin-dependent invagination of endocytic clathrin-coated pits. *eLife* **3**, e03311
 47. Bigay, J., Gounon, P., Robineau, S., and Antonny, B. (2003) Lipid packing sensed by ArfGAP1 couples COPI coat disassembly to membrane bilayer curvature. *Nature* **426**, 563–566
 48. Bigay, J., Casella, J. F., Drin, G., Mesmin, B., and Antonny, B. (2005) Arf-GAP1 responds to membrane curvature through the folding of a lipid packing sensor motif. *EMBO J.* **24**, 2244–2253
 49. Kung, L. F., Pagant, S., Futai, E., D'Arcangelo, J. G., Buchanan, R., Dittmar, J. C., Reid, R. J., Rothstein, R., Hamamoto, S., Snapp, E. L., Schekman, R., and Miller, E. A. (2012) Sec24p and Sec16p cooperate to regulate the GTP cycle of the COPII coat. *EMBO J.* **31**, 1014–1027
 50. Yorimitsu, T., and Sato, K. (2012) Insights into structural and regulatory roles of Sec16 in COPII vesicle formation at ER exit sites. *Mol. Biol. Cell* **23**, 2930–2942

Metformin Improves Metabolic Memory in High Fat Diet (HFD)-induced Renal Dysfunction^{*S}♦

Received for publication, April 15, 2016, and in revised form, August 4, 2016
 Published, JBC Papers in Press, August 22, 2016, DOI 10.1074/jbc.C116.732990

Kulbhushan Tikoo^{1,2}, Ekta Sharma¹, Venkateswara Rao Amara¹, Himani Pamulapati, and Vaibhav Shrirang Dhawale

From the Laboratory of Epigenetics and Diseases, Department of Pharmacology and Toxicology, National Institute of Pharmaceutical Education and Research, S.A.S. Nagar (Mohali), Punjab-160062, India

Recently, we have shown that high fat diet (HFD) *in vivo* and *in vitro* generates metabolic memory by altering H3K36me2 and H3K27me3 on the promoter of *FOXO1* (transcription factor of gluconeogenic genes) (Kumar, S., Pamulapati, H., and Tikoo, K. (2016) *Mol. Cell. Endocrinol.* 422, 233–242). Here we checked the hypothesis whether concomitant diet reversal and metformin could overcome HFD-induced metabolic memory and renal damage. Male adult Sprague-Dawley rats were rendered insulin-resistant by feeding high fat diet for 16 weeks. Then the rats were subjected to diet reversal alone and along with metformin for 8 weeks. Biochemical and histological markers of insulin resistance and kidney function were measured. Blood pressure and *in vivo* vascular reactivity to angiotensin II (200 ng kg⁻¹) were also checked. Diet reversal could improve lipid profile but could not prevent renal complications induced by HFD. Interestingly, metformin along with diet reversal restored the levels of blood glucose, triglycerides, cholesterol, blood urea nitrogen, and creatinine. In kidney, metformin increased the activation of AMP-activated protein kinase (AMPK) and decreased inflammatory markers (COX-2 and IL-1 β) and apoptotic markers (poly(ADP-ribose) polymerase (PARP) and caspase 3). Metformin was effective in lowering elevated basal blood pressure and acute change in mean arterial pressure in response to angiotensin II (Ang II). It also attenuated tubulointerstitial fibrosis and glomerulosclerosis induced by HFD feeding in kidney. Here we report, for the first time, that metformin treatment overcomes metabolic memory and prevents HFD-induced renal damage.

Long-term prospective clinical studies, the Diabetes Control and Complications Trial (DCCT), the follow-up study Epidemiology of Diabetes Interventions and Complications (EDIC), and the United Kingdom Prospective Diabetes Study (UKPDS),

stated that benefits of strict glycemic control following a period of poor glycemic control cannot be harnessed immediately following a period of poor glycemic control (1). This was called “metabolic or hyperglycemic memory” by the former two studies and “legacy effect” by the latter.

Various mechanisms have been used to explain the underlying cause of metabolic memory. There is a well established temporal relationship between the duration of poor glycemia and incidence of long-term complications. Glucotoxicity induces some irreversible mitochondrial or vascular changes, which then proceed to overt long-term complications. They also cause long-lasting activation of epigenetic changes in the promoter region of key inflammatory mediators as depicted in mice (2). Short-term hyperglycemia induces long-term changes in chromatin modifications (3). Poor glycemic control is associated with elevated oxidative stress and advanced glycation end products, which damage mitochondrial DNA in kidney, which in turn accentuates the risk of cardiovascular diseases (4). Dysregulated polyols, hexosamine fluxes, and activation of PKC isoforms induced by hyperglycemia also mediate renal damage (2).

Insulin resistance arises because of many reasons, such as increased Ser/Thr phosphorylation (rather than Tyr phosphorylation) of insulin receptor substrate (IRS) proteins, increased activity of Tyr phosphatases (*e.g.* SHIP2, phosphatase and tensin homolog (PTEN), and PTP-1B), and decreased activation of insulin receptor downstream signaling molecules (*e.g.* AKT, atypical PKC, etc.) (5). Chronic consumption of high fat diet leads to elevated levels of non-essential free fatty acids such as palmitic acid. In a series of classical experiments, Randle *et al.* (6) showed that fatty acids compete with glucose for substrate oxidation and hypothesized that increased fat oxidation is responsible for insulin resistance in obesity (reviewed in Ref. 7). Insulin resistance develops in type 2 diabetes and lipodystrophy because of the alteration in the partitioning of fat between the adipocyte and muscle or liver. This leads to the intracellular accumulation of triglycerides and fatty acid metabolites (diacylglycerol, ceramides, fatty acyl CoAs, etc.) in insulin-responsive tissues, leading to insulin resistance (8).

Insulin resistance and hyperinsulinemia due to the adipocytokine-induced activation of I κ B kinase (IKK), JNK, and PKC create an imbalance between PI3K/Akt and MAPK axes, decrease NO-dependent vasodilation, and increase vasoreactivity and angiogenesis, which are known to promote kidney dysfunction. A growing body of preclinical data suggests that glomerular hypertrophy appears within the period of hyperinsulinemia itself, before the onset of overt diabetes (2, 9). Profibrotic TGF- β damages glomerular function by overproduction of interstitial tissue matrix, thickening of the glomerular basement membrane, and the down-regulation of matrix metalloproteinase (MMP)-2, an enzyme involved in matrix degradation, eventually ending in albumin leakage into the urine (10).

Recently, it has been shown that *in vitro* chronic treatment of podocytes with palmitic acid leads to insulin resistance in podocytes (11), which comprise the major constituent of the glomer-

* This work was supported by grants from the National Institute of Pharmaceutical Education and Research, S.A.S. Nagar, India and the Department of Science & Technology, Government of India. The authors declare that they have no conflicts of interest with the contents of this article.

♦ This article was selected as a Paper of the Week.

S This article contains supplemental Figs. 1 and 2.

¹ These authors contributed equally to this work.

² To whom correspondence should be addressed: Laboratory of Epigenetics and Diseases, Dept. of Pharmacology and Toxicology, National Institute of Pharmaceutical Education and Research (NIPER), Sector 67, S.A.S. Nagar, Punjab-160062, India. Tel.: 91-172-2292049; E-mail: tikoo.k@gmail.com.

ular filtration barrier. Podocyte injury is mainly characterized by effacement of its foot process and the loss of key molecules such as nephrin and podocin (12).

Metformin, a biguanide oral hypoglycemic agent, has been the drug of choice for type 2 diabetes for many years, reducing morbidity and mortality. There is a growing body of evidence suggesting pleiotropic effects of metformin. Apart from its glucose- and free fatty acid-lowering action by sensitizing insulin, it was shown by Piwkowska *et al.* (13) to activate AMP-activated protein kinase (AMPK)³ and decrease NADPH oxidase (NOX) activity, thereby reducing reactive oxygen species production and TGF- β -induced epithelial-to-mesenchymal transition, a key event during the development of the tubulointerstitial fibrosis in diabetic nephropathy. Metformin reduces fat content by decreasing sterol regulatory element-binding protein 1 (SREBP-1), fatty acid synthetase (FAS), and acetyl-CoA carboxylase (ACC) expression in kidney (14).

Risk of kidney, eye, and nerve damage and nonfatal cardiovascular events can be reduced to almost ~50% only by following long-term intensive blood glucose control (15, 16). Because only long-term, intense blood glucose control (diet reversal) can reduce the risk of diabetic complications, there is a quest for a therapeutic agent that, along with intense blood glucose control, can harness the diet reversal within a short span of time (17). To achieve this objective, we used diet reversal along with metformin treatment for only 8 weeks to reverse the metabolic memory generated in 16 weeks.

Experimental Procedures

Animals—Adult male Sprague-Dawley rats of 180–200 g of body weight were procured from the Central Animal Facility, National Institute of Pharmaceutical Education and Research (NIPER), S.A.S. Nagar, India, and three rats/cage were housed under standard environmental conditions (temperature: 22 \pm 2 °C; humidity: 50 \pm 10%; and 12-h light/dark cycle) with access to food and water *ad libitum*. All protocols were approved by the Institutional Animal Ethics Committee (IAEC Approval Number 13/28, NIPER) and performed in accordance with the guidelines of the Committee for the Purpose of Control and Supervision of Experiments on Animals (CPCSEA), New Delhi, India.

Experimental Design—After 1 week of acclimatization, male Sprague-Dawley rats were randomly divided into two groups: CONTROL, which were fed with normal pellet diet (NPD), and HFD, which were fed with high fat diet (HFD). After 16 weeks, HFD-fed rats were further divided into three groups: (a) HFD-fed rats, which received HFD; (b) REV, which received NPD; and (c) REV+MET, which received NPD and metformin at 100 mg kg⁻¹ day⁻¹, p.o. for a period of 8 weeks (supplemental Fig. 1B). Crushed/pelleted NPD feed was standard chow from Pranav Agro Industries, New Delhi, India. Lard was procured from a

local slaughter house. HFD was prepared in-house as per the composition devised by Srinivasan *et al.* (18). Body weight was recorded every week right from the beginning to the end of the study. Blood biochemical parameters (glucose, lipid profile, and kidney function tests) were measured twice (8th and 16th weeks) during the insulin resistance model development (before diet reversal) and at the end (24th week) of the study (supplemental Fig. 1A).

Biochemical Parameters—After overnight fasting, blood was collected retro-orbitally under thiopentone anesthesia (50 mg kg⁻¹, i.p.), and then centrifuged at 4 °C, 2500 \times g for 10 min for separation of plasma. Different biochemical parameters were estimated as per the manufacturer's guidelines (Accurex Biomedical Pvt. Ltd., Mumbai, India). The parameters measured were: glucose (GOD-POD); lipid profile: triglycerides (LPL-GK-GPO-POD); cholesterol (CHE-CHO-POD); kidney function tests (KFT): BUN (urease); and creatinine (Jaffe's initial rate method).

Hemodynamic Parameters—Invasive blood pressure and change in mean arterial pressure (Δ MAP) in response to bolus infusion of angiotensin II (200 ng kg⁻¹, i.v.) were measured as described earlier (19). Hemodynamic data were acquired and analyzed using LabChart 7 (ADInstruments, Bella Vista, New South Wales, Australia).

Western Blotting—Immunoblotting was performed as described earlier (20). Briefly, frozen kidneys were thawed, minced, and homogenized in lysis buffer containing surfactants and protease and phosphatase inhibitors. Protein samples were resolved using sodium dodecyl sulfate-polyacrylamide gels depending on the molecular weight of the desired proteins. These were then transferred to nitrocellulose membranes and were probed with primary and secondary antibodies, developed using an HRP-chemiluminescent substrate solution, and captured onto x-ray film, which was then scanned using a flatbed scanner to an 8-bit grayscale image. The intensity of the band in each lane was quantified using the ImageJ (National Institutes of Health) software. All antibodies were procured from Santa Cruz Biotechnology, Inc., unless otherwise specified. The antibodies were: phospho-AMPK α 1/2 (sc-33524), AMPK α 1/2 (sc-25792), tubulin (sc-23948), Cox-2 (sc-7951), IL-1 β (sc-7884), caspase 3 (sc-7272), fibronectin (sc-9068, detects both cytoplasmic and extracellular fibronectin), α -SMA (sc-53015), and PARP (9542, Cell Signaling Technology).

Histopathology and Immunohistochemistry—Kidney was fixed in 10% v/v formal saline and embedded in paraffin, and 5- μ m transverse sections were prepared and mounted on slides previously coated with poly-L-lysine and stained with hematoxylin and eosin (for glomerular space and glomerular tuft area) and picosirius red (for extracellular matrix deposition). Coverslips were mounted and observed under OLYMPUS BX51 microscope. Immunostaining was performed for fibronectin and α -SMA on paraffin-embedded sections as described earlier (19).

TUNEL Assay—Terminal deoxynucleotidyl transferase-mediated dUTP nick end labeling (TUNEL) assay was performed on paraffin-embedded 5- μ m-thick kidney sections to assess the DNA fragmentation. The assay was conducted according to the manufacturer's instructions. The images were

³ The abbreviations used are: AMPK, AMP-activated protein kinase; PARP, poly(ADP-ribose) polymerase; ACC, acetyl-CoA carboxylase; α -SMA, α -smooth muscle actin; Ang II, angiotensin II; HFD, high fat diet; NPD, normal pellet diet; CHE-CHO-POD, cholesterol esterase-cholesterol oxidase-peroxidase; GOD-POD, glucose oxidase-peroxidase; LPL-GK-GPO-POD, lipoprotein lipase-glycerol kinase-glycerol phosphate oxidase-peroxidase; MAP, mean arterial pressure; Δ MAP, change in MAP; p.o., *per os* (orally); BUN, blood urea nitrogen; BW, body weight.

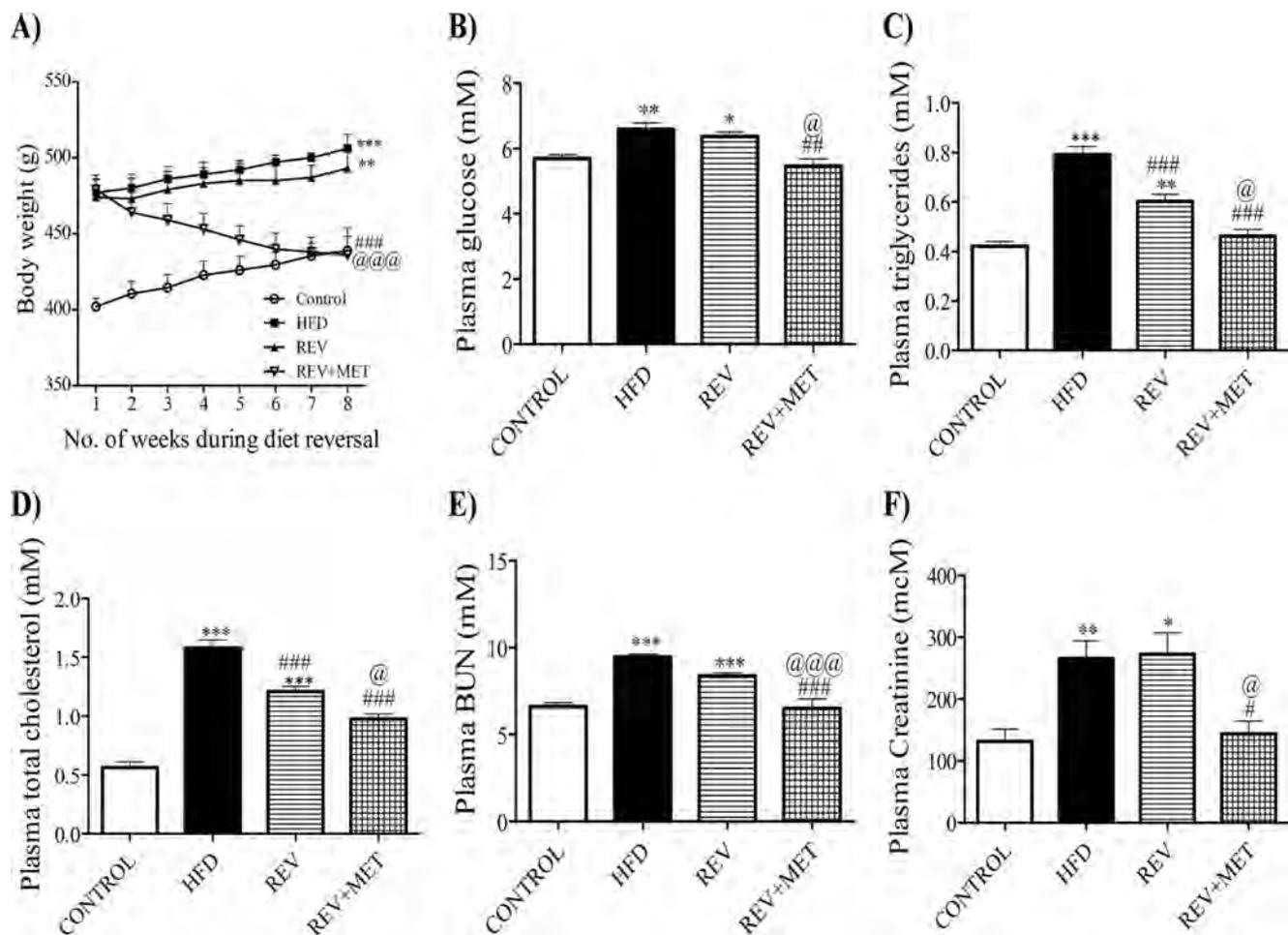


FIGURE 1. Parameters after 8 weeks of diet reversal, i.e. from the 17th week to the 24th week of the study. A, growth curve with body weights taken every week. B–F, glucose (B), triglycerides (C), cholesterol (D), blood urea nitrogen (E), and creatinine (F) measured at the end of the study after diet reversal. CONTROL, rats fed with normal pellet diet; HFD, rats fed with high fat diet; REV, reversal group rats fed with normal pellet diet for 8 weeks following 16 weeks of high fat diet feeding; REV+MET, reversal group rats fed with normal pellet diet and treated with metformin (100 mg kg⁻¹ day⁻¹, p.o.) after 16 weeks of high fat diet feeding. *, *p* < 0.05, **, *p* < 0.01, ***, *p* < 0.001 versus CONTROL; #, *p* < 0.05, ##, *p* < 0.01, ###, *p* < 0.001 versus HFD; @, *p* < 0.05, @@@, *p* < 0.001 versus REV; *n* = 6.

TABLE 1
Morphometric and hemodynamic parameters

Initial BW was taken after the 16-week period, i.e. at the 17th week, and final BW was taken on the day of sacrifice. CONTROL, control rats fed with normal pellet diet; HFD, rats fed with high fat diet throughout the study for 24 weeks; REV, rats fed with high fat diet for 16 weeks and switched to normal diet for the next 8 weeks; REV+MET, rats fed with high fat diet for 16 weeks and switched to normal diet and treated with metformin (100 mg kg⁻¹ day⁻¹, p.o.) for the next 8 weeks. BW, body weight; KW, kidney weight; HW, heart weight; LW, liver weight; WAT, white adipose tissue; MAP, mean arterial pressure; ΔMAP, acute change of MAP in response to a bolus of angiotensin II (200 ng kg⁻¹). *, *p* < 0.05, **, *p* < 0.01, and ***, *p* < 0.001 versus control; #, *p* < 0.05, ##, *p* < 0.01, and ###, *p* < 0.001 versus HFD; @, *p* < 0.05, @@@, *p* < 0.001 versus REV, *n* = 6.

Parameter	CONTROL	HFD	REV	REV+MET
Initial BW (g)	402.08 ± 5.24	473.55 ± 8.74***	469.45 ± 11.20***	472.75 ± 10.94***
Final BW (g)	438.8 ± 8.92	506.22 ± 4.79***	493.91 ± 11.04***	438.4 ± 11.98##,@@
Absolute kidney weight (g)	2.40 ± 0.07	2.93 ± 0.18**	2.83 ± 0.05**	2.54 ± 0.05#,@
KW/BW (mg g ⁻¹)	5.19 ± 0.09	5.83 ± 0.04**	5.79 ± 0.03**	5.38 ± 0.06##,@@
KW/tibial length (g cm ⁻¹)	0.57 ± 0.02	0.70 ± 0.03**	0.68 ± 0.02**	0.60 ± 0.01#,@
Absolute heart weight (g)	1.17 ± 0.02	1.31 ± 0.02**	1.30 ± 0.013**	1.20 ± 0.008#,@
HW/BW (mg g ⁻¹)	2.52 ± 0.07	2.89 ± 0.06**	2.90 ± 0.05**	2.57 ± 0.07#,@
HW/tibial length (g cm ⁻¹)	0.28 ± 0.005	0.32 ± 0.003***	0.32 ± 0.004***	0.29 ± 0.007##,@@
Absolute liver weight (g)	13.25 ± 0.48	18.47 ± 0.57***	16.97 ± 0.30***	12.99 ± 0.79###,@@@
LW/BW (mg g ⁻¹)	27.56 ± 1.25	36.12 ± 0.98***	34.14 ± 0.31**	27.70 ± 1.76###,@@
WAT (g)	6.55 ± 0.77	12.94 ± 0.64***	12.06 ± 0.10***	7.88 ± 0.98##,@@
WAT/BW (mg g ⁻¹)	13.54 ± 1.32	28.83 ± 1.58***	28.58 ± 3.04***	16.00 ± 2.80##,@@
MAP (mmHg)	82.74 ± 2.33	98.24 ± 1.30**	93.42 ± 1.30*	83.37 ± 0.50##,@@
ΔMAP in response to Ang II 200 mg kg ⁻¹	23.78 ± 1.27	35.08 ± 0.97**	32.87 ± 2.70*	25.14 ± 0.16##,@@

acquired using a charge-coupled device (CCD) camera. Finally, the average number of TUNEL-positive apoptotic bodies per glomerulus was calculated.

Statistical Analysis—All the data were expressed as mean ± S.E. Means of two groups were compared using unpaired Student's *t* test, and means of multiple groups were com-

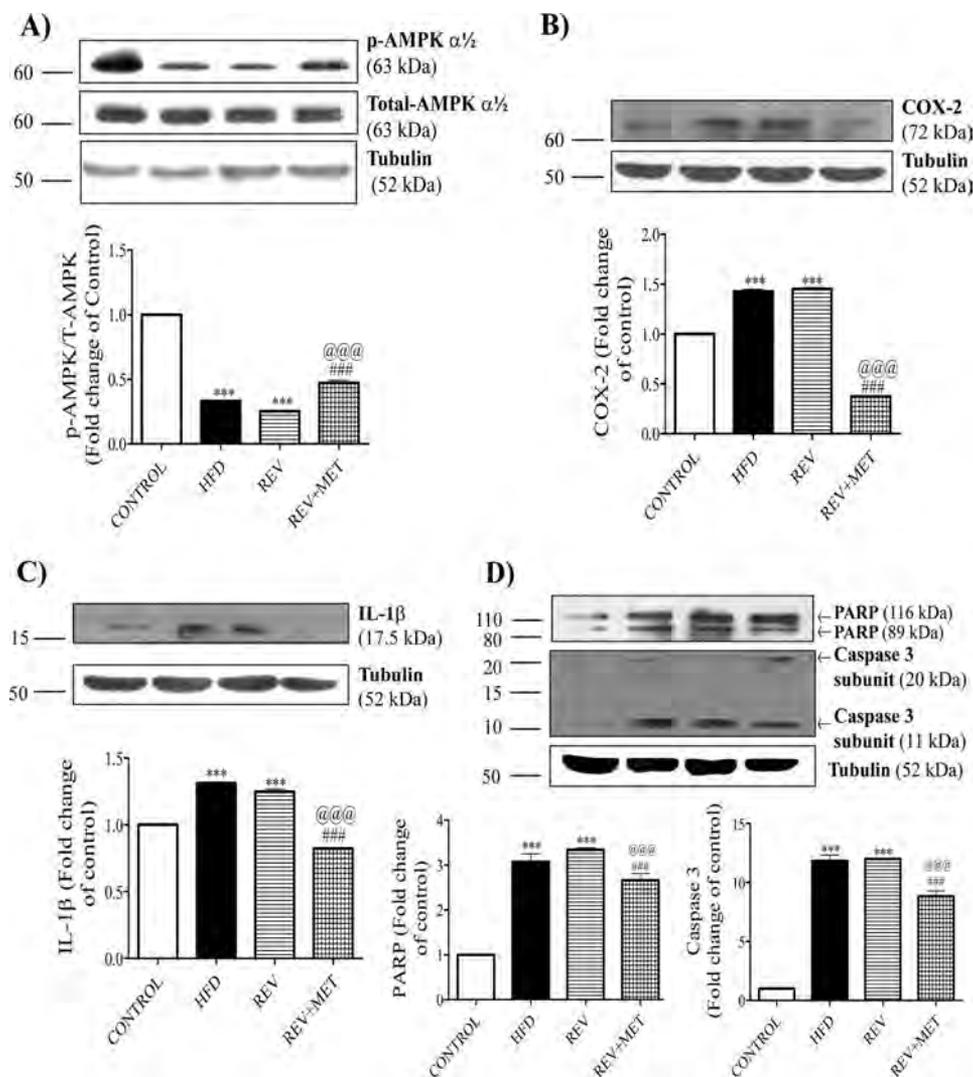


FIGURE 2. **Western blots of whole kidney lysates.** 60 μg of protein was loaded per each lane. 8, 10, and 12% SDS-PAGE gels were run according to the molecular weight of the protein of interest. To ensure the correct position of the protein of interest, a pre-stained protein marker (Invitrogen Novex[®] Sharp Pre-stained Protein Standard, catalog number LC5800, Thermo Fisher Scientific) was run along with the samples. Tubulin was used as internal control for normalization, and results are expressed as -fold change over control rats. *A*, Western blots and quantification for phospho-AMPK (*p*-AMPK), total AMPK, and tubulin. *B*, Western blots and quantification for inflammatory markers: COX-2 and tubulin. *C*, Western blots and quantifications for inflammatory cytokines: IL-1 β and tubulin. *D*, Western blots and quantifications for apoptotic markers: PARP, caspase 3, and tubulin. *CONTROL*, rats fed with normal pellet diet; *HFD*, rats fed with high fat diet; *REV*, reversal group rats fed with normal pellet diet for 8 weeks following 16 weeks of high fat diet feeding; *REV+MET*, reversal group rats fed with normal pellet diet and treated with metformin (100 mg kg⁻¹ day⁻¹, p.o.) after 16 weeks of high fat diet feeding. $n = 3$ blots. ***, $p < 0.001$ versus CONTROL; ###, $p < 0.001$ versus HFD; @@@, $p < 0.001$ versus REV.

pared using analysis of variance followed by Bonferroni's post hoc test. Values were considered statistically significant if $p < 0.05$. The statistical software used for analyzing the data was GraphPad Prism, version 5.01 (GraphPad Software, Inc.).

Results

High Fat Diet Feeding Induces Insulin Resistance—In accordance with earlier studies (19, 21), high fat diet-fed animals in our study exhibited significant increase in body weight, plasma glucose, triglycerides, and cholesterol levels at 8, 16, and 24 weeks of high fat diet feeding (supplemental Fig. 2). The higher rate of body weight gain was accentuated from 6 weeks onwards in HFD-fed rats. These results indicate the development of insulin resistance under hyperlipidemic conditions (elevated

levels of circulating saturated free fatty acids) was reached after feeding high fat diet.

Short-term Diet Reversal Displays Metabolic Memory, and Metformin Attenuates Metabolic Memory—The body weight gain of the reversal group was comparable with that of HFD even after 8 weeks of diet reversal, suggesting the footprints of the earlier poor glycemic control (high fat diet) period. Metformin was found to be effective in reducing the body weight gain from the first week of treatment itself (Fig. 1A). Diet reversal could not affect the hyperglycemia induced by high fat diet, indicating the presence of metabolic memory (Fig. 1B). However, metformin attenuated the elevated glucose, triglycerides, and cholesterol (Fig. 1, B–D).

Kidney Function Parameters—Plasma blood urea nitrogen and plasma creatinine rose from the 16th week of high fat diet

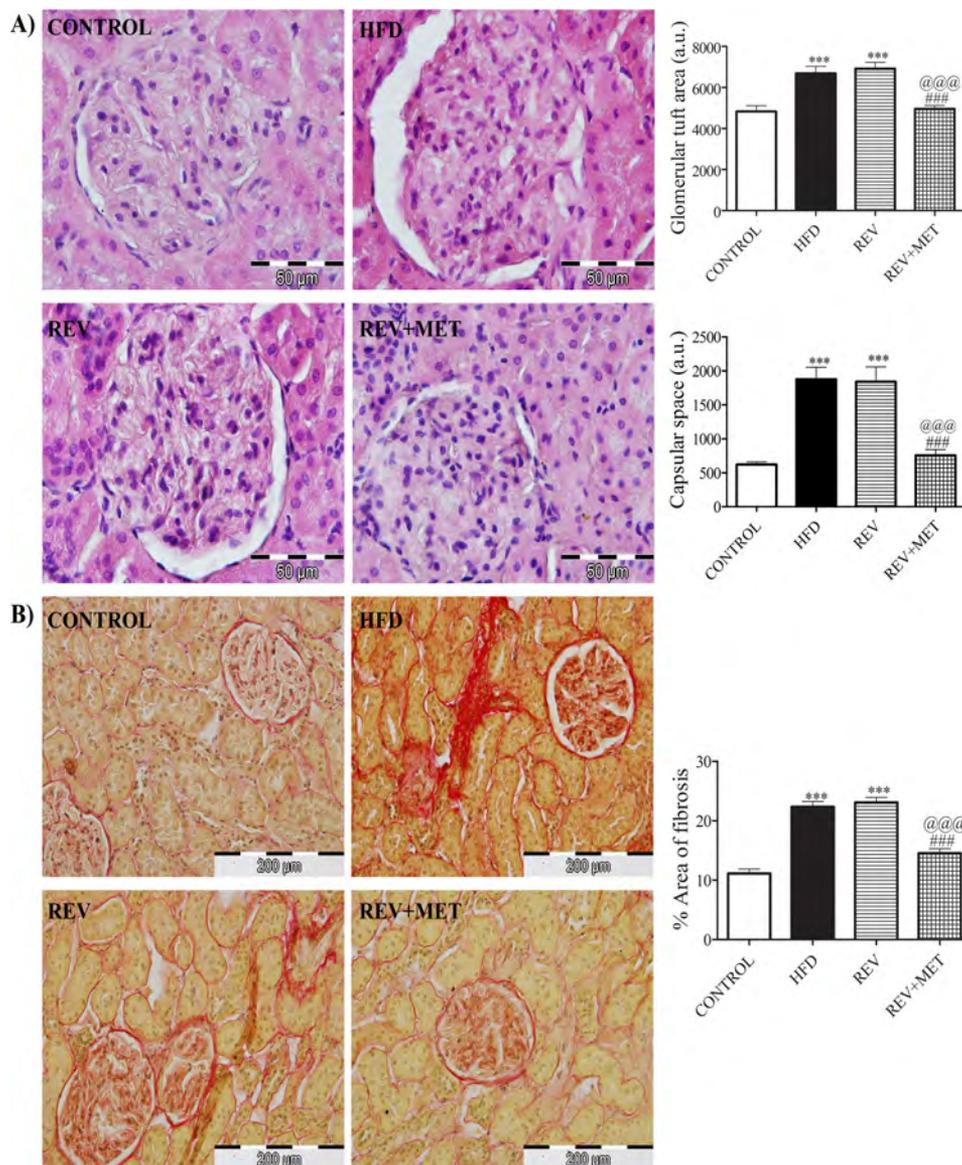


FIGURE 3. **H&E and picrosirius red staining of kidney.** A, H&E-stained images under 1000 \times magnification along with quantification of glomerular tuft area and capsular space. B, picrosirius red-stained images showing glomerular fibrosis under 400 \times magnification along with quantification of interstitial fibrosis. CONTROL, rats fed with normal pellet diet; HFD, rats fed with high fat diet; REV, reversal group rats fed with normal pellet diet for 8 weeks following 16 weeks of high fat diet feeding; REV+MET, reversal group rats fed with normal pellet diet and treated with metformin (100 mg kg⁻¹ day⁻¹, p.o.) after 16 weeks of high fat diet feeding. $n = 10$ –12 sections from each group. ^{***}, $p < 0.001$ versus CONTROL; ^{###}, $p < 0.001$ versus HFD; ^{@@@}, $p < 0.001$ versus REV.

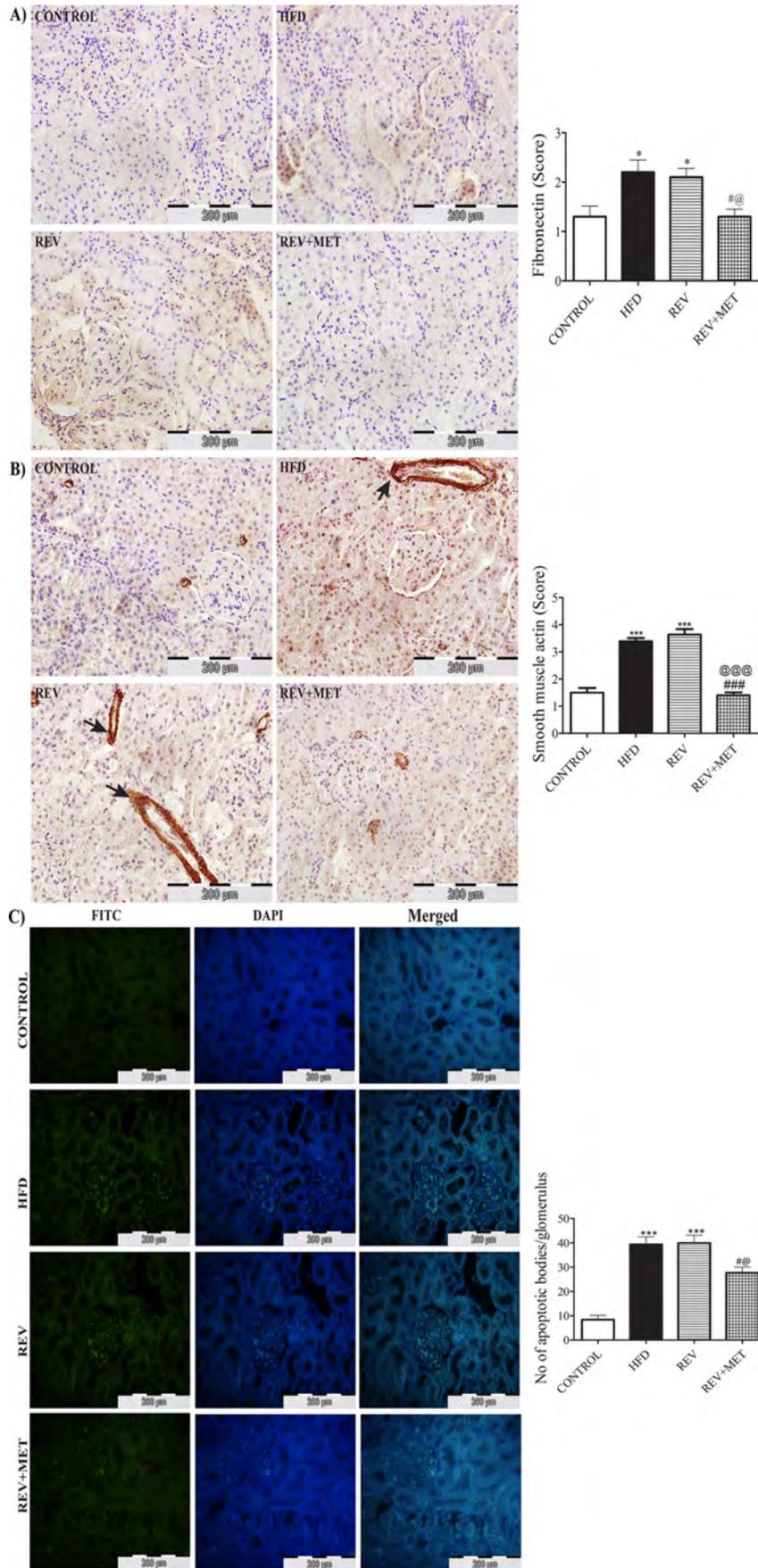
feeding, indicating the development of renal dysfunction. Diet switching did not improve the renal function indicators. However, metformin improved renal dysfunction by decreasing the plasma urea and creatinine levels (Fig. 1, E and F).

Metformin Improves Cardiovascular Function—It has been postulated that elevated basal MAP exacerbates nephropathy (22). To investigate this phenomenon, invasive blood pressure was measured in our study. HFD-fed rats had significantly higher basal MAP and increased response (change in MAP) to angiotensin II at a dose of 200 ng kg⁻¹ as compared with age-matched control rats. Diet reversal failed to restore both hemodynamic parameters. Metformin lowered the basal MAP as well as the Ang II-mediated change in MAP, comparable with the control group (Table 1).

Metformin Activates AMPK—High fat diet significantly reduced the level of activated AMPK, and diet reversal could

not restore it. We found that none of the diet/treatment manipulations changed the basal levels of AMPK. Metformin, a known AMPK activator (23), increased the activity of AMPK, as evident from the increased levels of the phosphorylated form of AMPK in Western blotting (Fig. 2A).

Metformin Reduces Renal Inflammation, Apoptosis, and Fibrosis—Clinical and experimental studies highlighted the association of high fat diet with inflammation and apoptosis (24, 25). Hence, we also checked the molecular markers of renal inflammation, apoptosis, and fibrosis. High fat diet feeding significantly augmented the levels of COX-2, a non-steroidal anti-inflammatory drug-sensitive pathological inflammation-mediating enzyme, and IL-1 β , an inflammatory cytokine (Fig. 2, B and C). HFD feeding also instigated apoptosis in kidney, as evident from the surge in the proapoptotic caspase 3 and PARP, an enzyme involved in DNA repair (Fig. 2D), and also the



increased number of apoptotic bodies as observed by TUNEL assay (see Fig. 4C). Diet reversal was unable to improve the renal inflammation and apoptosis, emphasizing the persistence of renal injury even after 8 weeks of dietary control. However, metformin treatment was able to reduce the levels of inflammatory and apoptotic markers.

Collagen and fibronectin are key components of the extracellular matrix during sclerosis (scar formation) following an episode of inflammatory injury in a tissue as a part of tissue repair (26). Interestingly, high fat diet also increased the deposition of α -SMA and fibronectin in the kidney, as evident from the immunostaining. Diet reversal could not bring down the levels of fibronectin and α -SMA, but metformin demonstrated its ability to ameliorate renal fibrosis (see Fig. 4, A and B). Kidney weight, both *per se* and when normalized with body weight or tibial length, crude markers of hypertrophy, increased significantly in animals fed with high fat diet, which could not be assuaged by switching to normal diet but could be normalized when treated with metformin (Table 1).

Metformin Ameliorates Insulin Resistance-induced Gross Morphological and Microscopic Histological Changes—High fat diet significantly amplified body weight gain and white adiposity index, indicating the presence of obesity. It also increased the weights of kidney, heart, and liver, either when considered alone or even after normalization with tibial length (Table 1). The latter is an excellent index of hypertrophy. Tibial length, but not body weight, is recommended as an index of hypertrophy of an organ, especially in cases where the body weight changes as in the case of obesity (27). Diet reversal proved inefficient in encountering the aberrations induced by HFD feeding, but diet reversal in combination with metformin could suppress them (Table 1).

Histological cross-sections of HFD-fed kidney observed under a microscope showed increased glomerular size (glomerular tuft area) and glomerular space as revealed by H&E staining (Fig. 3A), as well as collagenous extracellular matrix (tubulointerstitial fibrosis) as revealed by picosirius red staining (Fig. 3B). These aberrations caused by HFD were not alleviated by diet reversal alone but could be achieved in combination with metformin.

Discussion

We provide the first evidence that metformin prevents the metabolic memory responsible for the progression of HFD-induced renal dysfunction. Previously, we have reported the presence of metabolic memory in renal and vascular endothelial dysfunction induced by HFD (21, 28). Feeding rats with high fat diet for 16 weeks induced insulin resistance characterized by elevated blood glucose, triglycerides, and cholesterol, in agreement with the previous results of our lab (29, 30). Elevated

plasma BUN and creatinine showed the development of renal dysfunction after 16 weeks of HFD feeding. Subsequent diet reversal for 8 weeks reduced lipid levels but could not mitigate hyperglycemia and renal dysfunction. However, metformin treatment ($100 \text{ mg kg}^{-1} \text{ day}^{-1}$) along with diet reversal effectively restored insulin sensitivity and renal function.

Our model mimics the clinical situation in which high calorie/high fat diet induces renal complications in insulin resistance conditions. It is an early stage nephropathy because there was no change in the plasma levels of albumin (data not shown). The initial exposure to hyperglycemia and hypertriglyceridemia prevented the high fat-fed animals from responding to diet reversal at the biochemical, molecular, microscopic, and macroscopic levels, indicating the development of metabolic memory in our model. Here we show that 8 weeks of diet reversal, after 16 weeks of high fat diet feeding, could not restore the aberrations of insulin resistance, but diet reversal along with metformin treatment for a short duration (8 weeks) shows remarkable amelioration of anomalies associated with insulin resistance. Metformin treatment prevents HFD-induced metabolic memory and reverses renal complications.

Clinically, it has been shown that insulin resistance and hypertension are major contributors of renal dysfunction (22). In line with the above study, in our animal model, invasive blood pressure measurement showed an increase in MAP and Ang II-induced Δ MAP, which could not be reversed upon diet reversal. Metformin has been reported to have an anti-hypertensive effect in diabetes (31). Our study also found that metformin significantly mitigated the basal as well as the Ang II-induced MAP elevation. Enhanced angiotensin II induced acute contractile responses, and worsened endothelial dependent vasodilation in HFD-fed rats *ex vivo* has been shown to be associated with down-regulation of SIRT1/AMPK/eNOS (endothelial NOS) pathway (19, 29), suggesting to us that the improvement in hemodynamic parameters observed during metformin treatment may also involve up-regulation of this pathway.

AMPK is the master sensor of cellular energy metabolism and is initially activated during metabolic stress conditions (impaired ATP production and excess ATP consumption) such as diabetes and obesity, but is inhibited as the disease progresses (32). AMPK is suppressed in various organs, including kidney in high fat consumption, and its activation by 5-amino-4-imidazole carboxamide riboside (AICAR) or metformin was shown to be effective in reversing renal damage (33, 34). In line with this result, we observed a significant fall in the activity of renal AMPK in high fat-fed rats. Diet reversal could not restore the activity of AMPK unless combined with metformin treatment. AMPK activation inhibits inflammation in MRL/lpr

FIGURE 4. Immunohistochemistry of renal fibronectin, α -SMA, and TUNEL assay in kidney sections. A and B, 3,3'-diaminobenzidine-stained images of fibronectin and α -SMA under $400\times$ magnification (A) and quantification of immunohistochemistry by the indicated scoring method (B). Scores were measured as follows: 1 = no or very low color, 2 = mild color, 3 = moderate color, 4 = intense brown color. Arrows indicate the extracellular areas of α -SMA immunostaining. C, FITC and DAPI staining and merged images of TUNEL assay under $400\times$ magnification along with quantification of apoptotic bodies/glomerulus. *CONTROL*, rats fed with normal pellet diet; *HFD*, rats fed with high fat diet; *REV*, reversal group rats fed with normal pellet diet for 8 weeks following 16 weeks of high fat diet feeding; *REV+MET*, reversal group rats fed with normal pellet diet and treated with metformin ($100 \text{ mg kg}^{-1} \text{ day}^{-1}$, p.o.) after 16 weeks of high fat diet feeding. $n = 10$ – 12 sections from each group. *, $p < 0.05$ and ***, $p < 0.001$ versus *CONTROL*; #, $p < 0.05$, ##, $p < 0.001$ versus *HFD*; @, $p < 0.05$, @@@, $p < 0.001$ versus *REV*.

mouse mesangial cells and *in vivo* renal inflammation by decreasing MCP-1 (33, 35). In our study, renal COX-2 and IL-1 β , the markers of inflammation, were repressed by metformin, but not by diet reversal alone.

The beneficial effect of metformin treatment in reversing metabolic memory in a short duration (8 weeks) can be explained if we assume that metformin increases renal fatty acid oxidation by modulating AMPK/ACC pathway and in turn reducing renal lipotoxicity. ACC is a rate-limiting enzyme required for the generation of malonyl-coenzyme A (malonyl-CoA), which plays a critical role in the synthesis of fatty acids and inhibition of mitochondrial fatty acid oxidation. AMPK inhibits this ACC by phosphorylating it (34).

At the macroscopic level, the kidney index significantly increased in the HFD group and decreased only when subjected to diet reversal and metformin treatment. Crinigan *et al.* (36) has shown significant increase in the renal mass of Sprague-Dawley rats just after 6 weeks of high fat diet feeding. In addition, other studies have also shown that high fat diet feeding increases kidney weight (37, 38).

Histological sections of kidney showed characteristic structural alteration by HFD feeding, such as increased glomerular size and Bowman's capsular space as reported earlier (25). We also observed similar microscopic maladaptations in kidney even after diet reversal. During tissue repair and wound healing, the differentiation of tubular epithelial and progenitor cells, vascular pericytes and endothelial cells, and interstitial cells such as fibroblasts into myofibroblasts results in the deposition of extracellular matrix components such as collagen and fibronectin. However, excess deposition of extracellular matrix results in fibrosis. α -SMA is a selective marker of myofibroblasts (39). In our present study, collagen, fibronectin, and α -SMA increased in the HFD group and were attenuated only after treatment with metformin along with high fat diet reversal. Microscopic examination of kidney sections using picrosirius staining revealed the extent of renal injury as observed in terms of glomerular and interstitial fibrosis. The extent of renal damage after initial high fat consumption was comparable in both groups (HFD and REV), indicating no potential benefit of diet reversal. Metformin, an AMPK activator, significantly improved renal fibrosis. The immunohistochemical scores of fibronectin and α -SMA also confirmed the same results (Fig. 4). AMPK plays a crucial role in maintaining the structural integrity of kidney. AMPK activation by metformin restored the normal structure and physiology of kidney. Inhibition of AMPK has been reported to be associated with increased profibrotic markers (40). The decreased level of the phosphorylated form of AMPK coincided with increased collagen and extracellular matrix deposition in kidney of HFD-fed animals and its reversal by metformin treatment. This can very well explain the protection observed by metformin treatment but not by diet reversal.

Insulin resistance induces podocyte death, modulating the phosphorylation of PP2A and mTORc1 (41). High fat-mediated insulin resistance showed an increase in renal cell death and apoptotic bodies, which was confirmed with the increased protein expression of the apoptosis markers, PARP and caspase 3. Metformin effectively reduced renal apoptosis, as evident from the reduced number of TUNEL-positive cells.

Our data enable us to conclude that due to metabolic memory, short-term diet reversal cannot reverse the insulin resistance-induced complications. However, complications due to metabolic memory can very well be prevented by metformin, an AMPK activator. Recently, we have shown that fatty acid/high fat diet-induced metabolic memory involves changes in renal histone H3K36me2 and H3K27me3 (28). Further studies are required to ascertain the role of metformin in epigenetic alterations responsible for preventing metabolic memory.

Author Contributions—K. T. conceived the idea, designed and supervised the study, and approved the manuscript. E. S. prepared the high fat diet, treated animals, did biochemical and histological experiments and Western blotting, and wrote manuscript. V. R. A. performed hemodynamic experiments, histological experiments, and Western blotting and wrote manuscript. H. P. and V. S. D. helped during experiments.

Acknowledgments—We are grateful to Umashanker Navik for providing selfless help in capturing and quantification of histological images. We owe a lot to Pinakin Arun Karpe and Sandeep Kumar of our lab for their insightful discussion.

Note Added in Proof—In the version of this paper that was published as a Paper in Press on August 22, 2016, an incorrect figure was uploaded as Fig. 1. Additionally, Fig. 4C was labeled incorrectly. Errors arose while incorporating the figure/label in the manuscript. These errors have now been corrected and do not affect the results or conclusions of this work.

References

1. Tonna, S., El-Osta, A., Cooper, M. E., and Tikellis, C. (2010) Metabolic memory and diabetic nephropathy: potential role for epigenetic mechanisms. *Nat. Rev. Nephrol.* **6**, 332–341
2. Zick, Y. (2003) Role of Ser/Thr kinases in the uncoupling of insulin signaling. *Int. J. Obes. Relat. Metab. Disord.* **27**, Suppl. 3, S56–S60
3. Villeneuve, L. M., and Natarajan, R. (2010) The role of epigenetics in the pathology of diabetic complications. *Am. J. Physiol. Renal Physiol.* **299**, F14–F25
4. Drzewoski, J., Kasznicki, J., and Trojanowski, Z. (2009) The role of “metabolic memory” in the natural history of diabetes mellitus. *Pol. Arch. Med. Wewn.* **119**, 493–500
5. Zick, Y. (2004) Uncoupling insulin signalling by serine/threonine phosphorylation: a molecular basis for insulin resistance. *Biochem. Soc. Trans.* **32**, 812–816
6. Randle, P. J., Garland, P. B., Newsholme, E. A., and Hales, C. N. (1965) The glucose fatty acid cycle in obesity and maturity onset diabetes mellitus. *Ann. N.Y. Acad. Sci.* **131**, 324–333
7. Shulman, G. I. (2000) Cellular mechanisms of insulin resistance. *J. Clin. Invest.* **106**, 171–176
8. Roden, M., Price, T. B., Perseghin, G., Petersen, K. F., Rothman, D. L., Cline, G. W., and Shulman, G. I. (1996) Mechanism of free fatty acid-induced insulin resistance in humans. *J. Clin. Invest.* **97**, 2859–2865
9. Yang, J., Park, Y., Zhang, H., Xu, X., Laine, G. A., Dellsperger, K. C., and Zhang, C. (2009) Feed-forward signaling of TNF- α and NF- κ B via IKK- β pathway contributes to insulin resistance and coronary arteriolar dysfunction in type 2 diabetic mice. *Am. J. Physiol. Heart Circ. Physiol.* **296**, H1850–H1858
10. Peng, J., Li, X., Feng, Q., Chen, L., Xu, L., and Hu, Y. (2013) Anti-fibrotic effect of *Cordyceps sinensis* polysaccharide: inhibiting HSC activation, TGF- β 1/Smad signalling, MMPs and TIMPs. *Exp. Biol. Med. (Maywood)* **238**, 668–677
11. Lennon, R., Pons, D., Sabin, M. A., Wei, C., Shield, J. P., Coward, R. J., Tavaré, J. M., Mathieson, P. W., Saleem, M. A., and Welsh, G. I. (2009)

- Saturated fatty acids induce insulin resistance in human podocytes: implications for diabetic nephropathy. *Nephrol. Dial. Transplant.* **24**, 3288–3296
12. Deegens, J. K. J., Dijkman, H. B. P. M., Borm, G. F., Steenbergen, E. J., van den Berg, J. G., Weening, J. J., and Wetzels, J. F. M. (2008) Podocyte foot process effacement as a diagnostic tool in focal segmental glomerulosclerosis. *Kidney Int.* **74**, 1568–1576
 13. Piwkowska, A., Rogacka, D., Jankowski, M., Dominiczak, M. H., Stepínski, J. K., and Angielski, S. (2010) Metformin induces suppression of NAD(P)H oxidase activity in podocytes. *Biochem. Biophys. Res. Commun.* **393**, 268–273
 14. Wang, W., Guo, X. H., Wu, H. H., Wang, N. H., and Xu, X. S. (2006) [Effect of fenofibrate and metformin on lipotoxicity in OLETF rat kidney]. *Beijing Da Xue Xue Bao* **38**, 170–175
 15. The Diabetes Control and Complications Trial Research Group (1993) The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N. Engl. J. Med.* **329**, 977–986
 16. Nathan, D. M., Cleary, P. A., Backlund, J. Y., Genuth, S. M., Lachin, J. M., Orchard, T. J., Raskin, P., and Zinman, B.; Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study Research Group (2005) Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N. Engl. J. Med.* **353**, 2643–2653
 17. Zhang, L., Chen, B., and Tang, L. (2012) Metabolic memory: mechanisms and implications for diabetic retinopathy. *Diabetes Res. Clin. Pract.* **96**, 286–293
 18. Srinivasan, K., Patole, P. S., Kaul, C. L., and Ramarao, P. (2004) Reversal of glucose intolerance by pioglitazone in high fat diet-fed rats. *Methods Find. Exp. Clin. Pharmacol.* **26**, 327–333
 19. Karpe, P. A., and Tikoo, K. (2014) Heat shock prevents insulin resistance-induced vascular complications by augmenting angiotensin-(1–7) signaling. *Diabetes* **63**, 1124–1139
 20. Tikoo, K., Bhatt, D. K., Gaikwad, A. B., Sharma, V., and Kabra, D. G. (2007) Differential effects of tannic acid on cisplatin induced nephrotoxicity in rats. *FEBS Lett.* **581**, 2027–2035
 21. Tallapragada, D. S., Karpe, P. A., and Tikoo, K. (2015) Long-lasting partnership between insulin resistance and endothelial dysfunction: role of metabolic memory. *Br. J. Pharmacol.* **172**, 4012–4023
 22. El-Atat, F. A., Stas, S. N., McFarlane, S. I., and Sowers, J. R. (2004) The relationship between hyperinsulinemia, hypertension and progressive renal disease. *J. Am. Soc. Nephrol.* **15**, 2816–2827
 23. Meng, S., Cao, J., He, Q., Xiong, L., Chang, E., Radovick, S., Wondisford, F. E., and He, L. (2015) Metformin activates AMP-activated protein kinase by promoting formation of the $\alpha\beta\gamma$ heterotrimeric complex. *J. Biol. Chem.* **290**, 3793–3802
 24. Wu, Y., Liu, Z., Xiang, Z., Zeng, C., Chen, Z., Ma, X., and Li, L. (2006) Obesity-related glomerulopathy: insights from gene expression profiles of the glomeruli derived from renal biopsy samples. *Endocrinology* **147**, 44–50
 25. Deji, N., Kume, S., Araki, S., Soumura, M., Sugimoto, T., Isshiki, K., Chin-Kanasaki, M., Sakaguchi, M., Koya, D., Haneda, M., Kashiwagi, A., and Uzu, T. (2009) Structural and functional changes in the kidneys of high-fat diet-induced obese mice. *Am. J. Physiol. Renal Physiol.* **296**, F118–F126
 26. Sawicka, K. M., Seeliger, M., Musaev, T., Macri, L. K., and Clark, R. A. (2015) Fibronectin interaction and enhancement of growth factors: importance for wound healing. *Adv. Wound Care (New Rochelle)* **4**, 469–478
 27. Yin, F. C., Spurgeon, H. A., Rakusan, K., Weisfeldt, M. L., and Lakatta, E. G. (1982) Use of tibial length to quantify cardiac hypertrophy: application in the aging rat. *Am. J. Physiol.* **243**, H941–H947
 28. Kumar, S., Pamulapati, H., and Tikoo, K. (2016) Fatty acid induced metabolic memory involves alterations in renal histone H3K36me2 and H3K27me3. *Mol. Cell. Endocrinol.* **422**, 233–242
 29. Bendale, D. S., Karpe, P. A., Chhabra, R., Shete, S. P., Shah, H., and Tikoo, K. (2013) 17- β Oestradiol prevents cardiovascular dysfunction in postmenopausal metabolic syndrome by affecting SIRT1/AMPK/H3 acetylation. *Br. J. Pharmacol.* **170**, 779–795
 30. Karpe, P. A., Gupta, J., Marthong, R. F., Ramarao, P., and Tikoo, K. (2012) Insulin resistance induces a segmental difference in thoracic and abdominal aorta: differential expression of AT1 and AT2 receptors. *J. Hypertens.* **30**, 132–146
 31. Majithiya, J. B., and Balaraman, R. (2006) Metformin reduces blood pressure and restores endothelial function in aorta of streptozotocin-induced diabetic rats. *Life Sci.* **78**, 2615–2624
 32. Towler, M. C., and Hardie, D. G. (2007) AMP-activated protein kinase in metabolic control and insulin signaling. *Circ. Res.* **100**, 328–341
 33. Declèves, A. E., Mathew, A. V., Cunard, R., and Sharma, K. (2011) AMPK mediates the initiation of kidney disease induced by a high-fat diet. *J. Am. Soc. Nephrol.* **22**, 1846–1855
 34. Kim, D., Lee, J. E., Jung, Y. J., Lee, A. S., Lee, S., Park, S. K., Kim, S. H., Park, B. H., Kim, W., and Kang, K. P. (2013) Metformin decreases high-fat diet-induced renal injury by regulating the expression of adipokines and the renal AMP-activated protein kinase/acetyl-CoA carboxylase pathway in mice. *Int. J. Mol. Med.* **32**, 1293–1302
 35. Peairs, A., Radjavi, A., Davis, S., Li, L., Ahmed, A., Giri, S., and Reilly, C. M. (2009) Activation of AMPK inhibits inflammation in MRL/lpr mouse mesangial cells. *Clin. Exp. Immunol.* **156**, 542–551
 36. Crinigan, C., Calhoun, M., and Sweazea, K. L. (2015) Short-term high fat intake does not significantly alter markers of renal function or inflammation in young male Sprague-Dawley rats. *J. Nutr. Metab.* **2015**, 157520
 37. Nayak, B. K., Shanmugasundaram, K., Friedrichs, W. E., Cavaglieri, R. C., Patel, M., Barnes, J., and Block, K. (2016) HIF-1 mediates renal fibrosis in OVE26 type 1 diabetic mice. *Diabetes* **65**, 1387–1397
 38. Shevalye, H., Lupachyk, S., Watcho, P., Stavniichuk, R., Khazim, K., Aboud, H. E., and Obrosova, I. G. (2012) Prediabetic nephropathy as an early consequence of the high-calorie/high-fat diet: relation to oxidative stress. *Endocrinology* **153**, 1152–1161
 39. Grande, M. T., and López-Novoa, J. M. (2009) Fibroblast activation and myofibroblast generation in obstructive nephropathy. *Nat. Rev. Nephrol.* **5**, 319–328
 40. Kume, S., Uzu, T., Araki, S., Sugimoto, T., Isshiki, K., Chin-Kanasaki, M., Sakaguchi, M., Kubota, N., Terauchi, Y., Kadowaki, T., Haneda, M., Kashiwagi, A., and Koya, D. (2007) Role of altered renal lipid metabolism in the development of renal injury induced by a high-fat diet. *J. Am. Soc. Nephrol.* **18**, 2715–2723
 41. Kumar, S., and Tikoo, K. (2015) Independent role of PP2A and mTORc1 in palmitate induced podocyte death. *Biochimie* **112**, 73–84

Selenoprotein Gene Nomenclature*

Received for publication, August 30, 2016, Published, JBC Papers in Press, September 19, 2016, DOI 10.1074/jbc.M116.756155

Vadim N. Gladyshev,^{a,b1} Elias S. Arnér,^c Marla J. Berry,^d Regina Brigelius-Flohé,^e Elspeth A. Bruford,^f Raymond F. Burk,^g Bradley A. Carlson,^h Sergi Castellano,ⁱ Laurent Chavatte,^j Marcus Conrad,^k Paul R. Copeland,^l Alan M. Diamond,^m Donna M. Driscoll,ⁿ Ana Ferreira,^{o,p} Leopold Flohé,^{q,r} Fiona R. Green,^s Roderic Guigó,^{t,u} Diane E. Handy,^v Dolph L. Hatfield,^h John Hesketh,^{w,x,y} Peter R. Hoffmann,^d Arne Holmgren,^c Robert J. Hondal,^z Michael T. Howard,^{aa} Kaixun Huang,^{bb} Hwa-Young Kim,^{cc} Ick Young Kim,^{dd} Josef Köhrle,^{ee} Alain Krol,^{ff} Gregory V. Kryukov,^{gg} Byeong Jae Lee,^{hh} Byung Cheon Lee,^{dd} Xin Gen Lei,ⁱⁱ Qiong Liu,^{jj} Alain Lescure,^{ff,kk} Alexei V. Lobanov,^a Joseph Loscalzo,^{ll} Matilde Maiorino,^f Marco Mariotti,^a K. Sandeep Prabhu,^{mm} Margaret P. Rayman,ⁿⁿ Sharon Rozovsky,^{oo} Gustavo Salinas,^{pp} Edward E. Schmidt,^{qq} Lutz Schomburg,^{ee} Ulrich Schweizer,^{rr} Miljan Simonović,^{ss} Roger A. Sunde,^{tt} Petra A. Tsuji,^{uu} Susan Tweedie,^f Fulvio Ursini,^r Philip D. Whanger,^{vv} and Yan Zhang^{jj}

From the ^aDepartment of Medicine, Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, the ^bBroad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, the ^cDepartment of Medical Biochemistry and Biophysics (MBB), Division of Biochemistry, Karolinska Institutet, SE-171 77, Stockholm, Sweden, the ^dDepartment of Cell and Molecular Biology, John A. Burns School of Medicine, University of Hawaii at Manoa, Honolulu, Hawaii 96813, the ^eGerman Institute of Human Nutrition Potsdam-Rehbruecke, 14558 Nuthetal, Germany, the ^fHUGO Gene Nomenclature Committee (HGNC), European Bioinformatics Institute-European Molecular Biology Laboratory (EMBL-EBI), Hinxton CB10 1SD, United Kingdom, the ^gDepartment of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, the ^hMolecular Biology of Selenium Section, Mouse Cancer Genetics Program, Center for Cancer Research, National Institutes of Health, Bethesda, Maryland 20892, the ⁱDepartment of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany, the ^jCentre International de Recherche en Infectiologie, CIRI, INSERM U1111, and CNRS/ENS UMR5308, 69007 Lyon, France, the ^kHelmholtz Zentrum München, Institute of Developmental Genetics, 85764 Neuherberg, Germany, the ^lDepartment of Biochemistry and Molecular Biology, Rutgers-Robert Wood Johnson Medical School, Piscataway, New Jersey 08854, the ^mDepartment of Pathology, University of Illinois at Chicago, Chicago, Illinois 60607, the ⁿDepartment of Cellular and Molecular Medicine, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, Ohio 44195, the ^oPathophysiology of Striated Muscles Laboratory, Unit of Functional and Adaptive Biology (BFA), University Paris Diderot, Sorbonne Paris Cité, BFA, UMR CNRS 8251, 75250 Paris, France, the ^pAP-HP, Centre de Référence Maladies Neuromusculaires Paris-Est, Groupe Hospitalier Pitié-Salpêtrière, 75013 Paris, France, the ^qUniversidad de la República, Facultad de Medicina, Departamento de Bioquímica, 11800 Montevideo, Uruguay, the ^rDepartment of Molecular Medicine, University of Padova, I-35121 Padova, Italy, the ^sDivision of Cardiovascular Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom, the ^tCentre for Genomic Regulation (CRG), 08003 Barcelona, Spain, the ^uUniversitat Pompeu Fabra (UPF), 08002 Barcelona, Spain, the ^vDepartment of Medicine, Cardiovascular Division, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, the ^wInstitute for Cell and Molecular Biosciences, Newcastle University, Newcastle-upon-Tyne NE1 7RU, United Kingdom, the ^xHuman Nutrition Research Centre, Newcastle University, Newcastle-upon-Tyne NE1 7RU, United Kingdom, the ^yThe Medical School, Newcastle University, Newcastle-upon-Tyne NE2 4HH, United Kingdom, the ^zDepartment of Biochemistry, University of Vermont, Burlington, Vermont 05405, the ^{aa}Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112, the ^{bb}Hubei Key Laboratory of Bioinorganic Chemistry & Materia Medica, School of Chemistry and Chemical Engineering, Huazhong University of Science and Technology, Wuhan 430074, Peoples Republic of China, the ^{cc}Department of Biochemistry and Molecular Biology, Yeungnam University College of Medicine, Daegu 42415, South Korea, the ^{dd}College of Life Sciences and Biotechnology, Korea University, Seoul 02841, South Korea, the ^{ee}Institute for Experimental Endocrinology, Charité-Universitätsmedizin Berlin, D-13353 Berlin, Germany, the ^{ff}Architecture et Réactivité de l'ARN, Université de Strasbourg, Centre National de la Recherche Scientifique, Institut de Biologie Moléculaire et Cellulaire, 67084 Strasbourg, France, the ^{gg}KSQ Therapeutics, Cambridge, Massachusetts 02139, the ^{hh}School of Biological Sciences, Seoul National University, Seoul 151-742, South Korea, the ⁱⁱDepartment of Animal Science, Cornell University, Ithaca, New York 14853, the ^{jj}Shenzhen Key Laboratory of Marine Biotechnology and Ecology, College of Life Science, Shenzhen University, Shenzhen, 518060, Guangdong Province, Peoples Republic of China, the ^{kk}Centre National de la Recherche Scientifique, 75794 Paris, France, the ^{ll}Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, the ^{mm}Department of Veterinary and Biomedical Sciences, The Pennsylvania State University, University Park, Pennsylvania 16802, the ⁿⁿDepartment of Nutritional Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford GU2 7XH, United Kingdom, the ^{oo}Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, the ^{pp}Cátedra de Inmunología, Facultad de Química, Instituto de Higiene, CP11600 Montevideo, Uruguay, the ^{qq}Department of Microbiology and Immunology, Montana State University, Bozeman, Montana 59717, the ^{rr}Rheinische Friedrich-Wilhelms Universität Bonn, Institut für Biochemie und Molekularbiologie, 53115 Bonn, Germany, the ^{ss}Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, Illinois 60607, the ^{tt}Department of Nutritional Sciences, University of Wisconsin, Madison, Wisconsin 53706, the ^{uu}Department of Biological Sciences, Towson University, Towson, Maryland 21252, and the ^{vv}Department of Environmental and Molecular Toxicology, College of Agricultural Sciences, Oregon State University, Corvallis, Oregon 97331

Edited by Norma Allewell

The human genome contains 25 genes coding for selenocysteine-containing proteins (selenoproteins). These proteins are involved in a variety of functions, most notably redox homeostasis. Selenoprotein enzymes with known functions are designated according to these functions: TXNRD1, TXNRD2, and TXNRD3 (thioredoxin reductases), GPX1, GPX2, GPX3, GPX4, and GPX6 (glutathione peroxidases), DIO1, DIO2, and DIO3 (iodothyronine deiodinases), MSRB1 (methionine sulfoxide reductase B1), and SEPHS2 (selenophosphate synthetase 2). Selenoproteins without known functions have traditionally been denoted by SEL or SEP symbols. However, these symbols are sometimes ambiguous and conflict with the approved nomenclature for several other genes. Therefore, there is a need to implement a rational and coherent nomenclature system for selenoprotein-encoding genes. Our solution is to use the root symbol SELENO followed by a letter. This nomenclature applies to SELENOF (selenoprotein F, the 15-kDa selenoprotein, SEP15), SELENOH (selenoprotein H, SELH, C11orf31), SELENOI (selenoprotein I, SELI, EPT1), SELENOK (selenoprotein K, SELK), SELENOM (selenoprotein M, SELM), SELENON (selenoprotein N, SEPNI, SELN), SELENOO (selenoprotein O, SELO), SELENOP (selenoprotein P, SeP, SEPP1, SELP), SELENOS (selenoprotein S, SELS, SEPS1, VIMP), SELENOT (selenoprotein T, SELT), SELENOV (selenoprotein V, SELV), and SELENOW (selenoprotein W, SELW, SEPW1). This system, approved by the HUGO Gene Nomenclature Committee, also resolves conflicting, missing, and ambiguous designations for selenoprotein genes and is applicable to selenoproteins across vertebrates.

Selenium is an essential trace element in humans, which is present in proteins in the form of the 21st proteinogenic amino acid, selenocysteine (Sec).² Sec is co-translationally inserted into a polypeptide chain in response to in-frame UGA codons directed by the Sec insertion sequence element, a stem-loop structure in the 3'-UTRs of selenoprotein mRNAs. The human genome contains 25 selenoprotein genes (1), and selenoproteins are essential for embryo development and human health (2, 3). Among the selenoproteins, 13 have known functions; at least 12 of them serve as oxidoreductases, wherein Sec is the catalytic redox-active residue. The redox theme is also common for selenoproteins in other organisms (4).

The remaining 12 selenoproteins either have no known function, or their functions are only partially established. One of the selenoproteins, selenoprotein P (5), requires special mention as

it has more than one Sec. It is a major plasma selenoprotein that delivers selenium primarily from the liver to other organs (6, 7), and is involved in selenium transport and metabolism within organs. However, this protein also has an N-terminal Sec-containing thioredoxin domain similar to that found in most selenoproteins with known functions, which points to a potential redox function. Several other selenoproteins, including selenoproteins H, M, T, V, W, and Sep15, also possess thioredoxin-like domains, suggesting redox-related functions (8).

Selenoproteins are not all homologous, but are characterized by their incorporation of Sec. Historically they have been given designations by the groups that discovered them, e.g. because of its presence in plasma the respective selenoprotein was named selenoprotein P (9, 10), or because of its size another protein was called the 15-kDa selenoprotein or Sep15 (11). However, some selenoproteins were identified independently by two or more groups, which created confusion and discrepancies in the field. For example, the same protein was named selenoprotein R by one group (12), but discovered concurrently and designated by another group as selenoprotein X (13). This protein was then functionally characterized (14) and renamed MsrB1 (for methionine-*R*-sulfoxide reductase 1) (15), but all three designations persist in the literature and/or databases. Another problematic example is the nomenclature used for thioredoxin reductases. The names for the first thioredoxin reductase, which had been known decades before its selenoprotein nature was discovered (16), are generally internally consistent, although they differ in the abbreviations used, e.g. TR1 and TrxR1 (17). The second and third thioredoxin reductases discovered, however, were named inconsistently by the authors, wherein the mitochondrial thioredoxin reductase was designated as TrxR2 (18) and TR3 (19), and the testis-specific thioredoxin-glutathione reductase has been alternatively labeled as TR2 (19), TrxR3, or TGR.

Designations are also confusing for several other selenoproteins. For example, selenoprotein S was named SelS (1), but a later paper introduced the designation VIMP (20). Similarly, selenoprotein H was named SelH (1), but also C11orf31, and selenoprotein I was named SelI (1), but also called EPT1 (21). To avoid confusion, and at the instigation of the HUGO Gene Nomenclature Committee (HGNC), we describe a new standardized designation system for human (and other vertebrate) selenoproteins.

Results and Discussion

Resolving the Nomenclature of Selenoprotein Genes—Human gene designations are approved by the HUGO Gene Nomenclature Committee (HGNC), and genes in other mammals follow the same designations. Selenoproteins have traditionally been published using SEL or SEP symbols followed by a letter or a number. Unfortunately, for naming the genes encoding these proteins, the SEL root was not an option as it was already approved for the selectin gene family; for example, *SELP* is the approved gene symbol for selectin P (P-selectin) and not selenoprotein P. Some selenoprotein genes had been approved using the root SEP (i.e. *SEPNI*, *SEPP1*, and *SEPW1*) but this could not be utilized for all selenoproteins as selenoprotein T gene would then be *SEPT* or *SEPT1*, and *SEPT#* is already used

* This work was supported, in whole or in part, by National Institutes of Health Grant GM061603. The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

¹ To whom correspondence should be addressed. E-mail: vgladyshv@rics.bwh.harvard.edu.

² The abbreviations used are: Sec, selenocysteine; TXNRD, thioredoxin reductases; GPX, glutathione peroxidase; DIO, iodothyronine deiodinase; SEPHS2, selenophosphate synthetase 2; SELENOF, 15-kDa selenoprotein; SELENOH, selenoprotein H; SELENOI, selenoprotein I; SELENOK, selenoprotein K; SELENOM, selenoprotein M; SELENON, selenoprotein N; SELENOO, selenoprotein O; SELENOP, selenoprotein P; SELENOS, selenoprotein S; SELENOT, selenoprotein T; SELENOV, selenoprotein V; SELENOW, selenoprotein W; HGNC, HUGO Gene Nomenclature Committee.

Selenoprotein Gene Nomenclature

for the septin genes. HGNC does not use the same root for unrelated groups of genes (*e.g.* SEL for selectins and selenoproteins) and does not endorse the use of multiple root symbols for genes sharing a common name (*e.g.* SEP and SEL for selenoprotein). With a view to solving these issues, HGNC approached selenoprotein researchers to propose a new unifying root symbol for all selenoprotein genes.

Proposal for a New Nomenclature—We propose that all selenoproteins (except those that have been functionally characterized, *e.g.* with enzymatic activity) use the root symbol SELENO followed by a letter. This gene nomenclature is designed to highlight selenium, the key functional site in these proteins, and to provide a new and unambiguous root for these genes. The new nomenclature applies to 12 human selenoprotein genes as detailed in Table 1. Selenoproteins with known functions will continue to use the same designations (Table 2). Once functions are established for other selenoproteins, they may be renamed, as required. The proposed designations apply to the selenoprotein genes; although the same designations may be used for many of the encoded proteins, traditional names of selenoproteins, *e.g.* selenoprotein P, may also be used.

Selenoprotein Gene Designation in Other Species—The new HGNC nomenclature will automatically be used to designate orthologous selenoprotein genes in other vertebrates and extended to accommodate selenoprotein genes with no orthologs in human (22) (Table 3). Where vertebrate gene

duplications have occurred, the additional paralogs will be named in line with the human genes, but with suffixes on the symbols, *e.g.* zebrafish selenot1a, selenot1b, and selenot2. Selenoproteins are widespread in all three domains of life. Despite the fact that land plants, yeast, and some other species have lost selenoprotein biosynthesis pathways, a unifying nomenclature beyond vertebrates might be desirable. We suggest using the human nomenclature described in this paper for orthologs of vertebrate selenoprotein genes. This nomenclature may also be extended to accommodate additional selenoprotein genes as they are discovered. Although we use human designations in this paper, we note that most vertebrates use all uppercase letters for genes and proteins (italics for genes), rodents use title case for genes (uppercase for proteins), *Xenopus* and zebrafish use lowercase for genes and title case for proteins, and Anolis use lowercase for genes and uppercase for proteins.

Designations of Proteins That Do Not Contain Selenocysteine—There exists another class of selenium-containing proteins, those which contain a bound atom of selenium but do not contain a UGA-encoded Sec, for which there is also ambiguous nomenclature. For example, selenium-binding protein 1 (SBP1), also referred to as SELENBP1 or hSP56, is one such protein (23). The naming of such proteins will not be included in the new nomenclature as they lack Sec. Similarly, the machinery for Sec biosynthesis and insertion will not be renamed.

Implementation—The new selenoprotein gene nomenclature has been approved by the HGNC, can be found on their website (<http://www.genenames.org/cgi-bin/genefamilies/set/890>), and will be found in all major genomic resources in due course. We recommend that future publications primarily use

TABLE 1

Selenoprotein genes using the new SELENO root

New HGNC selenoprotein gene nomenclature is indicated in the column "symbol." Previous HGNC symbols (shown with *) will become synonyms, along with other previously used designations.

Symbol	Name	Synonyms	Refs.
SELENOF	Selenoprotein F	Selenoprotein 15, SEP15	11
SELENOH	Selenoprotein H	SELH, C11orf31*	1
SELENOI	Selenoprotein I	SELI, EPT1*	1, 21
SELENOK	Selenoprotein K	SELK	1
SELENO M	Selenoprotein M	SELM, SEPM	25
SELENON	Selenoprotein N	SEPN1*, SELN	13
SELENOO	Selenoprotein O	SELO	1
SELENO P	Selenoprotein P	SEPP1*, SeP, SELP, SEPP	26
SELENO S	Selenoprotein S	SELS, SEPS1, VIMP*	1
SELENOT	Selenoprotein T	SELT	12
SELENOV	Selenoprotein V	SELV	1
SELENOW	Selenoprotein W	SELW, SEPW1*	27

TABLE 2

Selenoprotein genes named based on encoded enzymatic activity

Note that the nomenclature of these genes will not be changing to use the SELENO root.

Symbol	Name	Synonyms	Refs.
TXNRD1	Thioredoxin reductase 1	TR1, TRXR1	16, 28, 29
TXNRD2	Thioredoxin reductase 2	TRXR2, TR3, mitochondrial thioredoxin reductase	18, 19
TXNRD3	Thioredoxin-glutathione reductase	TGR, TRXR3, TR2	19
GPX1	Glutathione peroxidase 1	Cytosolic glutathione peroxidase, GSHPX1	30–35
GPX2	Glutathione peroxidase 2	GSHPX-GI	36
GPX3	Glutathione peroxidase 3	Plasma glutathione peroxidase	37
GPX4	Glutathione peroxidase 4	Phospholipid hydroperoxide glutathione peroxidase, PHGPX	38, 39
GPX6	Glutathione peroxidase 6		
DIO1	Iodothyronine deiodinase 1	D1	40, 41
DIO2	Iodothyronine deiodinase 2	D2	42
DIO3	Iodothyronine deiodinase 3	D3	43
MSRB1	Methionine sulfoxide reductase B1	SELR, SELX, SEPX1	12–14
SEPHS2	Selenophosphate synthetase 2	SPS2	44

TABLE 3

Vertebrate selenoprotein genes absent in human and mouse

New selenoprotein gene nomenclature is indicated in the column "symbol."

Symbol	Name	Synonyms	Refs.
SELENOJ	Selenoprotein J	SELJ	45
SELENOU	Selenoprotein U	SELU	46
SELENO L	Selenoprotein L	SELL	47
SELENOE	Selenoprotein E, fish selenoprotein 15	FEP15	48
SELENO P2	Selenoprotein P2	SEPP2, SELPb	49, 24

the new SELENO designations, but supplement them (as secondary designations/synonyms) with the names previously used by the community. Once the new nomenclature is consistently used, the old designations will no longer be needed. We hope that other researchers in the field will join us in implementing this new nomenclature.

Author Contributions—The article was drafted by V. N. G. in consultation with other authors. All authors contributed to revisions and discussion.

References

- Kryukov, G. V., Castellano, S., Novoselov, S. V., Lobanov, A. V., Zehtab, O., Guigó, R., and Gladyshev, V. N. (2003) Characterization of mammalian selenoproteomes. *Science* **300**, 1439–1443
- Bösl, M. R., Takaku, K., Oshima, M., Nishimura, S., and Taketo, M. M. (1997) Early embryonic lethality caused by targeted disruption of the mouse selenocysteine tRNA gene (Trsp). *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5531–5534
- Schweizer, U., and Fradejas-Villar, N. (July 29, 2016) Why 21? the significance of selenoproteins for human health revealed by inborn errors of metabolism. *FASEB J.* fj201600414
- Fomenko, D. E., Xing, W., Adair, B. M., Thomas, D. J., and Gladyshev, V. N. (2007) High-throughput identification of catalytic redox-active cysteine residues. *Science* **315**, 387–389
- Burk, R. F., and Hill, K. E. (2015) Regulation of selenium metabolism and transport. *Annu. Rev. Nutr.* **35**, 109–134
- Carlson, B. A., Novoselov, S. V., Kumaraswamy, E., Lee, B. J., Anver, M. R., Gladyshev, V. N., and Hatfield, D. L. (2004) Specific excision of the selenocysteine tRNA[Ser]Sec (Trsp) gene in mouse liver demonstrates an essential role of selenoproteins in liver function. *J. Biol. Chem.* **279**, 8011–8017
- Schomburg, L., Schweizer, U., Holtmann, B., Flohé, L., Sendtner, M., and Köhrle, J. (2003) Gene disruption discloses role of selenoprotein P in selenium delivery to target tissues. *Biochem. J.* **370**, 397–402
- Dikiy, A., Novoselov, S. V., Fomenko, D. E., Sengupta, A., Carlson, B. A., Cerny, R. L., Ginalski, K., Grishin, N. V., Hatfield, D. L., and Gladyshev, V. N. (2007) SelT, SelW, SelH, and Rdx12: genomics and molecular insights into the functions of selenoproteins of a novel thioredoxin-like family. *Biochemistry* **46**, 6871–6882
- Motsenbocker, M. A., and Tappel, A. L. (1982) A selenocysteine-containing selenium-transport protein in rat plasma. *Biochim. Biophys. Acta* **719**, 147–153
- Burk, R. F., and Gregory, P. E. (1982) Some characteristics of 75Se-P, a selenoprotein found in rat liver and plasma, and comparisons of it with selenogluthione peroxidase. *Arch. Biochem. Biophys.* **213**, 73–80
- Gladyshev, V. N., Jeang, K. T., Wootton, J. C., and Hatfield, D. L. (1998) A new human selenium-containing protein. Purification, characterization, and cDNA sequence. *J. Biol. Chem.* **273**, 8910–8915
- Kryukov, G. V., Kryukov, V. M., and Gladyshev, V. N. (1999) New mammalian selenocysteine-containing proteins identified with an algorithm that searches for selenocysteine insertion sequence elements. *J. Biol. Chem.* **274**, 33888–33897
- Lescure, A., Gautheret, D., Carbon, P., and Krol, A. (1999) Novel selenoproteins identified *in silico* and *in vivo* by using a conserved RNA structural motif. *J. Biol. Chem.* **274**, 38147–38154
- Kryukov, G. V., Kumar, R. A., Koc, A., Sun, Z., and Gladyshev, V. N. (2002) Selenoprotein R is a zinc-containing stereo-specific methionine sulfoxide reductase. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4245–4250
- Kim, H. Y., and Gladyshev, V. N. (2004) Methionine sulfoxide reduction in mammals: characterization of methionine-R-sulfoxide reductases. *Mol. Biol. Cell* **15**, 1055–1064
- Holmgren, A. (1977) Bovine thioredoxin system: purification of thioredoxin reductase from calf liver and thymus and studies of its function in disulfide reduction. *J. Biol. Chem.* **252**, 4600–4606
- Arner, E. S., and Holmgren, A. (2000) Physiological functions of thioredoxin and thioredoxin reductase. *Eur. J. Biochem.* **267**, 6102–6109
- Lee, S. R., Kim, J. R., Kwon, K. S., Yoon, H. W., Levine, R. L., Ginsburg, A., and Rhee, S. G. (1999) Molecular cloning and characterization of a mitochondrial selenocysteine-containing thioredoxin reductase from rat liver. *J. Biol. Chem.* **274**, 4722–4734
- Sun, Q. A., Wu, Y., Zappacosta, F., Jeang, K. T., Lee, B. J., Hatfield, D. L., and Gladyshev, V. N. (1999) Redox regulation of cell signaling by selenocysteine in mammalian thioredoxin reductases. *J. Biol. Chem.* **274**, 24522–24530
- Ye, Y., Shibata, Y., Yun, C., Ron, D., and Rapoport, T. A. (2004) A membrane protein complex mediates retro-translocation from the ER lumen into the cytosol. *Nature* **429**, 841–847
- Horibata, Y., and Hirabayashi, Y. (2007) Identification and characterization of human ethanolaminephosphotransferase1. *J. Lipid Res.* **48**, 503–508
- Mariotti, M., Ridge, P. G., Zhang, Y., Lobanov, A. V., Pringle, T. H., Guigo, R., Hatfield, D. L., and Gladyshev, V. N. (2012) Composition and evolution of the vertebrate and mammalian selenoproteomes. *PLoS ONE* **7**, e33066
- Ansong, E., Yang, W., and Diamond, A. M. (2014) Molecular cross-talk between members of distinct families of selenium containing proteins. *Mol. Nutr. Food Res.* **58**, 117–123
- Sunde, R. A., Sunde, G. R., Sunde, C. M., Sunde, M. L., and Evenson, J. K. (2015) Cloning, sequencing, and expression of selenoprotein transcripts in the turkey (*Meleagris gallopavo*). *PLoS ONE* **10**, e0129801
- Korotkov, K. V., Novoselov, S. V., Hatfield, D. L., and Gladyshev, V. N. (2002) Mammalian selenoprotein in which selenocysteine (Sec) incorporation is supported by a new form of Sec insertion sequence element. *Mol. Cell. Biol.* **22**, 1402–1411
- Hill, K. E., Lloyd, R. S., and Burk, R. F. (1993) Conserved nucleotide sequences in the open reading frame and 3' untranslated region of selenoprotein P mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 537–541
- Vendeland, S. C., Beilstein, M. A., Yeh, J. Y., Ream, W., and Whanger, P. D. (1995) Rat skeletal muscle selenoprotein W: cDNA clone and mRNA modulation by dietary selenium. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8749–8753
- Gasdaska, P. Y., Gasdaska, J. R., Cochran, S., and Powis, G. (1995) Cloning and sequencing of a human thioredoxin reductase. *FEBS Lett.* **373**, 5–9
- Gladyshev, V. N., Jeang, K. T., and Stadtman, T. C. (1996) Selenocysteine, identified as the penultimate C-terminal residue in human T-cell thioredoxin reductase, corresponds to TGA in the human placental gene. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6146–6151
- Rotruck, J. T., Pope, A. L., Ganther, H. E., Swanson, A. B., Hafeman, D. G., and Hoekstra, W. G. (1973) Selenium: biochemical role as a component of glutathione peroxidase. *Science* **179**, 588–590
- Flohe, L., Günzler, W. A., and Schock, H. H. (1973) Glutathione peroxidase: a selenoenzyme. *FEBS Lett.* **32**, 132–134
- Forstrom, J. W., Zakowski, J. J., and Tappel, A. L. (1978) Identification of the catalytic site of rat liver glutathione peroxidase as selenocysteine. *Biochemistry* **17**, 2639–2644
- Chambers, L., Frampton, J., Goldfarb, P., Affara, N., McBain, W., and Harrison, P. R. (1986) The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the “termination” codon, TGA. *EMBO J.* **5**, 1221–1227
- Mills, G. C. (1957) Hemoglobin catabolism: I. glutathione peroxidase, an erythrocyte enzyme which protects hemoglobin from oxidative breakdown. *J. Biol. Chem.* **229**, 189–197
- Günzler, W. A., Steffens, G. J., Grossmann, A., Kim, S. M., Otting, F., Wendel, A., and Flohé, L. (1984) The amino acid sequence of bovine glutathione peroxidase. *Hoppe-Seyler's Z. Physiol. Chem.* **365**, 195–212
- Chu, F. F., Doroshov, J. H., and Esworthy, R. S. (1993) Expression, characterization, and tissue distribution of a new cellular selenium-dependent glutathione peroxidase, GSHPx-GI. *J. Biol. Chem.* **268**, 2571–2576
- Takahashi, K., Akasaka, M., Yamamoto, Y., Kobayashi, C., Mizoguchi, J., and Koyama, J. (1990) Primary structure of human plasma glutathione peroxidase deduced from cDNA sequences. *J. Biochem.* **108**, 145–148

Selenoprotein Gene Nomenclature

38. Ursini, F., Maiorino, M., and Gregolin, C. (1985) The selenoenzyme phospholipid hydroperoxide glutathione peroxidase. *Biochim. Biophys. Acta* **839**, 62–70
39. Brigelius-Flohé, R., Aumann, K. D., Blöcker, H., Gross, G., Kiess, M., Klöppel, K. D., Maiorino, M., Roveri, A., Schuckelt, R., Ursini, F., Wingender, E., and Flohé, L. (1994) Phospholipid hydroperoxide glutathione peroxidase: genomic DNA, cDNA and deduced amino acid sequence. *J. Biol. Chem.* **269**, 7342–7348
40. Berry, M. J., Banu, L., and Larsen, P. R. (1991) Type I iodothyronine deiodinase is a selenocysteine-containing enzyme. *Nature* **349**, 438–440
41. Behne, D., Kyriakopoulos, A., Meinhold, H., and Köhrle, J. (1990) Identification of type I iodothyronine 5'-deiodinase as a selenoenzyme. *Biochem. Biophys. Res. Commun.* **173**, 1143–1149
42. Davey, J. C., Becker, K. B., Schneider, M. J., St Germain, D. L., and Galton, V. A. (1995) Cloning of a cDNA for the type II iodothyronine deiodinase. *J. Biol. Chem.* **270**, 26786–26789
43. St Germain, D. L., Schwartzman, R. A., Croteau, W., Kanamori, A., Wang, Z., Brown, D. D., and Galton, V. A. (1994) A thyroid hormone-regulated gene in *Xenopus laevis* encodes a type III iodothyronine 5-deiodinase. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7767–7771
44. Guimarães, M. J., Peterson, D., Vicari, A., Cocks, B. G., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Ferrick, D. A., Kastelein, R. A., Bazan, J. F., and Zlotnik, A. (1996) Identification of a novel selD homolog from eukaryotes, bacteria, and archaea: is there an autoregulatory mechanism in selenocysteine metabolism? *Proc. Natl. Acad. Sci. U.S.A.* **93**, 15086–15091
45. Castellano, S., Lobanov, A. V., Chapple, C., Novoselov, S. V., Albrecht, M., Hua, D., Lescure, A., Lengauer, T., Krol, A., Gladyshev, V. N., and Guigó, R. (2005) Diversity and functional plasticity of eukaryotic selenoproteins: identification and characterization of the SelJ family. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16188–16193
46. Castellano, S., Novoselov, S. V., Kryukov, G. V., Lescure, A., Blanco, E., Krol, A., Gladyshev, V. N., and Guigó, R. (2004) Reconsidering the evolution of eukaryotic selenoproteins: a novel nonmammalian family with scattered phylogenetic distribution. *EMBO Rep.* **5**, 71–77
47. Shchedrina, V. A., Novoselov, S. V., Malinouski, M. Y., and Gladyshev, V. N. (2007) Identification and characterization of a selenoprotein family containing a diselenide bond in a redox motif. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13919–13924
48. Novoselov, S. V., Hua, D., Lobanov, A. V., and Gladyshev, V. N. (2006) Identification and characterization of Fep15, a new selenocysteine-containing member of the Sep15 protein family. *Biochem. J.* **394**, 575–579
49. Kryukov, G. V., and Gladyshev, V. N. (2000) Selenium metabolism in zebrafish: multiplicity of selenoprotein genes and expression of a protein containing 17 selenocysteine residues. *Genes Cells* **5**, 1049–1060

Accumulation of Peptidoglycan O-Acetylation Leads to Altered Cell Wall Biochemistry and Negatively Impacts Pathogenesis Factors of *Campylobacter jejuni*^{*[5]}

Received for publication, July 8, 2016 Published, JBC Papers in Press, July 29, 2016, DOI 10.1074/jbc.M116.746404

Reuben Ha^{†1}, Emilisa Firdich[‡], David Sychantha[§], Jacob Biboy[¶], Michael E. Taveirne^{||2}, Jeremiah G. Johnson^{||3}, Victor J. DiRita^{||4}, Waldemar Vollmer[¶], Anthony J. Clarke[§], and Erin C. Gaynor^{†5}

From the [†]Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada, the [§]Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada, the [¶]Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4AX, United Kingdom, and the ^{||}Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan 48109

Campylobacter jejuni is a leading cause of bacterial gastroenteritis in the developed world. Despite its prevalence, its mechanisms of pathogenesis are poorly understood. Peptidoglycan (PG) is important for helical shape, colonization, and host-pathogen interactions in *C. jejuni*. Therefore, changes in PG greatly impact the physiology of this organism. O-acetylation of peptidoglycan (OAP) is a bacterial phenomenon proposed to be important for proper cell growth, characterized by acetylation of the C6 hydroxyl group of *N*-acetylmuramic acid in the PG glycan backbone. The OAP gene cluster consists of a PG O-acetyltransferase A (*patA*) for translocation of acetate into the periplasm, a PG O-acetyltransferase B (*patB*) for O-acetylation, and an O-acetylpeptidoglycan esterase (*ape1*) for de-O-acetylation. In this study, reduced OAP in Δ *patA* and Δ *patB* had minimal impact on *C. jejuni* growth and fitness under the conditions tested. However, accumulation of OAP in Δ *ape1* resulted in marked differences in PG biochemistry, including O-acetylation, anhydromuropeptide levels, and changes not expected to result directly from Ape1 activity. This suggests that OAP may be a form of substrate level regulation in PG biosyn-

thesis. Ape1 acetyltransferase activity was confirmed *in vitro* using *p*-nitrophenyl acetate and O-acetylated PG as substrates. In addition, Δ *ape1* exhibited defects in pathogenesis-associated phenotypes, including cell shape, motility, biofilm formation, cell surface hydrophobicity, and sodium deoxycholate sensitivity. Δ *ape1* was also impaired for chick colonization and adhesion, invasion, intracellular survival, and induction of IL-8 production in INT407 cells *in vitro*. The importance of Ape1 in *C. jejuni* biology makes it a good candidate as an antimicrobial target.

Campylobacter jejuni is a leading bacterial cause of foodborne gastroenteritis in the developed world and the most common infectious antecedent to the autoimmune acute polyneuropathy Guillain-Barré syndrome (1, 2). As a commensal of the avian gut, it is a prevalent contaminant of uncooked poultry (3). Because of its high incidence rate, the costs of *C. jejuni* infection are a significant socioeconomic burden, making it both a health care concern and an economic issue (4). In addition, *C. jejuni* has been exhibiting alarming increases in resistance to ciprofloxacin and erythromycin, the most commonly used antibiotics for treatment of *C. jejuni* infection (5). Despite its prevalence, relatively little is known about *C. jejuni* pathogenesis in humans. Traditional virulence factors present in other gastrointestinal pathogens are either absent (*i.e.* dedicated type III secretion systems) or limited (*C. jejuni* possesses some stress-response elements such as the stringent response modulator SpoT, but it lacks several hallmark stress-response elements like RpoS and RpoE), or their role in pathogenicity is debated (*i.e.* the cytolethal distending toxin and a putative type IV secretion system on the pVIR plasmid) (6–11). However, factors considered to be fundamental aspects of bacterial physiology such as metabolism, stress response, and cell shape are hypothesized to play important roles in *C. jejuni* pathogenesis (12, 13).

The peptidoglycan (PG)⁶ sacculus is a heteropolymer of the bacterial cell wall composed of alternating β -1–4 *N*-acetylglu-

^{*} This work was supported in part by Canadian Institutes of Health Research Grants MOP-68981 (to E. C. G.) and TGC-114045 (to A. J. C.), Wellcome Trust Grant 101824/Z/13/Z (to W. V.), and National Institutes of Health NIAID Grant AI069383 (to V. J. D.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

[‡] Author's Choice—Final version free via Creative Commons CC-BY license.

^[5] This article contains supplemental text S1 and S2 and Tables S1 and S2.

¹ Supported by a Natural Sciences and Engineering Research Council Canada Graduate Scholarship.

² Supported in part with federal funds from the United States Department of Agriculture NIFA Grant 2012-67012-19704 and National Center for Advancing Translational Sciences Grant UL1TR000433. Present address: Dept. of Biological Sciences, North Carolina State University, Raleigh, NC 27695.

³ Supported by National Institutes of Health Molecular Mechanisms of Microbial Pathogenesis Training Program Grant T32 AI007528 and United States Department of Agriculture, National Institute for Food and Agriculture, Grants 2010-65201-20594 and 2013-67012-21136. Present address: Dept. of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824.

⁴ Present address: Dept. of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824.

⁵ To whom correspondence should be addressed: Dept. of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada. Tel.: 604-822-2710; Fax: 604-822-6041; E-mail: egaynor@mail.ubc.ca.

⁶ The abbreviations used are: PG, peptidoglycan; OAP, O-acetylation of peptidoglycan; MurNAc, *N*-acetylmuramic acid; LT, lytic transglycosylase; anHMP, anhydromuropeptide; pNPAC, *p*-nitrophenyl acetate; DICM, differential interference contrast microscopy; PCA, principal component analysis; MEM, minimum essential medium; DOC, deoxycholate; Gm, gentamicin; KS, Kolmogorov-Smirnov; MH, Mueller-Hinton; Cm, chloramphenicol; Km, kanamycin.

cosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) residues cross-linked to adjacent strands by peptides bound to the MurNAc residue. It is responsible for providing structural strength to the cell, enabling it to resist changes in osmotic pressure, and for maintenance of cell shape (14–16). The corkscrew motility of *C. jejuni* generated by its helical shape and polar flagella is thought to be important in enhancing its ability to move through viscous media, such as the mucous layer of the gastrointestinal tract (14, 17). Deletion of PG hydrolase enzymes Pgp1 and Pgp2 in *C. jejuni* has led to a change in morphology from helical to straight with accompanying defects in traits associated with pathogenesis, including motility in soft agar, biofilm formation, and chick colonization. PG isolated from Δ pgp1 and Δ pgp2 also exhibited altered stimulation of host cell NOD receptors, and Δ pgp1 elicits an enhanced proinflammatory IL-8 response from INT407 epithelial cells upon infection (18, 19). Changes in PG biosynthesis and composition as well as the release of PG products have long been known to affect physiological and pathogenic properties of many bacterial species (20), including *Listeria monocytogenes* (21, 22), *Helicobacter pylori* (23), and *Streptococcus pneumoniae* (24). Current research continues to support this concept (15). Some recent studies have shown that changes in morphology and PG structure in *Mycobacterium tuberculosis* affect its physiology and virulence in mice (25); changes in morphology in *Proteus mirabilis* affect its swarming motility (26); and changes in morphology in *Helicobacter pylori* alter its motility and colonization potential (27, 28).

Understanding PG biosynthetic mechanisms in *C. jejuni* may prove advantageous to the development of new antimicrobials. It has been suggested that O-acetylation of peptidoglycan (OAP) machinery may be an attractive target (29–33). OAP occurs in both Gram-positive and Gram-negative bacteria and is characterized by the addition of an acetyl group to the C6 hydroxyl group of MurNAc in the PG glycan backbone (Fig. 1A). This modification confers resistance to lysozyme (34, 35), which cleaves β -1,4-glycosidic bonds between MurNAc and GlcNAc (36). Despite the intrinsic resistance to lysozyme provided by the outer membrane of Gram-negative bacteria, lysozyme resistance was shown to be important in *H. pylori* using mutants defective in OAP addition and similar glycan modifications (37). These strains showed increased susceptibility to physiologically relevant concentrations of lysozyme in the presence of the host lactoferrin, which has membrane permeabilization properties (38, 39). O-Acetylated gonococcal PG is implicated in the development of arthritic symptoms and is hypothesized to be attributable to increased PG hydrolase resistance leading to large fragments of circulating PG (40). In addition, OAP is believed to be involved in the regulation of PG turnover by inhibiting endogenous lytic transglycosylase (LT) activity. LTs require an unmodified MurNAc C6 hydroxyl moiety to cleave β -1,4-glycosidic bonds between MurNAc and GlcNAc, generating anhydromuropeptides (anhMPs). LTs are therefore important for generating insertion sites for newly synthesized muropeptides during cell growth and division (41).

The OAP gene cluster was initially identified in *Neisseria gonorrhoeae* and was found to be responsible for OAP in many Gram-negative pathogens (42). It consists of a putative trans-

membrane protein, PG O-acetyltransferase A (PatA) predicted to be involved in the translocation of acetyl moieties from a cytoplasmic source into the periplasm, a periplasmic transferase, PG O-acetyltransferase B responsible for O-acetylation of MurNAc, and a periplasmic O-acetylpeptidoglycan esterase (Ape1) for MurNAc de-O-acetylation (Fig. 1A and B) (42–44). Since their discovery, PatB and Ape1 from *N. gonorrhoeae* and *Neisseria meningitidis* have been well characterized biochemically, including descriptions of the catalytic residues, mechanism of enzyme activity, and substrate specificity (30, 32, 43, 45). The availability of a recently solved crystal structure for *N. meningitidis* Ape1, recently identified Ape1 inhibitors, and studies showing reduced septicemia in mice infected with *N. meningitidis* ape1 mutants lend support for the application of Ape1 as an antimicrobial target (31, 33, 46).

C. jejuni encodes previously unstudied homologs of the OAP genes *patA*, *patB*, and *ape1*. In this work, the roles of these genes in PG O-acetylation and overall PG biosynthesis, as well as biological and pathogenic attributes were assessed via construction of strains deleted for each or all of these genes followed by biochemical and phenotypic analyses. Each mutant exhibited changes in PG O-acetylation consistent with predicted gene product functions. The accumulation of O-acetylated PG was found to be detrimental to *C. jejuni* fitness, whereas diminished O-acetylation had little to no effect. Δ ape1 exhibited defects in PG muropeptide composition, cell morphology, pathogenic attributes, and host-pathogen interactions, whereas Δ patA, Δ patB, and Δ oap mutants exhibited no or, in rare cases, only minimal defects for these phenotypes.

Results

C. jejuni OAP Genes Were Identified by BLAST and Mutant and Complemented Strains Were Generated—The OAP gene cluster was identified in *C. jejuni* 81-176 wild type by BLAST analysis using the *N. gonorrhoeae* OAP gene sequences. The loci identified were *cjj81176_0640*, *cjj81176_0639*, and *cjj81176_0638* for *patA*, *patB*, and *ape1*, respectively (Fig. 1B). Amino acid sequence identity and sequence similarity for these genes were 35/53%, 39/57%, and 26/44% identity/similarity to *N. gonorrhoeae* *patA*, *patB*, and *ape1*, respectively.

To investigate the role of OAP in *C. jejuni*, the *patA*, *patB*, and *ape1* homologs, as well as the entire gene cluster, were inactivated by deletion-insertion mutagenesis with the non-polar Km^R cassette (*aphA-3*) from pUC18K-2 lacking a transcriptional termination site (Fig. 1C; supplemental text S1) (47). Complementation was achieved using the pRRC integration vector (48). For complementation, the coding region of each OAP gene plus upstream sequence containing the ribosomal binding site was inserted into the genome of the corresponding mutant at ribosomal intergenic regions along with a Cm^R cassette (the Δ ape1 complemented strain, designated Δ ape1^C, is shown in Fig. 1D as an example). Expression of the wild-type genes at the rRNA site was driven from the promoter of the Cm^R cassette.

Growth rate experiments performed on all mutant strains indicated no significant difference in growth rates in broth cultures up to 36 h (data not shown) with the exception of Δ ape1^C, which grew at a slower rate (although this slower growth rate

Role of *C. jejuni* Peptidoglycan O-Acetylation

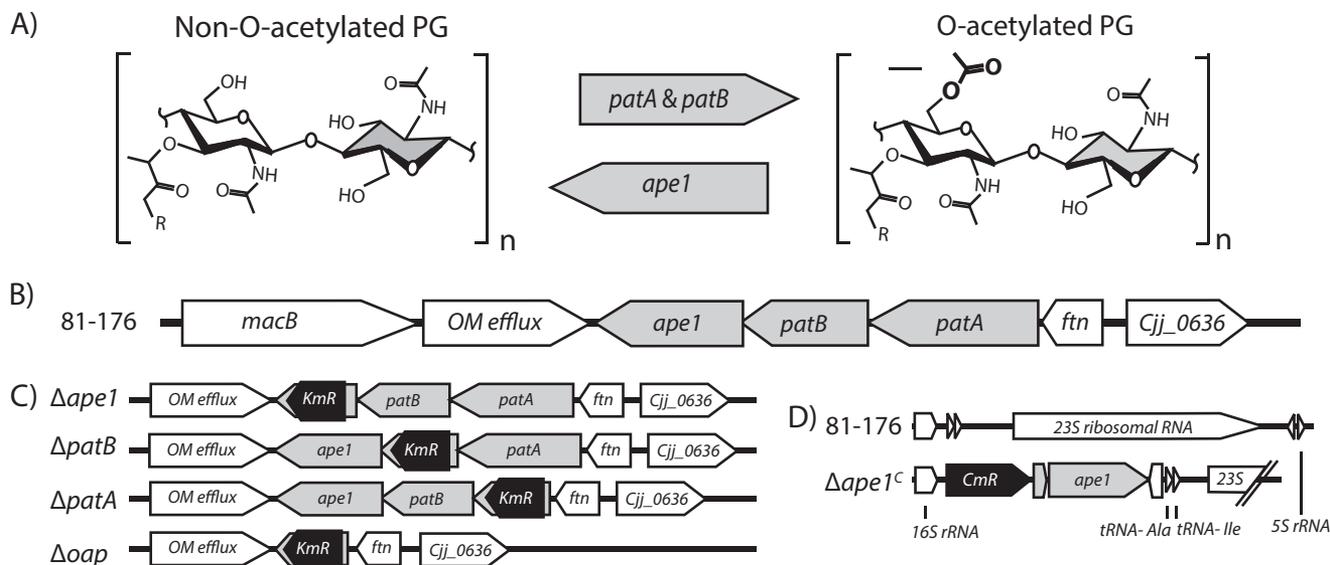


FIGURE 1. Location of O-acetyl groups on peptidoglycan subunits, organization of the *C. jejuni* OAP gene cluster, and description of deletion mutant and complement construction. A, structures of the disaccharide muropeptides showing non-O-acetylated PG and O-acetylated PG, location of O-acetylation (arrow), and the putative involvement of the *oap* genes. B, genomic organization of the *C. jejuni* OAP gene cluster in the 81-176 wild-type strain (gray). *cjj81176_0638*, *cjj81176_0639*, and *cjj81176_0640* are the *C. jejuni* homologs of Δ *ape1*, Δ *patB*, and Δ *patA* respectively, as identified by BLAST using the *N. gonorrhoeae* OAP genes sequences. C, OAP mutants were generated by homologous recombination with a mutated copy of the gene (or the entire cluster for Δ *oap*) in which a portion of the gene (or cluster) was deleted and replaced with a non-polar Km^R cassette (*aphA3*) (47). Resistance to Km was used as a selective marker for successful homologous recombination in *C. jejuni* with the mutated gene. D, complement construction (with Δ *ape1* used as an example, designated Δ *ape1*^C). Each OAP gene plus upstream sequence containing the ribosomal binding site was cloned into the pRRC vector that contains homologous regions to three ribosomal intergenic regions downstream of the Cm^R cassette for selection of successful *C. jejuni* transformants. Complement constructs were transformed into their respective mutant backgrounds. *MacB*, macrolide-specific efflux pump; *OM efflux*, outer membrane efflux; *ftn*, ferritin; 23S, 23S ribosomal RNA (48).

did not affect the ability of Δ *ape1*^C to complement Δ *ape1* phenotypes). Differences were observed in the long term survival properties of Δ *ape1* with more modest differences exhibited by Δ *patA*, Δ *patB*, Δ *oap*, and Δ *ape1*^C. At 48 h, a 1.0-log increase in recovery was observed for Δ *ape1*, but recovery fell at 72 h by 0.8-log relative to wild type. At 48 h, Δ *patA* and Δ *patB* exhibited a 0.3-log increase, and Δ *oap* showed a 0.4-log increase in recovery, relative to wild type. The recovery of all three OAP-deficient mutants was 0.3-log lower than wild type at 72 h. Δ *ape1*^C exhibited a 0.8-log increase in recovery at 48 h and a 0.3-log decrease at 72 h relative to wild type (data not shown).

O-Acetylation Levels of Purified PG from Deletion Mutants Reflect the Putative Functions of the *C. jejuni* OAP Gene Cluster—To determine whether the *C. jejuni* OAP gene homologs were involved in OAP, PG O-acetylation levels were determined for the mutants of the three putative OAP genes (Δ *patA*, Δ *patB*, and Δ *ape1*) and for the mutant lacking the entire cluster (Δ *oap*) (Fig. 1C). PG was isolated from strains using an established protocol that minimizes spontaneous O-linked acetate hydrolysis and was assessed for OAP levels by quantifying released acetate and MurNAc (49, 50).

The O-acetylation level for the wild-type strain 81-176 was determined to be $12.5 \pm 0.71\%$ relative to MurNAc content. O-acetylation levels among the mutants varied according to their predicted function (Table 1). Deletion of *patA* and *patB* resulted in a reduction in O-acetylation levels at 2.45 ± 0.14 and $3.05 \pm 0.22\%$ relative to MurNAc content, respectively. Deletion of the entire gene cluster in Δ *oap* resulted in a decrease in O-acetylation levels to $2.10 \pm 0.18\%$, similar to that of Δ *patA* and Δ *patB*. Previous studies using *Escherichia coli* (which lacks PG O-acetyl groups) showed undetectable levels of acetate

TABLE 1

O-Acetylation levels of *C. jejuni* 81-176, Δ *ape1*, Δ *ape1*^C (complemented Δ *ape1*), Δ *patB*, Δ *patA*, and Δ *oap* (a mutant in which the entire cluster was deleted: *ape1*, *patB*, and *patA*), as determined by base-catalyzed hydrolysis and release of acetate reported as a % O-acetylation relative to MurNAc content

Strain	% O-acetylation ^a (mean \pm S.D.)
81-176	12.5 \pm 0.71
Δ <i>ape1</i>	35.6 \pm 2.25
Δ <i>ape1</i> ^C	11.8 \pm 0.52
Δ <i>patB</i>	3.05 \pm 0.22
Δ <i>patA</i>	2.45 \pm 0.14
Δ <i>oap</i>	2.10 \pm 0.18

^a Results shown are of one representative biological replicate measured in triplicate \pm S.D.

using identical methods (50, 51). This suggests that *patA/B* contributes to PG O-acetylation in *C. jejuni*, but their absence is insufficient to abolish OAP. Deletion of *ape1* led to an increase in O-acetylation to $35.6 \pm 2.25\%$ relative to total MurNAc content. These results are in accordance with the functions described for homologs in *N. gonorrhoeae* and *N. meningitidis* (30, 42–44). O-Acetylation levels were restored to wild-type levels in the Δ *ape1* complement ($11.78 \pm 0.52\%$). Analysis of the O-acetylation levels for Δ *patA* and Δ *patB* complements were not performed as, unlike the Δ *ape1* mutant, phenotypic differences between these mutants, Δ *oap* and wild type, were in almost every case not statistically significant or were minimal (see below).

***C. jejuni* OAP Mutants Exhibit Altered PG Muropeptide Profiles, with Δ *ape1* Displaying the Most Dramatic Changes from Wild Type**—O-Acetylation has been described as a PG maturation event occurring after transglycosylation and transpeptidation (30). Cleavage of PG by bacterial LTs is inhibited by PG O-acetylation. Thus, O-acetylation may impact PG maturation

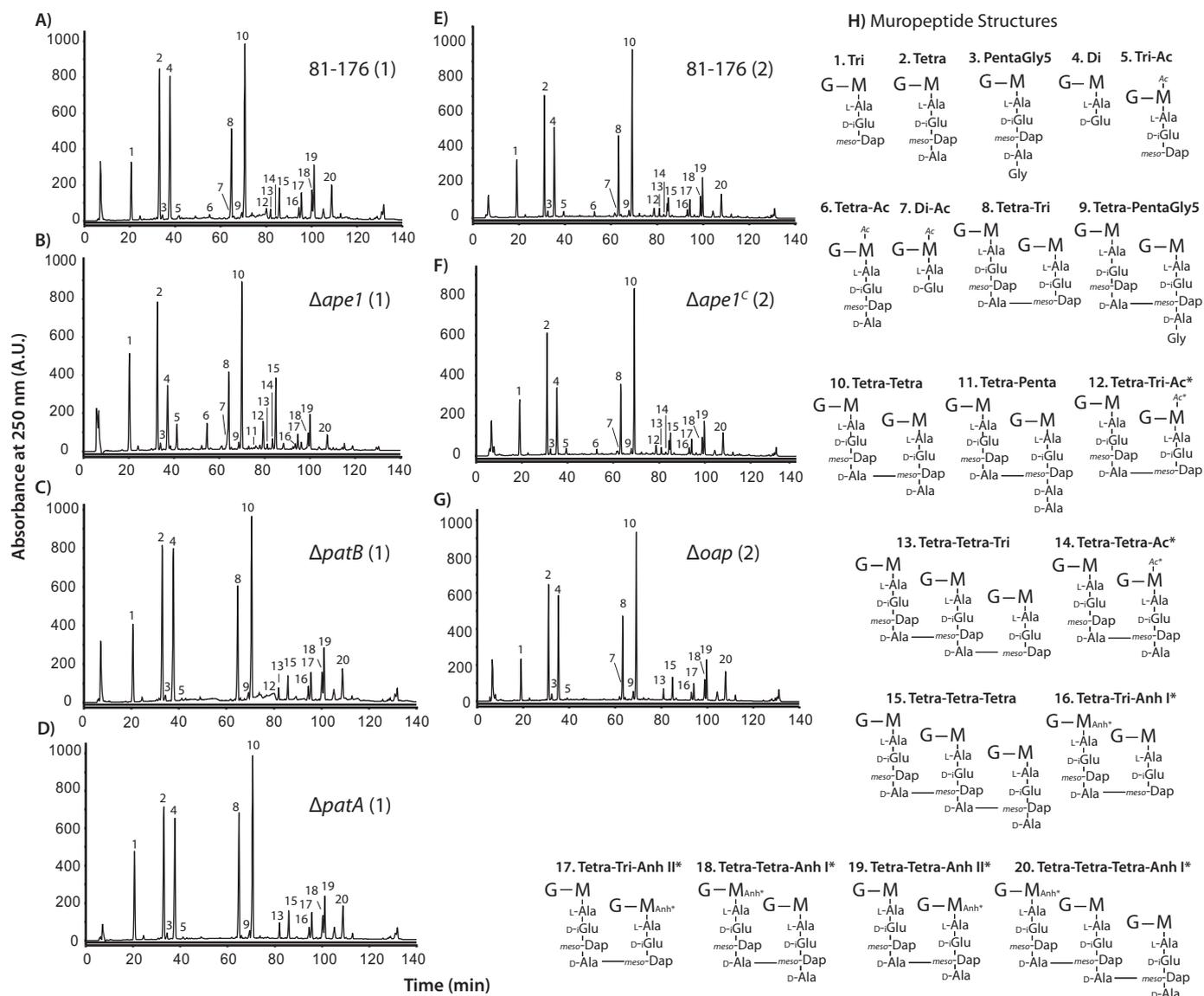


FIGURE 2. HPLC elution profile of *C. jejuni* mucopeptides and proposed mucopeptide structures. Purified PG was digested with cellosyl, and the resulting mucopeptides were reduced with sodium borohydride and separated on a Prontosil 120-3-C18 AQ reverse-phase column. HPLC profiles are shown for wild-type strain 81-176 (A and E), $\Delta ape1$ (B), $\Delta patB$ (C), $\Delta patA$ (D), $\Delta ape1^C$ (F), and Δoap (G). Mucopeptide profiles were generated in two sets of experiments indicated by (1) for Sample Set 1 and (2) for Sample Set 2. The mucopeptide structure represented by each peak corresponding to the peak number in the chromatogram are shown in H. The summary of the mucopeptide composition is shown in Table 3. G, N-acetylglucosamine; M, reduced N-acetylmuramic acid; L-Ala, L-alanine; D-Glu, D-isoglutamic acid; meso-Dap, meso-diaminopimelic acid; D-Ala, D-alanine; Ac, O-acetyl groups at MurNAC C6 position; Anh, 1,6-anhydro group of MurNAC; *, it is not known on which MurNAC residue the modification occurs.

events in *C. jejuni*, affecting aspects such as mucopeptide profiles and glycan chain length (52). To investigate this, PG was isolated from wild type and each of the mutant and complemented strains, and the mucopeptide composition was determined. HPLCs are shown in Fig. 2. Raw data (relative abundance of each mucopeptide) and summarized mucopeptide profiles can be found in Tables 2 and 3, respectively.

The method used for mucopeptide analysis results in the loss of some of the PG O-acetyl groups (due to the alkaline conditions for $NaBH_4$ reduction resulting in base-catalyzed hydrolysis of the O-linked acetate) and is thus less precise at determining O-acetylation levels than the methodology used above. Nonetheless, similar trends were observed, further supporting gene product function. PG O-acetylation levels were reduced in

$\Delta patA$, $\Delta patB$, and Δoap and increased in $\Delta ape1$ relative to wild-type 81-176. Monomeric O-acetylated tetrapeptide species and O-acetylated tetra-tetra dimeric species were absent in $\Delta patB$, $\Delta patA$, and Δoap . The abundance of all detectable O-acetylated mucopeptide species was increased in $\Delta ape1$.

$\Delta ape1$ exhibited a large decrease in total anhMP species and an increased average glycan chain length compared with wild type, similar to observations made in *N. meningitidis* (33), and is consistent with the observed O-acetylation levels, as de-O-acetylation must precede LT activity. Conversely, in the absence of *patA* or *patB*, the relative abundance of anhMP species did not vary strongly from wild type.

Changes were also observed in other mucopeptide species between wild type and $\Delta patA$, $\Delta patB$, Δoap , and $\Delta ape1$. The

Role of *C. jejuni* Peptidoglycan O-Acetylation

TABLE 2

Muropeptide composition of *C. jejuni* wild-type 81-176, Δ *ape1*, Δ *patB*, and Δ *patA*, Δ *ape1*^C, Δ *patB*^C, Δ *patA*^C, and Δ *oap* showing relative abundance of muropeptides corresponding to peaks in HPLCs (Fig. 2)

Peaks ^a	Muropeptide	Sample Set #1				Sample Set #2				
		81-176	Δ <i>ape1</i>	Δ <i>patB</i>	Δ <i>patA</i>	81-176	Δ <i>ape1</i> ^C	Δ <i>patB</i> ^C	Δ <i>patA</i> ^C	Δ <i>oap</i>
1	Tri	5.8	12.3	7.3	9.5	8.8	9.6	8.8	6.0	6.2
^b 2	Tetra-Gly-4					0.4	0.4	0.4	0.3	0.4
3	Tetra	16.1	16.6	15.8	15.2	16.1	16.9	16.2	15.7	15.1
4	Penta-Gly-5	0.6	0.8	0.7	0.7	0.8	1.0	1.2	0.9	1.0
5	Di	16.3	8.3	16.1	14.3	13.0	10.1	10.4	16.2	15.7
6	Tri-Ac	0.3	2.8	0.3	0.3	1.0	1.2	0.7	0.5	0.4
7	Tetra-Ac	0.6	2.6	0.0	0.0	0.6	0.7	0.5	0.0	0.0
8	Di-Ac	0.0	1.9	0.0	0.0	0.2	0.5	0.5	0.6	0.4
9	TetraTri	9.7	7.6	11.1	13.7	10.5	10.1	10.6	11.5	11.2
10	TetraPenta-Gly-5	0.4	0.7	0.6	1.0	0.8	0.9	1.0	0.8	1.2
11	TetraTetra	19.0	17.0	18.0	21.2	21.4	22.4	23.0	22.4	22.5
12	TetraPenta	0.0	0.1	0.0	0.0					
13	TetraTri-Ac	1.8	3.4	1.6	0.0	1.5	1.7	0.9	0.0	0.0
14	TetraTetraTri	0.8	0.5	0.9	1.5	1.0	0.9	1.3	1.3	1.4
15	TetraTetra-Ac	0.3	1.1	0.0	0.0	1.7	1.8	1.4	0.4	0.0
16	TetraTetraTetra	2.9	7.0	1.9	2.8	2.5	2.8	3.1	2.9	3.4
17	TetraTriAnh I	1.2	0.5	1.3	1.3	1.1	1.0	1.1	1.3	1.2
18	TetraTriAnh II	2.6	1.6	2.5	2.7	2.4	2.1	2.2	2.5	2.5
19	TetraTetraAnh I	3.0	1.8	2.8	2.7	2.8	2.8	3.0	3.1	3.0
20	TetraTetraAnh II	5.6	3.7	4.9	4.9	5.2	5.0	5.1	5.4	5.5
^b 20	TetraTetraTriAnh					1.1	0.9	1.3	1.4	1.7
1-20	All known ^c	4.3	1.9	3.9	4.3	4.1	4.2	4.9	4.5	5.3
		91.2	92.3	89.5	95.9	96.8	97.1	97.4	97.7	97.9

^a Peak numbers correspond to those from HPLCs in Fig. 2. Muropeptides are named according to Glauner *et al.* (88) and are depicted in Fig. 2H. Di, disaccharide dipeptide (disaccharide = β 1,4-linked *N*-acetylglucosamine-*N*-acetylmuramic acid); Tri, disaccharide tripeptide; Tetra, disaccharide tetrapeptide; Penta, disaccharide pentapeptide. Muropeptides 1-7 are monomeric, and muropeptides 8-20 are dimers and trimers connected by peptide cross-links. Modifications: Gly, glycine in position 5 of a peptide side chain; Ac, *O*-acetyl groups at the C-6 hydroxyl group of MurNAc; Anh, 1,6-anhydro-*N*-acetylmuramic acid.

^b Peak was previously unidentified in Sample Set #1.

^c The total abundance does not add up to 100% due to the presence of peaks for which a structure has not been assigned.

TABLE 3

Summary of PG muropeptide composition for *C. jejuni* 81-176, Δ *ape1*, Δ *patB*, Δ *patA*, Δ *ape1*^C, Δ *patB*^C, Δ *patA*^C, and Δ *oap*

Muropeptide species	Percentage of peak area ^a								
	Sample Set #1				Sample Set #2				
	81-176	Δ <i>ape1</i>	Δ <i>patB</i>	Δ <i>patA</i>	81-176	Δ <i>ape1</i> ^C	Δ <i>patB</i> ^C	Δ <i>patA</i> ^C	Δ <i>oap</i>
Monomers (total)	43.5	49.1	44.9	41.7	42.2	41.5	39.6	41.2	40.0
Dipeptide	17.8	11.0*	18.0	15.0	13.6	11.0	11.2	17.2	16.4
Tripeptide	6.7	16.4*	8.5	10.2*	10.1	11.1	9.7	6.6*	6.7*
Tetrapeptide	18.3	20.9	17.6	15.8	17.2	18.1	17.1	16.1	15.5
Pentapeptides-Gly	0.6	0.9*	0.7	0.7	0.8	1.0	1.3*	0.9	1.0
O-Acetylated	1.0	8.0*	0.3*	0.3*	1.9	2.4	1.7	1.1*	0.8*
Dimers (total)	47.7	40.7	47.6	49.4	48.8	49.4	49.5	48.5	48.1
TetraTri	16.8	14.2	18.4	18.5	16.0	15.5	15.2	15.7	15.2
TetraTetra	30.5	25.6	28.6	29.9	32.0	33.0	33.2	32.0	31.7
TetraPentaGly	0.4	0.9*	0.6*	1.0*	0.8	0.9	1.1*	0.8	1.2*
Anhydro-Dimers	13.5	8.2*	12.7	12.1	11.8	11.3	11.6	12.6	12.4
O-Acetylated	2.2	4.9*	1.7	0*	3.2	3.7	2.3	0.4*	0*
Trimers (total)	8.8	10.2	7.5	8.8	7.8	8.1	9.5	8.9	10.2*
TetraTetraTri	0.9	0.5*	1.0	1.5*	1.0	0.9	1.3*	1.3*	1.4*
TetraTetraTetra	8.0	9.7	6.5	7.3	6.8	7.2	8.2	7.6	8.8
Dipeptides (total)	17.8	11.0*	18.0	15.0	13.6	11.0	11.2	17.2	16.4
Tripeptides (total)	15.4	23.6*	18.0	19.9	18.4	19.1	17.7	14.9	14.7
Tetrapeptides (total)	65.9	64.0	62.9	63.9	65.1	67.1	67.6	64.8	65.1
Pentapeptides (total)	0.8	1.3*	1.0	1.2*	1.2	1.5	1.9	1.3	1.6*
O-Acetylated (total)	2.1	10.4*	1.2*	0.3*	3.5	4.3	2.9	1.3*	0.8*
Anhydromuropeptides	8.3	4.8*	7.8	7.5	7.7	7.4	7.9	8.3	8.6
Average chain length	12.0	20.8*	12.8	13.3	13.0	13.5	12.6	12.0	11.6
Degree of cross-linkage	29.7	27.2	28.8	30.6	29.6	30.1	31.1	30.2	30.9
% peptide cross-links	56.5	50.9	55.1	58.3	57.8	58.5	60.4	58.8	60.0

^a Values represent the percentage area of each muropeptide from raw data (found in Table 2) calculated to give a total of 100%. Boldface numbers represent a change in relative abundance of $\geq 20\%$ from wild type. Boldface numbers with an asterisk represent $\geq 30\%$ change from wild type. Percentages shown are calculated from values rounded to the nearest 0.1%. Muropeptide profiles were generated in two sets of experiments, Sample Set #1 and Sample Set #2. Muropeptide profiles were compared with the wild-type 81-176 muropeptide profile that was analyzed in the same sample set.

Δ *patA* mutant showed an increase in monomeric tripeptides, tetra-penta dipeptides, and tetra-tetra-tri tripeptides. For Δ *patB*, the monomeric tripeptides and dimeric tetra-penta species increased relative to wild type. Δ *oap* exhibited some differences from wild type that followed the same trend as with

Δ *patA*, an increase in tetra-penta dipeptides and tetra-tetra-tri tripeptides, and some that were unique to Δ *oap*, an increase in dipeptides and a decrease in tripeptides. Analyses of PG muropeptide profiles for Δ *patA*- and Δ *patB*-complemented strains Δ *patA*^C and Δ *patB*^C showed minimal changes from wild type

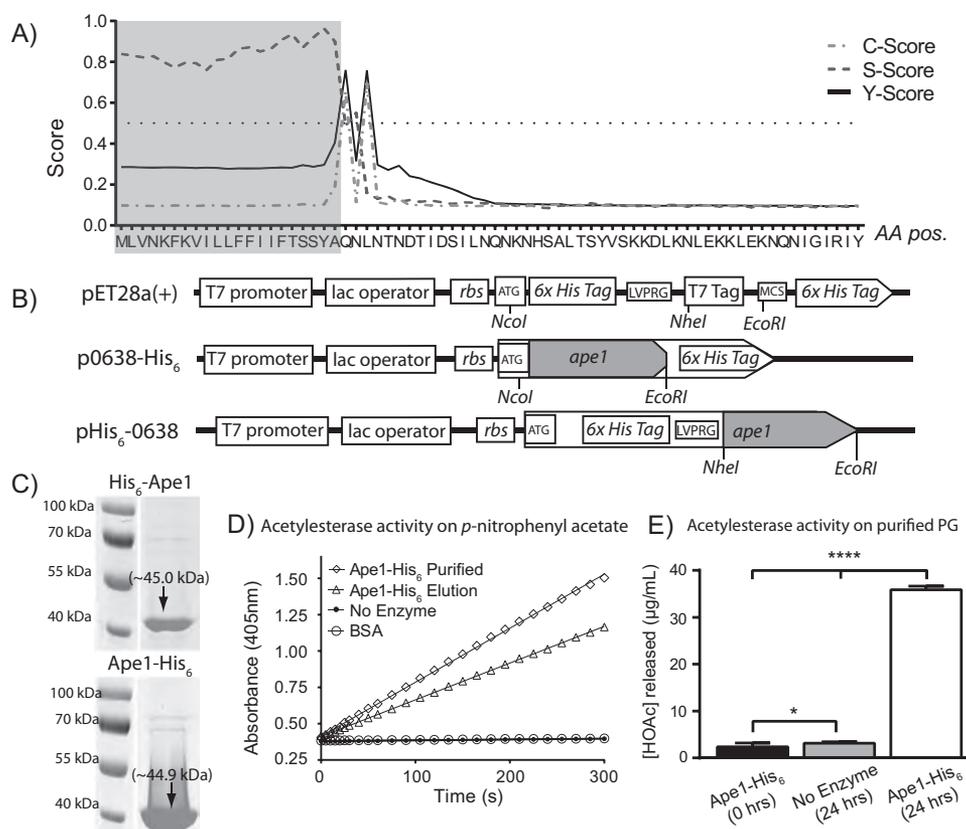


FIGURE 3. A, SignalP 4.1 server (83) output for signal peptide prediction of in-frame translation of *Cjj81-176_0638*. *C*-score (the predicted first amino acid of the mature protein), *S*-score (the likelihood that a particular amino acid is part of a signal peptide), and *Y*-score (amino acid with a high *C*-score exhibiting the greatest change in the *S*-score) predicted the cleavage site to be between the 21st and 22nd amino acids. B, *Cjj81-176_0638* was cloned in-frame without the signal peptide into pET28a(+) protein expression vectors. Top, map of cloning sites in pET28a(+) commercial expression vector. Middle, Ape1-His₆ expression construct. NcoI and EcoRI were used to produce a C-terminal His₆-tagged Ape1 protein that uses ATG start codon and TGA stop codon encoded in the vector. Bottom, His₆-Ape1 expression construct. NheI and EcoRI were used to produce an N-terminal His₆-tagged protein that uses AUG start codon encoded by the vector and the original stop codon from *Cjj81-176_0638*. rbs, ribosome-binding site; LVPRG, thrombin cleavage site; MCS, multiple cloning sites. Gene organization not to scale. C, His₆-tagged Ape1 after nickel-nitrilotriacetic acid-agarose purification shows protein of the predicted size (45.0 and 44.9 kDa for His₆-Ape1 and Ape1-His₆, respectively) in eluted fractions after SDS-PAGE analysis. D, purified Ape1-His₆ exhibits acetyltransferase/deacetylase activity using pNPAC as a substrate (45). Reactions were monitored over 5 min as a change in the absorbance at 405 nm (formation of *p*-nitrophenol) after cleavage of the acetyl group. No enzyme control and BSA control are overlapping and show no acetyltransferase activity. Results shown are from one protein purification experiment. Results are reproducible for each expression and purification experiment, and activity was routinely assessed before performing enzymatic assays on PG. E, Ape1-His₆ has acetyltransferase activity using PG mucopeptides as a substrate. Determination of acetic acid concentration after treatment of Δ ape1 PG with Ape1-His₆ for 24 h was performed using Megazyme acetic acid assay kit. Treatment and no enzyme control were compared with acetic acid concentration at 0 h of treatment using Student's *t* test with * and **** indicating *p* values of <0.05 and <0.0001, respectively. Results are from one representative experiment of two biological replicates performed in triplicate.

with restored *O*-acetylation levels for Δ pat^BC but not Δ pat^AC. The largest number of changes and greatest degree of change occurred in Δ ape1. In Δ ape1, total dipeptide species decreased and total tripeptides and pentapeptides increased relative to wild type. The total amount of dimers also appeared to be slightly lower in Δ ape1. The majority of the mucopeptide changes was restored to near wild-type levels in the Δ ape1 complement strain (Δ ape1^C).

Recombinant Ape1 Has in Vitro Acetyltransferase Activity on *p*-Nitrophenyl Acetate and *O*-Acetylated PG—Based on the mucopeptide data observed for Δ ape1 and phenotypic data shown below, the acetyltransferase activity of Ape1 was confirmed biochemically. Ape1 was expressed with a His₆ tag and minus the signal peptide at either the N or C terminus and purified (Fig. 3, A and B). Both recombinant proteins purified well (Fig. 3C), producing 3 ml of 0.9–1.2 mg/ml protein after dialysis from a starting culture of 100 ml. The expected sizes of the recombinant proteins are 45.0 and 44.9 kDa for His₆-Ape1 and Ape1-His₆, respectively. The specific activity of purified Ape1

was determined using *p*-nitrophenylacetate (*p*NPAC), a common substrate used to test esterase activity (31, 32). Specific activity for Ape1-His₆ ranged between 26.1 and 38.9 μmol/min/mg of protein (Fig. 3D), which was higher than the reported specific activity of ~10.4 μmol/min/mg for *N. gonorrhoeae* Ape1 measured under similar reaction conditions (32). This demonstrates that the recombinant protein exhibits acetyltransferase activity. His₆-Ape1 showed similar specific activity with *p*NPAC (data not shown).

Ape1-His₆ was also assayed for acetyltransferase activity on its native substrate (Fig. 3E), *O*-acetylated PG. PG isolated from Δ ape1 was used as the substrate due to the increased PG *O*-acetylation levels in this strain. Cleavage of *O*-acetyl groups was assessed using a commercial acetic acid assay kit (Megazyme) as an end point experiment. At 0 h, the average acetate concentration in the sample was 2.4 ± 0.3 μg/ml. Incubation of Δ ape1 PG for 24 h in the absence of enzyme resulted in an average acetate concentration of 3.1 ± 0.1 μg/ml, and an average acetate concentration of 35.9 ± 0.7 μg/ml after incubation

Role of *C. jejuni* Peptidoglycan O-Acetylation

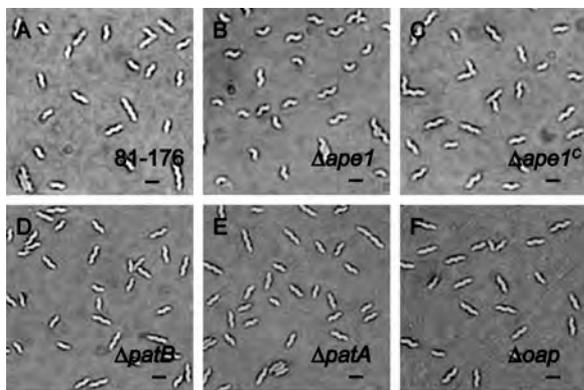


FIGURE 4. *C. jejuni* Δ ape1 mutant has a pleomorphic cell shape, and other OAP mutants display unaltered cell morphology. DICM showing the morphology of wild-type strain *C. jejuni* 81-176 (A), the differentially curved Δ ape1 strain (B), the complemented strain Δ ape1^C with restored morphology (C), Δ patB (D), Δ patA (E), and Δ oap (F). Cells were harvested from 7 h of growth in MH-TV broth at a mid-exponential phase of growth. Scale is 2 μ m (black bar).

with Ape1-His₆. Because of the insoluble nature of PG, data from this assay cannot be expressed in the classical definitions of enzyme kinetics using the native substrate.

Microscopy and CellTool Analyses of *C. jejuni* Δ ape1 Population Morphology Reveal Shape Pleomorphism—Because a number of changes were observed in the muropeptide profiles of the OAP mutants, it was hypothesized that these changes may result in changes in cell shape. The morphology was examined by DICM after 4 and 7 h of growth initiated at an A_{600} of 0.05 to obtain early- and mid-exponential phase cultures (Fig. 4). Whereas wild type exhibited the classical *C. jejuni* helical shape (Fig. 4A), Δ ape1 exhibited primarily “comma-shaped” and differentially curved cells (Fig. 4B). Wild-type helical morphology was restored upon complementation (Fig. 4C). A distinct change in morphology was not observed for the other OAP mutant populations (Fig. 4, D–F).

The open-source shape analysis program CellTool (53) was used to quantify the changes in shape in the OAP mutants (Fig. 5). The program contains a set of tools used to extract shapes from binary images that can then be used to assess and compare the population morphology using a variety of metrics. Extracted shapes from the wild-type population were aligned to one another, and principal component analysis (PCA) was performed to generate a baseline model for variation within the wild-type population. At mid-exponential phase, three shape modes (principal components that define specific changes in cell shape in the population) described 95% of the morphological variation in the *C. jejuni* wild-type population. Each shape mode represented an observable metric (Fig. 5, A and B). Shape mode 1 corresponded to variation in cell length, explaining >90% of the variance (as expected because a population likely exhibits asynchronous growth and division). Shape mode 2 explained 1.9% of the variance and appeared to have some relation to the curvature and wavelength of the cell. A third shape mode explained 1.7% of the variance and described differences in cell width. Contours of mutants were aligned to the wild-type shape model as a reference and Kolmogorov-Smirnov (KS) statistical tests were used to compare sample probability distributions. Based on the large population of bacterial cells assessed

and conditions required for KS analysis, a p value of 0.00001 was used as a cutoff for significance (Fig. 5B) (54).

No strains were significantly different from the wild type in shape mode 1 (cell length) or from each other, with the graphical output also showing that the population distributions overlap very closely. In shape mode 2 (cell curvature), some differences in population distribution between wild type and Δ ape1^C, Δ patA, Δ patB, and Δ oap were significant by the KS cutoff utilized. However, the graphical output showed that these strains were similar to wild type, whereas Δ ape1 exhibited a dramatic shift in the population distribution maximum (~2.2 standard deviations from the wild-type mean). Shape mode 3 (cell width) was significantly different in all strains compared with wild type (with the exception of Δ ape1^C), and each exhibited a shift of approximately 1 S.D. in the population maximum toward a reduced width compared with the wild-type mean as reflected in the graphical output. A 2D scatterplot of measurements of each individual contour of wild type, Δ ape1, and Δ ape1^C populations for shape modes 2 and 3 (Fig. 5C) likewise shows that there was a clear difference in shape for the Δ ape1 population compared with wild type and Δ ape1^C. Early-exponential phase bacteria exhibited similar population shifts as for mid-exponential phase bacteria (with the exception of shape mode 3, as cell width was not captured as a major contributor to the variance in shape for wild type at this time point). The most notable shift at early-exponential phase was observed for cell curvature (shape mode 2) in Δ ape1 (data not shown).

Phenotypic Analyses Reveal the Importance of O-Acetylpeptidoglycan Esterase Activity on Various Aspects of *C. jejuni* Physiology—The OAP mutants were assessed for different phenotypes serving as indicators of transmission and/or colonization efficiency: motility in soft agar, biofilm formation, hydrophobicity, and sensitivity to a variety of inhibitory compounds.

Motility is a major colonization determinant for *C. jejuni* (55). Although all strains exhibited defective halo formation compared with wild type in soft agar plates after point inoculation (Fig. 6A), the halo diameter of Δ ape1 was 70% of wild type, whereas Δ patB, Δ patA, and Δ oap were only modestly defective at 90, 90, and 87% of wild type. Complementation of Δ ape1 restored the halo formation of the mutant to 90% of wild type and was significantly different from that of Δ ape1. In addition, Δ ape1 formed aberrant halos on soft agar with rough perimeters as opposed to the circular halos formed by wild type. This halo phenotype was absent in the other mutants tested and was rescued by complementation.

The ability to form biofilms is important in *C. jejuni* persistence and transmission. *C. jejuni* has been shown to survive up to 28 days in a biofilm state and is a general stress response (56). The ability of our OAP mutants to form biofilms was assessed in borosilicate test tubes by crystal violet staining of standing cultures (57). The Δ ape1 mutant exhibited a hyper-biofilm phenotype, producing 5.5-fold more biofilm than wild type (Fig. 6B). Δ ape1 also developed flocs of bacteria suspended in the broth (58), which were not observed for wild type nor included in the crystal violet quantification of surface-adhered biofilm (data not shown). Complementation of Δ ape1 restored biofilm formation to wild-type levels (Fig. 6B). Biofilms formed

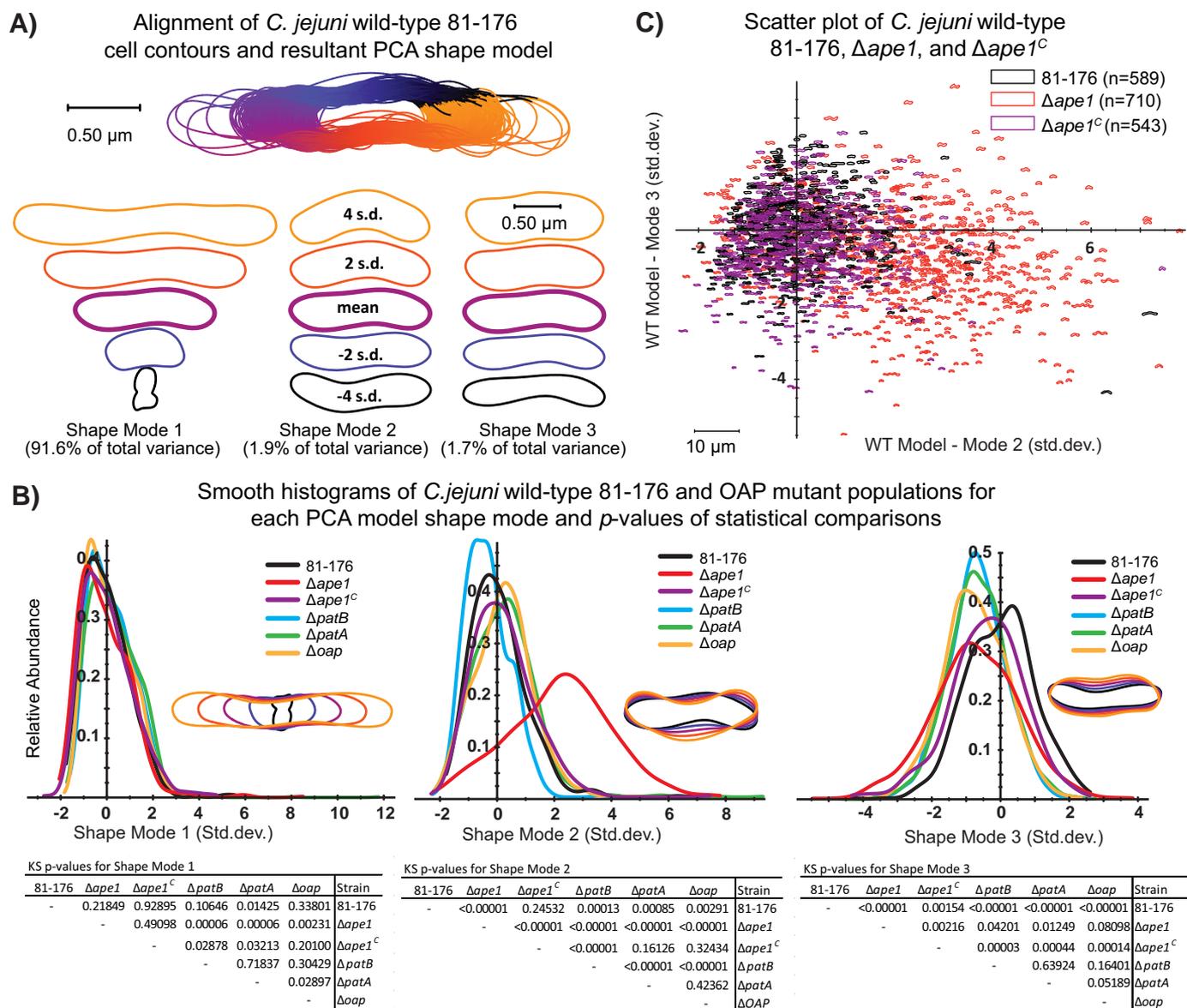


FIGURE 5. CellTool analysis of wild-type strain 81-176, $\Delta ape1$, $\Delta ape1^c$, $\Delta patB$, $\Delta patA$, and Δoap population morphology. Differential interference contrast images were taken of strains grown for 7 h in MH-TV broth at a starting A_{600} of 0.05 (to mid-exponential phase). Images were converted to binary format (white cells on a black background), and lumps and artifacts were manually removed before processing with CellTool “extract contours function” to generate contours representing each cell (53). *A*, contour extraction, alignment, and generation of the PCA shape model for *C. jejuni* wild-type strain 81-176. CellTool “align contours” function was used to align the contours of the wild-type population to one another. *B*, PCA was performed to generate a wild-type shape model that explains 95% variation in the population in principal components called “shape modes.” Shape modes 1, 2, and 3 represent variation in length, curvature/wavelength, and width, respectively. The extracted contours of the mutant populations were then aligned to the wild-type shape model, and a measurement representing the normalized standard deviation from the wild-type mean in each shape mode was generated and depicted graphically. KS tests were performed for each shape mode between each population and are summarized below the plots. *C*, measurements of wild type, $\Delta ape1$, and $\Delta ape1^c$ were plotted with shape mode 2 along the *x* axis and shape mode 3 along the *y* axis to create a scatterplot showing the variation in the different populations.

by $\Delta patB$ and $\Delta patA$ were indistinguishable from wild type, but Δoap produced ~ 2.5 -fold more biofilm than wild type. Characterization of $\Delta ape1$ biofilms by microscopy was unsuccessful as $\Delta ape1$ formed biofilms poorly on coverslips unlike wild type. This indicates altered cell surface properties in $\Delta ape1$. Cell surface hydrophobicity was assessed with hexadecane partitioning (Fig. 6C) (59). The percent hydrophobicity of $\Delta ape1$ was significantly higher (2.0-fold) than wild type and was restored to wild-type levels upon complementation.

The sensitivity of the OAP mutants to detergents, salts, and antimicrobial compounds was tested by determining the mini-

mum inhibitory concentration that reduces growth by 50% relative to a control as measured by A_{600} (MIC_{50}) (Table 4). Only $\Delta ape1$ exhibited an increased susceptibility to any of the compounds tested as follows: the amphipathic bile salt sodium deoxycholate (DOC) and $MgCl_2$. For $\Delta ape1$, an MIC_{50} range for DOC of 0.16 to 0.31 mg/ml was observed, whereas the MIC_{50} for wild type was greater than the highest concentration of DOC tested (> 10 mg/ml). $\Delta ape1$ exhibited a 4–8-fold reduction in MIC_{50} for $MgCl_2$ compared with wild type. Complementation of $\Delta ape1$ restored wild type sensitivity profiles.

Role of *C. jejuni* Peptidoglycan O-Acetylation

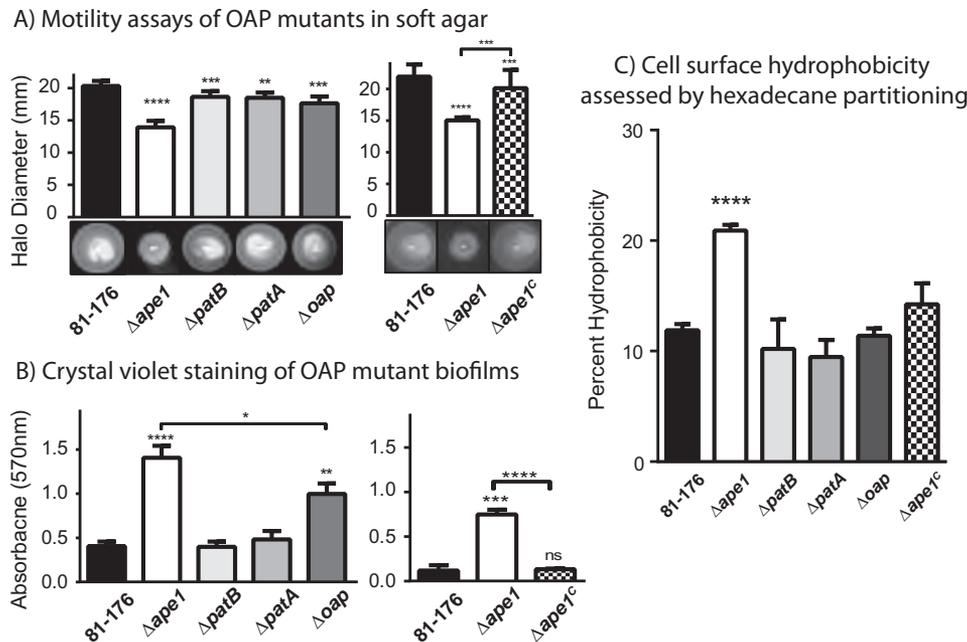


FIGURE 6. Motility in soft agar, biofilm formation, and cell surface hydrophobicity of OAP mutants and wild-type strain 81-176. *A*, $\Delta ape1$ exhibits a 30% decrease in halo diameter and abnormal halo formation (rough edges). Motility in soft agar was assessed by measuring the halo diameter after 24 h of strains point-inoculated in 0.4% semi-solid agar. Representative images of halos are shown *below* each graph. Results shown are representative of one of three independent experiments with 6 replicates. Each strain was compared with wild-type using a paired Student's *t* test, with **, ***, and **** indicating $p < 0.01$, $p < 0.001$, and $p < 0.0001$. *B*, $\Delta ape1$ and Δoap exhibit 5.5- and 2.5-fold enhanced biofilm formation, respectively, at 24 h. Biofilm formation was assessed after 24 h by crystal violet staining of standing cultures in borosilicate tubes and spectrophotometric quantification of dissolved crystal violet at 570 nm. Results shown for the mutants (*left*) are representative of one of three independent experiments carried out in triplicate. The results for $\Delta ape1^c$ (*right*) are representative of one of two experiments performed in triplicate. *ns*, not significant. *C*, $\Delta ape1$ exhibited a 2.0-fold increase hydrophobicity relative to wild type, as assessed by hexadecane partitioning. Results are representative of one of three independent experiments performed in triplicate. For biofilm and hydrophobicity, strains were compared using an unpaired Student's *t* test, with *, **, ***, and **** indicating $p < 0.05$, $p < 0.01$, $p < 0.001$, and $p < 0.0001$. Error bars represent standard deviation.

TABLE 4

MIC₅₀ of *C. jejuni* OAP mutants determined by broth dilution

Measurements indicated with a "-" have not been tested. Measurements in boldface were consistently ≥ 4 -fold different from wild type over three experiments. MIC₅₀, minimum inhibitory concentration to reduce growth by 50% as assessed by optical density.

Compound	MIC ₅₀					
	81-176	$\Delta ape1$	$\Delta patB$	$\Delta patA$	Δoap	$\Delta ape1^c$
Detergents						
Deoxycholate (mg/ml)	>10	0.16–0.31	5->10	1.3->10	1.3->10	1.3->10
SDS (mg/ml)	10–12.5	2.5–6.25	10–12.5	2.5–12.5	5–6.25	12.5
Triton (% v/v)	0.05	0.02–0.005	0.02–0.05	0.02–0.05	0.05	0.05
Antimicrobials						
Ampicillin (μ g/ml)	2.4–4.9	1.2–4.9	4.9	4.9	2.4–4.9	2.4
Lysozyme (mg/ml)	>5	>5	>5	>5	>5	>5
Polymyxin B (μ g/ml)	20	10	10	10	10	10–20
Protamine (μ g/ml)	31.3	15.6–31.3				31.3–62.5
Chelating agent						
EDTA (μ M)	78–156	1.2–156	156	156	156	156
Salts						
NaCl (mM)	62.5–250	31.3–62.5	62.5–125	125	62.5–125	62.5–125
MgCl ₂ (mM)	62.5–125	15.6	62.5–125	62.5–125	62.5–125	62.5–125
CaCl ₂ (mM)	125	125–250	125–250	125–500	500	125
KCl (mM)	62.5	31.3–62.5	62.5	31.3–62.5	125	31.3–62.5

Ape1 Is Required for *C. jejuni* Bacteria-Host Interactions—The contribution of OAP to *C. jejuni* host interactions was examined by determining recovery of the mutants after chick colonization and host cell infections, as well as the ability to elicit IL-8 secretion *in vitro* in human epithelial infections.

Chickens are an avian reservoir for *C. jejuni* and a common source of human infection. The $\Delta ape1$ mutant exhibited a significant 4.4-log decrease in colonization (Fig. 7A),

whereas $\Delta patB$, $\Delta patA$, and Δoap mutants were not defective for chick colonization. The defects in long term survival in broth for the OAP-deficient mutants were modest compared with $\Delta ape1$, so the defect in chick colonization could be related to the altered long term survival properties of $\Delta ape1$.

The ability of a *C. jejuni* strain to invade and survive in non-phagocytic epithelial cell lines has been shown to correlate with virulence (11, 60, 61). The ability of the *C. jejuni* OAP mutants to

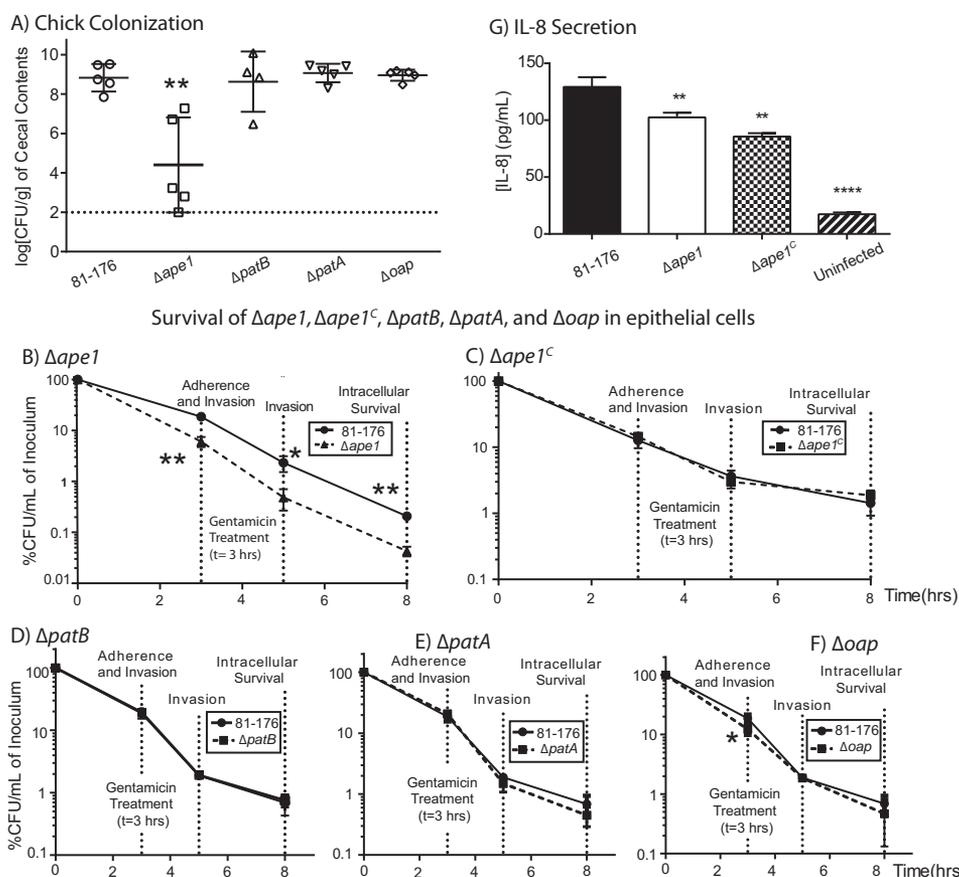

 Survival of Δ ape1, Δ ape1^C, Δ patB, Δ patA, and Δ oap in epithelial cells

FIGURE 7. Effect of OAP levels on *C. jejuni* host-bacteria interactions. *A*, Δ ape1 shows reduced chick colonization compared with wild-type strain 81-176, whereas Δ patB, Δ patA, and Δ oap mutants display wild-type colonization. Each point represents the recovery of *C. jejuni* strains in log CFU/g of cecal contents from individual day-old chicks 6 days post-colonization with 1×10^4 CFU/ml of the indicated strain. The geometric mean is denoted by a black bar. Error bars represent 95% confidence intervals. Adherence, invasion, and intracellular survival of *C. jejuni* in INT407 epithelial cells were assessed by a Gm protection assay and OAP mutant strains. Δ ape1 (*B*) shows a reduced ability to adhere to, invade, and survive in INT407 epithelial cells that were restored upon complementation (*C*). Δ patB (*D*), Δ patA (*E*), and Δ oap (*F*) exhibit near wild-type adherence, invasion, and intracellular survival properties. INT407 cells were infected with *C. jejuni* at a multiplicity of infection of ~ 80 . Adherence and invasion were quantified at 3 h post-infection. At this point, the media in the remaining wells were replaced with MEM containing gentamicin (150 μ g/ml) and incubated for 2 h, after which the amount of bacterial cells that had invaded the epithelial cells was measured (5-h invasion time point). The Gm in the remaining wells was washed off, and the cells were incubated with fresh MEM containing 3% FBS and a low dose of Gm (10 μ g/ml) for an additional 3 h (8-h intracellular survival time point). CFU/ml was determined for each well by lysing the cells with water and plating the dilutions onto MH-TV plates. Results for *B* and *C* are representative of three independent experiments performed in biological triplicate. The data in *D*, *E*, and *F* are representative of two independent experiments performed with three biological replicates. *G*, INT407 epithelial cells secrete less IL-8 upon infection with Δ ape1 than wild type. Results are from one representative experiment of three independent experiments performed in triplicate. Error bars represent the standard deviation. *, denotes statistically significant difference using the unpaired Student's *t* test, with *, **, and **** indicating $p < 0.05$, $p < 0.01$, and $p < 0.00001$ respectively.

adhere to, invade, and survive inside the human epithelial cell line INT407 was assessed by a gentamicin (Gm) protection assay. Recovery of Δ ape1 was significantly reduced at the adherence, invasion, and intracellular survival time points in comparison with wild type (Fig. 7*B*). Δ ape1^C, Δ patB, Δ patA, and Δ oap displayed wild-type INT407 infection profiles (Fig. 7, *C–F*).

As Δ ape1 was the only OAP mutant to demonstrate reduced invasion of the INT407 cells, its ability to elicit IL-8 secretion from INT407 cells was assessed by ELISA. Cells infected with Δ ape1 reproducibly exhibited statistically significant lower levels (60–79%) of IL-8 secretion compared with cells infected with wild type (Fig. 7*G*). Δ ape1^C did not complement IL-8 induction defects.

Discussion

PG plays roles in multiple facets of bacterial physiology. PG modifications have been shown to influence pathogenic properties in several bacterial species (15, 20, 36). Here, the OAP genes in *C. jejuni* were shown to contribute to PG O-acetyla-

tion/de-O-acetylation, consistent with their predicted functions. These genes were also important for several key physiological and pathogenic properties. This was most notable for *ape1*, which was involved in PG de-O-acetylation and the only OAP gene significantly required for every phenotype examined.

Deletion of *patA* or *patB*, which act to O-acetylate MurNAc, was non-lethal as found in several other bacterial species (33, 37, 44). This suggests that O-acetyl groups added by PatA/B play a non-essential role for growth of *C. jejuni* in the laboratory. Unlike in *N. meningitidis* where OAP is exclusively mediated by *patA/B* (33), the O-acetylation levels were not reduced to 0% in *C. jejuni*, indicating the presence of alternative PG O-acetylation machinery or compensation by alternative mechanisms, as was observed with *E. coli* WeCH that acted as an acetate transporter (62). Expression of *N. gonorrhoeae* PatB in *E. coli* increased OAP levels from <0.05 to 1%, which was detrimental to the cells (62). The effect of low levels of OAP on *E. coli* biology provides support that the residual O-acetylation

Role of *C. jejuni* Peptidoglycan O-Acetylation

in *C. jejuni* may be sufficient to mask mutant phenotypes in $\Delta patA/B/oap$ that would otherwise be observed if PG O-acetylation were completely absent. There are conflicting results for the essentiality of *ape1* in *N. gonorrhoeae* (43, 44). In *C. jejuni*, *ape1* was not essential. Deletion of *ape1* resulted in increased O-acetylation levels almost triple that of wild type, supporting the role of Ape1 in *C. jejuni* PG de-O-acetylation. Ape1 acetyltransferase activity was also confirmed *in vitro* using the artificial substrate pNPAC as well as its natural substrate, O-acetylated PG from $\Delta ape1$.

It should be noted that the OAP levels of wild-type *C. jejuni* 81-176 described here were lower than those reported in a previous study for ATCC 700819 and NCTC 11168 (42). However, the strains and growth conditions used differed between the studies. A direct comparison of how these and other potential factors might affect *C. jejuni* PG O-acetylation has not yet been assessed but will be the topic of future work.

As with the OAP analyses described above, the muropeptide profiles also showed differences in PG O-acetylation levels for the *C. jejuni* OAP mutants. Although these data are not truly quantitative for O-acetylation, they offer additional qualitative support for the role of these OAP genes in PG O-acetylation. *N. meningitidis* $\Delta ape1$ showed an increase in only O-acetylated muropeptides with a tri-peptide stem in comparison with wild type (33), although the muropeptide analysis of *C. jejuni* $\Delta ape1$ suggests that Ape1 in *C. jejuni* may be regulated differently, as this specificity was not observed. Previous observations with *N. gonorrhoeae* PatB O-acetyltransferase using *in vitro* assays showed specificity of PatB toward O-acetylation of tetrapeptides (30). A decrease in O-acetylated tetrapeptide species was observed for the *C. jejuni* $\Delta patA$ and $\Delta patB$ mutants (Table 2); however, as this could have been due to hydrolysis during the preparation procedure, further biochemical analysis will be required.

As expected, $\Delta ape1$ also exhibited a 42.1% decrease in relative anhMP levels (presumably due to impaired LT activity) and a greater average chain length. Although chain length was not directly measured, these data support previous findings that Ape1 regulates PG chain length as in *N. meningitidis* (33). The anhMP levels changed only marginally in $\Delta patA$, $\Delta patB$, and Δoap supporting a putative compensation of O-acetylation by a yet unknown mechanism consistent with our OAP analyses and/or O-acetylation itself may not be an essential or the only control mechanism for LT activity in *C. jejuni*. LTs in *E. coli* and *Pseudomonas aeruginosa* have been found in complexes with peripheral membrane-bound lipoproteins and PBPs and are thought to be controlled spatially as well as coupled with synthesis to prevent autolysis (52, 63–65). For *C. jejuni*, the observations here do suggest a role for OAP in regulating LT activity, but other control mechanisms likely exist.

Differences in the muropeptide composition could be possible if O-acetyl groups influence substrate recognition by PG remodeling enzymes. Care must be taken in interpreting how differences in relative abundance actually affect overall PG composition. For instance, small changes in muropeptides of low abundance can result in changes $\geq 20\%$ (i.e. total penta-Gly-5 species, which were 0.8% in wild type and 1.3% for $\Delta ape1$; Table 3). Conversely, larger changes in muropeptides of high abundance can produce changes $< 20\%$ yet may still be consid-

ered significant. For example, dimeric species constituted 47.7% of the muropeptides in wild type and 40.7% in $\Delta ape1$; this degree of change may be meaningful, as it affects 7% of the total muropeptides, is unique compared with other mutants tested, and would be considered significant using the 10% cutoff described for *H. pylori* (27). Regardless, it is clear that the absence of *ape1* affects the PG muropeptide profile more so than the absence of *patA/B* (Table 3 and Fig. 2). These changes could be a result of increased O-acetylation affecting substrate recognition by PG remodeling enzymes or missing protein-protein interactions in the absence of Ape1 and will require more extensive analysis in later studies.

N. meningitidis Ape1 showed preference for O-acetylated tripeptide substrates *in vivo* (33), as mentioned above. The crystal structure for *N. meningitidis* Ape1 has recently been solved (46), confirming its classification as a member of the Ser-Gly-Asn-His (SGNH) hydrolase superfamily based on active site catalytic residues (32). These residues are also conserved in *C. jejuni* Ape1. The putative PG binding domain in the N-terminal lobe of *N. meningitidis* Ape1 and its interaction with PG have yet to be described (46). As *N. meningitidis* Ape1 was active against various O-acetylated muropeptides *in vitro*, specificity may be due to regulation of activity through unknown interaction partners. The putative PG binding domain at the N terminus may confer substrate specificity (33). There is only 26/42% amino acid sequence identity/similarity between the N-terminal domains of *N. meningitidis* Ape1 and *C. jejuni* Ape1, so the two enzymes may possess different regulatory regions. Another possibility is that *C. jejuni* lacks the Ape1 interaction partners present in *N. meningitidis* conferring substrate specificity.

One of *C. jejuni*'s defining characteristics is its helical shape, a trait defined by the cytoskeleton-like components that coordinate the PG biosynthetic machinery (66). The muropeptide composition was altered in *C. jejuni*/*H. pylori* periplasmic PG hydrolase mutants, i.e. $\Delta pgp1/\Delta csd4$ and $\Delta pgp2/\Delta csd6$, and exhibited a straight rod versus helical morphology (18, 19, 28, 67). Deletion of *C. jejuni* *ape1* also resulted in altered muropeptide composition and shape, but the change in shape was not as dramatic as in the abovementioned straight-rod mutants. Thus, CellTool was employed for shape quantification. This analysis showed that $\Delta ape1$ was significantly different from the wild-type population in curvature in that it had both an average shape with a larger wavelength compared with wild type and a greater variance of curvature within the population. Ape1 was shown to affect cell size in *N. meningitidis* (33). In this study, there was a significant increase in total area of $\Delta ape1$ cells when compared with wild type at early-exponential phase but not at mid-exponential phase (data not shown). One explanation could be that Ape1 activity varies at different growth stages in *C. jejuni*.

Multiprotein flagellar complexes span the PG layer with some proteins of the complex proposed to directly interact with PG. These proteins include FlgI, which makes up the P-ring of the periplasmic rod-structure in the hook-basal body (55, 68, 69), and MotB in *H. pylori* that makes up part of the flagellar stator responsible for generating torque (70). In *Salmonella enterica*, the switch protein FlgG of the C-ring, which acts as the rotary component of the flagella, responds to chemotactic sig-

nals and interacts with MotA of the stator that in turn interacts with PG-bound MotB (71). In *E. coli*, CheY is the response regulator that interacts with FliG to alter rotational direction of the flagella (72). Homologs of all these flagellar components are found in *C. jejuni* (55).

In *H. pylori*, the loss of the membrane-bound LT, MltD, affected motility without affecting the localization or number of flagella; this was hypothesized to result from the inability of MotA/B to generate torque due to impaired PG-MotB interactions (75). Similarly, the accumulation of O-acetylation in Δ *ape1* and the subsequent effect on LT activity could affect motility. An improperly assembled or unstable stator may impair the ability of the flagella to alter rotational direction in response to chemotactic signals. Mutants in *C. jejuni cheY* appear completely immotile on soft agar but not by microscopy (73, 74). Thus, the halo morphology does not support a complete loss in the ability to alter rotational direction but may still suggest an impaired response. The Δ *pgp1* and Δ *pgp2* straight mutants were also defective for motility in soft agar, so the changes in Δ *ape1* morphology could also account for the observed defects.

Biofilm formation in *C. jejuni* requires flagellum-mediated motility and attachment to a surface, lysis, and release of extracellular DNA to form the biofilm matrix (76). Δ *ape1* was defective for motility in soft agar but was not immotile. Despite this, Δ *ape1* exhibited a hyper- rather than hypo-biofilm formation phenotype. Envelope stress was recently shown to be a trigger for *C. jejuni* biofilm formation. A mutant exhibiting envelope stress was hyper-biofilm, and with DOC at 0.5 mg/ml, *C. jejuni* 81-176 wild type also exhibited enhanced biofilm formation (76). In this study, DOC concentrations below 0.5 mg/ml inhibit growth of Δ *ape1* (MIC₅₀; Table 4). This, together with the hyper-biofilm phenotype and increased cell surface hydrophobicity, suggests that the accumulation of OAP results in altered membrane properties and may be contributing to membrane stress in Δ *ape1*. However, because Δ *oap* exhibited hyper-biofilm formation as well and did not show evidence of membrane stress (*i.e.* no change in surface hydrophobicity or DOC sensitivity), the hyper-biofilm property may be partially independent of Ape1 activity and could be due to the loss of the Ape1 protein itself.

In *E. coli*, PG-associated lipoprotein (Pal) is often found in PG-protein complexes that are proposed to maintain envelope integrity. The phenotypes of some *E. coli pal* deletion mutants share similarities to *C. jejuni* Δ *ape1*, including increased sensitivity to bile salts and signs of altered motility (77). In addition, the PG binding domain of *E. coli* MotB and *E. coli* Pal are interchangeable (78) and both interact with MurNAc; our observations in Δ *ape1* could be a result of the presence of excess of O-acetyl groups on the PG MurNAc residues preventing stabilizing interactions between multiprotein structures and the PG sacculus.

Given the proposed role of O-acetylation in lysozyme resistance, it was expected that *C. jejuni* OAP mutants would demonstrate differential resistance to lysozyme. However, no differences were observed in lysozyme sensitivity (MIC₅₀). Attempts to destabilize the outer membrane by adding EDTA (at MIC₅₀ and concentrations down to 4-fold less than MIC₅₀) or lacto-

ferrin (at physiological concentration of 3 mg/ml, as described previously (37)) to the lysozyme incubations also failed to result in differential lysozyme sensitivity (data not shown). Lysozyme turbidometric assays were also unsuccessful due to the low yield of PG, resulting in an initial absorbance reading too low to accurately detect a response (data not shown).

Chick colonization by Δ *ape1* was significantly impaired compared with wild type and Δ *patA*, Δ *patB*, and Δ *oap* colonized to wild-type levels (Fig. 6A). Motility and chemotaxis are important for colonization (79); thus, this could be a potential explanation for the Δ *ape1* chick data. Alternatively, and/or additionally, the altered morphology and PG structure (18, 19), increased DOC susceptibility, differential long term survival properties, and other as-yet unknown factors could also contribute to the Δ *ape1* colonization defect.

Of the OAP mutants, only Δ *ape1* was impaired in adherence, invasion, and intracellular survival in INT407 epithelial cells. Whether these observations represent defects at each time point, defects in adherence that in turn affect recovery at later time points, or if Δ *ape1* is very rapidly killed upon invasion (as the 3-h “adherence” time point will also reflect invaded bacteria) will require further experimentation. Infection of human INT407 epithelial cells by Δ *ape1* also led to a decrease in IL-8 secretion. This may correlate with its reduced invasion properties.

Somewhat surprisingly, the reduction in PG O-acetylation had no significant effects on colonization, host cell interactions, or any other phenotype examined except for marginal decreases in halo formation, suggesting that under these conditions OAP by PatA and PatB offers no fitness advantage in host survival, which is perplexing and leaves the role of PG O-acetylation in *C. jejuni* yet to be determined. In contrast, the increase in O-acetylated PG in Δ *ape1* was detrimental to *C. jejuni* in multiple aspects important for pathogenesis. Future studies will focus on finding a direct link between PG O-acetylation and the observed changes in physiology, identifying other potential mechanisms of PG O-acetylation and de-O-acetylation, and revealing the underlying cause(s) of the impaired host interactions for Δ *ape1*.

Experimental Procedures

Strains and Growth Conditions—A list of bacterial strains and plasmids used in this study can be found in Table 5. Construction of mutant and complemented strains is described in [supplemental text S1](#) using primers listed in [supplemental Table S1](#). *C. jejuni* strains, unless otherwise stated, were grown in Mueller-Hinton (MH; Oxoid) broth or agar (1.7% w/v) supplemented with vancomycin (10 μg/ml) and trimethoprim (5 μg/ml) and when appropriate kanamycin (Km; 50 μg/ml) and chloramphenicol (Cm; 25 μg/ml). Standard laboratory conditions for *C. jejuni* growth were 38 °C under microaerophilic conditions (12% CO₂, 6% O₂, in N₂) in a Sanyo tri-gas incubator for MH agar or for standing MH broth cultures. For shaking MH broth cultures (hereafter referred to as broth cultures), *C. jejuni* were cultured in airtight jars using the Oxoid CampyGen Atmosphere generation system with shaking at 200 rpm. Experiments were performed using cultures initiated at A₆₀₀ 0.002 and grown in shaking broth for 16–18 h to reach exponential phase. For plasmid construction and protein purifica-

TABLE 5
Bacterial strains or plasmids used in this study

Strain or plasmid	Genotype or description	Source
<i>C. jejuni</i> strains		
81-176	Wild-type isolated from diarrheic patient	87
Δ <i>ape1</i>	81-176 <i>ape1::aphA3;Km^R</i>	This study
Δ <i>patB</i>	81-176 <i>patB::aphA3;Km^R</i>	This study
Δ <i>patA</i>	81-176 <i>patA::aphA3;Km^R</i>	This study
Δ <i>oap</i>	81-176 <i>oap::aphA3;Km^R</i>	This study
Δ <i>ape1^C</i>	81-176 Δ <i>ape1 rrr::ape1</i> (from pRRC-0638)	This study
Δ <i>patB^C</i>	81-176 Δ <i>patB rrr::patB</i> (from pRRC-0639)	This study
Δ <i>patA^C</i>	81-176 Δ <i>patA rrr::patA</i> (from pRRC-0640)	This study
<i>E. coli</i> strains		
DH5- α	F ⁻ , ϕ 80d <i>deoR lacZΔM15 endA1 recA1 hsdR17(r_K-m_K+) supE44 thi-1 gyrA96 relA1 Δ(lacZYA-argF) U169</i>	Invitrogen
BL21(λ DE3)	F ⁻ <i>ompT hsdS_B(r_B⁻, m_B⁻) gal dcm</i> [λ DE3]	Novagen
Plasmids		
pGEM-T	High copy, linearized, T-tailed, Blue/White, Ap ^R	Promega
pUC18-K2	Source of non-polar <i>aphA3</i> cassette; Ap ^R Km ^R	47
pGEM-T-0638	pGEM-T ligated to 0638 amplified with 0637-2 and 0639-5 (2113 bp); Ap ^R	This study
pGEM-0638:: <i>aphA-3</i>	pGEM-T-0638 inverse PCR amplified with 0638-3 and 0638-2 (4098 bp) and ligated to <i>aphA-3</i> (KpnI, HincII); Ap ^R , Km ^R	This study
pGEM-T-0639	pGEM-T ligated to 0639 amplified with 0639-1 and 0639-2 (2146 bp); Ap ^R	This study
pGEM-0639:: <i>aphA-3</i>	pGEM-T-0639 inverse PCR amplified with 0639-3 and 0639-4 (4267 bp) and ligated to <i>aphA-3</i> (KpnI, HincII); Ap ^R , Km ^R	This study
pGEM-T-0640	pGEM-T ligated to 0640 amplified with 0639-6 and 0641-4 (2396 bp); Ap ^R	This study
pGEM-0640:: <i>aphA-3</i>	pGEM-T-0640 inverse PCR amplified with 0640-1 and 0640-2 (4168 bp) and ligated to <i>aphA-3</i> (KpnI, HincII); Ap ^R , Km ^R	This study
pGEM-0638-40:: <i>aphA-3</i>	pGEM-T ligated to 0637 fragment amplified with 0637-1 and 0638-1 (1483 bp), 0641 fragment amplified with 0641-1 and 0641-2 (822 bp), and <i>aphA-3</i> (KpnI, HincII); Ap ^R , Km ^R	This study
pRRC	<i>C. jejuni</i> rRNA spacer integration vector; Cm ^R	48
pRRC-0638	pRRC ligated to 0638 amplified with 0638-C1(NheI) and 0638-C2(MfeI) (1347 bp); Cm ^{RR}	This study
pRRC-0639	pRRC ligated to 0639 amplified with 0639-C1(NheI) and 0639-C2(MfeI) (1276 bp); Cm ^{RR}	This study
pRRC-0640	pRRC ligated to 0640 amplified with 0640-C1(NheI) and 0640-C2(MfeI) (1616 bp); Cm ^{RR}	This study
pET28a(+)	Commercial vector for expression of recombinant His ₆ -tagged protein	Novagen
p0638-His ₆	pET28a(+) ligated to <i>ape1</i> amplified with 0638-eCF (NcoI) and 0638-eCR (EcoRI) (1121 bp) for expression of C-terminal His ₆ -tagged 0638; Km ^R	This study
pHis ₆ -0638	pET28a(+) ligated to <i>ape1</i> amplified with 0638-eNF (NheI) and 0638-eNR (EcoRI) (1116 bp); for expression of N-terminal His ₆ -tagged 0638 Km ^R	This study

tion, *E. coli* (DH5- α or BL21) strains were grown at 37 °C in Luria-Bertani (LB; Sigma) broth or LB agar (7.5% w/v) supplemented with ampicillin (100 μ g/ml), Km (25 μ g/ml), or Cm (15 μ g/ml) as required.

PG Isolation and Assessment of O-Acetylation Levels—PG isolation for O-acetylation analysis was performed as described previously with minor modifications (32, 51). Each strain was grown on ~60 MH-T agar plates (supplemented with Km or Cm as necessary) for ~18–20 h. The cells were harvested from the plate with 1 ml of ice-cold MH broth per plate and added to a conical tube. Strains were assessed by DICM to examine for contamination and the presence of coccoid cells, ensuring that the cultures had not entered stationary phase. *C. jejuni* cells transition from a helical to coccoid form in stationary phase. The cells were collected by centrifugation, resuspended in 50 ml of 25 mM sodium phosphate buffer, pH 6.5, and boiled in an equal volume of 8% w/v SDS in 25 mM sodium phosphate buffered at pH ~6.5 for 3 h under reflux with stirring (final concentration 4% SDS w/v). SDS-insoluble PG was washed with sterile double distilled H₂O (methylene blue/chloroform tests were performed to detect SDS) (80), frozen, and lyophilized. Lyophilized PG was resuspended in a minimal volume of buffer containing 10 mM Tris-HCl, pH 6.5, and 10 mM NaCl and sonicated (Misonix XL 2020, Mandel Scientific) on ice with a microtip for 2 min. The suspension was treated with 100 μ g/ml α -amylase (Fluka Biochemika), 10 μ g/ml DNase I (Invitrogen), 50 μ g/ml RNase A (ThermoScientific), and 20 mM MgSO₄ overnight at 37 °C. Protease (from *Streptomyces griseus*, Sigma), pre-incu-

bated at 60 °C for 2 h, was added to 200 μ g/ml and incubated overnight at 37 °C. Samples were then re-extracted in SDS, purified as above, lyophilized, and stored at -20 °C. O-Acetylation levels of lyophilized PG were evaluated as a ratio of total saponified O-linked acetate relative to total MurNAc content using mild base-catalyzed release of O-linked acetate (0.1 M NaOH, 40 °C for 4 h) and acid-catalyzed hydrolysis (6 N HCl, 100 °C for 1.5 h) for the complete liberation of PG monosaccharides, acetate, and muramic acid content. Components were quantified by HPLC as described previously (49, 50).

PG Isolation and Muropeptide Analysis—Each strain was grown on ~20–25 MH-T plates (supplemented with Km as required) for ~18–20 h to standardize growth phase and harvested with ice-cold MH-TV broth. Strains were assessed by DICM for contamination and coccoid cells to ensure that cultures had not grown into stationary phase. Cells were lysed using the boiling SDS technique as described previously (19). PG was further purified from the cell lysate and digested with the muramidase cellosyl (kindly provided by Hoechst, Frankfurt, Germany). The muropeptides were reduced with sodium borohydride, and subsequently separated by HPLC, all as described previously (81). Muropeptide structures were assigned based on (i) comparison with retention times of known muropeptides from *C. jejuni* and (ii) by mass spectrometry (Fig. 2) (18, 19, 82).

Expression, Purification, and Biochemical Assays of *C. jejuni* *Ape1*-His₆-Cjj81176_0638 (encoding *ape1*) was PCR-amplified without the predicted 21-amino acid signal peptide (as

identified by SignalP 4.1 Server) (83) and cloned into the pET28a(+) (Novagen) expression vector in-frame with the encoded His₆ tag at either the N or C termini of the gene forming pHis₆-0638 and p0638-His₆, respectively (Fig. 3, A and B). Expression constructs were transformed into *E. coli* BL21(ΔDE3), selected for Km^R, and confirmed by PCR and sequencing. Expression and isolation of recombinant Ape1 are described in the supplemental text S2.

Acetyltransferase activity was determined using pNPAC as a colorimetric substrate, as described previously (45). Assays were performed with 2.5 μg/ml purified Ape1-His₆ in 50 mM sodium phosphate buffer, pH 6.5, and 2 mM pNPAC. Reactions were monitored over 5 min, and specific activity was calculated with an experimentally determined molar absorptivity of 3.42 mM⁻¹ cm⁻¹ for *p*-nitrophenol (room temperature at pH 6.5).

PG-*O*-acetyltransferase activity was tested using purified PG retaining *O*-acetyl groups (as described under “PG Isolation and Assessment of *O*-Acetylation Levels”). Lyophilized PG was resuspended in sodium phosphate buffer, pH 6.5, to a concentration of 5 mg/ml and sonicated on ice with a microtip for 2 min (10 s on/10 s off). 500 μl of PG suspension was aliquoted into Eppendorf tubes to which, Ape1-His₆ was added to a concentration of 10 μg/ml (buffer only for negative control). Samples were incubated at 37 °C in a water bath for 24 h, after which samples were centrifuged (10 000 × *g*, 10 min, 4 °C) to pellet PG. Acetate content in supernatants were assessed using a commercial acetic acid assay kit (Megazyme International) as directed by the manufacturer.

Microscopy and CellTool Shape Analysis—Overnight MH-TV log-phase broth cultures were standardized to A₆₀₀ 0.05 and incubated for 4 or 7 h at 38 °C to generate early-exponential phase and mid-exponential phase cultures, respectively. The samples were processed for DICM. Live cells were imaged on agarose slabs on a Nikon Eclipse TE2000-U microscope equipped with a Hamamatsu C4742-95 digital camera.

For CellTool analysis (53), DICM images from multiple fields (yielding ≥400 cells per strain) were taken for each sample and processed by thresholding to generate binary images. Artifacts and cells that were clumped or ill-represented based on lighting were manually removed. The contours of the wild-type population were aligned to generate an average shape, and PCA was performed to generate a “shape model” based off principal components called “shape modes” that, together, describe at least 95% of the variation in the wild-type population. Contours of other strains were then aligned to the wild-type PCA shape model as a reference. Kolmogorov-Smirnov tests were used on each shape mode to determine whether there was a statistically significant difference in population distribution between the strains based on this wild-type shape model (53).

Phenotypic Characterization of OAP Mutants: Motility, Biofilm, Minimum Inhibitory/Bactericidal Concentrations and Cell Surface Hydrophobicity—Motility and biofilm formation assays were performed on log-phase bacterial broth cultures as described previously (19).

MIC₅₀ was determined in a 96-well plate as standing culture as described previously (84). Inocula of log-phase overnight cultures (100 μl) standardized to A₆₀₀ 0.0002 (10⁶

CFU/ml) in MH-TV and 11 μl of 10× concentrated test compound (in 2× serial dilutions) were added to each well. A₆₀₀ was measured for each well using the Varioskan Flash Multimode Plate Reader (Thermo Scientific), and MIC₅₀ was recorded as the lowest concentration of compound that reduced growth by 50% (by turbidity) relative to a positive control after 24 h.

Cell surface hydrophobicity was assessed using exponential phase bacterial broth cultures as described previously (85) with the following adjustments. Cultures were harvested at 8 000 × *g* for 10 min and washed three times with PBS. Cells were resuspended to an A₆₀₀ ~0.5 in PBS, and the absorbance was recorded. Hexadecane was added to the standardized cultures in a ratio of 1:4 hexadecane/culture by volume, vortexed for 5 min, and incubated at 38 °C for 30 min. The aqueous layer was isolated, aerated by bubbling N₂ gas through the aqueous layer for 30 s, and left open to the air for 10 min to ensure removal of all traces of hexadecane, and the A₆₀₀ was measured. Cell surface hydrophobicity was expressed as follows, where A_{600i} and A_{600f} refer to the optical densities before and after extraction, respectively.

$$\% \text{ Hydrophobicity} = \left(\frac{A_{600i} - A_{600f}}{A_{600i}} \right) \times 100\% \quad (\text{Eq. 1})$$

Chick Colonization—Chick colonization was assessed under protocol 10462 approved by the University of Michigan Committee on Care and Use of Animals, as described previously (19, 79).

Gentamicin Protection Assay for Host-Cell Infection—Gm protection assays were performed essentially as described previously (86). INT407 human epithelial cells were seeded into 24-well tissue culture plates at ~1.25 × 10⁵ cells in minimum essential medium (MEM) supplemented with 10% (v/v) FBS and 1× penicillin/streptomycin (Gibco, Life Technologies, Inc.) 24 h prior to infection. Infections were initiated by adding log-phase bacterial cultures standardized to A₆₀₀ 0.002 in MEM (1 × 10⁷ CFU/ml) and added to INT407 cells previously washed twice with 1 ml of MEM to give a multiplicity of infection of ~80. Adherence/invasion after 3 h of infection, invasion following a 2-h Gm treatment (150 μg/ml) to kill extracellular bacteria, and intracellular survival following removal of the high Gm concentration and incubation of cells in fresh medium with 10 μg/ml Gm and 3% FBS for an additional 3 h were assessed as described (86).

Interleukin-8 Quantification—The concentration of IL-8 secreted by INT407 cells either left uninfected or infected with *C. jejuni* wild-type strain 81-176, Δ*a*pe1, or Δ*a*pe1^C for 24 h was assayed using the human IL-8 ELISA kit (Thermo Fisher Scientific) as described previously (18).

Author Contributions—Initial conceptualization of the project details was by E. F. and E. C. G. R. H. conducted most of the experiments in the study and prepared the manuscript with E. C. G. and E. F. D. S. conducted the *O*-acetylation assays on purified PG from OAP mutants. J. B. conducted experiments on the muropeptide profiles of the OAP mutants. M. E. T. and J. G. J. conducted the chick colonization assays with OAP mutants. Oversight of the project was provided by E. C. G., A. J. C., W. V., and V. J. D. All authors reviewed and approved the final version of the manuscript.

Acknowledgment—We thank Jenny Vermeulen for technical assistance in PG sample preparations.

References

1. Yuki, N., and Hartung, H. P. (2012) Guillain-Barre syndrome. *N. Engl. J. Med.* **366**, 2294–2304
2. Jacobs, B. C., Rothbarth, P. H., van der Meché, F. G., Herbrink, P., Schmitz, P. I., de Klerk, M. A., and van Doorn, P. A. (1998) The spectrum of antecedent infections in Guillain-Barre syndrome: a case-control study. *Neurology* **51**, 1110–1115
3. Man, S. M. (2011) The clinical importance of emerging *Campylobacter* species. *Nat. Rev. Gastroenterol. Hepatol.* **8**, 669–685
4. Silva, J., Leite, D., Fernandes, M., Mena, C., Gibbs, P. A., and Teixeira, P. (2011) *Campylobacter* spp. as a foodborne pathogen: a review. *Front. Microbiol.* **2**, 200
5. Luangtongkum, T., Jeon, B., Han, J., Plummer, P., Logue, C. M., and Zhang, Q. (2009) Antibiotic resistance in *Campylobacter*: emergence, transmission and persistence. *Future Microbiol.* **4**, 189–200
6. Dasti, J. I., Tareen, A. M., Lugert, R., Zautner, A. E., and Gross, U. (2010) *Campylobacter jejuni*: a brief overview on pathogenicity-associated factors and disease-mediating mechanisms. *Int. J. Med. Microbiol.* **300**, 205–211
7. Gundogdu, O., Bentley, S. D., Holden, M. T., Parkhill, J., Dorrell, N., and Wren, B. W. (2007) Re-annotation and re-analysis of the *Campylobacter jejuni* NCTC11168 genome sequence. *BMC Genomics* **8**, 162
8. Parkhill, J., Wren, B. W., Mungall, K., Ketley, J. M., Churcher, C., Basham, D., Chillingworth, T., Davies, R. M., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, A. V., Moule, S., Pallen, M. J., Penn, C. W., et al. (2000) The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403**, 665–668
9. Louwen, R. P., van Belkum, A., Wagenaar, J. A., Doorduyn, Y., Achterberg, R., and Endtz, H. P. (2006) Lack of association between the presence of the pVir plasmid and bloody diarrhea in *Campylobacter jejuni* enteritis. *J. Clin. Microbiol.* **44**, 1867–1868
10. Tracz, D. M., Keelan, M., Ahmed-Bentley, J., Gibreel, A., Kowalewska-Grochowska, K., and Taylor, D. E. (2005) pVir and bloody diarrhea in *Campylobacter jejuni* enteritis. *Emerg. Infect. Dis.* **11**, 838–843
11. Bacon, D. J., Alm, R. A., Burr, D. H., Hu, L., Kopecko, D. J., Ewing, C. P., Trust, T. J., and Guerry, P. (2000) Involvement of a plasmid in virulence of *Campylobacter jejuni* 81-176. *Infect. Immun.* **68**, 4384–4390
12. Gaynor, E. C., Wells, D. H., MacKichan, J. K., and Falkow, S. (2005) The *Campylobacter jejuni* stringent response controls specific stress survival and virulence-associated phenotypes. *Mol. Microbiol.* **56**, 8–27
13. Szymanski, C. M., and Gaynor, E. C. (2012) How a sugary bug gets through the day: recent developments in understanding fundamental processes impacting *Campylobacter jejuni* pathogenesis. *Gut Microbes* **3**, 135–144
14. Dworkin, J. (2010) Form equals function? Bacterial shape and its consequences for pathogenesis. *Mol. Microbiol.* **78**, 792–795
15. Frirdich, E., and Gaynor, E. C. (2013) Peptidoglycan hydrolases, bacterial shape, and pathogenesis. *Curr. Opin. Microbiol.* **16**, 767–778
16. Vollmer, W. (2008) Structural variation in the glycan strands of bacterial peptidoglycan. *FEMS Microbiol. Rev.* **32**, 287–306
17. Young, K. D. (2006) The selective value of bacterial shape. *Microbiol. Mol. Biol. Rev.* **70**, 660–703
18. Frirdich, E., Biboy, J., Adams, C., Lee, J., Ellermeier, J., Giolda, L. D., Dirita, V. J., Girardin, S. E., Vollmer, W., and Gaynor, E. C. (2012) Peptidoglycan-modifying enzyme Pgp1 is required for helical cell shape and pathogenicity traits in *Campylobacter jejuni*. *PLoS Pathog.* **8**, e1002602
19. Frirdich, E., Vermeulen, J., Biboy, J., Soares, F., Taveirne, M. E., Johnson, J. G., DiRita, V. J., Girardin, S. E., Vollmer, W., and Gaynor, E. C. (2014) Peptidoglycan LD-carboxypeptidase Pgp2 influences *Campylobacter jejuni* helical cell shape and pathogenic properties and provides the substrate for the DL-carboxypeptidase Pgp1. *J. Biol. Chem.* **289**, 8007–8018
20. Boneca, I. G. (2005) The role of peptidoglycan in pathogenesis. *Curr. Opin. Microbiol.* **8**, 46–53
21. Cabanes, D., Dussurget, O., Dehoux, P., and Cossart, P. (2004) Auto, a surface associated autolysin of *Listeria monocytogenes* required for entry into eukaryotic cells and virulence. *Mol. Microbiol.* **51**, 1601–1614
22. Lenz, L. L., Mohammadi, S., Geissler, A., and Portnoy, D. A. (2003) SecA2-dependent secretion of autolytic enzymes promotes *Listeria monocytogenes* pathogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12432–12437
23. Viala, J., Chaput, C., Boneca, I. G., Cardona, A., Girardin, S. E., Moran, A. P., Athman, R., Mémet, S., Huerre, M. R., Coyle, A. J., DiStefano, P. S., Sansonetti, P. J., Labigne, A., Bertin, J., Philpott, D. J., and Ferrero, R. L. (2004) Nod1 responds to peptidoglycan delivered by the *Helicobacter pylori* cag pathogenicity island. *Nat. Immunol.* **5**, 1166–1174
24. Vollmer, W., and Tomasz, A. (2002) Peptidoglycan N-acetylglucosamine deacetylase, a putative virulence factor in *Streptococcus pneumoniae*. *Infect. Immun.* **70**, 7176–7178
25. Schoonmaker, M. K., Bishai, W. R., and Lamichhane, G. (2014) Nonclassical transpeptidases of *Mycobacterium tuberculosis* alter cell size, morphology, the cytosolic matrix, protein localization, virulence, and resistance to beta-lactams. *J. Bacteriol.* **196**, 1394–1402
26. Strating, H., Vandenende, C., and Clarke, A. J. (2012) Changes in peptidoglycan structure and metabolism during differentiation of *Proteus mirabilis* into swarmer cells. *Can. J. Microbiol.* **58**, 1183–1194
27. Sycuro, L. K., Pincus, Z., Gutierrez, K. D., Biboy, J., Stern, C. A., Vollmer, W., and Salama, N. R. (2010) Peptidoglycan crosslinking relaxation promotes *Helicobacter pylori*'s helical shape and stomach colonization. *Cell* **141**, 822–833
28. Sycuro, L. K., Wyckoff, T. J., Biboy, J., Born, P., Pincus, Z., Vollmer, W., and Salama, N. R. (2012) Multiple peptidoglycan modification networks modulate *Helicobacter pylori*'s cell shape, motility, and colonization potential. *PLoS Pathog.* **8**, e1002603
29. Moynihan, P. J., and Clarke, A. J. (2013) Assay for peptidoglycan O-acetyltransferase: a potential new antibacterial target. *Anal. Biochem.* **439**, 73–79
30. Moynihan, P. J., and Clarke, A. J. (2014) Substrate specificity and kinetic characterization of peptidoglycan O-acetyltransferase B from *Neisseria gonorrhoeae*. *J. Biol. Chem.* **289**, 16748–16760
31. Pfeffer, J. M., and Clarke, A. J. (2012) Identification of the first known inhibitors of O-acetylpeptidoglycan esterase: a potential new antibacterial target. *Chembiochem* **13**, 722–731
32. Pfeffer, J. M., Weadge, J. T., and Clarke, A. J. (2013) Mechanism of action of *Neisseria gonorrhoeae* O-acetylpeptidoglycan esterase, an SGNH serine esterase. *J. Biol. Chem.* **288**, 2605–2613
33. Veyrier, F. J., Williams, A. H., Mesnage, S., Schmitt, C., Taha, M. K., and Boneca, I. G. (2013) De-O-acetylation of peptidoglycan regulates glycan chain extension and affects *in vivo* survival of *Neisseria meningitidis*. *Mol. Microbiol.* **87**, 1100–1112
34. Rosenthal, R. S., Blundell, J. K., and Perkins, H. R. (1982) Strain-related differences in lysozyme sensitivity and extent of O-acetylation of gonococcal peptidoglycan. *Infect. Immun.* **37**, 826–829
35. Bera, A., Biswas, R., Herbert, S., and Götz, F. (2006) The presence of peptidoglycan O-acetyltransferase in various staphylococcal species correlates with lysozyme resistance and pathogenicity. *Infect. Immun.* **74**, 4598–4604
36. Moynihan, P. J., Sychantha, D., and Clarke, A. J. (2014) Chemical biology of peptidoglycan acetylation and deacetylation. *Bioorg. Chem.* **54**, 44–50
37. Wang, G., Lo, L. F., Forsberg, L. S., and Maier, R. J. (2012) *Helicobacter pylori* peptidoglycan modifications confer lysozyme resistance and contribute to survival in the host. *mBio* **3**, e00409–00412
38. Ellison, R. T., 3rd, and Giehl, T. J. (1991) Killing of Gram-negative bacteria by lactoferrin and lysozyme. *J. Clin. Invest.* **88**, 1080–1091
39. Ellison, R. T., 3rd, Giehl, T. J., and LaForce, F. M. (1988) Damage of the outer membrane of enteric Gram-negative bacteria by lactoferrin and transferrin. *Infect. Immun.* **56**, 2774–2781
40. Fleming, T. J., Wallsmith, D. E., and Rosenthal, R. S. (1986) Arthropathic properties of gonococcal peptidoglycan fragments: implications for the pathogenesis of disseminated gonococcal disease. *Infect. Immun.* **52**, 600–608

41. Moynihan, P. J., and Clarke, A. J. (2011) O-Acetylated peptidoglycan: controlling the activity of bacterial autolysins and lytic enzymes of innate immune systems. *Int. J. Biochem. Cell Biol.* **43**, 1655–1659
42. Weadge, J. T., Pfeffer, J. M., and Clarke, A. J. (2005) Identification of a new family of enzymes with potential O-acetylpeptidoglycan esterase activity in both Gram-positive and Gram-negative bacteria. *BMC Microbiol.* **5**, 49
43. Weadge, J. T., and Clarke, A. J. (2006) Identification and characterization of O-acetylpeptidoglycan esterase: a novel enzyme discovered in *Neisseria gonorrhoeae*. *Biochemistry* **45**, 839–851
44. Dillard, J. P., and Hackett, K. T. (2005) Mutations affecting peptidoglycan acetylation in *Neisseria gonorrhoeae* and *Neisseria meningitidis*. *Infect. Immun.* **73**, 5697–5705
45. Weadge, J. T., and Clarke, A. J. (2007) *Neisseria gonorrhoeae* O-acetylpeptidoglycan esterase, a serine esterase with a Ser-His-Asp catalytic triad. *Biochemistry* **46**, 4932–4941
46. Williams, A. H., Veyrier, F. J., Bonis, M., Michaud, Y., Raynal, B., Taha, M. K., White, S. W., Haouz, A., and Boneca, I. G. (2014) Visualization of a substrate-induced productive conformation of the catalytic triad of the *Neisseria meningitidis* peptidoglycan O-acetyltransferase reveals mechanistic conservation in SGNH esterase family members. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 2631–2639
47. Ménard, R., Sansonetti, P. J., and Parsot, C. (1993) Nonpolar mutagenesis of the *ipa* genes defines IpaB, IpaC, and IpaD as effectors of *Shigella flexneri* entry into epithelial cells. *J. Bacteriol.* **175**, 5899–5906
48. Karlyshev, A. V., and Wren, B. W. (2005) Development and application of an Insertional system for gene delivery and expression in *Campylobacter jejuni*. *Appl. Environ. Microbiol.* **71**, 4004–4013
49. Clarke, A. J. (1993) Compositional analysis of peptidoglycan by high-performance anion-exchange chromatography. *Anal. Biochem.* **212**, 344–350
50. Clarke, A. J. (1993) Extent of peptidoglycan O-acetylation in the tribe proteaeae. *J. Bacteriol.* **175**, 4550–4553
51. Dupont, C., and Clarke, A. J. (1991) Dependence of lysozyme-catalysed solubilization of *Proteus mirabilis* peptidoglycan on the extent of O-acetylation. *Eur. J. Biochem.* **195**, 763–769
52. Scheurwater, E., Reid, C. W., and Clarke, A. J. (2008) Lytic transglycosylases: bacterial space-making autolysins. *Int. J. Biochem. Cell Biol.* **40**, 586–591
53. Pincus, Z., and Theriot, J. A. (2007) Comparison of quantitative methods for cell-shape analysis. *J. Microsc.* **227**, 140–156
54. Lin, M. F., Lucas, H. C., and Shmueli, G. (2013) Too big to fail: large samples and the *p*-value problem. *Inform. Syst. Res.* **24**, 906–917
55. Lertsethtakarn, P., Ottemann, K. M., and Hendrixson, D. R. (2011) Motility and chemotaxis in *Campylobacter* and *Helicobacter*. *Annu. Rev. Microbiol.* **65**, 389–410
56. Buswell, C. M., Herlihy, Y. M., Lawrence, L. M., McGuigan, J. T., Marsh, P. D., Keevil, C. W., and Leach, S. A. (1998) Extended survival and persistence of *Campylobacter* spp. in water and aquatic biofilms and their detection by immunofluorescent-antibody and -rRNA staining. *Appl. Environ. Microbiol.* **64**, 733–741
57. McLennan, M. K., Ringoir, D. D., Fridrich, E., Svensson, S. L., Wells, D. H., Jarrell, H., Szymanski, C. M., and Gaynor, E. C. (2008) *Campylobacter jejuni* biofilms up-regulated in the absence of the stringent response utilize a calcofluor white-reactive polysaccharide. *J. Bacteriol.* **190**, 1097–1107
58. Joshua, G. W., Guthrie-Irons, C., Karlyshev, A. V., and Wren, B. W. (2006) Biofilm formation in *Campylobacter jejuni*. *Microbiology* **152**, 387–396
59. Rosenberg, M., Gutnick, D., and Rosenberg, E. (1980) Adherence of bacteria to hydrocarbons—a simple method for measuring cell-surface hydrophobicity. *FEMS Microbiol. Lett.* **9**, 29–33
60. Everest, P. H., Goossens, H., Butzler, J. P., Lloyd, D., Knutton, S., Ketley, J. M., and Williams, P. H. (1992) Differentiated Caco-2 cells as a model for enteric invasion by *Campylobacter jejuni* and *C. coli*. *J. Med. Microbiol.* **37**, 319–325
61. Bacon, D. J., Szymanski, C. M., Burr, D. H., Silver, R. P., Alm, R. A., and Guerry, P. (2001) A phase-variable capsule is involved in virulence of *Campylobacter jejuni* 81-176. *Mol. Microbiol.* **40**, 769–777
62. Moynihan, P. J., and Clarke, A. J. (2010) O-Acetylation of peptidoglycan in Gram-negative bacteria: identification and characterization of peptidoglycan O-acetyltransferase in *Neisseria gonorrhoeae*. *J. Biol. Chem.* **285**, 13264–13273
63. Romeis, T., and Höltje, J. V. (1994) Specific interaction of penicillin-binding proteins 3 and 7/8 with soluble lytic transglycosylase in *Escherichia coli*. *J. Biol. Chem.* **269**, 21603–21607
64. Vollmer, W., von Rechenberg, M., and Höltje, J. V. (1999) Demonstration of molecular interactions between the murein polymerase PBP1B, the lytic transglycosylase MltA, and the scaffolding protein MipA of *Escherichia coli*. *J. Biol. Chem.* **274**, 6726–6734
65. Clarke, A. J., and Dupont, C. (1992) O-Acetylated peptidoglycan: its occurrence, pathobiological significance, and biosynthesis. *Can. J. Microbiol.* **38**, 85–91
66. Mattei, P. J., Neves, D., and Dessen, A. (2010) Bridging cell wall biosynthesis and bacterial morphogenesis. *Curr. Opin. Struct. Biol.* **20**, 749–755
67. Scyuro, L. K., Rule, C. S., Petersen, T. W., Wyckoff, T. J., Sessler, T., Nagarkar, D. B., Khalid, F., Pincus, Z., Biboy, J., Vollmer, W., and Salama, N. R. (2013) Flow cytometry-based enrichment for cell shape mutants identifies multiple genes that influence *Helicobacter pylori* morphology. *Mol. Microbiol.* **90**, 869–883
68. Hizukuri, Y., Kojima, S., Yakushi, T., Kawagishi, I., and Homma, M. (2008) Systematic Cys mutagenesis of FlgI, the flagellar P-ring component of *Escherichia coli*. *Microbiology* **154**, 810–817
69. Jones, C. J., Homma, M., and Macnab, R. M. (1989) L-, P-, and M-ring proteins of the flagellar basal body of *Salmonella typhimurium*: gene sequences and deduced protein sequences. *J. Bacteriol.* **171**, 3890–3900
70. Roujeinikova, A. (2008) Crystal structure of the cell wall anchor domain of MotB, a stator component of the bacterial flagellar motor: implications for peptidoglycan recognition. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10348–10353
71. Morimoto, Y. V., Nakamura, S., Hiraoka, K. D., Namba, K., and Minamino, T. (2013) Distinct roles of highly conserved charged residues at the MotA-FlgI interface in bacterial flagellar motor rotation. *J. Bacteriol.* **195**, 474–481
72. Roman, S. J., Meyers, M., Volz, K., and Matsumura, P. (1992) A chemotactic signaling surface on CheY defined by suppressors of flagellar switch mutations. *J. Bacteriol.* **174**, 6247–6255
73. Kanungpean, D., Kakuda, T., and Takai, S. (2011) False-positive responses of *Campylobacter jejuni* when using the chemical-in-plug chemotaxis assay. *J. Vet. Med. Sci.* **73**, 389–391
74. McGee, D. J., Langford, M. L., Watson, E. L., Carter, J. E., Chen, Y. T., and Ottemann, K. M. (2005) Colonization and inflammation deficiencies in Mongolian gerbils infected by *Helicobacter pylori* chemotaxis mutants. *Infect. Immun.* **73**, 1820–1827
75. Roue, S., Bonis, M., Chaput, C., Ecobichon, C., Mattox, A., Barrière, C., Geldmacher, N., Guadagnini, S., Schmitt, C., Prévost, M. C., Labigne, A., Backert, S., Ferrero, R. L., and Boneca, I. G. (2012) Peptidoglycan maturation enzymes affect flagellar functionality in bacteria. *Mol. Microbiol.* **86**, 845–856
76. Svensson, S. L., Pryjma, M., and Gaynor, E. C. (2014) Flagella-mediated adhesion and extracellular DNA release contribute to biofilm formation and stress tolerance of *Campylobacter jejuni*. *PLoS ONE* **9**:e106063
77. Godlewska, R., Wiśniewska, K., Pietras, Z., and Jagusztyn-Krynicka, E. K. (2009) Peptidoglycan-associated lipoprotein (Pal) of Gram-negative bacteria: function, structure, role in pathogenesis and potential application in immunoprophylaxis. *FEMS Microbiol. Lett.* **298**, 1–11
78. Hizukuri, Y., Morton, J. F., Yakushi, T., Kojima, S., and Homma, M. (2009) The peptidoglycan-binding (PGB) domain of the *Escherichia coli* pal protein can also function as the PGB domain in *E. coli* flagellar motor protein MotB. *J. Biochem.* **146**, 219–229
79. Hendrixson, D. R., and DiRita, V. J. (2004) Identification of *Campylobacter jejuni* genes involved in commensal colonization of the chick gastrointestinal tract. *Mol. Microbiol.* **52**, 471–484
80. Hayashi, K. (1975) A rapid determination of sodium dodecyl sulfate with methylene blue. *Anal. Biochem.* **67**, 503–506
81. Glauner, B. (1988) Separation and quantification of mucopeptides with high-performance liquid-chromatography. *Anal. Biochem.* **172**, 451–464

Role of *C. jejuni* Peptidoglycan O-Acetylation

82. Bui, N. K., Gray, J., Schwarz, H., Schumann, P., Blanot, D., and Vollmer, W. (2009) The peptidoglycan sacculus of *Myxococcus xanthus* has unusual structural features and is degraded during glycerol-induced myxospore development. *J. Bacteriol.* **191**, 494–505
83. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786
84. Hancock, R. E. (1999) *Hancock Laboratory Methods*, Department of Microbiology and Immunology, University of British Columbia, British Columbia, Canada
85. Ben Abdallah, F., Lagha, R., Said, K., Kallel, H., and Gharbi, J. (2014) Detection of cell surface hydrophobicity, biofilm and fimbriae genes in *Salmonella* isolated from tunisian clinical and poultry meat. *Iran J. Public Health* **43**, 423–431
86. Pryjma, M., Apel, D., Huynh, S., Parker, C. T., and Gaynor, E. C. (2012) FdhTU-modulated formate dehydrogenase expression and electron donor availability enhance recovery of *Campylobacter jejuni* following host cell infection. *J. Bacteriol.* **194**, 3803–3813
87. Korlath, J. A., Osterholm, M. T., Judy, L. A., Forfang, J. C., and Robinson, R. A. (1985) A point-source outbreak of campylobacteriosis associated with consumption of raw milk. *J. Infect. Dis.* **152**, 592–596
88. Glauner, B., Hölzle, J. V., and Schwarz, U. (1988) The composition of the murein of *Escherichia coli*. *J. Biol. Chem.* **263**, 10088–10095

A Germline Variant in the *PANX1* Gene Has Reduced Channel Function and Is Associated with Multisystem Dysfunction*[♦]

Received for publication, January 27, 2016, and in revised form, April 12, 2016. Published, JBC Papers in Press, April 15, 2016, DOI 10.1074/jbc.M116.717934

Qing Shao^{†1}, Kristin Lindstrom^{§1}, Ruoyang Shi^{||1,2}, John Kelly[‡], Audrey Schroeder^{**}, Jane Juusola^{‡‡}, Kara L. Levine^{‡‡}, Jessica L. Esseltine[‡], Silvia Penuela[‡], Michael F. Jackson^{||}, and Dale W. Laird^{‡‡3}

From the [†]Department of Anatomy and Cell Biology, The University of Western Ontario, London, Ontario N6A 5C1, Canada, the [§]Division of Genetics and Metabolism, Phoenix Children's Hospital, Phoenix, Arizona 85016, the ^{||}Department of Pharmacology and Therapeutics, University of Manitoba, Winnipeg, Manitoba R3E 0Z3, Canada, the ^{||}Kleyns Institute for Advanced Medicine, Health Sciences Centre, Winnipeg, Manitoba R3E 0Z3, Canada, the ^{**}Division of Genetics, University of Rochester Medical Center, Rochester, New York 14642, and ^{‡‡}GeneDx, Gaithersburg, Maryland 20877

Pannexin1 (PANX1) is probably best understood as an ATP release channel involved in paracrine signaling. Given its ubiquitous expression, *PANX1* pathogenic variants would be expected to lead to disorders involving multiple organ systems. Using whole exome sequencing, we discovered the first patient with a homozygous *PANX1* variant (c.650G→A) resulting in an arginine to histidine substitution at position 217 (p.Arg217His). The 17-year-old female has intellectual disability, sensorineural hearing loss requiring bilateral cochlear implants, skeletal defects, including kyphoscoliosis, and primary ovarian failure. Her consanguineous parents are each heterozygous for this variant but are not affected by the multiorgan syndromes noted in the proband. Expression of the p.Arg217His mutant in HeLa, N2A, HEK293T, and Ad293 cells revealed normal PANX1 glycosylation and cell surface trafficking. Dye uptake, ATP release, and electrophysiological measurements revealed p.Arg217His to be a loss-of-function variant. Co-expression of the mutant with wild-type PANX1 suggested the mutant was not dominant-negative to PANX1 channel function. Collectively, we demonstrate a *PANX1* missense change associated with human disease in the first report of a “*PANX1*-related disorder.”

Pannexins are a new class of large-pore channels that were discovered early in the new millennium (1, 2). Members of the gene family (*PANX1*, *PANX2*, and *PANX3*) are expressed in numerous organs, tissues, and cells with *PANX1* being the most prevalent (3). Rodent Panx1 is an ~41–48-kDa protein with its broad range in size due to the fact that it is post-translationally modified in what is now referred to as Gly0, Gly1, and Gly2 species to reflect the degree of glycosylation (4–6). *PANX1* oligomerizes into a hexamer that contains a large pore functioning at the cell surface to allow the passage of small molecules below 1000 daltons in size (7, 8). Although the scope of

molecules that pass through PANX1 pores is likely broad (9, 10), the functional consequence of ATP release via these channels is best understood (11). For instance, PANX1 channels have been shown to release ATP in apoptotic immune cells as “find me” signals for the clearing of dying cells (12).

In the last decade, PANX1 channels have become intimately linked to disease because of the fact that they are expressed in the vast majority of human cell types (13). Until this study, the link to disease has been associated with basal or elevated functional levels of PANX1, but the mechanisms involved remain poorly understood (13). In the first reported association with disease, PANX1 was linked to neuronal cell death in models of ischemia and stroke followed later by clear linkages to seizure severity and duration (14–16). The abundant expression of PANX1 in enteric neurons led to the discovery that these channels played vital roles in inflammatory bowel diseases, including ulcerative colitis and Crohn's disease (17). Surprisingly, PANX1 channels can also be hijacked by viruses to facilitate infection as documented for HIV-1 (18). Furthermore, in mouse models, high levels of Panx1 in melanomas have been shown to facilitate disease progression, although Panx1 overexpression has been shown to be tumor suppressive in glioblastomas, suggesting that pannexins are likely to have tumor-specific effects in cancer (19–21). The list of connections between PANX1 and disease is extensive and continues to grow as there are elegant studies supporting a link between PANX1 and epilepsy (22, 23), glaucoma (24), migraines (25), Alzheimer disease (26), and diabetes (27).

Although no disease-linked germline *PANX1* variants have been identified prior to this study, Kwak and co-workers (28), including a member from our team, discovered through sequencing of 96 healthy patients that a single nucleotide polymorphism (400A→C) existed with a frequency of approximately one-third 400A allele and two-thirds 400C allele. Although none exhibited overt disease, those homozygous for the 400C allele exhibited greater collagen-induced platelet aggregation, suggesting the possibility that there may be some variability in platelet reactivity among healthy individuals (28).

In this study we report on the first patient with a *PANX1* homozygous germline variant resulting in an arginine at position 217 being replaced with a histidine (p.Arg217His). This young female patient clinically presents with extensive disease that includes intellectual disabilities, severe hearing loss, and

* This work was supported in part by Canadian Institutes of Health Research Grants 130530 (to D. W. L. and S. P.) and 125901 (to M. F. J.). The authors declare that they have no conflicts of interest with the contents of this article.

♦ This article was selected as a Paper of the Week.

¹ These authors contributed equally to this work.

² Recipient of a postdoctoral fellowship from Research Manitoba.

³ To whom correspondence should be addressed: Dept. of Anatomy and Cell Biology, University of Western Ontario, London, Ontario N6A-5C1, Canada. Tel.: 519-661-2111, ex. 86827; E-mail: dale.laird@schulich.uwo.ca.

multiple other multisystem defects. Her unaffected parents and sibling are heterozygous for the c.650G→A *PANX1* variant. Generation and characterization of the R217H mutant revealed that it is a loss-of-function variant as assessed by ATP release, dye uptake, and electrophysiological evaluation of the channel properties. Although functional levels of PANX1 have been correlated to the onset and/or progression of over 10 diseases (13), disease-linked germline variants in the *PANX1* gene have not been previously reported. This study represents the first report of a patient harboring a disease-associated *PANX1* variant.

Experimental Procedures

Sequencing—Genomic DNA was extracted from whole blood from the proband and her parents. Exome sequencing was performed on exon targets captured using the Agilent SureSelect Human All Exon V4 (50 Mb) kit (Agilent Technologies, Santa Clara, CA). The sequencing methodology and variant interpretation protocol has been previously described (29). Briefly, whole exome sequencing (WES)⁴ produced 18.2 Gb of sequencing data for the proband. Mean coverage of captured regions was ~251× for the proband's sample, with ~99.17% covered with at least 10× coverage, an average of 89.82% of base call quality of Q30 or greater, and an overall average mean quality score of Q35. Filtering of common single nucleotide polymorphisms (>10% frequency present in the 1000 Genomes database) resulted in ~4187 variants in the proband sample. After automated filtering of variants with a minor allele frequency of >10%, manual curation was performed to filter less common variants with a minor allele frequency of 1–10% and single variants in genes inherited from unaffected parents and to evaluate predicted effects of rare variants and associated human conditions.

Cells and Reagents—Normal rat kidney (NRK), mouse Neuro2A (N2A), and human embryonic kidney (HEK293T) cells were obtained from the American Type Culture Collection (Manassas, VA). Ad293 cells (derivative of human embryonic kidney 293 cells with improved cell adherence) were obtained from Agilent Technologies (Palo Alto, CA). All cells were grown in high glucose (4500 mg of glucose/liter) DMEM (Invitrogen) supplemented with 10% fetal bovine serum, 100 units/ml penicillin, 100 μg/ml streptomycin, and 2 mM L-glutamine. Cells were cultured within a humidified environment that maintained 5% CO₂ and a temperature of 37 °C. Probenecid (catalog no. P36400) was purchased from Molecular Probes, and ATP (catalog no. PV3227) was obtained from ThermoFisher Scientific. Carbenoxolone (catalog no. C4790) and Dulbecco's phosphate-buffered saline (catalog no. D1283) were obtained from Sigma.

Mutant DNA Constructs—In humans, the gene encoding pannexin 1 is referred to as *PANX1*. *Panx1* is used when referring to the equivalent mouse gene, and the encoded proteins are not italicized. An expression vector encoding human PANX1 was purchased from InvivoGen (pUNO1-hPanx1). The R217H

PANX1 mutant encoding expression vector was generated by special order to Norclone Biotech Industries (London, Ontario, Canada). The fidelity of the PANX1- and R217H-encoding vectors was confirmed by DNA sequencing at the Robarts Research Institute DNA Sequencing Facility (London, Ontario, Canada) using an Applied Biosystems (Foster City, CA) 3730 analyzer.

Transfections and Immunofluorescent Labeling—Transient transfections were performed using Lipofectamine2000 (Life Technologies, Inc.) according to the manufacturer's instructions. For some experiments, a GFP-encoding vector was co-transfected with the PANX1/R217H-encoding vector. For stable transfections, cells were kept under G418/blasticidin selection pressure.

For electrophysiological experiments, pUNO1-PANX1 (2 μg) or pUNO1-R217H (2 μg) plasmid DNA was co-transfected with pLB-GFP DNA (0.6 μg) and pCDNA3.1 (7 μg; carrier DNA) using JetPRIME® (Polyplus-transfection Inc., New York) according to the manufacturer's instructions. In some cases, equal concentrations of plasmids encoding PANX1 and R217H were used in the transfection to determine whether the mutant affected the function of wild-type PANX1. For these experiments, pUNO1-PANX1 (2 μg) plus pUNO1-R217H (2 μg) plasmid DNA was co-transfected with pLB-GFP DNA (0.6 μg) and pCDNA3.1 (7 μg; carrier DNA) using JetPRIME®. HEK293T cells not transfected with either PANX1 or R217H served as negative controls (untransfected). In this case, cells were mock-transfected with only pLB-GFP DNA (0.6 μg) and pCDNA3.1 (7 μg; carrier DNA).

HEK293T cells, maintained as described above but with no antibiotics, were plated onto 100-mm dishes and transfected when reaching 60–80% confluence. Transfected cells were subsequently resuspended and seeded onto 35-mm dishes for electrophysiological recordings. The remaining sister cells were used for immunoblotting, as described below.

For immunolabeling studies, cells were fixed in 80% methanol, 20% acetone for 20 min at 4 °C and blocked for 30 min with 2% BSA. For PANX1, we used an affinity-purified custom-made rabbit anti-human PANX1 polyclonal antibody (PANX1 CT-412, 0.5 μg/ml), generated by Genemed Synthesis (San Francisco) against the C-terminal sequence of human PANX1 (⁴¹²NGEKNARQRLDSSC⁴²⁶) (30). An anti-Cx43 monoclonal antibody (1:50 dilution; P4G9 from Dr. Paul D. Lampe, Fred Hutchinson Cancer Research Center, Seattle) was also used. Primary antibody labeling was followed by fluorescently tagged secondary antibodies Alexa Fluor® 555 and Alexa Fluor® 488 (Life Technologies, Inc.; diluted 1:500) for 1 h at room temperature. Cell nuclei were stained for 5 min with TO-PRO®-3 Iodide (642/661) (Life Technologies, Inc.) and then rinsed in distilled H₂O before mounting. All labeling was visualized with an LSM 510 META inverted confocal microscope equipped with a 63× oil objective (Carl Zeiss, Jena, Germany).

Western Blot—Cells were solubilized using cell lysis buffer (1% Triton X-100, 150 mM NaCl, 10 mM Tris/HCl (pH 7.4), 1 mM EDTA, 1 mM EGTA, and 0.5% Nonidet P-40, supplemented with protease inhibitor mixture (CompleteMini, Roche Applied Science)) and phosphatase inhibitors (50 mM sodium fluoride and 0.5 mM sodium orthovanadate). Protein concen-

⁴The abbreviations used are: WES, whole exome sequencing; ANOVA, analysis of variance; NRK, normal rat kidney; BFA, brefeldin A; CBX, carbenoxolone.

Disease-associated Loss-of-function Panx1 Variant

trations were determined using a BCA protein determination kit (Pierce). Cleared lysates were subjected to SDS-PAGE and immunoblotted with a rabbit anti-PANX1 antibody (PANX1 CT-412, 0.25 $\mu\text{g}/\text{ml}$) or a mouse anti- β -tubulin (0.4 $\mu\text{g}/\text{ml}$, Sigma, catalog no. T8328) antibody at 4 °C overnight. Primary antibodies were detected using the fluorescently conjugated anti-rabbit Alexa Fluor® 680 (1:10,000 dilution, LI-COR Biosciences) or anti-mouse IRDye 800 (1:10,000 dilution, Rockland Immunochemicals, Inc.) antibodies and scanned using the Odyssey Infrared Imaging System (LI-COR Biosciences).

Methods used by the University of Manitoba group were similar to that of the London group. Briefly, cells were solubilized in lysis buffer containing 1% Triton X-100, 0.1% SDS, 0.5% sodium deoxycholate, 150 mM NaCl, 50 mM Tris/HCl (pH 7.4), 10 mM EDTA. An antibody recognizing β -actin (HRP conjugated anti- β -actin, 1:10,000 dilution, Sigma) was used as a loading control. Immunoblots were imaged using a ChemiDoc MP system (Bio-Rad).

Cell Surface Biotinylation—Ad293 or N2A cells were transiently transfected with the cDNAs encoding wild-type PANX1 or the R217H mutant. Forty eight hours after transfection, cells were treated for 40 min with either DMEM + 5% fetal bovine serum (FBS) (control) or DMEM/FBS containing 140 mM K^+Glu and 100 mM K^+Cl^- . Cells were then placed on ice, washed in ice-cold Hanks' balanced salt solution, and cell surface proteins labeled with 1.5 mg/ml EZLink Sulfo-NHS-SS-biotin (ThermoFisher catalog no. 21331) for 1 h. Biotin was subsequently quenched with 100 mM glycine for 30 min, and cells were lysed. Cleared supernatant containing 150–250 μg of protein was incubated with 25 μl of NeutrAvidin affinity beads (ThermoFisher catalog no. 29200) overnight, rotating at 4 °C to precipitate biotin-labeled proteins. Following incubation, the beads were washed twice with PBS, and precipitated proteins were separated by SDS-PAGE, transferred to a nitrocellulose membrane, and immunoblotted to identify biotinylated PANX1 proteins using a primary rabbit anti-human PANX1 antibody (1:1000 dilution). Total PANX1 and GAPDH levels were determined by immunoblotting 20 μg of protein from each cell lysate used for biotinylation and the input.

ATP Release Assay—N2A cells were transfected to express wild-type PANX1 or the R217H mutant and subcultured into 24-well dishes. Forty eight hours post-transfection, the cell medium was exchanged to medium with 5% FBS (heat-inactive) plus 20 μM ARL67156 (Santa Cruz Biotechnology, Santa Cruz, CA). ATP release was stimulated with 140 mM potassium gluconate and 100 mM KCl ("High K^+ " solution) for 20 min. In addition, some cells were pre-treated with 70 μM carbenoxolone (CBX, Sigma) to block the opening of PANX1 channels (31, 32). Samples were taken and processed for ATP content using a bioluminescence ATP determination kit (Life Technologies, Inc.). Briefly, 10 μl of the cell supernatants were transferred to 96-well dishes with 90 μl of a luciferase standard reaction solution, according to the manufacturer's protocol. ATP released into the medium was quantitatively measured by luminescence. Although cells were plated at equal numbers, to account for any unexpected differences in cell numbers, luminescence values were normalized to the total amount of protein for each well (as determined by a BCA assay, Pierce), and ATP

concentrations were then determined from an ATP standard curve. Cell-free solutions were tested in the presence of different concentrations of ATP as a control. Plotted values are means \pm S.E. * = <0.05, ** = <0.01, and *** = <0.001 as determined by one-way ANOVA analysis followed by a Bonferroni corrected post hoc test ($n = 3$). All statistical analyses were performed using GraphPad version 4.1.

Dye Uptake Assay—Ad293 cells were transiently and/or stably engineered to express GFP together with wild-type PANX1 or the R217H mutant. Briefly, cells were sparsely subcultured on 35-mm glass-bottom dishes (MatTek Corp., Ashland, MA) for 1 day prior to the medium being replaced with 5% FBS (heat-inactive), DMEM containing 140 mM potassium gluconate, 100 mM KCl, and 10 μM ethidium bromide (EtBr) (7, 33, 34). To block PANX1 channels, cells were treated with 1 mM probenecid (35). Rapid time-lapse imaging was used to detect EtBr uptake and GFP fluorescence every 5 min for up to 40 min using a Zeiss LSM 510 META imaging system. Mean fluorescence intensity was measured using Zen lite (Zeiss). For each treatment, 10–50 cells were analyzed for mean fluorescence intensity at 0, 5, 30, and 40 min after exposure to the dye. Data are presented as means \pm S.E. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$ PANX-expressing cells compared with wild-type or R217H mutant-expressing cells as determined by a one-way ANOVA followed by Tukey's Multiple Comparison Test. Statistical analysis was performed using GraphPad version 4.1.

Electrophysiology—Tight-seal whole-cell recordings were conducted at room temperature 24–72 h after transfection using an Axon MultiClamp 700A patch clamp amplifier and Digidata 1322A data acquisition system (Molecular Devices). On the day of experiments, recordings were performed using an interleaved design from 293T cells expressing wild-type PANX1, R217H, or both prepared in parallel. Transfected cells were visually selected for recording on the basis of GFP fluorescence. Using a gravity-driven multi-barreled perfusion system, cells were continuously superfused with bath solution (extracellular fluid) containing 140 mM NaCl, 5.4 mM KCl, 2 mM CaCl_2 , 1 mM MgCl_2 , 33 mM D-glucose, and 25 mM HEPES, adjusted to pH 7.4 (with NaOH), and osmolarity between 310 and 320 mosm/liter. Patch pipettes were pulled from borosilicate glass (World Precision Instruments, Inc., Sarasota, FL) and had a resistance of 4–5 megohms when filled with a pipette solution (intracellular fluid) containing 150 mM cesium gluconate, 2 mM MgCl_2 , and 10 mM HEPES, adjusted to pH 7.3 (with CsOH), and osmolarity between 290 and 300 mosm/liter. For the high potassium treatment, a modified bath solution was applied containing 90 mM NaCl, 50 mM potassium gluconate, 5.4 mM KCl, 2 mM CaCl_2 , 1 mM MgCl_2 , 33 mM D-glucose, and 25 mM HEPES, adjusted to pH 7.4 (with NaOH), and osmolarity between 310 and 320 mosm/liter.

In cells voltage clamped at -60 mV, PANX1 and R217H currents were recorded in response to 500-ms voltage ramps (± 100 mV, 1/10 s). PANX1 and R217H ramp currents were first recorded in normal bath solution (or high potassium solution) for 5 min and then inhibited by a bath solution (normal or high potassium) containing 100 μM carbenoxolone. Signals, filtered at 2 kHz and sampled at 10 kHz, were collected and analyzed using Clampex 9.2 and Clampfit 9.2 software (Molecular

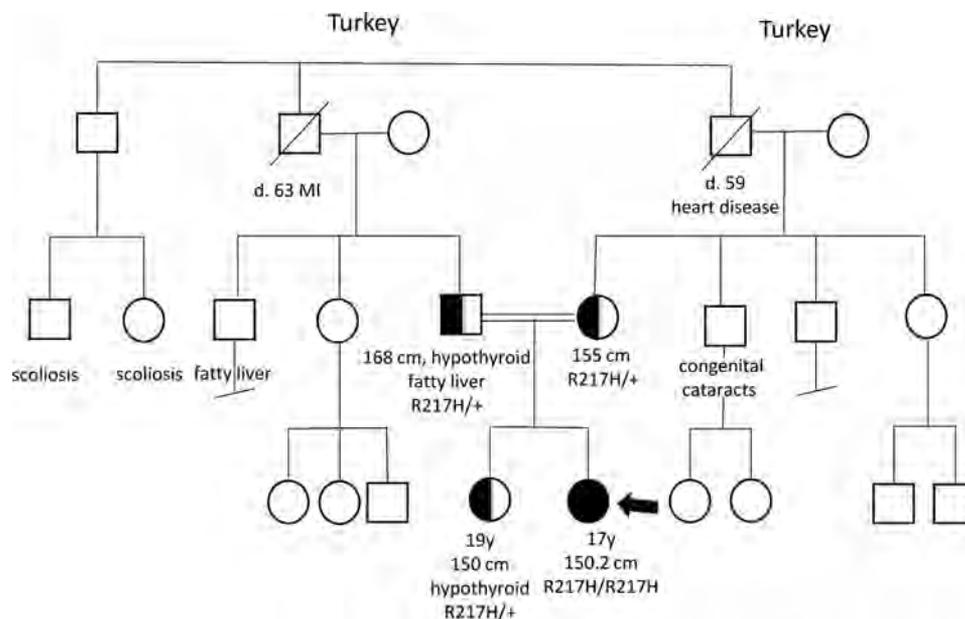


FIGURE 1. **Three generation pedigree demonstrating consanguinity, with the proband's parents as first cousins.** Selected clinical conditions are noted. Full shading represents the homozygous R217H variants, and half-shading represents heterozygous R217H PANX1 variants. Arrow indicates proband; "+" indicates wild-type allele.

Devices). Data are presented as means \pm S.E. Statistical analyses were conducted using GraphPad Prism 5.0 software (GraphPad Software, Inc., La Jolla, CA). Comparisons between PANX1 and R217H groups were made using two-way ANOVA analyses with Bonferroni corrected post hoc test. Comparisons between PANX1, R217H, and the co-expression groups were made using one-way ANOVA analyses with Bonferroni corrected post hoc test. A difference was considered significant at *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

In other experiments on the day of recordings, 293T cells expressing wild-type PANX1 or R217H were first pretreated with brefeldin A (BFA) (5 μ g/ml; Tocris Bioscience, UK)-containing medium for 6 h at 37 $^{\circ}$ C, a treatment regimen shown previously to prevent ion channel forward trafficking (36). Later, the culture medium was removed and replaced by a bath solution containing 5 μ g/ml BFA right before the recording.

Results

Clinical Presentation—We report on a 17-year-old female who initially presented for clinical genetics evaluation at 15 years of age with multiple concerns, including primary ovarian failure, intellectual disability, sensorineural hearing loss, and kyphosis. She had significant speech delay that led to the diagnosis of severe sensorineural hearing loss at 15 months of age, requiring bilateral cochlear implants at age 6. Despite the implants, her receptive and expressive language remains delayed with nearly absent speech and the ability to use only basic signs. She is unable to read or to count past 30. Formal psychological testing at 16 years of age revealed both verbal comprehension index and perceptual reasoning indices of 45 (less than 2nd percentile) on the Wechsler Intelligence Scale for Children (WISC-IV), commensurate with a child who is 6 years old, and academic skills at a first grade level, although it was noted that this test is not specifically designed for assessing students who are deaf or hard-of-hearing. She is mainly inde-

pendent with basic activities of daily living but does require some help with bathing, cooking, oral care, managing medications, and, occasionally, toileting. Primary ovarian failure was diagnosed via elevated follicle-stimulating hormone (108.6 mIU/ml), luteinizing hormone (23.7 mIU/ml), and low estradiol (< 5 pg/ml), in the setting of primary amenorrhea with normal pelvic ultrasound and secondary sexual characteristics. Further endocrine evaluation revealed that she had a delayed bone age, hypothyroidism, and ultrasound evidence of fatty liver disease associated with mild transaminitis of unclear etiology. Kyphosis was first noticed in early adolescence and improved, but did not resolve, with physical therapy. Other investigations included a normal echocardiogram, renal ultrasound, and ophthalmic exam, although her family reports some suspected issues with night vision. Physical examination at the age of 16 years revealed mild short stature with a height of 150.2 cm (3rd percentile) and weight of 46.8 kg (12th percentile). She was not facially dysmorphic and generally resembled her immediate family, although she did have kyphosis and bilateral 2–3 toe syndactyly. The family declined the inclusion of patient photos in this report.

Family History and Gene Sequencing—Family history was notable for one sister, 2 years her elder, with mild short stature (height 150 cm) and hypothyroidism, her mother with mild short stature (height 155 cm), and her father having hypothyroidism and fatty liver disease (Fig. 1). The proband's parents are known to be first cousins, originally from Turkey. The family is not aware of any other consanguineous relationships further back in the pedigree. Genetic evaluation of the proband demonstrated a normal chromosome analysis. Chromosome microarray revealed no microdeletions or duplications but did show 139.4 Mb of homozygosity spread over seven chromosomes, consistent with the known consanguinity. Given her complex phenotype, the possibility of a new genetic syndrome

Disease-associated Loss-of-function Panx1 Variant

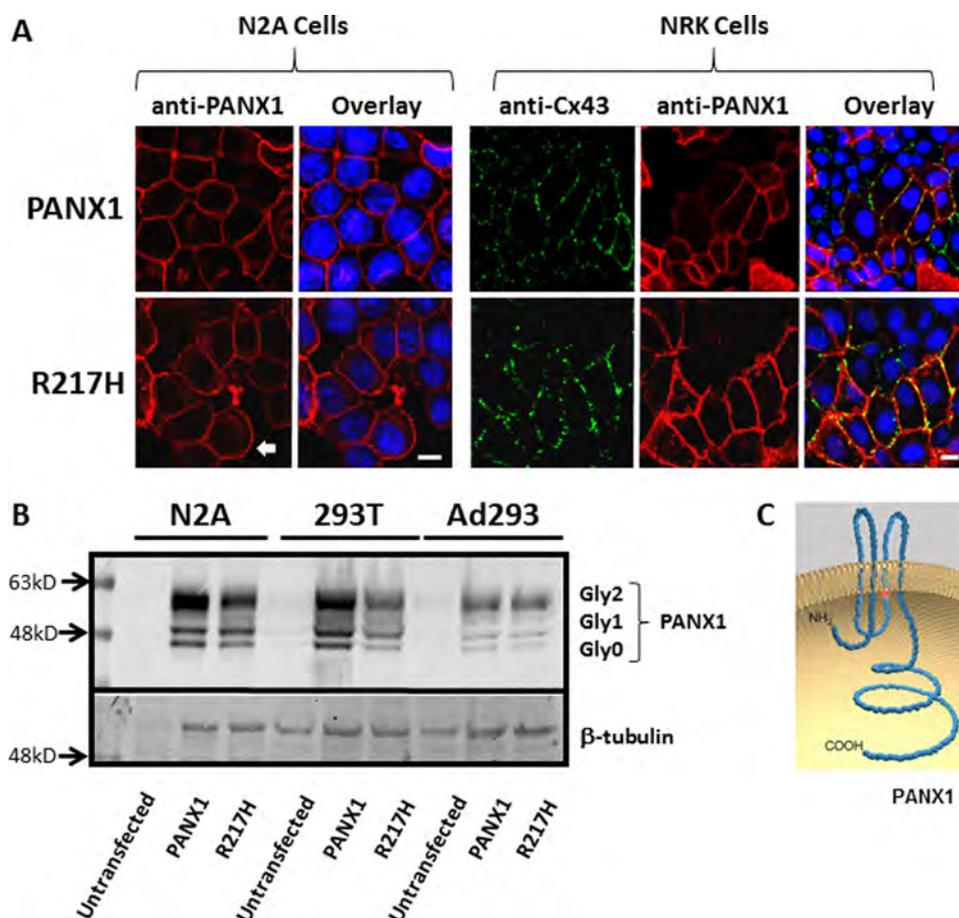


FIGURE 2. Disease-linked R217H mutant exhibits no characteristic difference in cellular localization or glycosylated isoforms in comparison with wild-type PANX1. *A*, N2A and NRK cells were engineered to express wild-type PANX1 or the R217H mutant and immunostained for the location of PANX1 (red) or the gap junction protein, Cx43 (green). Arrow indicates the R217H mutant at the cell surface with no apposing cells. Nuclei were stained with TO-PRO[®]-3 (blue). Bars, 10 μ m. *B*, untransfected N2A, 293T, and Ad293 cells or cells expressing wild-type PANX1 or the R217H mutant were immunoblotted for PANX1 or the gel loading control β -tubulin. Note that all glycosylated species of wild-type PANX1 and the R217H mutant (Gly0, Gly1, and Gly2) are expressed. Molecular mass markers are shown, and the experiments were repeated across a minimum of three cell lines to eliminate any cell type differences that might exist. *C*, model of PANX1 illustrating the approximate topological position of the R217H variant (red sphere).

was considered, and whole exome sequencing of the patient and her parents was performed. A homozygous missense variant in *PANX1*, c.650G→A (R217H) was identified, with each parent and her sister found to be heterozygous. These findings were confirmed by Sanger sequencing. This variant is predicted to be deleterious by PhyloP, PolyPhen2, SIFT, CADD, and MutationTaster and is present in the Exome Aggregation Consortium (ExAC) Browser with an allele frequency of <1/10,000 (6.65e-05). The *PANX1* gene is within one of her regions of homozygosity, on chromosome 11.

Expression and Localization of the R217H Mutant—Given that arginine at position 217 is conserved in vertebrates, and *in silico* analysis predicts the motif containing this residue is likely of structural and/or functional importance, we engineered and sequence-confirmed this mutation into the pUNO1-hPANX1 expression vector. Expression of the R217H mutant in mouse neuroblastoma (N2A) and NRK cells revealed that the mutant trafficked to the cell surface (as revealed by the location of the Cx43 gap junction protein) with seemingly identical efficacy as wild-type PANX1 (Fig. 2A). The R217H mutant was glycosylated to the well documented Gly1 and Gly2 species (4, 30, 37) of PANX1 when expressed in N2A, 293T, and Ad293-human

embryonic kidney cells (Fig. 2B). This was not unexpected given that the topological location of the mutation is within the intracellular loop of PANX1 (Fig. 2C), whereas glycosylation occurs on the second extracellular loop (4, 30). Thus, through the use of several cell lines to eliminate any potential cell type differences, these studies strongly suggest that the R217H mutant is trafficking-competent and appropriately glycosylated.

Functional Analysis of the R217H Mutant—Because PANX1 channels are large-pore channels (7, 9, 32) suitable for the uptake of small fluorescent dyes (38, 39), we assessed whether channels formed from the R217H mutant were functionally compromised. Ad293 cells expressing wild-type PANX1 or the R217H mutant (Fig. 3A) were assessed for ethidium bromide (EtBr) dye uptake (9) after channels were induced open in high potassium medium (Fig. 3B) (7, 34, 40). Quantitative assessment of EtBr uptake revealed that cells expressing the R217H mutant exhibited significantly reduced dye uptake over a time course of 5–40 min (Fig. 3B). Confirming the involvement of PANX1-based channels, the pannexin channel blocker probenecid (35) was found to block dye uptake in cells expressing PANX1. Moreover, in untransfected Ad293 cells, high potassium treatment did not stimulate dye uptake. These studies

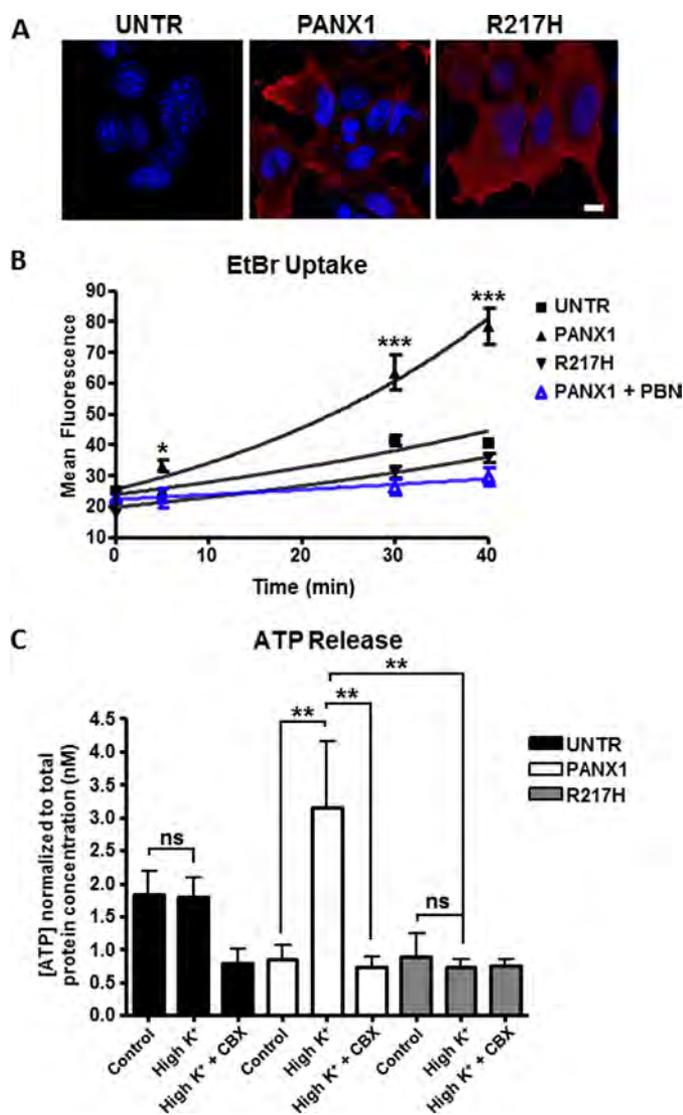


FIGURE 3. R217H mutant exhibits defective dye uptake and ATP release. A, untransfected Ad293 cells or cells expressing wild-type PANX1 or the R217H mutant were immunolabeled for PANX1 (red), and all cells were counterstained for TO-PRO³ (blue). Bar, 10 μ m. B, untransfected (UNTR) Ad293 cells or cells expressing wild-type PANX1 or R217H, or expressing wild-type PANX1 and treated with probenecid were subjected to ethidium bromide (EtBr) dye uptake over a period of 40 min. Mean fluorescent measurements revealed that wild-type PANX1-expressing cells were capable of dye uptake, whereas R217H-expressing, untreated cells, or cells expressing wild-type PANX1 and treated with probenecid exhibited significantly reduced dye uptake. *, $p < 0.05$; ***, $p < 0.001$, $n = 3$. C, untransfected (UNTR) N2A cells or N2A cells expressing wild-type PANX1 or the R217H mutant were assayed for ATP release upon treatment with a high potassium (K^+) medium containing or lacking the channel blocker CBX. ns, not significant. **, $p < 0.01$, $n = 3$.

strongly suggest that channels formed from the R217H mutant had impaired ability to uptake a small fluorescent dye.

As PANX1 channels are probably best known as ATP release channels (7, 12, 41), we next determined whether channels assembled from the R217H mutant had reduced ability to release ATP. To eliminate any confounding problems that might be due to ATP release from connexin-based hemichannels (42), we expressed the R217H mutant and wild-type PANX1 in connexin- and pannexin-deficient N2A cells (43) and assessed channel activity under physiological levels of calcium. As predicted, upon exposure to high potassium, there

was a surge of ATP release in N2A cells expressing wild-type PANX1 but not when cells expressed the R217H mutant (Fig. 3C). To confirm that cell surface channels were indeed responsible for ATP release, the channel blocker CBX (12, 44) eliminated the surge of ATP release induced by high potassium. Variability in ATP release in control cells with and without PANX1 or mutant may represent depletion of intracellular stores of ATP due to baseline activity of PANX1 channels. Nevertheless, these studies support the dye uptake data and strongly suggest that the R217H variant greatly attenuates channel function.

To determine whether the R217H variant would alter ionic current flow through the channel, we performed whole-cell voltage clamp recordings. For these experiments, we made use of 293T cells, devoid of endogenously expressed CBX-sensitive currents (45, 46), and we assessed pannexin function by recording membrane currents generated by voltage ramps from -100 to $+100$ mV. In 293T cells expressing R217H, ramp currents recorded at $+100$ mV were reduced by $\sim 50\%$ compared with wild-type PANX1 (Fig. 4, A and B). The reversal potentials of R217H (-71.77 ± 4.415 , $n = 6$) and PANX1 currents (-61.83 ± 3.508 , $n = 6$) were comparable, suggesting that reduced currents in R217H-expressing cells could not be attributed to a change in ionic permeability. Consistent with previous reports (7, 34, 40), PANX1-mediated currents were augmented by treatment with high potassium, most notably at negative holding potentials (Fig. 4C). High potassium augmented ramp currents in R217H-expressing cells were reduced by $\sim 50\%$ when compared with wild-type PANX1 (Fig. 4C). Ramp currents recorded from R217H- and PANX1-expressing cells, in control or high potassium-containing solutions, were suppressed by CBX (Fig. 4, A–C). The absence of comparable CBX-sensitive currents in mock-transfected 293T cells (Fig. 4B) confirms the specific involvement of pannexin channels under our recording conditions. To exclude the possibility that the loss-of-channel function was due to a dramatic reduction in the expression of the R217H mutant, PANX1 immunoblotting confirmed that both the mutant and wild-type PANX1 were expressed at equal levels in 293T cells prepared for electrophysiological measurements (Fig. 4D).

To further assess whether the R217H mutant was dominant to the functional channel properties of PANX1, we co-expressed equal quantities of plasmids encoding both wild-type and mutant PANX1. The ramp currents recorded at $+100$ mV were reduced by $\sim 50\%$ for the mutant compared with wild-type PANX1, although currents were not significantly reduced when both mutant and wild-type PANX1 were co-expressed (Fig. 5, A and B). Similar findings were observed when channel function was augmented by high potassium (Fig. 5C) suggesting that the R217H mutant was not dominantly inhibiting the function of PANX1. As before, carbenoxolone was effective in blocking PANX1 channel function as well as the residual mutant channel function (Fig. 5C).

Collectively, our functional assays monitoring dye, ATP, and ion flux support that R217H represents a loss-of-channel-function variant. Loss-of-function was evident in cells maintained under basal (*i.e.* reduced current amplitude in “normal” potassium) and high potassium-stimulated conditions. Loss-of-function with variants targeting channels that operate at the

Disease-associated Loss-of-function Panx1 Variant

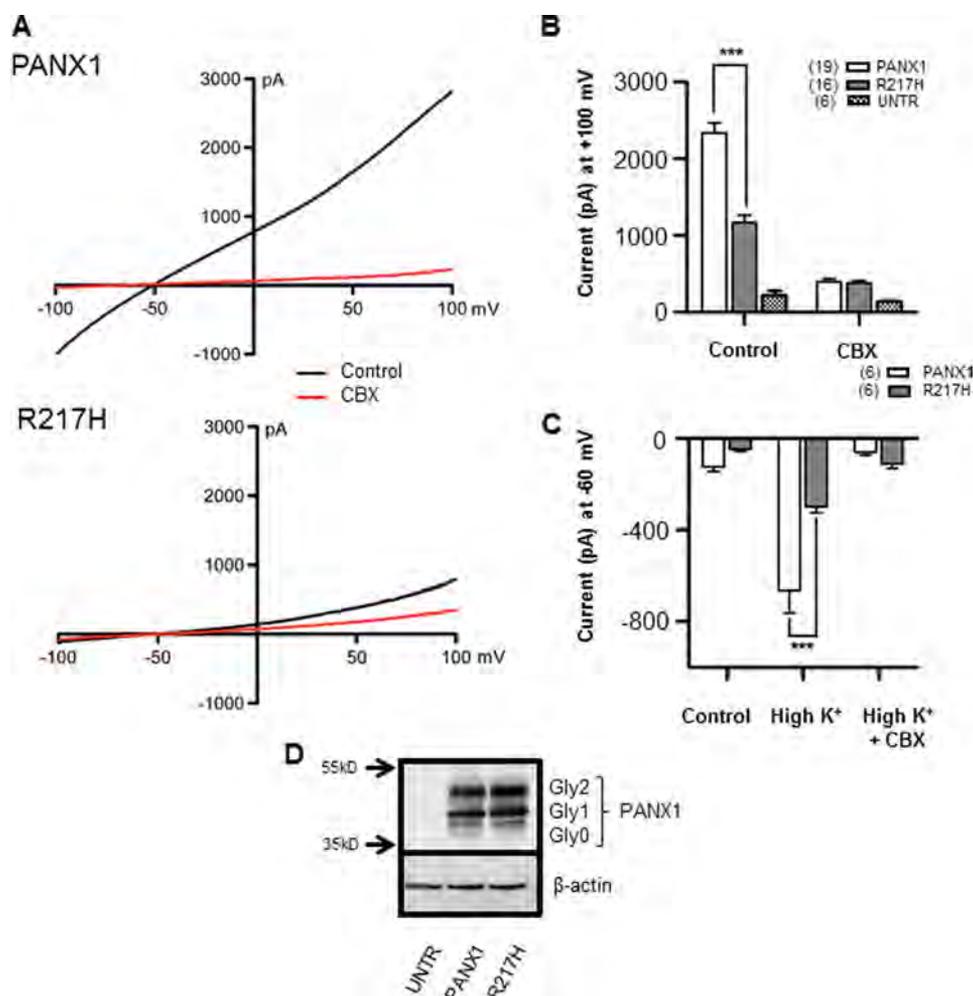


FIGURE 4. R217H mutant shows a dramatic reduction in channel function when expressed in HEK293T cells. *A*, representative current-voltage relationships recorded in the absence or presence of CBX in cells expressing wild-type PANX1 (*top*) and the R217H mutant (*bottom*). *B*, summary of ramp currents recorded at +100 mV from wild-type PANX1 and R217H mutant expressing or mock-transfected (UNTR) HEK293T cells. *C*, ramp currents at -60 mV recorded in HEK293T cells expressing wild-type PANX1 and the R217H mutant, before (*control*) and after treatment with high potassium (*High K⁺*) or high potassium and CBX. *D*, representative Western blot reveals comparable cellular expression of wild-type PANX1 and R217H channels in whole-cell lysates prepared from sister HEK293T cells used in electrophysiological recordings (position of molecular mass standards are shown). For all panels, the number of cells recorded is indicated in parentheses. *p* values were calculated in comparison with PANX1 groups using two-way ANOVA analyses with Bonferroni post-tests (*B* and *C*), *p* < 0.001.

cell surface can result from changes in channel function and/or defects in folding and cell surface expression. Evidence demonstrating that R217H expression, glycosylation, and cell surface trafficking is unaltered suggests that reduced current amplitudes observed in the absence of high potassium stimulation can be attributed to reduced channel function. However, with extended high potassium stimulation an additional mechanism could be recruited, namely reduced stimulated delivery of pannexin channels to the cell surface. Although high potassium is reported to allosterically augment pannexin function in a rapidly reversible manner (7, 47), a parallel increase in pannexin surface expression via increased forward trafficking could also contribute to augmented pannexin function. In this context, reduced R217H function could be attributed in part to a deficit in stimulated forward trafficking despite unaltered steady state surface expression. To address this possibility, we used cell surface biotinylation to first examine whether high potassium treatment altered the surface expression of R217H relative to that of PANX1. As might be expected, high potassium did not

increase the amount of wild-type PANX1 or the R217H mutant at the cell surface of Ad293 or N2A cells (Fig. 6A). Furthermore, we undertook a series of whole-cell recordings from 293T cells expressing PANX1 or R217H, which were pretreated with BFA, to block protein transport from the endoplasmic reticulum to the cell surface. Although current amplitude in cells expressing either R217H or wild-type PANX1 channels was reduced by BFA treatment, likely reflecting reduced surface expression, R217H functional deficits at rest and in the presence of high potassium were maintained (Fig. 6, *B* and *C*).

Discussion

The pannexin family of channel-forming glycoproteins has grown in importance given their documented roles in ischemia, stroke, overactive bladder, HIV infections, Crohn's disease, platelet aggregation, and over a half-dozen other diseases (13, 28, 45, 48, 49). In most of these cases, PANX1 was identified as being the key pannexin linked to the disease, but the causal mechanism of how PANX1 large-pore channels are associated

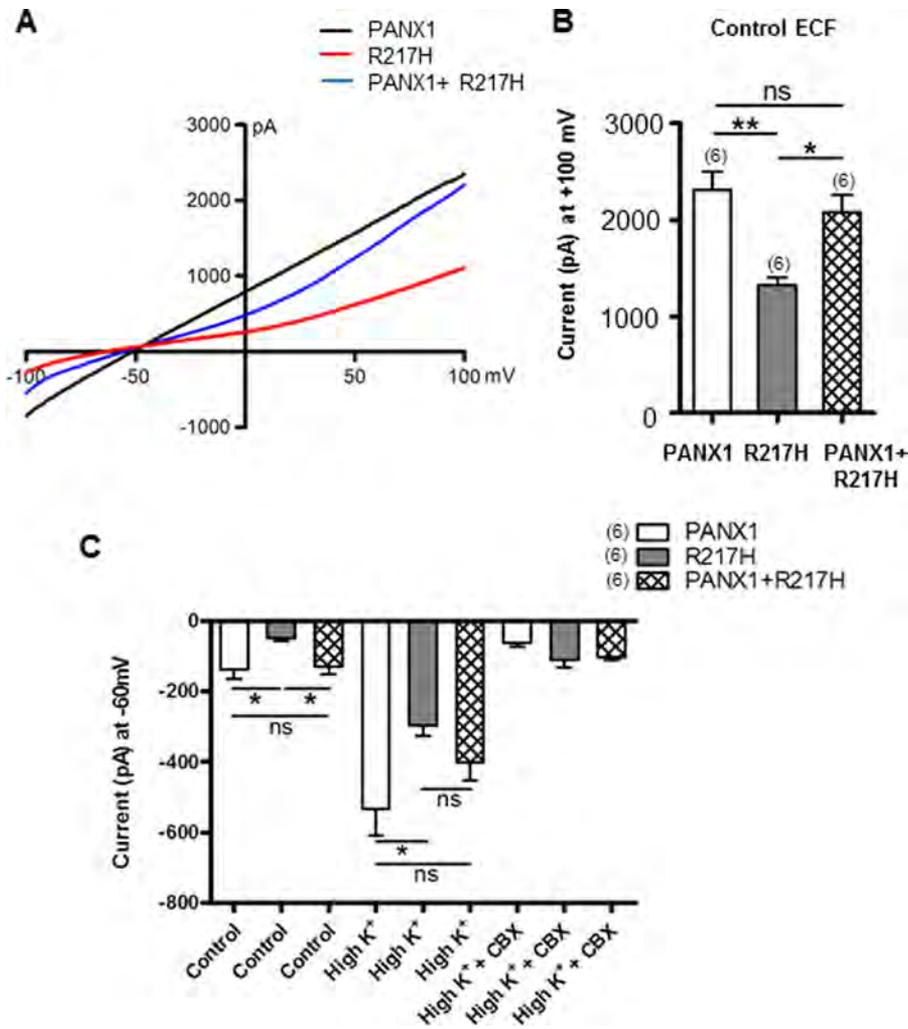


FIGURE 5. **R217H mutant is not dominant-negative to co-expressed PANX1.** *A*, representative current-voltage relationships recorded in HEK293T cells expressing PANX1, R217H, or both. *B*, summary of ramp currents recorded at +100 mV from wild-type PANX1, R217H, or PANX1/R217H co-expressing cells. *C*, ramp currents at -60 mV recorded in HEK293T cells expressing wild-type PANX1, the R217H mutant, or both, before (control) and after treatment with high potassium (*High K⁺*) or high potassium and CBX. For all panels, the number of cells recorded is indicated in parentheses. *, $p < 0.05$; **, $p < 0.01$. ns, not significant.

with disease susceptibility remains poorly understood (13). Given the breadth of diseases linked to PANX1 and its ubiquitous distribution, we predicted that loss- or gain-of-function pathogenic variants in the *PANX1* gene would likely be causally related to disease, and the disease may involve many organs. To the best of our knowledge, this report is the first and only germline variant identified in any of the three pannexin-encoding genes that has been associated with inherited human disabilities and disease. As anticipated, the proband has extensive and severe multi-organ involvement that has affected her cognition, hearing, skeleton, and reproductive organs. These clinical presentations are consistent with the current literature on Panx1 expression and function in animal studies (3). One of the organs with the highest expression of Panx1 is the brain (21, 50). Neurons and glial cells of the central nervous system have been shown to express Panx1 at early stages of development (51, 52), and therefore, the expression of a functionally impaired channel may have an effect on neuronal development and differentiation. Similarly, Panx1 expression in the cochlea, primarily in epithelial cells of the organ of Corti, and neurons of the spiral

ganglia (49, 53), would predict a potentially important function for these channels in hearing. Consistent with severe hearing loss found in the proband, hearing loss was recently reported in a *Panx1* conditional knock-out (KO) mouse from the cochlea (54, 55). At present, there is evidence of Panx1 expression in osteoblasts (56), but because Panx3 has been the most studied pannexin in bone and cartilage (57), our knowledge of the role of Panx1 in skeletal development remains limited. Likewise, although Panx1 has been reported in the ovaries and male reproductive organs, its functional importance is poorly understood (50, 58). However, based on the proband in this study, we suggest that PANX1 plays a role in both the skeleton and reproductive organs, although it likely is not accountable for the proband's short stature as her mother and sister have short stature as well. Also, in the absence of additional patients harboring homozygous *PANX1* gene mutations, we cannot definitively assign causal relationships to all the clinical presentations of the proband to *PANX1* as this awaits further confirmation.

The severity of developmental abnormalities is in striking contrast to the lack of overt phenotype noted in homozygous

Disease-associated Loss-of-function *Panx1* Variant

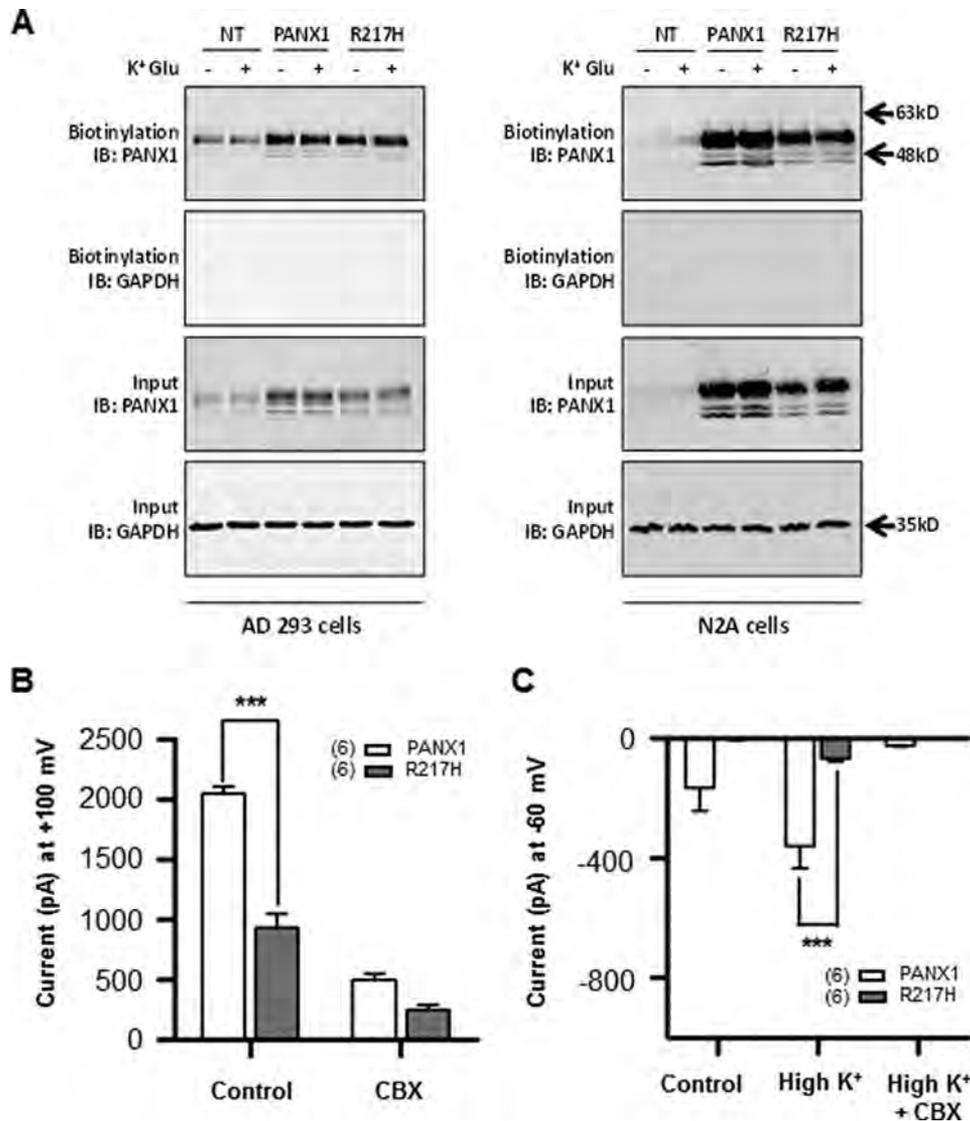


FIGURE 6. *A*, cell surface biotinylation of wild-type PANX1 and the R217H mutant in the presence and absence of high potassium. Cells were not transfected (NT) or transiently transfected with cDNA encoding either wild-type PANX1 or the R217H mutant. Forty eight hours after transfection, cells were untreated (–) or treated (+) with 140 mM K⁺ Glu in DMEM for 40 min. Cells were placed on ice, washed in ice-cold Hanks' balanced salt solution, and cell surface proteins biotinylated. Biotinylated proteins were precipitated using neutravidin-conjugated beads, subjected to SDS-PAGE, transferred to nitrocellulose, and immunoblotted (IB) for PANX1 or GAPDH. Note the abundance of cell surface biotinylated PANX1 and the R217H mutant, although the lack of biotinylated GAPDH indicates that only cell surface proteins were biotinylated. Position of molecular mass standards are shown. *B*, R217H mutant showed a dramatic reduction in channel function when expressed in HEK293T cells pretreated with BFA for 6 h while CBX effectively blocked channel function. *C*, after pre-treatment with BFA for 6 h, ramp currents at –60 mV were recorded in HEK293T cells expressing wild-type PANX1 or the R217H mutant, before (control) and after treatment with high potassium (High K⁺) or high potassium and CBX. *n* = 6 for each panel. ***, *p* < 0.001.

and heterozygous global *Panx1* KO mice (3, 59) suggesting that reduced expression by 50% or more does not phenocopy the human R217H variant. However, the lack of overt phenotypes in global *Panx1* KO mice may be due, at least in part, to compensation by other pannexins. For example, when *Panx1* was constitutively deleted in a *Panx1* KO mouse, there was an increase in Panx3 expression in dorsal skin (60). A similar up-regulation of Panx3 expression was observed in the wall of thoracodorsal arteries of *Panx1* KO mice compared with C57Bl/6 controls (61). In another case, both Panx1 and Panx2 needed to be deleted to observe neuroprotective effects in a mouse model of ischemia because it appeared that one pannexin compensated for the other (62). It should also be noted that the expression of the loss-of-function PANX1 variant that reduces chan-

nel function does not equate to the loss of PANX1 expression. For example, we have observed in the connexin channel field that genetically modified mice heterozygous for the *GJA1* (Cx43) gene do not exhibit disease, whereas heterozygous mice harboring the I130T mutant, reducing overall Cx43 function to 50% of controls, phenocopy the human disease known as oculodentodigital dysplasia (63–65).

Our multidimensional molecular analysis of the R217H mutant revealed that PANX1 channels formed from the mutant have an overall functional capacity of only 50% compared with controls. In this context, it is relevant to consider whether the developmental abnormalities are strictly attributed to loss-of-function or whether, alternatively, a gain-of-toxic-function (irrespective of reduced channel flux) could contribute. Of

note, expression of the R217H mutant in five distinct cell lines was well tolerated with no apparent detrimental effect on cell viability or morphology, suggesting that this latter scenario is unlikely. Thus, we propose that the R217H mutant disrupts some essential function of pannexin channels critical for development in a variety of tissue types. Because the proband's parents, as well as her sister, are heterozygous for R217H and exhibit no symptoms that can be linked to the variant, it would suggest that the disease is autosomal recessive in nature. Although the allele frequency of the R217H variant is not specifically known in the Turkish population, it has been reported in 3/8596 European alleles in the Exome Sequencing Project and 7/66,102 European alleles and 1/16,412 South Asian alleles in the Exome Aggregation Consortium. Consistent with the proband's parent and sister not exhibiting the clinical symptoms of the proband, we found that the mutant was not dominant to the function of *PANX1* when co-expressed in cultured cells.

Given that this patient is the first reported individual with biallelic *PANX1* gene variants, and that WES in any individual will reveal variants in numerous genes, we assessed the evidence that the *PANX1* variants are responsible for this individual's features. A review of the full WES findings reveals variants in three other genes known to be related to a human disease phenotype as follows: a single pathogenic variant in the autosomal recessive *RDX* gene associated with a form of nonsyndromic hearing loss, inherited from her unaffected mother; a missense variant in *FGFR3*, an autosomal dominant gene responsible for various forms of skeletal dysplasias, inherited from her unaffected father; and a single missense variant in the autosomal recessive *POR* gene, associated with disorders of steroidogenesis, inherited from her asymptomatic mother. None of the phenotypes associated with these genes are similar to the proband's presentation, and two of the three genes associated with autosomal recessive inheritance are not located in our patient's regions of loss of heterozygosity, and have only one variant found. In addition, despite the large runs of homozygosity seen on chromosome microarrays, WES only identified three other genes of interest with homozygous variants in addition to *PANX1* (R217H): *PHF12* (R765W), *MRPL49* (R88C), and *NRXN2* (L53F). *PHF12* encodes a PHD finger protein that has been proposed to be associated with the regulation of intraocular pressure, which is not associated with our proband's phenotype (66). *MRPL49* encodes the mitochondrial ribosomal protein L49, which is involved in protein synthesis (67). This gene is associated with glossopharyngeal neuralgia, a condition characterized by repeated episodes of severe pain in the tongue, throat, ear, and tonsils, conditions not experienced by the proband in this study. The *NRXN2* gene is expressed in the brain and encodes a synaptic organizing protein that helps to mediate the differentiation of certain synapses. At least one loss-of-function truncating mutation of this gene has been identified in the heterozygous state in a family with autism spectrum disorder, language delay, and a family history of schizophrenia, suggesting autosomal dominant inheritance (68). The autosomal dominant mode of inheritance proposed by this and other studies does not fit in this family as both parents are carriers of the missense change and are unaffected. Furthermore, the *NRXN2* proband did not present with the scope of organ involvement

noted in the proband of this study. Although this provides a compelling argument to rule out the involvement of *NRXN2*, protein structure/function studies also provide supportive evidence against a pivotal role for *NRXN2*. In brief, *NRXN2* encodes a type I transmembrane protein with six conserved interacting LNS domains (laminin/neurexin/sex hormone) (69–71). The LNS domains are key for interactions between neurexin and several postsynaptic binding partners that have been described (e.g. neuroligins, leucine-rich repeat transmembrane protein, and dystroglycan). Whereas LNS6 is the most important domain in terms of binding, the proband's variant is positioned in the first LNS region, with no known interacting partners. Therefore, LNS1 is predicted to not be a functionally relevant domain.

This analysis leaves us with *PANX1* as the best gene candidate for the proband's multi-organ involvement, although given the fact that other homozygous and heterozygous variants exist in several genes, we await substantiation of this case study by the identification of other patients that harbor *PANX1* variants. A *PANX1* gene search of the Online Mendelian Inheritance in Man (OMIM) database did not reveal any *PANX1* gene mutations linked to any known disease. In addition, OMIM was searched for conditions that include all of the major and most striking features of the proband's phenotype, namely intellectual disability, primary ovarian insufficiency, and sensorineural hearing loss, to make sure that the genes associated with these other conditions had good coverage on WES and did not contain any variants of uncertain significance. The main conditions that matched the proband's phenotype on this search included the Perrault syndromes and Woodhouse-Sataki syndrome. No suspicious variants were found in any of these genes associated with these conditions on WES. Overall, our findings fit well with the broad spectrum of *PANX1* expression in all of the proband's affected organs (13) and to the fact that a new conditional mouse model of *Panx1* ablation in the cochlea revealed severe hearing loss (54, 55). We suggest that *PANX1* be considered as a rare cause of intellectual disability, particularly when accompanied by other systemic dysfunction, especially sensorineural hearing loss and premature ovarian failure. Finally, as this is the first report of a disease-associated *PANX1* germline variant, it is possible that a broader contribution of *PANX1* to human disease has previously been overlooked. Accordingly, this study may raise awareness of *PANX1* as a candidate gene contributing to the genetic basis of disease.

Author Contributions—K. L. and A. S. identified the patient, obtained informed consent, and worked closely with the family to acquire all the clinical information. J. J. and K. L. L. evaluated all the gene sequencing. Q. S., J. K., J. L. E., and S. P. performed the expression, ATP release, biotinylation, and dye uptake studies. R. S. and M. F. J. performed all the electrophysiology studies. D. W. L. oversaw the project and wrote the manuscript with assistance from all authors, especially K. L., J. J., and M. F. J.

Acknowledgments—We thank the patient and family for their participation. We also thank Natalie Lavine for technical assistance and Jamie Simek for the artwork in Fig. 1C. Informed consent was obtained from the patient and her family to use her information for publication.

References

- Panchin, Y., Kelmanson, I., Matz, M., Lukyanov, K., Usman, N., and Lukyanov, S. (2000) A ubiquitous family of putative gap junction molecules. *Curr. Biol.* **10**, R473–474
- Sosinsky, G. E., Boassa, D., Dermietzel, R., Duffy, H. S., Laird, D. W., MacVicar, B., Naus, C. C., Penuela, S., Scemes, E., Spray, D. C., Thompson, R. J., Zhao, H. B., and Dahl, G. (2011) Pannexin channels are not gap junction hemichannels. *Channels* **5**, 193–197
- Bond, S. R., and Naus, C. C. (2014) The pannexins: past and present. *Front. Physiol.* **5**, 58
- Boassa, D., Ambrosi, C., Qiu, F., Dahl, G., Gaietta, G., and Sosinsky, G. (2007) Pannexin1 channels contain a glycosylation site that targets the hexamer to the plasma membrane. *J. Biol. Chem.* **282**, 31733–31743
- Penuela, S., Simek, J., and Thompson, R. J. (2014) Regulation of pannexin channels by post-translational modifications. *FEBS Lett.* **588**, 1411–1415
- Gehi, R., Shao, Q., and Laird, D. W. (2011) Pathways regulating the trafficking and turnover of pannexin1 protein and the role of the C-terminal domain. *J. Biol. Chem.* **286**, 27639–27653
- Wang, J., Ambrosi, C., Qiu, F., Jackson, D. G., Sosinsky, G., and Dahl, G. (2014) The membrane protein Pannexin1 forms two open-channel conformations depending on the mode of activation. *Sci. Signal.* **7**, ra69
- Ambrosi, C., Gassmann, O., Pranskevich, J. N., Boassa, D., Smock, A., Wang, J., Dahl, G., Steinem, C., and Sosinsky, G. E. (2010) Pannexin1 and Pannexin2 channels show quaternary similarities to connexons and different oligomerization numbers from each other. *J. Biol. Chem.* **285**, 24420–24431
- Pelegri, P., and Surprenant, A. (2006) Pannexin-1 mediates large pore formation and interleukin-1 β release by the ATP-gated P2X7 receptor. *EMBO J.* **25**, 5071–5082
- Pelegri, P., and Surprenant, A. (2007) Pannexin-1 couples to maitotoxin- and nigericin-induced interleukin-1 β release through a dye uptake-independent pathway. *J. Biol. Chem.* **282**, 2386–2394
- Dubyak, G. R. (2009) Both sides now: multiple interactions of ATP with pannexin-1 hemichannels. Focus on “a permeant regulating its permeation pore: inhibition of pannexin 1 channels by ATP”. *Am. J. Physiol. Cell Physiol.* **296**, C235–C241
- Chekeni, F. B., Elliott, M. R., Sandilos, J. K., Walk, S. F., Kinchen, J. M., Lazarowski, E. R., Armstrong, A. J., Penuela, S., Laird, D. W., Salvesen, G. S., Isakson, B. E., Bayliss, D. A., and Ravichandran, K. S. (2010) Pannexin 1 channels mediate ‘find-me’ signal release and membrane permeability during apoptosis. *Nature* **467**, 863–867
- Penuela, S., Harland, L., Simek, J., and Laird, D. W. (2014) Pannexin channels and their links to human disease. *Biochem. J.* **461**, 371–381
- Thompson, R. J. (2015) Pannexin channels and ischaemia. *J. Physiol.* **593**, 3463–3470
- Thompson, R. J., Jackson, M. F., Olah, M. E., Rungta, R. L., Hines, D. J., Beazely, M. A., MacDonald, J. F., and MacVicar, B. A. (2008) Activation of pannexin-1 hemichannels augments aberrant bursting in the hippocampus. *Science* **322**, 1555–1559
- Thompson, R. J., Zhou, N., and MacVicar, B. A. (2006) Ischemia opens neuronal gap junction hemichannels. *Science* **312**, 924–927
- Gulbransen, B. D., Bashashati, M., Hirota, S. A., Gui, X., Roberts, J. A., MacDonald, J. A., Muruve, D. A., McKay, D. M., Beck, P. L., Mawe, G. M., Thompson, R. J., and Sharkey, K. A. (2012) Activation of neuronal P2X7 receptor-pannexin-1 mediates death of enteric neurons during colitis. *Nat. Med.* **18**, 600–604
- S  ror, C., Melki, M. T., Subra, F., Raza, S. Q., Bras, M., Saïdi, H., Nardacci, R., Voisin, L., Paoletti, A., Law, F., Martins, I., Amendola, A., Abdul-Sater, A. A., Ciccocanti, F., Delelis, O., et al. (2011) Extracellular ATP acts on P2Y2 purinergic receptors to facilitate HIV-1 infection. *J. Exp. Med.* **208**, 1823–1834
- Penuela, S., Gyenis, L., Ablack, A., Churko, J. M., Berger, A. C., Litchfield, D. W., Lewis, J. D., and Laird, D. W. (2012) Loss of pannexin 1 attenuates melanoma progression by reversion to a melanocytic phenotype. *J. Biol. Chem.* **287**, 29184–29193
- Bao, B. A., Lai, C. P., Naus, C. C., and Morgan, J. R. (2012) Pannexin1 drives multicellular aggregate compaction via a signaling cascade that remodels the actin cytoskeleton. *J. Biol. Chem.* **287**, 8407–8416
- Lai, C. P., Bechberger, J. F., Thompson, R. J., MacVicar, B. A., Bruzzone, R., and Naus, C. C. (2007) Tumor-suppressive effects of pannexin 1 in C6 glioma cells. *Cancer Res.* **67**, 1545–1554
- Mylvaganam, S., Ramani, M., Krawczyk, M., and Carlen, P. L. (2014) Roles of gap junctions, connexins, and pannexins in epilepsy. *Front. Physiol.* **5**, 172
- Carlen, P. L. (2012) Curious and contradictory roles of glial connexins and pannexins in epilepsy. *Brain Res.* **1487**, 54–60
- Kurtenbach, S., Kurtenbach, S., and Zoidl, G. (2014) Emerging functions of pannexin 1 in the eye. *Front. Cell Neurosci.* **8**, 263
- Sarrouilhe, D., Dejean, C., and Mesnil, M. (2014) Involvement of gap junction channels in the pathophysiology of migraine with aura. *Front. Physiol.* **5**, 78
- Orellana, J. A., Froger, N., Ezan, P., Jiang, J. X., Bennett, M. V., Naus, C. C., Giaume, C., and S  ez, J. C. (2011) ATP and glutamate released via astroglial connexin 43 hemichannels mediate neuronal death through activation of pannexin 1 hemichannels. *J. Neurochem.* **118**, 826–840
- Schenk, U., Westendorf, A. M., Radaelli, E., Casati, A., Ferro, M., Fumagalli, M., Verderio, C., Buer, J., Scanziani, E., and Grassi, F. (2008) Purinergic control of T cell activation by ATP released through pannexin-1 hemichannels. *Sci. Signal.* **1**, ra6
- Molica, F., Morel, S., Meens, M. J., Denis, J. F., Bradfield, P. F., Penuela, S., Zufferey, A., Monyer, H., Imhof, B. A., Chanson, M., Laird, D. W., Fontana, P., and Kwak, B. R. (2015) Functional role of a polymorphism in the Pannexin1 gene in collagen-induced platelet aggregation. *Thromb. Haemost.* **114**, 325–336
- Tanaka, A. J., Cho, M. T., Millan, F., Juusola, J., Retterer, K., Joshi, C., Niyazov, D., Garnica, A., Gratz, E., Deardorff, M., Wilkins, A., Ortiz-Gonzalez, X., Mathews, K., Panzer, K., Brilstra, E., et al. (2015) Mutations in SPATA5 are associated with microcephaly, intellectual disability, seizures, and hearing loss. *Am. J. Hum. Genet.* **97**, 457–464
- Penuela, S., Bhalla, R., Gong, X. Q., Cowan, K. N., Celetti, S. J., Cowan, B. J., Bai, D., Shao, Q., and Laird, D. W. (2007) Pannexin 1 and pannexin 3 are glycoproteins that exhibit many distinct characteristics from the connexin family of gap junction proteins. *J. Cell Sci.* **120**, 3772–3783
- Benfenati, V., Caprini, M., Nicchia, G. P., Rossi, A., Dovizio, M., Cervetto, C., Nobile, M., and Ferroni, S. (2009) Carbenoxolone inhibits volume-regulated anion conductance in cultured rat cortical astroglia. *Channels* **5**, 323–336
- Bruzzone, R., Barbe, M. T., Jakob, N. J., and Monyer, H. (2005) Pharmacological properties of homomeric and heteromeric pannexin hemichannels expressed in *Xenopus* oocytes. *J. Neurochem.* **92**, 1033–1043
- Bao, L., Locovei, S., and Dahl, G. (2004) Pannexin membrane channels are mechanosensitive conduits for ATP. *FEBS Lett.* **572**, 65–68
- Silverman, W. R., de Rivero Vaccari, J. P., Locovei, S., Qiu, F., Carlsson, S. K., Scemes, E., Keane, R. W., and Dahl, G. (2009) The pannexin 1 channel activates the inflammasome in neurons and astrocytes. *J. Biol. Chem.* **284**, 18143–18151
- Silverman, W., Locovei, S., and Dahl, G. (2008) Probenecid, a gout remedy, inhibits pannexin 1 channels. *Am. J. Physiol. Cell Physiol.* **295**, C761–C767
- Zadeh, A. D., Cheng, Y., Xu, H., Wong, N., Wang, Z., Goonasekara, C., Steele, D. F., and Fedida, D. (2009) Kif5b is an essential forward trafficking motor for the Kv1.5 cardiac potassium channel. *J. Physiol.* **587**, 4565–4574
- Boassa, D., Qiu, F., Dahl, G., and Sosinsky, G. (2008) Trafficking dynamics of glycosylated pannexin 1 proteins. *Cell Commun. Adhes.* **15**, 119–132
- Hansen, D. B., Ye, Z. C., Calloe, K., Braunstein, T. H., Hofgaard, J. P., Ransom, B. R., Nielsen, M. S., and MacAulay, N. (2014) Activation, permeability, and inhibition of astrocytic and neuronal large pore (hemi-)channels. *J. Biol. Chem.* **289**, 26058–26073
- Penuela, S., Bhalla, R., Nag, K., and Laird, D. W. (2009) Glycosylation regulates pannexin intermixing and cellular localization. *Mol. Biol. Cell* **20**, 4313–4323
- Scemes, E., and Spray, D. C. (2012) Extracellular K⁺ and astrocyte signaling via connexin and pannexin channels. *Neurochem. Res.* **37**, 2310–2316
- Sandilos, J. K., Chiu, Y. H., Chekeni, F. B., Armstrong, A. J., Walk, S. F., Ravichandran, K. S., and Bayliss, D. A. (2012) Pannexin 1, an ATP release channel, is activated by caspase cleavage of its pore-associated C-terminal

- autoinhibitory region. *J. Biol. Chem.* **287**, 11303–11311
42. Sáez, J. C., and Leybaert, L. (2014) Hunting for connexin hemichannels. *FEBS Lett.* **588**, 1205–1211
 43. Bunse, S., Schmidt, M., Hoffmann, S., Engelhardt, K., Zoidl, G., and Dermietzel, R. (2011) Single cysteines in the extracellular and transmembrane regions modulate pannexin 1 channel function. *J. Membr. Biol.* **244**, 21–33
 44. Ma, W., Hui, H., Pelegrin, P., and Surprenant, A. (2009) Pharmacological characterization of pannexin-1 currents expressed in mammalian cells. *J. Pharmacol. Exp. Ther.* **328**, 409–418
 45. Romanov, R. A., Bystrova, M. F., Rogachevskaya, O. A., Sadovnikov, V. B., Shestopalov, V. I., and Kolesnikov, S. S. (2012) The ATP permeability of pannexin 1 channels in a heterologous system and in mammalian taste cells is dispensable. *J. Cell Sci.* **125**, 5514–5523
 46. Billaud, M., Chiu, Y. H., Lohman, A. W., Parpaite, T., Butcher, J. T., Mutchler, S. M., DeLalio, L. J., Artamonov, M. V., Sandilos, J. K., Best, A. K., Somlyo, A. V., Thompson, R. J., Le, T. H., Ravichandran, K. S., Bayliss, D. A., and Isakson, B. E. (2015) A molecular signature in the pannexin1 intracellular loop confers channel activation by the alpha1 adreno-receptor in smooth muscle cells. *Sci. Signal.* **8**, ra17
 47. Dahl, G. (2015) ATP release through pannexon channels. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 1672
 48. Prochnow, N., Abdulazim, A., Kurtenbach, S., Wildförster, V., Dvoriantchikova, G., Hanske, J., Petrasch-Parwez, E., Shestopalov, V. I., Dermietzel, R., Manahan-Vaughan, D., and Zoidl, G. (2012) Pannexin1 stabilizes synaptic plasticity and is needed for learning. *PLoS ONE* **7**, e51767
 49. Tang, W., Ahmad, S., Shestopalov, V. I., and Lin, X. (2008) Pannexins are new molecular candidates for assembling gap junctions in the cochlea. *Neuroreport* **19**, 1253–1257
 50. Baranova, A., Ivanov, D., Petrash, N., Pestova, A., Skoblov, M., Kelmanson, I., Shagin, D., Nazarenko, S., Geraymovych, E., Litvin, O., Tiunova, A., Born, T. L., Usman, N., Staroverov, D., Lukyanov, S., and Panchin, Y. (2004) The mammalian pannexin family is homologous to the invertebrate innexin gap junction proteins. *Genomics* **83**, 706–716
 51. Vogt, A., Hormuzdi, S. G., and Monyer, H. (2005) Pannexin1 and Pannexin2 expression in the developing and mature rat brain. *Brain Res. Mol. Brain Res.* **141**, 113–120
 52. Wicki-Stordeur, L. E., Dzugal, A. D., Swansburg, R. M., Suits, J. M., and Swayne, L. A. (2012) Pannexin 1 regulates postnatal neural stem and progenitor cell proliferation. *Neural Dev.* **7**, 11
 53. Wang, X. H., Streeter, M., Liu, Y. P., and Zhao, H. B. (2009) Identification and characterization of pannexin expression in the mammalian cochlea. *J. Comp. Neurol.* **512**, 336–346
 54. Chen, J., Zhu, Y., Liang, C., Chen, J., and Zhao, H. B. (2015) Pannexin1 channels dominate ATP release in the cochlea ensuring endocochlear potential and auditory receptor potential generation and hearing. *Sci. Rep.* **5**, 10762
 55. Zhao, H. B., Zhu, Y., Liang, C., and Chen, J. (2015) Pannexin 1 deficiency can induce hearing loss. *Biochem. Biophys. Res. Commun.* **463**, 143–147
 56. Penuela, S., Celetti, S. J., Bhalla, R., Shao, Q., and Laird, D. W. (2008) Diverse subcellular distribution profiles of pannexin 1 and pannexin 3. *Cell Commun. Adhes.* **15**, 133–142
 57. Moon, P. M., Penuela, S., Barr, K., Khan, S., Pin, C. L., Welch, I., Attur, M., Abramson, S. B., Laird, D. W., and Beier, F. (2015) Deletion of Panx3 prevents the development of surgically induced osteoarthritis. *J. Mol. Med.* **93**, 845–856
 58. Turmel, P., Dufresne, J., Hermo, L., Smith, C. E., Penuela, S., Laird, D. W., and Cyr, D. G. (2011) Characterization of pannexin1 and pannexin3 and their regulation by androgens in the male reproductive tract of the adult rat. *Mol. Reprod. Dev.* **78**, 124–138
 59. Penuela, S., Gehi, R., and Laird, D. W. (2013) The biochemistry and function of pannexin channels. *Biochim. Biophys. Acta* **1828**, 15–22
 60. Penuela, S., Kelly, J. J., Churko, J. M., Barr, K. J., Berger, A. C., and Laird, D. W. (2014) Panx1 regulates cellular properties of keratinocytes and dermal fibroblasts in skin development and wound healing. *J. Invest. Dermatol.* **134**, 2026–2035
 61. Lohman, A. W., and Isakson, B. E. (2014) Differentiating connexin hemichannels and pannexin channels in cellular ATP release. *FEBS Lett.* **588**, 1379–1388
 62. Bargiotas, P., Krenz, A., Hormuzdi, S. G., Ridder, D. A., Herb, A., Barakat, W., Penuela, S., von Engelhardt, J., Monyer, H., and Schwaninger, M. (2011) Pannexins in ischemia-induced neurodegeneration. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20772–20777
 63. Laird, D. W. (2014) Syndromic and non-syndromic disease-linked Cx43 mutations. *FEBS Lett.* **588**, 1339–1348
 64. Stewart, M. K., Gong, X. Q., Barr, K. J., Bai, D., Fishman, G. I., and Laird, D. W. (2013) The severity of mammary gland developmental defects is linked to the overall functional status of Cx43 as revealed by genetically modified mice. *Biochem. J.* **449**, 401–413
 65. Kalcheva, N., Qu, J., Sandeep, N., Garcia, L., Zhang, J., Wang, Z., Lampe, P. D., Suadicani, S. O., Spray, D. C., and Fishman, G. I. (2007) Gap junction remodeling and cardiac arrhythmogenesis in a murine model of oculodentodigital dysplasia. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20512–20516
 66. Chen, F., Klein, A. P., Klein, B. E., Lee, K. E., Truitt, B., Klein, R., Iyengar, S. K., and Duggal, P. (2015) Exome array analysis identifies CAV1/CAV2 as a susceptibility locus for intraocular pressure. *Invest. Ophthalmol. Vis. Sci.* **56**, 544–551
 67. Jung, S. J., Seo, Y., Lee, K. C., Lee, D., and Roe, J. H. (2015) Essential function of Aco2, a fusion protein of aconitase and mitochondrial ribosomal protein bL21, in mitochondrial translation in fission yeast. *FEBS Lett.* **589**, 822–828
 68. Gauthier, J., Siddiqui, T. J., Huashan, P., Yokomaku, D., Hamdan, F. F., Champagne, N., Lapointe, M., Spiegelman, D., Noreau, A., Lafrenière, R. G., Fathalli, F., Joobor, R., Krebs, M. O., DeLisi, L. E., Mottron, L., et al. (2011) Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. *Hum. Genet.* **130**, 563–573
 69. Tanaka, H., Nogi, T., Yasui, N., Iwasaki, K., and Takagi, J. (2011) Structural basis for variant-specific neuroligin-binding by alpha-neurexin. *PLoS ONE* **6**, e19411
 70. Sheckler, L. R., Henry, L., Sugita, S., Südhof, T. C., and Rudenko, G. (2006) Crystal structure of the second LNS/LG domain from neurexin 1α: Ca²⁺ binding and the effects of alternative splicing. *J. Biol. Chem.* **281**, 22896–22905
 71. Rudenko, G., Hohenester, E., and Muller, Y. A. (2001) LG/LNS domains: multiple functions—one business end? *Trends Biochem. Sci.* **26**, 363–368

Structure and Dynamics of Single-isoform Recombinant Neuronal Human Tubulin^{*S}♦

Received for publication, April 5, 2016, and in revised form, April 19, 2016
 Published, JBC Papers in Press, April 25, 2016, DOI 10.1074/jbc.C116.731133

Annappurna Vemu^{‡1}, Joseph Atherton^{§1}, Jeffrey O. Spector^{‡1},
 Agnieszka Szyk[‡], Carolyn A. Moores^{§2}, and Antonina Roll-Mecak^{‡¶3}

From the [‡]Cell Biology and Biophysics Unit, NINDS, and [¶]Biophysics Center, NHLBI, National Institutes of Health, Bethesda, Maryland 20892 and the [§]Institute of Structural and Molecular Biology, Department of Biological Sciences, Birkbeck College, University of London, London WC1E, United Kingdom

Microtubules are polymers that cycle stochastically between polymerization and depolymerization, *i.e.* they exhibit “dynamic instability.” This behavior is crucial for cell division, motility, and differentiation. Although studies in the last decade have made fundamental breakthroughs in our understanding of how cellular effectors modulate microtubule dynamics, analysis of the relationship between tubulin sequence, structure, and dynamics has been held back by a lack of dynamics measurements with and structural characterization of homogeneous isotypically pure engineered tubulin. Here, we report for the first time the cryo-EM structure and *in vitro* dynamics parameters of recombinant isotypically pure human tubulin. $\alpha 1A/\beta III$ is a purely neuronal tubulin isoform. The 4.2-Å structure of post-translationally unmodified human $\alpha 1A/\beta III$ microtubules shows overall similarity to that of heterogeneous brain microtubules, but it is distinguished by subtle differences at polymerization interfaces, which are hot spots for sequence divergence between tubulin isoforms. *In vitro* dynamics assays show that, like mosaic brain microtubules, recombinant homogeneous microtubules undergo dynamic instability, but they polymerize slower and have fewer catastrophes. Interestingly, we find that epitaxial growth of $\alpha 1A/\beta III$ microtubules from heterogeneous brain seeds is inefficient but can be fully rescued by incorporating as little as 5% of brain tubulin into the homogeneous $\alpha 1A/\beta III$ lattice. Our study establishes a system to examine the struc-

ture and dynamics of mammalian microtubules with well defined tubulin species and is a first and necessary step toward uncovering how tubulin genetic and chemical diversity is exploited to modulate intrinsic microtubule dynamics.

Microtubules cycle stochastically between periods of polymerization and depolymerization, *i.e.* they exhibit “dynamic instability” (1). This behavior is crucial in cell division, motility, and differentiation. Despite the discovery of dynamic instability more than 30 years ago (1) and fundamental breakthroughs in our understanding of microtubule dynamics modulation by cellular effectors (2, 3), analysis of the relationship between tubulin sequence, structure, and dynamics has been held back by a lack of structural and *in vitro* dynamics data with homogeneous isotypically pure engineered tubulin. Eukaryotes have multiple tubulin genes (humans have eight α - and eight β -tubulin isotypes) with tissue-specific distributions (4). Some microtubules are isotype mixtures, and others are formed from a predominant single isotype (5). Moreover, tubulin is subject to abundant and chemically diverse post-translational modifications that include acetylation, detyrosination, phosphorylation, glutamylation, glycylation, and amination (6, 7). Virtually all biochemical studies have used tubulin purified from mammalian brain tissue through multiple cycles of *in vitro* depolymerization and polymerization (8). Although tubulin is abundant in this source, the resulting material is highly heterogeneous, being composed of multiple tubulin isotypes bearing chemically diverse and abundant post-translational modifications (9–11). More than 22 different charge variants are repolymerized in random fashion for *in vitro* polymerization assays (12). Thus, microtubules used for *in vitro* dynamics assays have been mosaic, with random distributions of isoforms and post-translational modifications. Moreover, this purification procedure selects tubulin subpopulations that polymerize robustly while discarding those that do not. Efforts to reduce metazoan tubulin heterogeneity exploited differences in isoform compositions between various tissues or cell lines (*e.g.* avian erythrocytes (13) and HeLa cells (14)) or the use of isoform-specific antibodies for immunopurification (15). However, neither of these approaches yielded homogeneous single-isoform tubulin. Here, we report for the first time the expression and purification of recombinant isotypically pure unmodified human tubulin competent for *in vitro* dynamics assays and report its dynamic parameters as well as cryo-EM structure at 4.2 Å resolution. We find that isotypically pure unmodified $\alpha 1A/\beta III$ -tubulin exhibits subtle differences in dynamics when compared with heterogeneous brain tubulin, consistent with the small conformational rearrangements at tubulin polymerization interfaces revealed by our near-atomic resolution structure of $\alpha 1A/\beta III$ microtubules. Our study establishes a system to examine the structure and dynamics of mammalian microtubules with well defined α and β -tubulin species and is a first and necessary step toward exploring the biophysical correlates

* This work was supported by grants from the Medical Research Council, United Kingdom (to J. A. and C. A. M.), and by the Intramural Programs of the NINDS and the NHLBI, National Institutes of Health (to A. V., J. O. S., A. S., and A. R.-M.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

♦ This article was selected as a Paper of the Week.

✎ Author's Choice—Final version free via Creative Commons CC-BY license.

§ This article contains supplemental Table 1 and supplemental Movies 1–3. The atomic coordinates and structure factors (code 5JCO) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

The EMDB accession code for the GMPCPP $\alpha 1A/\beta III$ microtubule reconstruction is 8150.

¹ These authors contributed equally to this work.

² To whom correspondence may be addressed: Institute of Structural and Molecular Biology, Birkbeck College, London WC1E 7HX, United Kingdom. Tel.: 020-7631-6858; E-mail: c.moores@mail.cryst.bbk.ac.uk.

³ To whom correspondence may be addressed: Cell Biology and Biophysics Unit, Porter Neuroscience Research Center, National Institutes of Health, Bldg. 35, Rm. 3B-203, 35 Convent Dr., MSC 3700, Bethesda, MD 20892-3700. Tel.: 301-814-8119; E-mail: Antonina@mail.nih.gov.

between sequence, structure, and dynamics for mammalian microtubules.

Experimental Procedures

Expression and Purification of Human Recombinant Tubulin Constructs—Codon-optimized genes for human α 1A-tubulin (NP_001257328) with an internal His tag in the acetylation loop and a PreScission protease-cleavable C-terminally FLAG-tagged β III-tubulin (NM_006077) were custom-synthesized by Integrated DNA Technologies and cloned into a pFastBacTM-Dual vector as described previously (16, 17). The internal His tag in α -tubulin allowed production of an α -tubulin ending in its natural C-terminal tyrosine (17, 18). Without an affinity-based selection for α -tubulin, the final sample contains ~30% contamination with endogenous insect α -tubulin species that can be variable from construct to construct. The Bac-to-Bac System (Life Technologies, Inc.) was used to generate bacmids for baculovirus protein expression. High-Five or SF9 cells were grown to a density between 1.3 and 1.6×10^6 cells/ml and infected with viruses at the multiplicity of infection of 1. Cultures were grown in suspension for 48 h, and cell pellets were collected, washed in PBS, and flash-frozen. Cells were lysed by gentle sonication in $1 \times$ BRB80 buffer (80 mM PIPES, pH 6.9, 1 mM MgCl₂, 1 mM EGTA) with addition of: 0.5 mM ATP, 0.5 mM GTP, 1 mM PMSF, and 25 units/ μ l benzamide nuclease. The lysate was supplemented with 500 mM KCl and cleared by centrifugation (15 min at $400,000 \times g$). The crude supernatant (supplemented with 25 mM imidazole, pH 8.0) was loaded on a nickel-nitrilotriacetic acid column (Qiagen) equilibrated with high salt buffer (BRB80, 500 mM KCl, 25 mM imidazole). His-tagged tubulin was eluted with 250 mM imidazole in BRB80 buffer. The eluate was further purified on anti-FLAG G1 affinity resin (Gen Script). FLAG-tagged tubulin was eluted by incubation with FLAG peptide (GenScript) at 0.25 g/liter concentration followed by removal of the tag by PreScission protease. A final purification step was performed on a Resource Q anion exchange column (GE Healthcare) with a linear gradient from 100 mM to 1 M KCl in BRB80 buffer. Peak fractions were pooled and buffer-exchanged on a PD10 desalting column (GE Healthcare) equilibrated with BRB80, 20 μ M GTP. Small aliquots of tubulin were frozen in liquid nitrogen and stored at -80°C until use. The purified tubulin was subjected to ESI-TOF LC-MS analysis and detected no endogenous tubulin or post-translational modifications (Fig. 1A). The sensitivity of our mass spectrometric analyses is high enough to detect as little as 1% contaminating post-translationally modified tubulin species (17). The final yield is ~1 mg of >99% recombinant isotypically pure $\alpha\beta$ -tubulin per liter of SF9 cells.

Cryo-EM Sample Preparation and Data Collection—Recombinant human α 1A/ β III-tubulin was polymerized at a final concentration of 2.5 mg/ml in BRB80 buffer (80 mM PIPES, 2 mM MgCl₂, 1 mM EGTA, 1 mM DTT) with 1 mM GMPCPP⁴ or 2 mM GTP at 37°C for 1 h. GMPCPP-bound microtubules were double-cycled by depolymerizing on ice and then repolymerized at

37°C for 1 h with an additional 2 mM GMPCPP. Stabilized α 1A/ β III microtubules were diluted in BRB20 (20 mM PIPES, 2 mM MgCl₂, 1 mM EGTA, 1 mM DTT) to a final concentration of 2.5 μ M. Human kinesin-3 motor domain (Kif1A, residues 1–361) (19) was diluted to 20 μ M in BRB20 with 2 mM AMPNP. The microtubules and motor were applied sequentially to glow-discharged C-flatTM holey carbon grids (Prochips), and the sample was vitrified using a Vitrobot (FEI Co.). The presence of the kinesin motor domain allowed differentiation between α - and β -tubulin during processing. Images were collected with a DE20 direct electron detector (Direct Electron) on a FEI Tecnai G2 Polara operating at 300 kV with a calibrated magnification of $\times 52,117$ corresponding to a final sampling of 1.22 $\text{\AA}/\text{pixel}$. A total electron dose of $\sim 50 e^-/\text{\AA}^2$ over a 1.5-s exposure and a frame rate of 15 frames/s was used, giving a total of 23 frames at ~ 2.2 electrons/frame. Dynamic microtubules grown from GMPCPP seeds were polymerized at 2 mg/ml for 30 min, kept at 37°C throughout, and vitrified as above. Images were collected on a FEI Tecnai T12 operating at 120 kV using a 4096×4096 -pixel CCD camera (Gatan Inc.).

Data Processing for Three-dimensional Reconstruction—Individual $\sim 2.2 e^-/\text{\AA}^2$ frames were globally aligned using IMOD scripts (20) then locally aligned using the Optical Flow approach (21) implemented in Xmipp (22). The full dose of $\sim 50 e^-/\text{\AA}^2$ was used for particle picking and CTF determination in CTFFind3 (23), whereas $\sim 25 e^-/\text{\AA}^2$ was used in particle processing to center particles and determine their Euler angles. Euler angles and shifts determined using $\sim 25 e^-/\text{\AA}^2$ dose were used to generate reconstructions from either the first ~ 25 or $\sim 12 e^-/\text{\AA}^2$ of the exposure. Kinesin-3 microtubules were manually boxed in Eman Boxer (24), serving as input for a set of custom-designed semi-automated single-particle processing scripts utilizing Spider and FREALIGN as described previously (25) with minor modifications. 10,164 particles or 142,296 asymmetric units were used in the final reconstruction, which was assessed for overfitting using a high resolution noise-substitution test (26). Using local resolution estimates determined with the blocres program in Bsoft, the reconstruction was sharpened with a B factor of -180 up to a resolution of 5.5 or 4 \AA for visualization of kinesin or tubulin densities, respectively. The overall resolution of the reconstruction is 4.2 \AA (FSC_{true}, 0.143 criteria) (26) encompassing a resolution range of ~ 3.5 – 5.5\AA . The best regions of the reconstruction are within the tubulin portion of the complex (Figs. 1B and 2) from which we built an α 1A/ β III microtubule model. The quality of our reconstruction was sufficient to confirm that GMPCPP was found in the E-site (Fig. 1C) and GTP in the N-site.

Model Building and Refinement—The polypeptide model of the unmodified α 1A/ β III-tubulin GMPCPP microtubule was built directly into density in Coot (27) using PDB 3JAT (28) as a starting model. The structure was refined under symmetry restraints in REFMAC version 5.8 (29). Secondary structure and reference restraints based on the high resolution tubulin crystal structure PDB 4DRX (30) were generated with ProSMART (31). Model building in Coot and refinement in REFMAC were repeated iteratively until the quality of the model and fit were optimized (supplemental Table 1).

⁴The abbreviations used are: GMPCPP, guanylyl-(α,β)-methylene-diphosphate; AMPNP, 5'-adenylyl- β,γ -imidodiphosphate; PDB, Protein Data Bank.

In Vitro Microtubule Dynamics Assays—GMPCPP stabilized seeds were prepared as described (32). The GMPCPP seeds were immobilized in flow chambers using neutravidin as described previously (33). The final imaging buffer contained $1\times$ BRB80 supplemented with 1 mM GTP, 100 mM KCl, 1% pluronic F-127, and oxygen scavengers prepared as described (34). An objective heater (Bioprotechs) was used to warm the chamber to 30 °C. All chambers were sealed and allowed to equilibrate on the microscope stage for 5 min prior to imaging. Dark field images were acquired every 5 s for 30 min. For depolymerization rate measurements, the frame rate used was 40 frames/s. Imaging was performed on a Nikon Eclipse Ti-E equipped with a high NA dark field condenser, a $\times 100$ adjustable iris objective and a Hamamatsu Flash4.0 version 2 camera with 2×2 binning. The final pixel size was 108 nm. Dark field illumination was provided by a Lumencor SOLA SE-II light engine. A Nikon GIF filter was used to protect the seeds from excessive photodamage.

Dynamic Parameter Measurements—Using ImageJ, kymographs were generated from dark field images. Kymographs were traced by hand, and dynamic parameters were calculated. Growth and depolymerization rates were determined from the slope of the growing or depolymerizing microtubule in the kymographs. Catastrophe frequency was determined as the number of observed catastrophes divided by the total time spent in the growth phase. Extremely rare rescue events were observed under our experimental conditions and thus were not quantified. Mean microtubule lifetime was calculated as the average time a microtubule spent in the growth phase before a catastrophe. Mean microtubule length was calculated as the average length a microtubule reached before a catastrophe. The probability of nucleation was determined by determining the percentage of seeds that nucleated in 30 min in a field of view. Dynamicity was determined as defined in Toso *et al.* (35) as the sum of total growth and shortening lengths divided by total time.

Results

Near Atomic Resolution Structure of Single-isoform Human α 1A/ β III Microtubules—We selected for our study α 1A/ β III-tubulin. β III is a neuronal isoform that constitutes 25% of purified brain tubulin (10). It is expressed in non-neuronal tissues only during tumorigenesis (36, 37). It is also the most divergent of all β -tubulin isotypes. It is highly overexpressed in non-neuronal cells upon transformation and has been identified as a strong prognosticator of poor clinical outcomes (37). We expressed human α 1A/ β III-tubulin in insect cells (16). Through a new double-selection strategy using affinity tags on both α - and β -tubulin, we produced $>99\%$ homogeneous, modification-free, single-isotope human $\alpha\beta$ -tubulin, free of contamination from endogenous insect tubulins (Fig. 1A and see under “Experimental Procedures”) that is assembly-competent in the absence of stabilizing drugs like taxol and thus suitable for *in vitro* dynamics assays. Our tagging scheme generates an α -tubulin with a native C terminus and thus this recombinant tubulin is suitable for the investigation of the effects of the tubulin detyrosination/tyrosination cycle on intrinsic microtubule dynamics and those mediated by the modification-dependent recruitment of cellular effectors (38, 39).

To gain insight into the assembly properties of α 1A/ β III recombinant tubulin, we determined the structure of α 1A/ β III microtubules in complex with the GTP analog GMPCPP at near-atomic resolution using cryo-electron microscopy and single-particle image reconstruction (Figs. 1B and 2) (25). There is a resolution gradient in the reconstruction, with the best resolution (~ 3.5 Å) within the body of the microtubule (encompassing a resolution range of ~ 3.5 – 4.5 Å, Fig. 2A). The resolution range of the kinesin motor domain, used to facilitate reconstruction, is ~ 4.5 – 5.5 Å. Overall, the reconstruction has a resolution of 4.2 Å (Fourier shell correlation, 0.143 criterion (26), encompassing a resolution range of ~ 3.5 – 5.5 Å) (Fig. 2, B and C). The reconstruction shows clearly resolved β -sheets and α -helical pitch (Fig. 2, D–F). The majority (92%) of human α 1A/ β III GMPCPP microtubules have 14 protofilaments, similar to brain GMPCPP microtubules (40). The tubulin monomer consists of a well folded globular core and highly negatively charged and flexible C-terminal tails (41). The C-terminal tails are the locus of the greatest chemical heterogeneity in tubulin. They appear disordered in all microtubule structures to date either because (i) they have no unique well defined conformation or (ii) defined conformations unique to particular isoforms or post-translationally modified forms are lost during the iterative averaging used in EM reconstructions due to the high heterogeneity of these tails in brain tubulin samples. Despite the chemical homogeneity of our sample, there is no density attributable to them, indicating that they are intrinsically disordered unless engaged by an effector as seen for the tubulin tyrosine ligase like 7 glutamylase or the NDC80 complex (42–44).

Consistent with the high sequence conservation of the tubulin body, our structure is similar to that of heterogeneous mosaic mammalian brain GMPCPP microtubules, and the overall conformation of the tubulin dimers in our reconstruction is consistent with a GTP-like extended conformation (Fig. 1C) (28). The backbone root mean square deviation of our tubulin dimer model overlaid on that of the recently published structure of mammalian heterogeneous brain GMPCPP 14 protofilament microtubules is <2 Å. A difference in the tubulin repeat distance is observed between α 1A/ β III and brain microtubules as follows: 82.7 ± 0.2 versus 83.1 ± 0.0 Å measured from the EM reconstruction (*i.e.* model-independent); 82.6 versus 83.2 Å measured by comparing models, for α 1A/ β III and brain microtubules, respectively (28, 45). However, the tubulin repeat distance for the recombinant α 1A/ β III microtubules (~ 82.7 Å) is roughly comparable with the repeat distance for heterogeneous brain GMPCPP microtubules (~ 83 Å), which is more extended than that of the GDP state (~ 81.5 Å) (28, 45). Nevertheless, we find two subtle differences that have the potential to impact polymerization dynamics. First, the loop connecting helices H1 and H1' in β -tubulin shifts ~ 3 Å away from the H1'-S2 loop, which makes lateral contacts with the M-loop (microtubule loop) of the neighboring dimer (Fig. 1, D and E). The M-loop is a sequence element crucial to lateral contacts between adjacent protofilaments. Strikingly, the H1'-S2, H2-S3, and M-loops are a hot spot of sequence variation across β -tubulin isoforms (Fig. 1F), consistent with the structural plasticity we observe at this interface. Second, when one α -protomer each of brain GMPCPP and recombinant α 1A/ β III

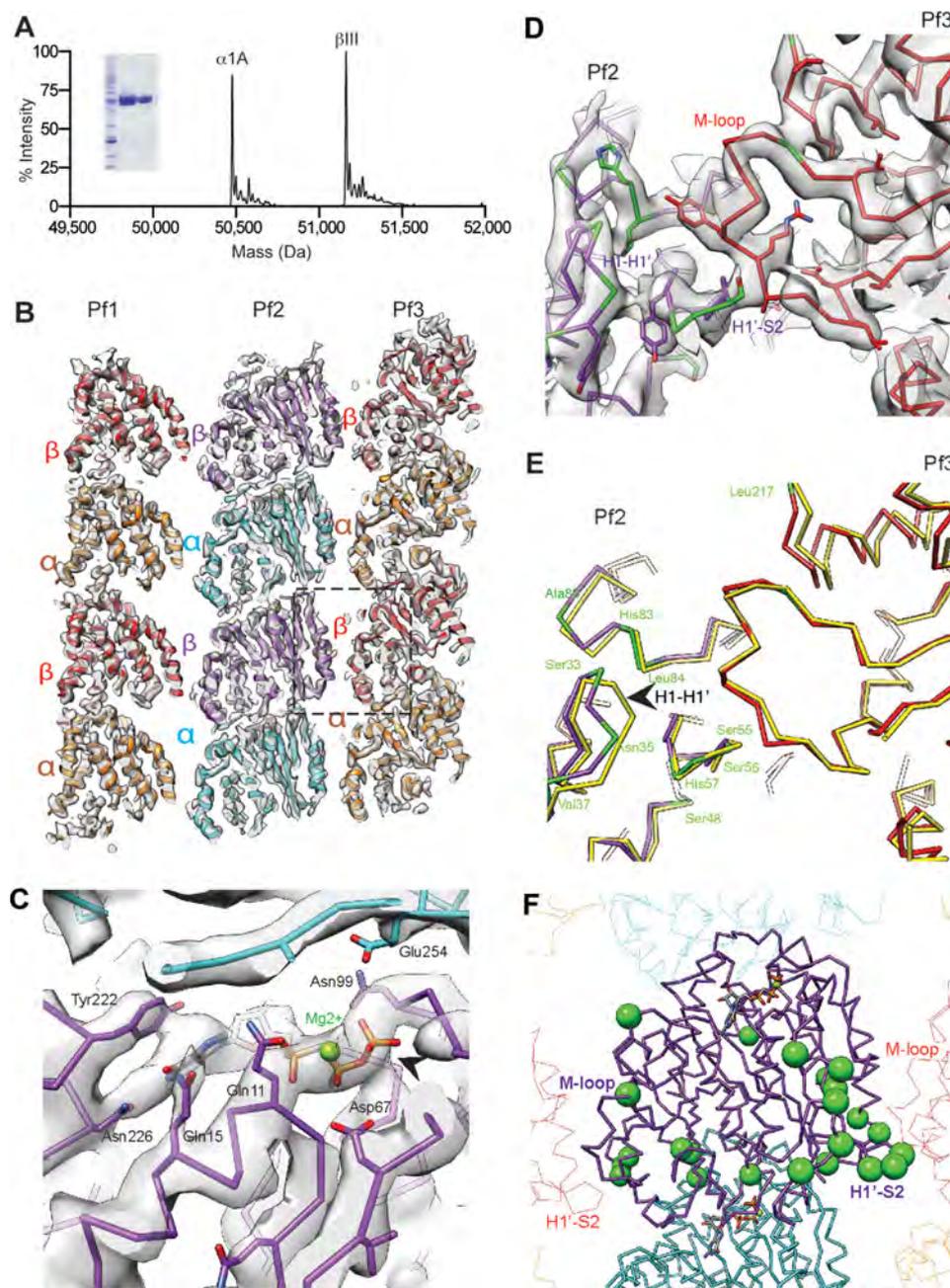


FIGURE 1. Structure of unmodified single-isoform human α 1A/ β III microtubules. *A*, mass spectra and SDS-polyacrylamide gel (*inset*) of recombinant human α 1A/ β III-tubulin purified to >99% homogeneity. The experimentally determined masses for α 1A- and β III-tubulin were 50,477.8 and 51,163.6 Da, respectively. The theoretical masses for α 1A- and β III-tubulin are 50,476.8 and 51,162.4 Da, respectively. *B*, cryo-EM map (4.2 Å resolution, 2.8 σ contour) and model of GMPCPP recombinant human α 1A/ β III microtubules viewed from the lumen (three protofilaments shown). A central protofilament (Pf2) makes lateral contacts with adjacent protofilaments (Pf1 and Pf3); α -tubulin, orange; β -tubulin, red (Pf1, Pf3); α -tubulin, cyan; β -tubulin, purple (Pf2). *C*, E-site in β III-tubulin shows clear density for GMPCPP and its three phosphate groups. *D*, model and map of the β III-tubulin lateral interface (boxed and colored as in *B*). β III-specific residues are in green. *E*, superposition of the α 1A/ β III (colored as in *B*) and brain (PDB, 3JAT; atomistic models of brain microtubules use the β III isotype sequence because it constitutes ~50% of these preparations (28, 44); yellow) microtubule structures; residues specific to β III are in green. *F*, β III sequence variability concentrates at the lateral interface. Green spheres denote residues that are different between the β III and β II isotypes, the most abundant tubulin isoforms in brain tubulin preparations (10).

GMPCPP microtubule protofilaments are superimposed, a clear displacement of successive recombinant α 1A/ β III dimers becomes apparent (Fig. 3*A*). This propagates from the exchangeable GTP-site (E-site) and β III-tubulin longitudinal interface and results in a progressive stagger that increases with each dimer along the protofilament, such that the first neighboring dimer is offset by 1.7 Å (all C α root mean square deviation), the second by 3.4 Å, and so on. Together, these rel-

atively subtle structural differences could contribute to differences in dynamic properties. Interestingly, we find that at 6 μ M α 1A/ β III-tubulin, 92% of α 1A/ β III GMPCPP seeds nucleate microtubules but only 33% brain seeds nucleate α 1A/ β III microtubules (Fig. 3*B*), suggestive of lattice mismatch effects between the brain microtubule seed and the lattice parameters of the growing α 1A/ β III microtubule. This is consistent with the subtle structural differences between α 1A/ β III and

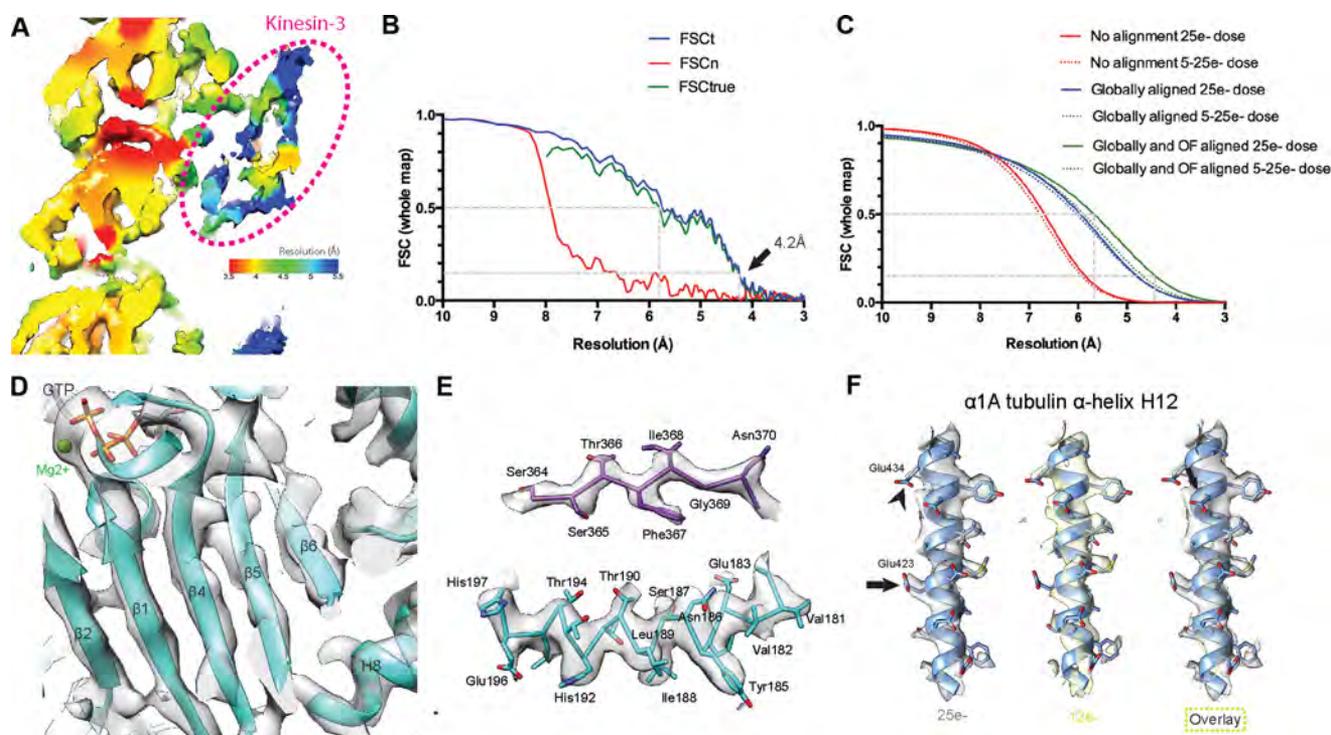


FIGURE 2. Data processing, map quality, and resolution determination for cryo-EM reconstruction of recombinant human α 1A/ β III microtubules. *A*, local resolution estimates calculated using the Bsoft program blocres (51) were used to color the unfiltered whole reconstruction density. Red density corresponds to 3.5 Å resolution, with a continuum of colors indicating the resolution gradient, ending with blue at 5.5 Å resolution. Tubulin is at a higher resolution, ranging from ~3.5 Å in central regions to ~4.5 Å in a more flexible peripheral surface-exposed region. Although used for the initial alignment, kinesin-3 is less ordered (resolution of ~5.5 Å) and is excluded from display items. *B*, Fourier shell correlation (FSC) curves. The gold standard noise-substitution test (26) on the whole microtubule + kinesin-3 map indicates no overfitting at high resolution and an overall resolution of 4.2 Å (FSC_{true} at 0.143 cutoff). *C*, R_{measure} (52) fitted curves give the same resolution estimate. Global alignment of whole movie frames improved resolution dramatically, whereas local alignment using an optical flow technique (21) yielded further improvements, especially for frames from early dosing of the data most susceptible to beam-induced motion. *D*, higher resolution (<4 Å) in the tubulin dimer core is supported by clear density for the backbone and most side chains (see also *E*). *E*, representative density for a β -strand in β -tubulin (top) and an α -helix in α -tubulin (bottom). *F*, reconstructions from the first 12 $e^{-}/\text{Å}^2$ dose data (yellow) showed improved density for some side chains when compared with the 25 $e^{-}/\text{Å}^2$ dose data (gray), regardless of whether they were acidic. The highly negatively charged helix H12 of α -tubulin is shown. Arrowheads indicate acidic side chains that are notable for their different appearance in 12 and 25 $e^{-}/\text{Å}^2$ maps.

heterogeneous brain microtubules that we identified (Figs. 1, *D* and *E*, and 3A). Unexpectedly, robust growth off brain seeds at 6 μM α 1A/ β III could be rescued (from 33 to 91%) if as little as 5% brain tubulin was added (Fig. 3B). Thus, a small level of tubulin heterogeneity can alleviate the nucleation defect that arises from the potential mismatch between the lattices of the two microtubule types. Our finding has intriguing consequences for the nucleation *in vivo* of microtubules composed of mixtures of tubulin isoforms.

In Vitro Dynamics of Single-isoform α 1A/ β III-tubulin—To determine dynamic parameters of single-isoform α 1A/ β III-tubulin, we performed label-free *in vitro* dynamic assays using dark field microscopy (Fig. 4 and supplemental Movies 1 and 2) (46) so that our dynamic parameters are not confounded by effects arising from the addition of fluorescently labeled brain tubulin to the otherwise homogeneous microtubules. The α 1A/ β III microtubules have the typical end appearance observed for brain microtubules consisting of a mixture of short sheet-like and blunter structures (Fig. 4B) (47). To quantify their dynamics, we generated kymographs from time-lapse imaging of dynamic microtubule assays (Fig. 4C). The growth rates at the plus-end are 35% slower when compared with those of heterogeneous brain tubulin, whereas the minus-end growth rates are statistically indistinguishable. Consistent with this, the on-rate of α 1A/ β III-tubulin at the plus-end is 1.8 dimers $s^{-1} \mu\text{M}^{-1}$

compared with the 3.6 dimers $s^{-1} \mu\text{M}^{-1}$ for brain tubulin (our measurements for brain microtubules are similar to those reported in Ref. 48). Dark field imaging allows data collection at the high frame rates needed to determine microtubule depolymerization rates with high accuracy (“Experimental Procedures” and see supplemental Movie 3). These measurements revealed that α 1A/ β III microtubules depolymerize slower than brain microtubules ($30.5 \pm 1.3 \mu\text{m}/\text{min}$ versus $39.9 \pm 1.5 \mu\text{m}/\text{min}$; Fig. 4D). This suggests that microtubules with different chemical compositions (isoform or post-translational modifications) have the potential to generate different end depolymerization forces that could be harnessed to move cargo in the cell, such as chromosomes during cell division (49).

The catastrophe (the transition between growth and shrinkage) frequency of recombinant microtubules is slightly reduced by 20 and 44% at the plus- and minus-ends, respectively, when compared with heterogeneous brain tubulin (Fig. 4, *E* and *F*). Interestingly, although 46% of brain microtubule exhibit growth at their minus ends, fewer than 7% of recombinant microtubules display minus-end dynamics under our assay conditions. Early studies reported faster polymerization rates for α β III-tubulin (α denotes here an unknown mixture of α -tubulin isoforms) immunopurified from brain tubulin preparations than for brain tubulin (15). Those studies also found that α β III-tubulin immunopurified from brain tubulin preparations

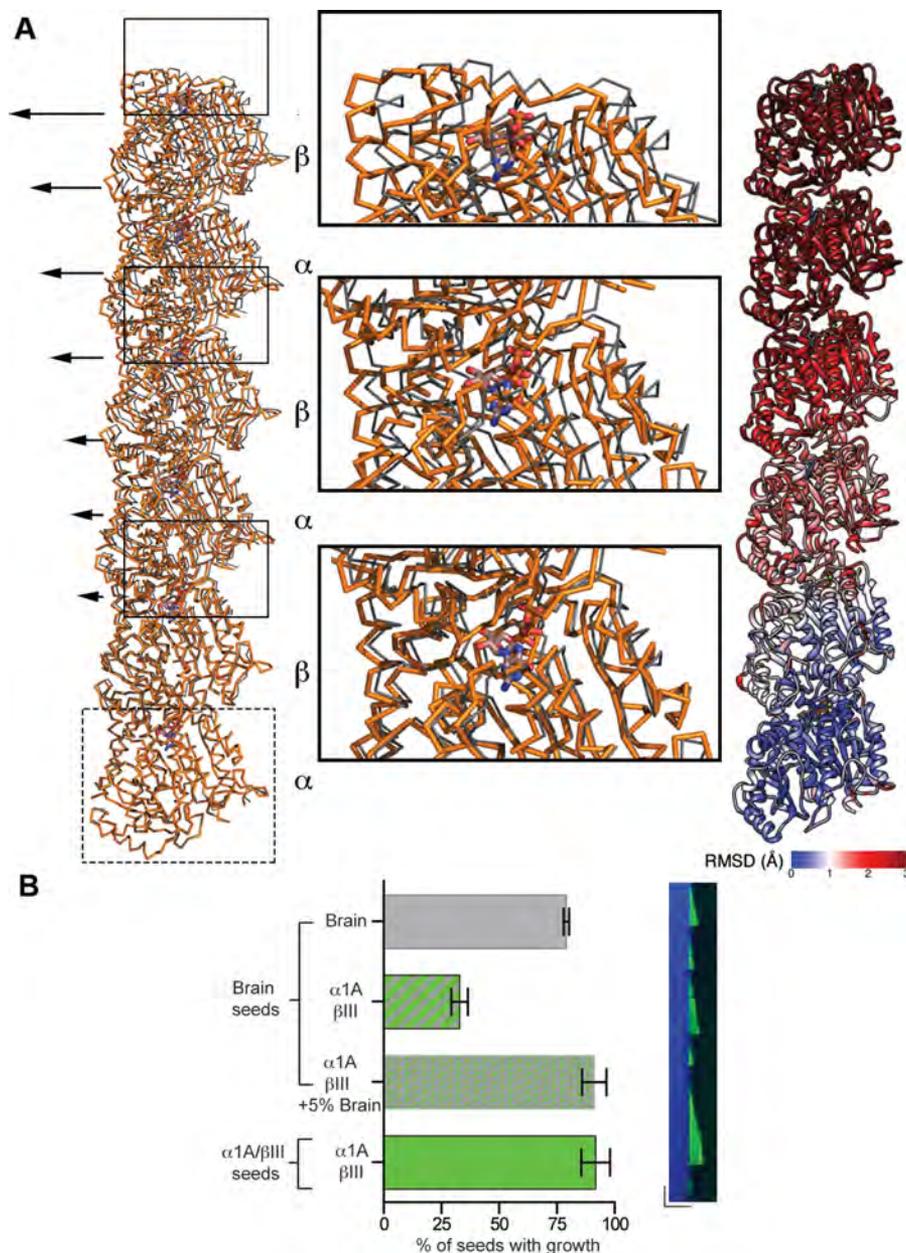


FIGURE 3. Comparison between α 1A/ β III and mosaic brain 14 protofilament microtubule structures. *A*, left panel, dimer displacement compared with the structure of mosaic brain microtubules PDB 3JAT (28) as viewed from the microtubule lumen. The boxed α 1A-tubulin protomer from the α 1A/ β III structure (orange $C\alpha$ trace) was superimposed on the α -tubulin protomer from the brain microtubule structure (gray $C\alpha$ trace). Arrows indicate the gradual increase in displacement of the α 1A/ β III heterodimers as one advances toward the plus-end of the protofilament. The GTP and GMPCPP in the N-site of α -tubulin and the E-site of β -tubulin are shown as ball-and-stick. Middle panel, zoomed in view of regions highlighted by boxes in the left panel showing details of the displacement between the dimers from the recombinant α 1A/ β III and brain microtubule structures; Right panel, three α 1A/ β III heterodimers within one protofilament colored according to main chain displacement from the brain microtubule structure. *B*, left panel, percentage of seeds that nucleate microtubules at $6 \mu\text{M}$ tubulin. Brain, α 1A/ β III, α 1A/ β III + 5% brain tubulin elongated from brain seeds and α 1A/ β III-tubulin elongated from α 1A/ β III seeds. More than 100 seeds across multiple chambers were counted for these measurements. Right panel, kymograph of microtubule growth for recombinant α 1A/ β III at $5.7 \mu\text{M}$ supplemented with 5% Hilyte 488 brain tubulin ($0.3 \mu\text{M}$) from brain GMPCPP seeds showing incorporation of the brain tubulin into the α 1A/ β III lattice. Horizontal and vertical scale bar, $5 \mu\text{m}$ and 2 min, respectively.

had higher dynamicity than brain tubulin, although our measurements with recombinant α 1A/ β III show lower dynamicity for this species than for brain microtubules ($1.31 \pm 0.05 \mu\text{m}/\text{min}$ versus $2.30 \pm 0.07 \mu\text{m}/\text{min}$ for α 1A/ β III and brain, respectively; “Experimental Procedures”). However, it is important to note that the tubulin used in these earlier studies had an unknown α -tubulin composition and a poorly defined mixture of diverse post-translational modifications, unlike our recombinant tubulin, which contains a single α - and β -tubulin isoform and is unmodified (Fig. 1A and “Experimental Proce-

dures”). It is unclear at this point whether the subtle differences in dynamics we observe between the recombinant α 1A/ β III microtubules and heterogeneous mosaic brain microtubules are due to isoform differences, purification method, and/or the abundant and diverse post-translational modifications found on brain microtubules. Future studies with recombinantly expressed isoforms and quantitatively defined post-translationally modified tubulin using the expression and purification system described here will shed light on their individual contributions to dynamic instability parameters.

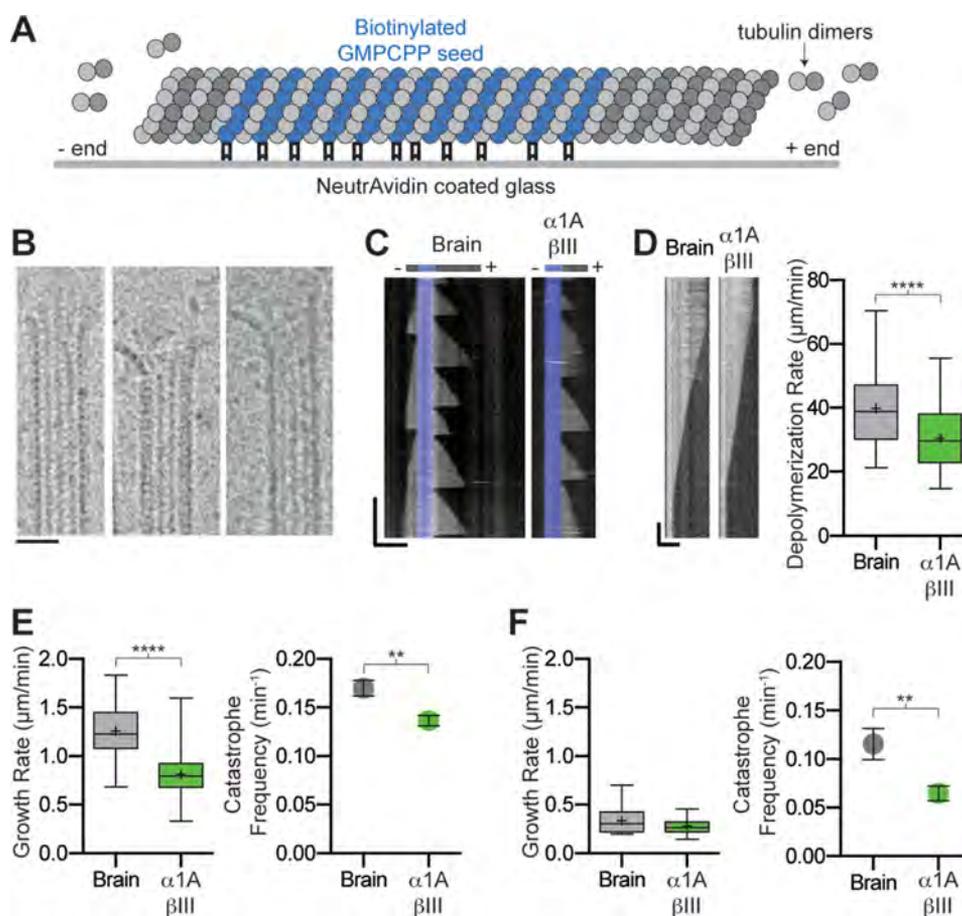


FIGURE 4. **Dynamic parameters of recombinant human α 1A/ β III microtubules.** *A*, schematic of assay design (see under “Experimental Procedures”). *B*, micrographs of representative dynamic α 1A/ β III microtubule ends. Scale bar, 20 nm. *C*, kymographs showing typical microtubule growth for brain and recombinant α 1A/ β III-tubulin at 9 μ m. Blue marks the GMPCPP seed. Horizontal and vertical scale bars, 5 μ m and 5 min, respectively. *D*, left panel, kymographs showing a typical depolymerization event for brain and α 1A/ β III microtubules. Horizontal and vertical scale bar, 5 μ m and 2 s, respectively. Right panel, Tukey plot showing plus-end depolymerization rates at 9 μ m tubulin; $n = 55$ and 58 events for brain and α 1A/ β III microtubules, respectively. *E*, plus-end dynamics of brain and α 1A/ β III-tubulin at 9 μ m tubulin. Left panel, box-whisker plot (whiskers indicate minimum and maximum) showing growth rates; $n = 255$ and 504 events for brain and α 1A/ β III-tubulin, respectively. Right panel, catastrophe frequencies; $n = 48$ and 167 microtubules for brain and α 1A/ β III-tubulin, respectively. *F*, minus-end dynamics of brain and α 1A/ β III-tubulin at 9 μ m tubulin. Left panel, box-whisker plot (whiskers indicate minimum and maximum) showing growth rates; $n = 32$ and 25 events for brain and α 1A/ β III-tubulin, respectively. Right panel, catastrophe frequencies; $n = 7$ and 16 microtubules for brain and α 1A/ β III-tubulin, respectively. Error bars represent S.E. ** and ****, p values < 0.01 and < 0.0001, respectively determined by unpaired t test.

Discussion

Using our dual-tag purification system for recombinant tubulin, we report for the first time the structure and *in vitro* dynamics parameters for isotopically pure human unmodified microtubules, an essential and important initial step in quantitatively establishing the correlates between sequence and dynamics for mammalian microtubules. The dual-tag selection system is necessary as a single-tag purification strategy results in significant levels of contamination with endogenous tubulin (~30% of insect α -tubulin if α -tubulin is not selected via an affinity tag). Thus, our tagging and purification strategy allows the characterization of both α - and β -tubulin engineered constructs. The majority of *in vitro* dynamics studies presently performed use heterogeneous mosaic brain microtubules with isoform composition and post-translational modifications different from those found *in vivo*, for example in an epithelial cell or the axonal or dendritic compartment of a neuron. A recent study revealed different activities of the *Saccharomyces cerevisiae* Stu2p on yeast microtubules compared with heterogeneous brain microtubules (50), indicating the importance

of examining the effects of regulators with the physiologically relevant tubulin substrate. Our study establishes a system to examine the dynamics of mammalian microtubules with well defined tubulin species and opens the way to study tubulin isoform-specific effects of microtubule-associated proteins and motors and to uncover the tubulin sequence elements critical for their recruitment and activation.

Author Contributions—A. R.-M. conceived the project. A. V. and J. O. S. performed and analyzed the dynamics assays. J. A. determined EM structure, and A. S. purified recombinant tubulin. All authors interpreted data. A. R.-M. wrote the manuscript with contributions from A. V., J. O. S., J. A., and C. A. M.

References

- Mitchison, T., and Kirschner, M. (1984) Dynamic instability of microtubule growth. *Nature* **312**, 237–242
- Bieling, P., Laan, L., Schek, H., Munteanu, E. L., Sandblad, L., Dogterom, M., Brunner, D., and Surrey, T. (2007) Reconstitution of a microtubule plus-end tracking system *in vitro*. *Nature* **450**, 1100–1105

3. Brouhard, G. J., Stear, J. H., Noetzel, T. L., Al-Bassam, J., Kinoshita, K., Harrison, S. C., Howard, J., and Hyman, A. A. (2008) XMAP215 is a processive microtubule polymerase. *Cell* **132**, 79–88
4. Leandro-García, L. J., Leskelä, S., Landa, I., Montero-Conde, C., López-Jiménez, E., Letón, R., Cascón, A., Robledo, M., and Rodríguez-Antona, C. (2010) Tumoral and tissue-specific expression of the major human β -tubulin isoforms. *Cytoskeleton* **67**, 214–223
5. Miller, K. E., and Joshi, H. C. (1996) Tubulin transport in neurons. *J. Cell Biol.* **133**, 1355–1366
6. Yu, I., Garnham, C. P., and Roll-Mecak, A. (2015) Writing and reading the tubulin code. *J. Biol. Chem.* **290**, 17163–17172
7. Verhey, K. J., and Gaertig, J. (2007) The tubulin code. *Cell Cycle* **6**, 2152–2160
8. Weisenberg, R. C. (1972) Microtubule formation *in vitro* in solutions containing low calcium concentrations. *Science* **177**, 1104–1105
9. Sullivan, K. F., and Cleveland, D. W. (1986) Identification of conserved isotype-defining variable region sequences for four vertebrate beta tubulin polypeptide classes. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4327–4331
10. Banerjee, A., Roach, M. C., Wall, K. A., Lopata, M. A., Cleveland, D. W., and Ludueña, R. F. (1988) A monoclonal antibody against the type II isotype of β -tubulin. Preparation of isotypically altered tubulin. *J. Biol. Chem.* **263**, 3029–3034
11. Garnham, C. P., and Roll-Mecak, A. (2012) The chemical complexity of cellular microtubules: tubulin post-translational modification enzymes and their roles in tuning microtubule functions. *Cytoskeleton* **69**, 442–463
12. Zambito, A. M., Knipping, L., and Wolff, J. (2002) Charge variants of tubulin, tubulin S, membrane-bound and palmitoylated tubulin from brain and pheochromocytoma cells. *Biochim. Biophys. Acta* **1601**, 200–207
13. Trinczek, B., Marx, A., Mandelkow, E. M., Murphy, D. B., and Mandelkow, E. (1993) Dynamics of microtubules from erythrocyte marginal bands. *Mol. Biol. Cell* **4**, 323–335
14. Newton, C. N., DeLuca, J. G., Himes, R. H., Miller, H. P., Jordan, M. A., and Wilson, L. (2002) Intrinsically slow dynamic instability of HeLa cell microtubules *in vitro*. *J. Biol. Chem.* **277**, 42456–42462
15. Panda, D., Miller, H. P., Banerjee, A., Ludueña, R. F., and Wilson, L. (1994) Microtubule dynamics *in vitro* are regulated by the tubulin isotype composition. *Proc. Natl. Acad. Sci.* **91**, 11358–11362
16. Minoura, I., Hachikubo, Y., Yamakita, Y., Takazaki, H., Ayukawa, R., Uchimura, S., and Muto, E. (2013) Overexpression, purification, and functional analysis of recombinant human tubulin dimer. *FEBS Lett.* **587**, 3450–3455
17. Valenstein, M. L., and Roll-Mecak, A. (2016) Graded control of microtubule severing by tubulin glutamylation. *Cell* **164**, 911–921
18. Sirajuddin, M., Rice, L. M., and Vale, R. D. (2014) Regulation of microtubule motors by tubulin isoforms and post-translational modifications. *Nat. Cell Biol.* **16**, 335–344
19. Atherton, J., Farabella, I., Yu, I. M., Rosenfeld, S. S., Houdusse, A., Topf, M., and Moores, C. A. (2014) Conserved mechanisms of microtubule-stimulated ADP release, ATP binding, and force generation in transport kinesins. *Elife* **3**, e03680
20. Kremer, J. R., Mastrorade, D. N., and McIntosh, J. R. (1996) Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76
21. Abrishami, V., Vargas, J., Li, X., Cheng, Y., Marabini, R., Sorzano, C. Ó., and Carazo, J. M. (2015) Alignment of direct detection device micrographs using a robust Optical Flow approach. *J. Struct. Biol.* **189**, 163–176
22. de la Rosa-Trevín, J. M., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J. M., and Sorzano, C. O. (2013) Xmipp 3.0: an improved software suite for image processing in electron microscopy. *J. Struct. Biol.* **184**, 321–328
23. Mindell, J. A., and Grigorieff, N. (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347
24. Ludtke, S. J., Baldwin, P. R., and Chiu, W. (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97
25. Sindelar, C. V., and Downing, K. H. (2007) The beginning of kinesin's force-generating cycle visualized at 9-Å resolution. *J. Cell Biol.* **177**, 377–385
26. Chen, S., McMullan, G., Faruqi, A. R., Murshudov, G. N., Short, J. M., Scheres, S. H., and Henderson, R. (2013) High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35
27. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
28. Zhang, R., Alushin, G. M., Brown, A., and Nogales, E. (2015) Mechanistic origin of microtubule dynamic instability and its modulation by EB proteins. *Cell* **162**, 849–859
29. Brown, A., Long, F., Nicholls, R. A., Toots, J., Emsley, P., and Murshudov, G. (2015) Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 136–153
30. Pecqueur, L., Duellberg, C., Dreier, B., Jiang, Q., Wang, C., Plückthun, A., Surrey, T., Gigant, B., and Knossow, M. (2012) A designed ankyrin repeat protein selected to bind to tubulin caps the microtubule plus end. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12011–12016
31. Nicholls, R. A., Long, F., and Murshudov, G. N. (2012) Low-resolution refinement tools in REFMAC5. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 404–417
32. Gell, C., Bormuth, V., Brouhard, G. J., Cohen, D. N., Diez, S., Friel, C. T., Helenius, J., Nitzsche, B., Petzold, H., Ribbe, J., Schäfer, E., Stear, J. H., Trushko, A., Varga, V., Widlund, P. O., Zanic, M., and Howard, J. (2010) Microtubule dynamics reconstituted *in vitro* and imaged by single-molecule fluorescence microscopy. *Methods Cell Biol.* **95**, 221–245
33. Szyk, A., Deaconescu, A. M., Spector, J., Goodman, B., Valenstein, M. L., Ziolkowska, N. E., Kormendi, V., Grigorieff, N., and Roll-Mecak, A. (2014) Molecular basis for age-dependent microtubule acetylation by tubulin acetyltransferase. *Cell* **157**, 1405–1415
34. Ziolkowska, N. E., and Roll-Mecak, A. (2013) in *Adhesion Protein Protocols* (Coutts, A. S., ed) pp. 323–334, Humana Press, New York
35. Toso, R. J., Jordan, M. A., Farrell, K. W., Matsumoto, B., and Wilson, L. (1993) Kinetic stabilization of microtubule dynamic instability *in vitro* by vinblastine. *Biochemistry* **32**, 1285–1293
36. Burgoyne, R. D., Cambray-Deakin, M. A., Lewis, S. A., Sarkar, S., and Cowan, N. J. (1988) Differential distribution of β -tubulin isoforms in cerebellum. *EMBO J.* **7**, 2311–2319
37. Kavallaris, M. (2010) Microtubules and resistance to tubulin-binding agents. *Nat. Rev. Cancer* **10**, 194–204
38. Peris, L., They, M., Fauré, J., Saoudi, Y., Lafanechère, L., Chilton, J. K., Gordon-Weeks, P., Galjart, N., Bornens, M., Wordeman, L., Wehland, J., Andrieux, A., and Job, D. (2006) Tubulin tyrosination is a major factor affecting the recruitment of CAP-Gly proteins at microtubule plus ends. *J. Cell Biol.* **174**, 839–849
39. Peris, L., Wagenbach, M., Lafanechère, L., Brocard, J., Moore, A. T., Kozielski, F., Job, D., Wordeman, L., and Andrieux, A. (2009) Motor-dependent microtubule disassembly driven by tubulin tyrosination. *J. Cell Biol.* **185**, 1159–1166
40. Vale, R. D., Coppin, C. M., Malik, F., Kull, F. J., and Milligan, R. A. (1994) Tubulin GTP hydrolysis influences the structure, mechanical properties, and kinesin-driven transport of microtubules. *J. Biol. Chem.* **269**, 23769–23775
41. Nogales, E., Wolf, S. G., and Downing, K. H. (1998) Structure of the $\alpha\beta$ tubulin dimer by electron crystallography. *Nature* **391**, 199–203
42. Garnham, C. P., Vemu, A., Wilson-Kubalek, E. M., Yu, I., Szyk, A., Lander, G. C., Milligan, R. A., and Roll-Mecak, A. (2015) Multivalent microtubule recognition by tubulin tyrosine ligase-like family glutamylases. *Cell* **161**, 1112–1123
43. Roll-Mecak, A. (2015) Intrinsically disordered tubulin tails: complex tuners of microtubule functions? *Semin. Cell Dev. Biol.* **37**, 11–19
44. Alushin, G. M., Ramey, V. H., Pasqualato, S., Ball, D. A., Grigorieff, N., Musacchio, A., and Nogales, E. (2010) The Ndc80 kinetochore complex forms oligomeric arrays along microtubules. *Nature* **467**, 805–810
45. Alushin, G. M., Lander, G. C., Kellogg, E. H., Zhang, R., Baker, D., and Nogales, E. (2014) High-resolution microtubule structures reveal the structural transitions in $\alpha\beta$ -tubulin upon GTP hydrolysis. *Cell* **157**, 1117–1129
46. Horio, T., and Hotani, H. (1986) Visualization of the dynamic instability

- ity of individual microtubules by dark-field microscopy. *Nature* **321**, 605–607
47. Mandelkow, E. M., Mandelkow, E., and Milligan, R. A. (1991) Microtubule dynamics and microtubule caps: a time-resolved cryo-electron microscopy study. *J. Cell Biol.* **114**, 977–991
48. Gardner, M. K., Charlebois, B. D., János, I. M., Howard, J., Hunt, A. J., and Odde, D. J. (2011) Rapid microtubule self-assembly kinetics. *Cell* **146**, 582–592
49. Grishchuk, E. L., Molodtsov, M. I., Ataulkhanov, F. I., and McIntosh, J. R. (2005) Force production by disassembling microtubules. *Nature* **438**, 384–388
50. Podolski, M., Mahamdeh, M., and Howard, J. (2014) Stu2, the budding yeast XMAP215/Dis1 homolog, promotes assembly of yeast microtubules by increasing growth rate and decreasing catastrophe frequency. *J. Biol. Chem.* **289**, 28087–28093
51. Cardone, G., Heymann, J. B., and Steven, A. C. (2013) One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.* **184**, 226–236
52. Sousa, D., and Grigorieff, N. (2007) Ab initio resolution measurement for single particle structures. *J. Struct. Biol.* **157**, 201–210

REPORT: *Structure and Dynamics of Human α 1A/ β III-tubulin*

Brain-derived Neurotrophic Factor in Megakaryocytes*[‡]

Received for publication, February 7, 2016, and in revised form, March 21, 2016 Published, JBC Papers in Press, March 22, 2016, DOI 10.1074/jbc.M116.720029

Pedro Chacón-Fernández[‡], Katharina Säuberli[‡], Maria Colzani[§], Thomas Moreau[§], Cedric Ghevaert[§], and Yves-Alain Barde^{‡1}

From the [‡]School of Biosciences, Cardiff University, Cardiff CF10 3AX, Wales and the [§]Department of Haematology, University of Cambridge, Cambridge CB2 0PT, United Kingdom

The biosynthesis of endogenous brain-derived neurotrophic factor (BDNF) has thus far been examined in neurons where it is expressed at very low levels, in an activity-dependent fashion. In humans, BDNF has long been known to accumulate in circulating platelets, at levels far higher than in the brain. During the process of blood coagulation, BDNF is released from platelets, which has led to its extensive use as a readily accessible biomarker, under the assumption that serum levels may somehow reflect brain levels. To identify the cellular origin of BDNF in platelets, we established primary cultures of megakaryocytes, the progenitors of platelets, and we found that human and rat megakaryocytes express the *BDNF* gene. Surprisingly, the pattern of mRNA transcripts is similar to neurons. In the presence of thapsigargin and external calcium, the levels of the mRNA species leading to efficient BDNF translation rapidly increase. Under these conditions, pro-BDNF, the obligatory precursor of biologically active BDNF, becomes readily detectable. Megakaryocytes store BDNF in α -granules, with more than 80% of them also containing platelet factor 4. By contrast, BDNF is undetectable in mouse megakaryocytes, in line with the absence of BDNF in mouse serum. These findings suggest that alterations of BDNF levels in human serum as reported in studies dealing with depression or physical exercise may primarily reflect changes occurring in megakaryocytes and platelets, including the ability of the latter to retain and release BDNF.

BDNF² is a secretory protein regulating the development and function of neural circuits (1, 2). The functional relevance of BDNF in humans is firmly established following the discovery of polymorphisms and loss-of-allele mutations associated with deficits ranging from subtle memory alterations (3) to severe symptoms early in life (4). The cDNA sequence of *BDNF* predicts that like other cystine-knot proteins, BDNF is first synthesized as a precursor protein, referred to as pro-BDNF ensuring the proper formation of disulfide bridges and of a biologically

active, mature neurotrophin (5, 6). Numerous experiments with various artificial expression systems have confirmed this view, in line with the results of early experiments with nerve growth factor (7). So far, only a very small number of studies have been performed addressing the question of the biosynthesis, storage, and secretion of endogenous BDNF (8, 9). As a result of the scarcity of the protein in neurons, most studies used instead overexpression paradigms, leading to uncertainties as to whether the processing of pro-BDNF takes place within neurons or also in the extracellular space following the secretion of pro-BDNF. Human platelets contain between 100–1,000-fold more BDNF than brain tissue when brain and platelets are compared on a protein basis (10–12). As it appears unlikely that the biosynthesis of BDNF takes place in platelets, we established primary cultures of megakaryocytes (Mks), the progenitors of platelets. Beyond questions related to the biosynthesis of endogenous BDNF and to the productive expression of its gene in non-neuronal cells, the question of the origin of BDNF in human blood and serum is of wider interest. Indeed, BDNF levels in human serum are widely used as a biomarker speculated to somehow reflect brain levels. Thus, countless studies have reported decreased BDNF levels in serum in mood disorders, including depression (13), although by contrast physical exercise has been found to increase them (14).

Experimental Procedures

Reagents—For recombinant proteins, BDNF produced in *Escherichia coli* was from Amgen/Regeneron partners (Tarrytown, NY). Cleavage-resistant mouse pro-BDNF and the BDNF pro-peptide were produced in COS-7 cells transfected with the corresponding cDNA (15). Stem cell factor and thrombopoietin (TPO) were from R&D Systems (Abingdon, UK). Interleukin-1 β (IL1- β) was from Miltenyi Biotec (Bisley, Surrey, UK). Human recombinant fibrinogen was from EMD-Millipore (Calbiochem; Darmstadt, Germany). The mouse monoclonal anti-BDNF 3C11 was from Icosagen (Tartu, Estonia). BDNF antibodies directed against mature BDNF recognize the mouse, rat, and human protein with equal affinity as the corresponding amino acid sequence is identical. The mouse monoclonal anti-pro-BDNF H100G was from GeneCopoeia Inc. (Rockville, MD). The monoclonal anti-BDNF#9 was used as described (15). The chicken anti- β -actin Ab13822, the rabbit anti-platelet factor (PF4) Ab129183, and the goat anti-chicken HRP-conjugated secondary antibody Ab97135 were from Abcam (Cambridge, UK). The donkey anti-mouse HRP-conjugated secondary antibody was from Promega (Madison, WI). Alexa Fluor-488 donkey anti-mouse IgG secondary antibody, Alexa Fluor-594 donkey anti-rabbit IgG

* This work was supported by the Life Science Research Network Wales, an initiative funded through the Welsh Government's Ser Cymru program, and National Institute for Health Research Grant RP-PG-0310-1002 (to C. G.).

[‡] This article was selected as a Paper of the Week.

[‡] Author's Choice—Final version free via Creative Commons CC-BY license.

¹ To whom correspondence should be addressed: Sir Martin Evans Bldg., School of Biosciences, Cardiff University, Cardiff CF10 3AX, Wales, United Kingdom. Tel.: 442920870987; E-mail: bardey@cardiff.ac.uk.

² The abbreviations used are: BDNF, brain-derived neurotrophic factor; TPO, thrombopoietin; TRITC, tetramethylrhodamine isothiocyanate; Mk, megakaryocyte.

secondary antibody, and phalloidin-TRITC were purchased from Invitrogen. Hanks' balanced salt solution, Iscove's modified Dulbecco's medium, HEPES, fetal bovine serum (FBS), L-glutamine, and penicillin/streptomycin were from Invitrogen. Cellgro SCGM medium was from Cellgenic (Freiburg, Germany). All other reagents, including thapsigargin and ionomycin, were of analytical grade and were purchased from Sigma (Dorset, UK).

Cell Culture and Isolation—Approximately 10-week-old CD1 or C57BL/6 mice and Wistar rats were sacrificed by rising CO₂ inhalation and blood subsequently drawn by cardiac puncture using acid citric dextrose-B tubes (BD Vacutainer, BD; Plymouth, UK). Femurs and tibiae of CD1 or C57BL/6 mice and Wistar rats were removed, and bone marrow was extracted by flushing the bones with Hanks' balanced salt solution containing 0.38% sodium citrate, 1 mM adenosine, 2 mM theophylline, and 5% heat-inactivated FBS. After lysis of the red cells, the suspension was passed through a 40- μ m strainer, and the pellet was resuspended in culture medium (Iscove's modified Dulbecco's medium with L-glutamine, 25 mM HEPES, 5% heat-inactivated FBS, 100 units/ml penicillin, and 100 μ g/ml streptomycin) containing 25 ng/ml stem cell factor and TPO. After 6–7 days of culture *in vitro*, mature Mks were purified on a 1.5–3% bovine serum albumin gradient and cultured in Cellgro SCGM medium for up to 16 h. Hippocampi and lungs were dissected and kept frozen at –80 °C until use. Human samples, both neonatal cord blood and adult peripheral blood, were obtained after securing informed consent following a protocol approved by the National Research Ethics Service. CD34-positive cells (\geq 98%) were isolated by magnetic cell sorting, cultured for 10 days in Cellgro SCGM medium containing TPO and IL1- β , and analyzed by flow cytometry with 70–90% CD41a⁺ and 20–60% CD42a⁺ cells representing committed progenitors and mature Mks, respectively. Platelet-rich plasma was obtained from either animal or human samples by centrifugation of acid citric dextrose-B tubes at 200 \times g for 20 min in the presence of prostaglandin-E1 (1 μ M) and apyrase (1 unit/ml) to prevent cellular activation. Platelets were pelleted from platelet-rich plasma by centrifugation at 1,100 \times g for 10 min and lysed immediately.

Western Blot and Densitometric Analyses—Platelets and Mks were lysed for 30 min on ice in a buffer containing 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, and 0.2% sodium deoxycholate, supplemented with protease and phosphatase inhibitor mixture mix, 10 μ M 1,10-phenanthroline monohydrate, 10 mM 6-aminohexanoic acid and 10 μ g/ml aprotinin. After sonication (1 pulse, 50% amplitude), insoluble debris was removed by centrifugation. Proteins were separated on 4–12% NuPAGE gradient gels containing SDS (Invitrogen) and transferred to nitrocellulose membranes using the semi-dry Trans-Blot unit (Bio-Rad, Hertfordshire, UK). To allow the detection of the BDNF propeptide (15), the membranes were fixed after transfer with 2.5% glutaraldehyde for 30 min at room temperature and subsequently incubated for 2 h with blocking solution (3% blocking reagent (Invitrogen) and 3% BSA in TBS-T) and then probed overnight at 4 °C with anti-BDNF 3C11

(1:2,000), anti-pro-BDNF H1001G (1:1,000), or chicken anti- β -actin (1:2,000). Binding of primary antibodies was visualized with donkey anti-mouse HRP-conjugated secondary antibodies (1:10,000) or goat anti-chicken HRP-conjugated secondary antibody (1:5,000). Chemiluminescence was developed using the Lumi Glo Reverse Western blotting substrate (Cell Signaling Technology, Danvers, MA). Care was taken in all experiments to ensure that the signal was well within the linear range, and densitometry was carried out using ImageJ software.

BDNF ELISA Determinations—BDNF sandwich ELISA was performed using a combination of monoclonal antibodies as described previously (16), with the following modifications. Streptavidin high binding capacity coated white plates (Pierce) were incubated overnight at room temperature with 10 μ g/ml mouse anti-BDNF antibody 1 conjugated with biotin (Sulfo-NHS-LC-biotinylation kit EZ-Link, Pierce) in coating buffer (25 mM Tris, 150 mM NaCl, 0.1% BSA, 0.05% Tween 20, pH 7.2). Following an overnight incubation, the plates were washed with 0.1% Tween 20 in PBS (washing buffer) and then blocked for 2 h with 4% BSA in PBS. To each well 150 μ l of incubation buffer (0.1 M KH₂PO₄, 0.1 M Na₂HPO₄, pH 7.6) were added, followed by 50 μ l of sample or BDNF standard. Mk lysates or recombinant BDNF were incubated overnight at 4 °C with 10 μ g/ml mouse anti-BDNF antibody 9 conjugated with horseradish peroxidase (activated peroxidase kit, EZ-Link Plus, Pierce). Mk lysates were used at 1:3 in 0.1% Triton X-100 PBS (PBT). To test for a potential release of BDNF from Mks into their incubation medium, purified Mks were cultured on a 24-well plate and incubated for 2 days at 37 °C with 5% CO₂, in CellGro SCGM medium containing 10 μ g/ml horseradish peroxidase-conjugated anti-BDNF antibody 9. The standard curve was established in a parallel plate using various concentrations of recombinant BDNF incubated for 2 days in the same culture medium without cells. The limit of detection is 1.25 pg of BDNF per well. After 2 days, the corresponding media (Mks and standard curve) were collected and incubated for 3 h on a rotating platform followed by a rinse with washing buffer. BM chemiluminescence ELISA substrate POD (Roche Applied Science, Mannheim, Germany) was added, and luminescence was measured with a microplate reader (FLUOstar Omega, BMG labtech). Both standards and samples were determined in triplicate.

Immunocytochemistry and Confocal Analysis—Mks were cultured on 200 μ g/ml human fibrinogen-coated coverslips for 6 h (mouse and rat Mks) or 36 h (human MKs), fixed for 30 min in 4% paraformaldehyde in PBS, and permeabilized for 15 min at room temperature with 0.5% Triton X-100 in PBS. Mks were then blocked with blocking solution (10% donkey serum in PBT) for 1 h. Primary antibodies were diluted in blocking solution at the following final concentrations/dilutions: 1:500 rabbit anti-PF4 and 10 μ g/ml mouse anti-BDNF antibody 9. After 2 h of incubation at room temperature, coverslips were washed three times with PBT and incubated for 1 h with the secondary antibodies used at a 1:500 dilution in blocking solution or phalloidin/TRITC (at 1:50 dilution in blocking solution). After further washing with PBT, labeled coverslips were mounted onto glass slides with DAPI-containing mounting medium. Images

TABLE 1
Primers used in conventional PCR experiments

m is mouse; r is rat; and h is human.

Name of oligonucleotide	Sequence (5'–3')	Product size
mr <i>BdnfI</i> sense	GTGTGACCTGAGCAGTGGGCAAAGGA	803 bp
mr <i>BdnfII</i> sense	GGAAGTGGAAAGAAACCGTCTAGAGCA	469 bp (IIa) 682 bp (IIb) 765 bp (IIc)
m <i>BdnfIII</i> sense	GCTTTCTATCATCCCTCCCCGAGAGT	425 bp
r <i>BdnfIII</i> sense	CCTTTCTATTTTCCTCCCCGAGAGT	427 bp
mr <i>BdnfIV</i> sense	CTCTGCCTAGATCAAATGGAGCTTC	553 bp
mr <i>BdnfV</i> sense	CTCTGTGTAGTTTCATTGTGTGTTTC	364 bp
m <i>BdnfVI</i> sense	GCTGGCTGTGCGACGGTCCCATT	542 bp
r <i>BdnfVI</i> sense	GCTGGCTGTGCGACGGTCCCATT	543 bp
mr <i>BdnfVII</i> sense	CCTGAAAGGGTCTGCGAACTCCA	420 bp
mr <i>BdnfVIII</i> sense	GTGTGTGTCTCTGCGCCTCAGTGGA	362 bp
m <i>BdnfIXA</i> sense	CCCAAAGCTGCTAAAGCGGGAGGAAG	
r <i>BdnfIXA</i> sense	CCAGAGCTGCTAAAGTGGGAGGAAG	525 bp
mr <i>Bdnf</i> antisense	GAAGTGTACAAGTCCGCTCTCTTA	
h <i>BDNF</i> sense	GATGCCAGTTGCTTGTCTCTGTAG	471 bp
h <i>BDNFII</i> sense	GGGCGATAGGAGTCCATTCAGCACC	311 bp (IIa) 526 bp (IIb) 609 bp (IIc)
h <i>BDNFIII</i> sense	AGTTTCGGGGCTGGCTTAGAG	346 bp
h <i>BDNFIV</i> sense	GCTGCAGAACAGAGGAGTACA	411 bp
h <i>BDNFV</i> sense	TCGCGTTCGCAAGCTCCGTAGTG	272 bp (Va) 282 bp (Vb) 565 bp (V-VIII-VIIIh) 682 bp (V-VIII)
h <i>BDNFVh</i> sense	GGCTGGAACACCCCTCGAA	339 bp
h <i>BDNFVI</i> sense	GGCTTTAATGAGACACCCACCGC	368 bp (VIa) 386 bp (VIb) 493 bp (VI-IXb)
h <i>BDNFVII</i> sense	GAAGTAAAGGGTCTGCGACACTCT	328 bp (VIIa) 428 bp (VIIb)
h <i>BDNFIX</i> sense	TTTCTCGTGACAGCATGAGCAG	352 bp (IXabd) 536 bp (IXabcd)
h <i>BDNF</i> antisense	GTCCTCATCCAACAGCTCTTCTATC	
hmr <i>HPRT</i> sense	GATGATGAACCAGGTATGAC	469 bp
hmr <i>HPRT</i> antisense	GTCCTTTTCACCAGCAAGCTTG	

were acquired at ×63 magnification using a confocal microscope (LSM 780; Carl Zeiss). Co-localization coefficient between BDNF and PF4 was determined using the Zeiss co-localization coefficient software (ZEN black 2011), which utilizes the Manders overlap coefficient equation to quantify overlapping pixels. At least 15 cells per sample were analyzed.

RNA Extraction, Retrotranscription, and Conventional and Real Time Quantitative PCRs—Total RNA was extracted from animal tissue (dissected hippocampi or lungs) using TRIzol reagent (Invitrogen) and cells (mouse, rat, and human MKs) using RNeasy kit (Qiagen, Valencia, CA) according to the manufacturer's instructions, including a DNase treatment. Human RNA samples from hippocampus or lung were obtained from Clontech. cDNA was prepared from 1,250 ng of total RNA using random hexamers (Promega) and SuperScript III first-strand synthesis system (Invitrogen). To analyze expression of mouse, rat, and human *BDNF* exon-specific transcripts, cDNA was amplified with 35–40 cycles of PCR using an annealing temperature of 57–60 °C for all primer combinations (primers listed in Table 1). Real time quantitative PCR was performed on the StepOne system (Applied Biosystems, Invitrogen) using TaqMan probes and primers for mouse, rat, and human *BDNF* exon-specific transcripts or *BDNF* coding sequence along with the housekeeping genes *GAPDH* and *rRNA18S* (Applied Biosystem, Invitrogen, primers listed in Table 2). To compare the expression levels of the *BDNF* mRNA coding sequence between species, SYBR® master

TABLE 2
Primers with references used in real time quantitative PCR experiments
m is mouse; r is rat; and h is human.

Name of oligonucleotide	Reference or sequence (5'–3')	Product size
m <i>Bdnf</i> (coding sequence)	Mm04230607_s1	92 bp
m <i>BdnfI</i>	Mm01334047_m1	105 bp
m <i>BdnfIV</i>	Mm00432069_m1	145 bp
m <i>BdnfVI</i>	Mm01334042_m1	108 bp
m <i>BdnfIXa</i>	Mm04230616_s1	77 bp
m <i>Gapdh</i>	Mm99999915_g1	109 bp
m <i>rRNA18S</i>	Mm03928990_g1	61 bp
r <i>Bdnf</i> (coding sequence)	Rn02531967_s1	142 bp
r <i>BdnfI</i>	Rn01484924_m1	106 bp
r <i>BdnfIV</i>	Rn01484927_m1	120 bp
r <i>BdnfVI</i>	Rn01484928_m1	109 bp
r <i>BdnfIXA</i>	Rn04230568_s1	95 bp
r <i>Gapdh</i>	Rn01775763_g1	174 bp
r <i>rRNA18S</i>	Rn03928990_g1	61 bp
h <i>BDNF</i> (coding sequence)	Hs02718934_s1	74 bp
h <i>BDNEI</i>	Hs00538277_m1	104 bp
h <i>BDNFIV</i>	Hs00380947_m1	116 bp
h <i>BDNFVIa</i>	Hs04188535_m1	119 bp
h <i>BDNFVib</i>	Hs00156058_m1	143 bp
h <i>BDNFIXabcd</i>	Hs00542425_s1	81 bp
h <i>GAPDH</i>	Hs02758991_g1	93 bp
h <i>rRNA18S</i>	Hs03003631_g1	69 bp
mrh <i>BDNF</i> sense (coding sequence)	GAGCTGAGCGTGTGTGACAG	256 bp
mrh <i>BDNF</i> antisense (coding sequence)	CGCCAGCCAATTCTCTTTTGTG	
mrh <i>rRNA18S</i> sense	GTCTGTGATGCCCTTAGATG	176 bp
mrh <i>rRNA18S</i> antisense	AGCTTATGACCCGCACTTAC	

mix (Applied Biosystems, Invitrogen) was used together with the corresponding primers (Table 2). Relative *BDNF* gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method

and the housekeeping genes for normalization. The expression of proprotein convertases in rat Mks was analyzed using primers as described previously (17).

Statistical Analysis—All values are expressed as mean \pm S.E. All statistical tests were paired *t* test, one-tailed.

Results

To determine whether BDNF can be detected in Mks, we established primary cultures of mouse, rat, and human Mks and analyzed their content by Western blotting using a BDNF monoclonal antibody specifically recognizing BDNF and not NGF or NT4, under conditions where both are recognized by their corresponding antibodies (data not shown). A strong signal corresponding to the expected size of the monomer of mature BDNF was detected in both rat and human but not in mouse Mks (Fig. 1A). Note that pro-BDNF is barely detectable under these conditions, with only a faint band detected in human Mks. As a control, we also analyzed the BDNF content of platelets purified from the corresponding species and confirmed the presence of considerable amounts of BDNF in human platelets, with substantial levels also detected in rat but not in mouse platelets (Fig. 1B). This result with platelets is in agreement with previous conclusions using immunoassay determinations with the serum prepared from these three species (18). We then explored whether BDNF would be localized in α -granules, the storage compartment of a number of growth factors in the Mk lineage. A BDNF-specific signal was found to extensively co-localize with PF4, also designated CXCL4, one of the most abundant α -granule proteins (Fig. 1C). A distinct BDNF signal could also be seen in the tips of the proplatelet-forming human Mks (Fig. 1D, *arrows*), consistent with the notion that BDNF is transferred from Mks to platelets. As a control for the specificity of the BDNF signal in these immunostaining experiments, we used mouse Mks. When incubated with the corresponding antibodies under the same experimental conditions as rat and human Mks, PF4, but not BDNF, was detected (Fig. 1C). The lack of any detectable BDNF in mouse platelets is not a feature specific to CD1 mouse strain that we used in most of our experiments, as the C57BL/6 strain led to identically negative results, both in Mks and platelets. We note that in a recent highly sensitive and quantitative proteomic analysis of C57BL/6 platelet extracts, no BDNF could be detected (19). Quantification of the BDNF levels in rat Mks by ELISA indicated that their lysates contain 1.40 ± 0.13 ng/mg protein (mean \pm S.E., $n = 6$). We also attempted to determine whether proplatelet-forming Mks release BDNF into the medium by incubating the cells with HRP-conjugated BDNF monoclonal antibodies (see "Experimental Procedures"). After 2 days of incubation with the BDNF capture antibody, we failed to detect any release of BDNF into the Mk-conditioned medium, suggesting that the bulk of BDNF is transferred into platelets and not released into the medium. This result is consistent with the previous work indicating that both in rats and humans the levels of BDNF are far higher in serum than in plasma, suggesting that the bulk of BDNF in serum results from platelet degranulation (18). To determine whether the BDNF biosynthetic machinery is expressed in Mks, we isolated RNA from mature platelet-forming mouse, rat, and human Mks. Sig-

nificant levels of *BDNF* mRNAs were detected in both rat and in human cells (Fig. 2A). In four separate experiments, the total mRNA levels in both species were found to be about 200-fold higher in rat and human, compared with mouse Mks. All known *BDNF* transcripts present in RNA extracted from Mks of the three species were analyzed and compared with RNA extracted from the hippocampus and the lung as neuronal and non-neuronal reference tissues of the corresponding species (Fig. 2A). The results of these experiments revealed a neuronal pattern of mRNA expression in both rat and human Mks, with a prominent inclusion of exon I and IV transcripts. Notably, exon I-containing transcripts have recently been shown to markedly increase *Bdnf* mRNA translatability (20). By contrast, these typical neuronal transcripts were undetectable in the mouse (Fig. 2A). The levels of all main transcripts were also assessed in the three species by real time PCR (Fig. 2B).

In neurons, increased levels of cytoplasmic calcium have long been known to activate *Bdnf* transcription by activating promoters I and IV in particular (21). As Mks are devoid of voltage-activated calcium channels, the addition of thapsigargin to non-electrically excitable cells such as Mks offered an opportunity to test whether the corresponding *Bdnf* promoters are also responsive to calcium levels in Mks. Thapsigargin is a selective inhibitor of the sarco/endoplasmic reticulum Ca^{2+} -ATPase, secondarily leading to the opening of stored-operated calcium channels at the cell surface and increased levels of cytoplasmic calcium (22). We found that at nanomolar concentrations (Fig. 3A), thapsigargin massively activates transcription in a time-dependent manner (Fig. 3B). This increase was primarily accounted for by contributions from exons I, IV, and IXa and to a lesser extent exon VI (Fig. 3C). We then tested whether complexing extracellular calcium would decrease the thapsigargin-induced transcriptional activation of *Bdnf* and found that 2.5 mM EGTA completely blocked the inductive effects of thapsigargin as assessed both by primers corresponding to the protein coding sequence or by exon-specific primers (Fig. 3C). To further test the notion that intracellular calcium levels regulate *Bdnf* transcription, we tested the effects of the calcium ionophore ionomycin. When added to mature Mk cultures for 4 h, it increased the levels of *Bdnf* mRNA by 10.50 ± 1.85 -fold ($n = 3$, mean \pm S.E.). As still very little is known about the biosynthesis of endogenous BDNF in any cell type, we were then curious to see how the Mk translation and processing machinery would cope with the massive increase in *Bdnf* mRNA levels caused by thapsigargin. We found that at 10 nM, thapsigargin led not only to a marked increase of processed (or mature) BDNF but also to readily detectable levels of pro-BDNF, suggesting that the thapsigargin-induced increased transcription may temporarily saturate the pro-BDNF processing capacity of Mks (Fig. 4, A–C). The 3C11 BDNF monoclonal antibody recognizes not only mature but also (as expected) unprocessed and partially processed forms of BDNF as indicated in Fig. 4, A and B. In previous experiments using heterologous expression systems, we noted that the replacement of an arginine residue in position -1 by lysine at the furin cleavage site of pro-BDNF led to the accumulation of an *N*-glycosylated intermediate product corresponding to the size indicated by the *arrows* in Fig. 4, A and B. Amino-terminal sequencing of this product indicated that the

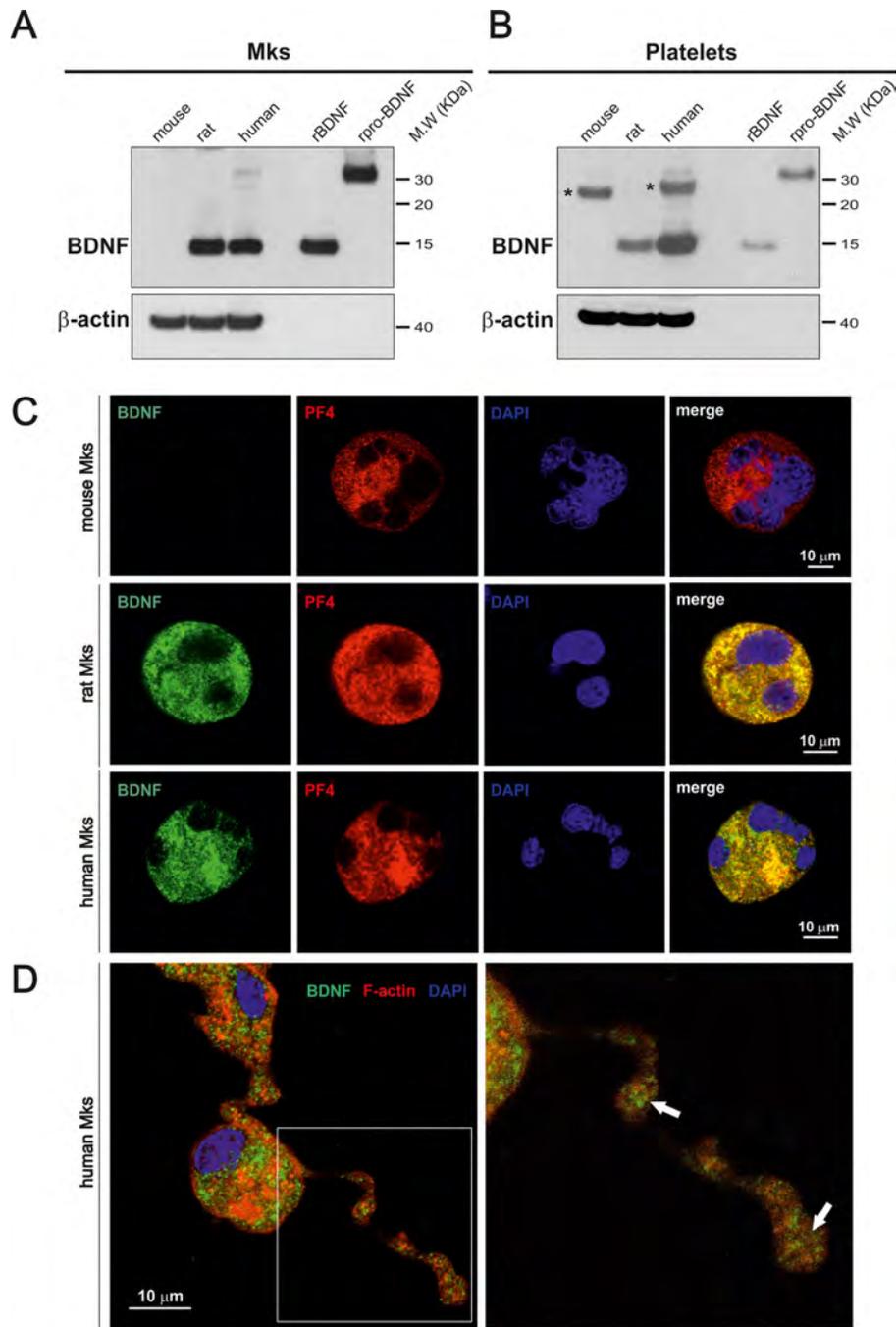


FIGURE 1. Differential BDNF protein levels in mouse, rat, and human megakaryocytes and platelets. Western blot lysates of cultured Mks (A) and blood platelets (B) are shown. Eighty micrograms of protein per lane were loaded, and the blotting membrane was incubated with the mouse monoclonal antibody 3C11 developed by Icosagen (Tartu, Estonia). Recombinant BDNF and pro-BDNF were used as molecular mass markers and antibodies to β -actin as loading controls. Asterisks (top right panels) point to a band unrelated to BDNF likely corresponding to immunoglobulin light chains in the mouse sample. Note the absence of BDNF in mouse Mks and platelets. C, antibodies to BDNF 9 (green) (15) and PF4 (red) reveal expression of both antigens in mature rat and human Mks. Note that unlike PF4, BDNF is not detectable in mouse Mks. The co-localization of BDNF with PF4 in rat and human Mks was quantified using the pixel intensity specifically generated by each channel. In humans, 83% and in rats 86% BDNF-positive granules were also PF4-positive. Blue, DAPI staining. D, immunofluorescence staining of F-actin (red) and BDNF (green) in proplatelet-forming cultured human Mks. Arrows indicate BDNF accumulation in proplatelet buds.

use of an alternative cleavage site generated a product with a 15-residue addition to the amino terminus of mature BDNF (23). This product was shown to be *N*-glycosylated (21). The use of the pro-BDNF monoclonal antibody H1001G independently confirmed the identity of pro-BDNF and also allowed the detection of the BDNF cleaved pro-peptide (Fig. 4C). We note that in the absence of thapsigargin stimulation, the steady state levels

of pro-BDNF are even lower than in neurons where we estimated them to represent about 1 molecule for 10 molecules of mature BDNF (8, 15). We also tested the expression of the most likely protease candidate thought to be involved in the generation of mature BDNF, namely furin and the proprotein convertase 7, the latter having been recently shown to be necessary for the processing of BDNF in neurons (24). The results of con-

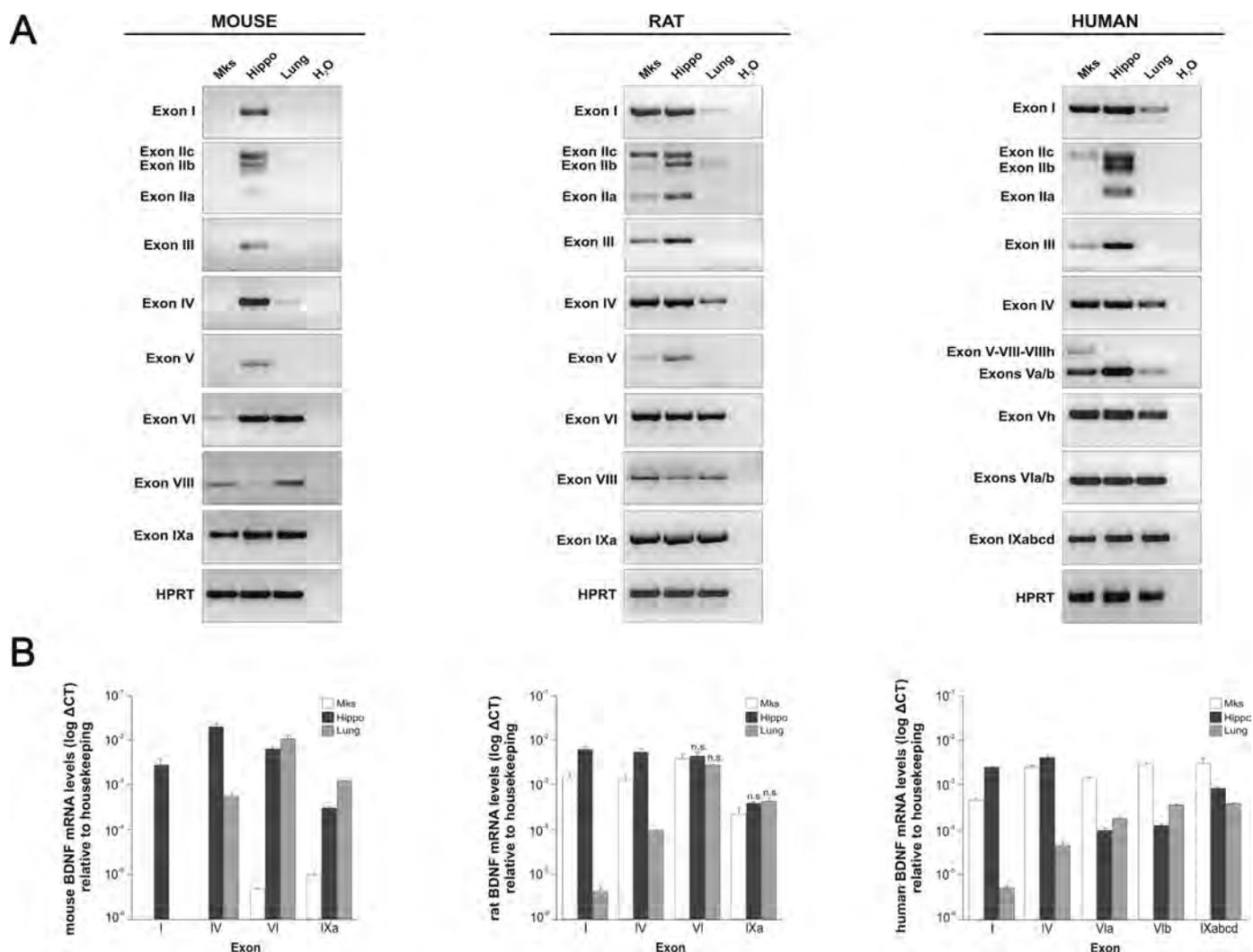


FIGURE 2. Transcriptional analysis of BDNF in mouse, rat, and human megakaryocytes. Conventional (A) and real time quantitative (B) PCR using exon-specific primers with RNA extracted from mature cultured Mks, adult hippocampus (*Hippo*), and lung are shown. Note that in the mouse, the neuron-specific transcripts, including exon I and IV, are not detected and that by contrast the transcript pattern resembles the non-neuronal pattern observed in lung tissue. The converse is the case with RNA extracted from rat and human Mks with transcript patterns, including exon I and IV, that are characteristic of a neuronal pattern as illustrated with the hippocampus. Unless indicated as non-significant (*n.s.*), all values are mean values \pm S.E. in triplicates and based on three independent experiments, at $p < 0.001$ (paired *t* test).

ventional (Fig. 5A) and real time quantitative PCR (Fig. 5B) experiments indicated that both enzymes are present in rat Mks, at levels comparable with those found in the rat hippocampus (Fig. 5). In line with this, the processing of pro-BDNF was largely prevented by the addition of the convertase inhibitor decanoyl-Arg-Val-Lys-Arg-chloromethyl ketone (data not shown).

Discussion

The results obtained with primary cultures of Mks suggest that these cells represent the main source of BDNF in platelets as well as in serum. First, Mks contain readily detectable levels of BDNF protein. Second, BDNF is stored in α -granules in Mks, long known to also represent the storage compartment of various growth factors and cytokines in platelets (25). Third, BDNF can be visualized in proplatelets (Fig. 1D) suggesting that platelets contain BDNF by the time they begin to separate from Mks. Fourth, the *BDNF* gene is expressed at relatively high levels in rat and human Mks, although by contrast, the levels are

much lower in mice. Notably, the mouse transcripts do not include exon I, which allows efficient translation of BDNF mRNA (20). This negative result with mouse Mks is significant as it correlates with the lack of detectable levels of BDNF in mouse serum, unlike the case with rat and human sera (18). Taken together, it would seem that circulating platelets, once filled up with BDNF packaged in α -granules inherited from Mks, represent the only significant source of BDNF in serum. Other sources such as endothelial or immune cells that do express the *BDNF* gene at low levels (26, 27) do not seem to make significant contributions to circulating levels of BDNF as the levels are undetectable in mouse serum (18). Also, the very low levels of BDNF found in human and rat plasma may in fact be accounted for by microparticles or exosomes released from platelets (28). It is intriguing to note that human platelets and sera contain BDNF levels that are about 10 times higher than in the rat. The reasons for this difference are unclear at this point, as is the function of BDNF in platelets. The lack of BDNF in mouse platelets suggests that whatever the biological role of

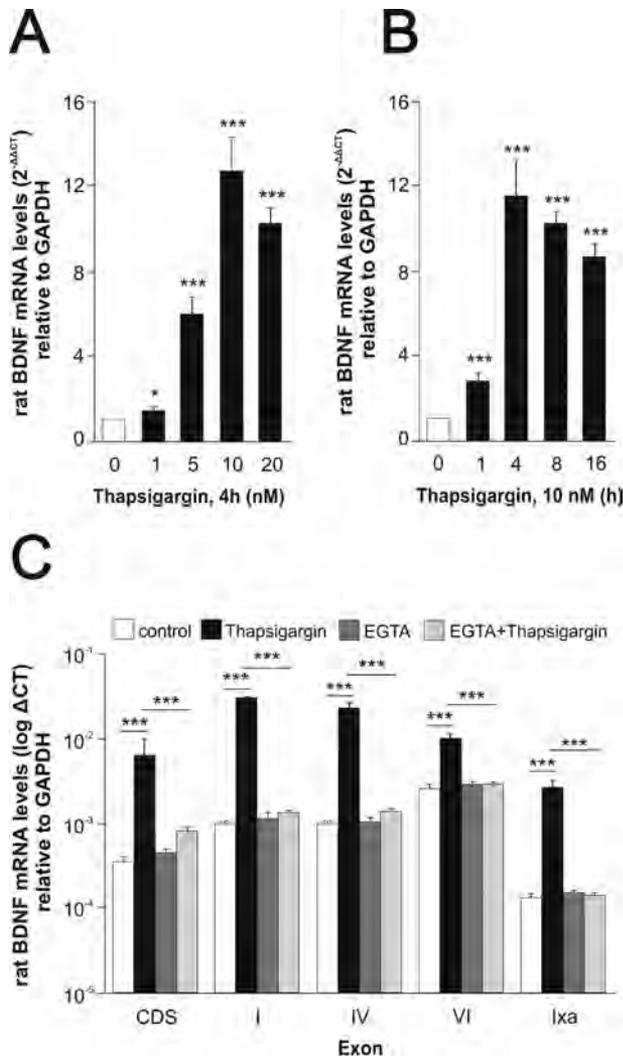


FIGURE 3. Up-regulation of Bdnf mRNA by thapsigargin. Effect of extracellular calcium. Dose response (A) and time course (B) of Bdnf mRNA expression by rat Mks after thapsigargin treatment. Purified mature rat Mks were cultured in the presence or absence of thapsigargin or vehicle (DMSO) used at the indicated concentrations (A) and for different lengths of times (B). Total mRNA was extracted and reverse-transcribed, and the resulting cDNA was amplified by real time quantitative PCR using specific primers for the coding sequence of Bdnf. C, extracellular calcium dependence of thapsigargin-induced Bdnf mRNA increase. Rat Mks were preincubated with 2.5 mM EGTA for 1.5 h at 37 °C followed by 10 nM thapsigargin for 4 h. mRNA expression was analyzed by real time quantitative PCR using specific primers for the coding sequence of Bdnf (CDS) or exon-specific primers. All values are mean values ± S.E. in triplicates and based on three independent experiments. Unless indicated, all the statistical values are compared with the control. *, $p < 0.05$; ***, $p < 0.001$ (paired *t* test).

platelet-derived BDNF may be, it could be redundant in the mouse with the function of other platelet-derived growth factors. A signaling system based on the circulation of small and ubiquitous cellular fragments, including exosomes loaded up with a powerful neurotrophic factor, has the potential to be functionally relevant in the context of human brain function. As blood flow is tightly regulated by neuronal activity (29), it is conceivable that in humans exosomes loaded with BDNF (30) may be delivered to the brain in activity-dependent fashion, possibly explaining the beneficial effects of physical exercise. Although this is a matter of speculation at this point, the possibility of a functional role for BDNF in platelets can now be

tested by engineering the mouse genome so as to replicate the situation in humans. Alternatively, it is also conceivable that the functional significance of BDNF in human platelets may remain as mysterious as that of NGF and EGF in the adult male mouse submandibular gland (31).

It may seem surprising that the cellular source of BDNF in rat and human serum has not been previously uncovered, especially in view of the very extensive use of BDNF as a biomarker in human blood (32). An analysis of the corresponding literature reveals that negative results were obtained early on in experiments specifically addressing the question of BDNF expression in human Mks (16). These experiments were performed with the megakaryocyte lines DAMI and Meg-01 and led to the conclusion that the BDNF gene is not expressed in the cells (16). Although we confirmed these results, it appears plausible that these tumor lines fail to faithfully replicate late aspects of Mk maturation, as is not rarely the case with readily expandable tumor cells. Following this negative result, the presence of BDNF in platelets has been speculated to results from a hypothetical uptake from sources such as the brain. However, this notion has not been substantiated by plausible mechanisms, unlike in the case of serotonin, a neurotransmitter long known to accumulate in the dense granules of platelets following its uptake by specific transporters located in the membrane of platelets.

The identification of Mks as the source of BDNF in platelets invites a revision of the widely held view that in humans the serum levels of BDNF reflect its levels in the brain. In addition to our findings, it has long been established that radiolabeled BDNF does not reach the brain when injected into the peripheral circulation (33). It appears then that the variations in the levels of BDNF reported in various conditions, including the increase after physical exercise (14) or decrease during the course of depressive episodes (13), are in need of alternative plausible explanations, and it is conceivable that these variations may reflect different degrees of platelet activation (34). In addition, there is emerging evidence that the hematopoietic niche where Mks develop (35) is innervated by the peripheral nervous system and that hematopoietic cells may respond to nerve-derived signals (36). However, whether or not these stimuli change the expression levels of BDNF in Mks remains unclear at this point.

With regard to the biosynthesis of endogenous BDNF, our results suggest that Mks could represent an alternative cellular model to neurons, which have been so difficult to study in the face of the very low levels of expression of BDNF levels in these cells. By comparison with BDNF levels in the brain (11), the levels of BDNF in human platelets are significantly higher, 100–1,000-fold on a per mg of protein basis when brain extracts and purified platelets are compared (12). In particular, Mks offer a new opportunity to examine the biosynthesis of endogenous BDNF and the possible role of pro-BDNF. Transcription activation by thapsigargin leads to clearly detectable levels of pro-BDNF without the need for prior enrichment by immunoprecipitation. In view of the current interest related to the Val → Met substitution in pro-BDNF (3), human Mks of the corresponding genotype may represent an interesting cellular

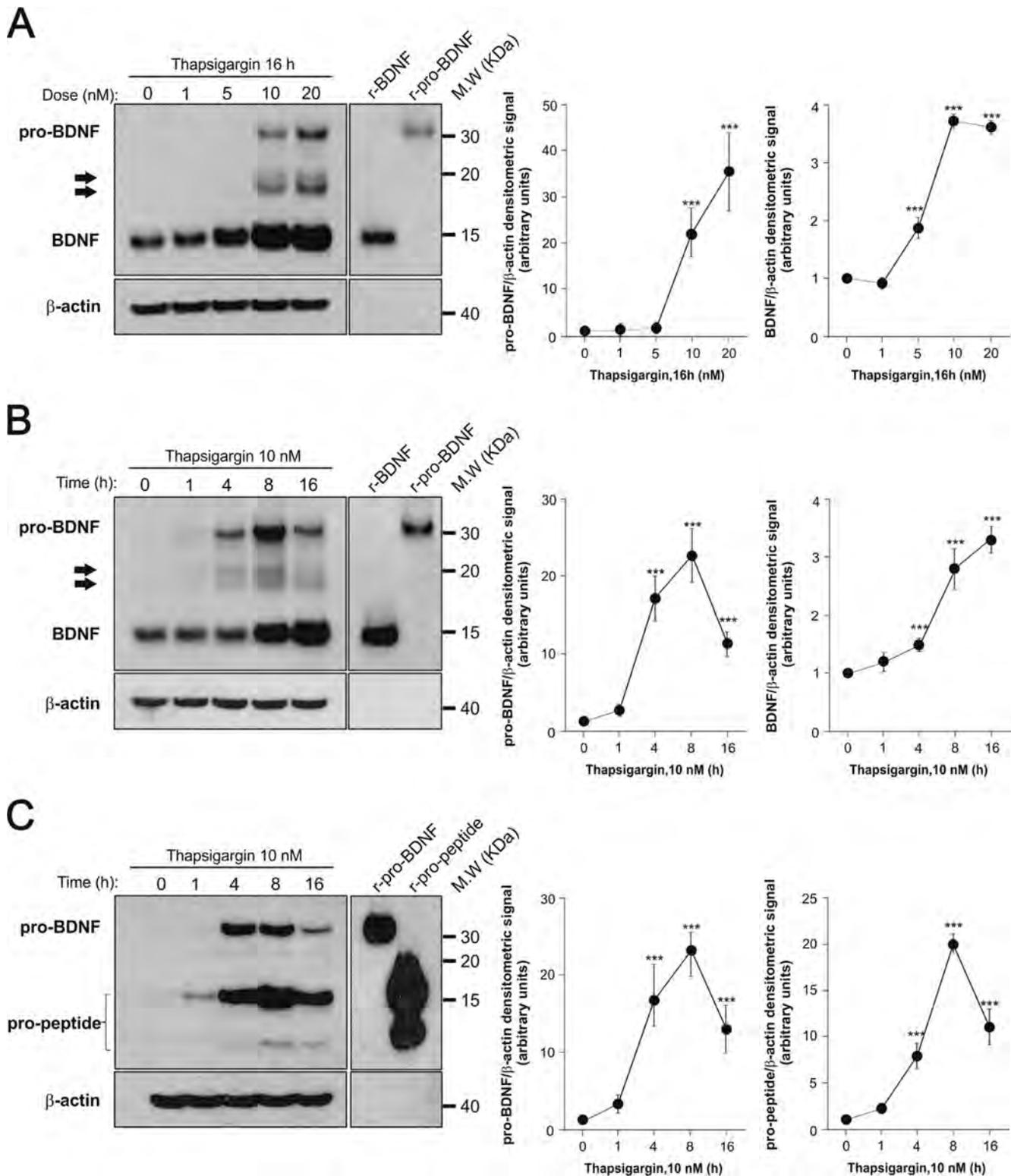


FIGURE 4. Effect of thapsigargin on pro-BDNF, mature BDNF, and pro-peptide in rat Mks. Dose response (A) and time course (B) of pro-BDNF and mature BDNF proteins by rat Mks after thapsigargin treatment are shown. Mature Mks were cultured for 16 h at the indicated doses of thapsigargin (A) or 10 nM thapsigargin for the indicated times (B). Forty micrograms of protein per lane were loaded, and the blotting membrane was incubated with the mouse monoclonal antibody 3C11 developed by Icosagen. Arrows indicate intermediate proteolytic products of pro-BDNF (C). Time course of pro-BDNF and pro-peptide proteins generated by rat Mks incubated with 10 nM thapsigargin for the indicated time periods. Eighty micrograms protein per lane were loaded, and the blotting membrane was incubated with the mouse monoclonal antibody H1001G developed by GeneCopeia, Inc. The blots shown are representative of three independent experiments with similar results. Graphs show mean \pm S.E. of the densitometric values quantified from the blots of the three separate experiments. $***, p < 0.001$ (paired *t* test compared the corresponding controls). Recombinant BDNF (150–300 pg), cleavage-resistant recombinant pro-BDNF (0.5–1 ng), and recombinant pro-peptide (1–10 ng) were used as molecular mass markers and antibodies to β -actin as loading controls.

BDNF and Megakaryocytes

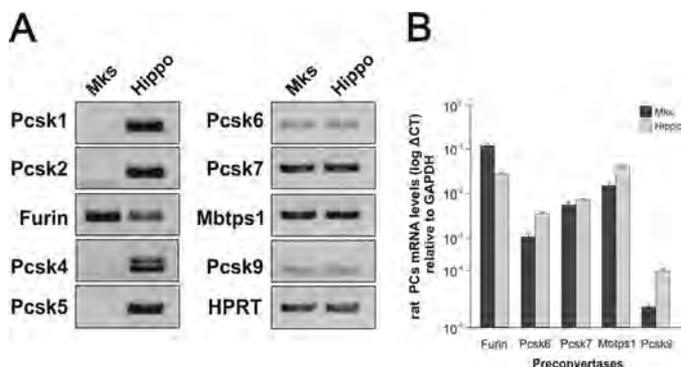


FIGURE 5. Differential expression of proprotein convertases in primary MKs. Conventional (A) and real time quantitative (B) PCRs using specific primers and RNA extracted from mature cultured rat Mks and adult hippocampus are shown. Note that although transcripts, including *Pcsk1*, *Pcsk2*, *Pcsk4*, and *Pcsk5*, are expressed in hippocampal tissue, they are not detected in Mks (A). Comparative expression levels between the two tissues for the expressed proprotein convertases are shown in B. All values are mean values \pm S.E. in triplicates and are based on three independent experiments.

model to understand the biochemical consequences of this amino acid replacement.

In conclusion, our results contribute to clarify the cellular origin of BDNF in human blood; and they describe a tractable cellular system to study the biosynthesis of endogenous BDNF.

Author Contributions—P. C. F. designed, performed, and analyzed the experiments illustrated in Figs. 1–5 and K. S. was also involved. T. M., M. C., and C. G. were involved in all aspects of the work related to human, rat, and mouse Mks and in the interpretation of the results. Y. A. B. helped with the initiation of the project, the design of the experiments, and with the interpretation of the results. All have read the manuscript and discussed its content. The final version was approved by all.

Acknowledgments—We thank Mike Greenberg, Michael Frotscher, and Hayley Dingsdale for useful discussions.

References

- Park, H., and Poo, M. M. (2013) Neurotrophin regulation of neural circuit development and function. *Nat. Rev. Neurosci.* **14**, 7–23
- Zagrebelsky, M., and Korte, M. (2014) Form follows function: BDNF and its involvement in sculpting the function and structure of synapses. *Neuropharmacology* **76**, 628–638
- Egan, M. F., Kojima, M., Callicott, J. H., Goldberg, T. E., Kolachana, B. S., Bertolino, A., Zaitsev, E., Gold, B., Goldman, D., Dean, M., Lu, B., and Weinberger, D. R. (2003) The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell* **112**, 257–269
- Gray, J., Yeo, G. S., Cox, J. J., Morton, J., Adlam, A. L., Keogh, J. M., Yanovski, J. A., El Gharbawy, A., Han, J. C., Tung, Y. C., Hodges, J. R., Raymond, F. L., O'rahilly, S., and Farooqi, I. S. (2006) Hyperphagia, severe obesity, impaired cognitive function, and hyperactivity associated with functional loss of one copy of the brain-derived neurotrophic factor (BDNF) gene. *Diabetes* **55**, 3366–3371
- Leibrock, J., Lottspeich, F., Hohn, A., Hofer, M., Hengerer, B., Masiaowski, P., Thoenen, H., and Barde, Y. A. (1989) Molecular cloning and expression of brain-derived neurotrophic factor. *Nature* **341**, 149–152
- McDonald, N. Q., and Hendrickson, W. A. (1993) A structural superfamily of growth factors containing a cystine knot motif. *Cell* **73**, 421–424
- Suter, U., Heymach, J. V., Jr., and Shooter, E. M. (1991) Two conserved domains in the NGF propeptide are necessary and sufficient for the biosynthesis of correctly processed and biologically active NGF. *EMBO J.* **10**,

- 2395–2400
- Matsumoto, T., Rauskolb, S., Polack, M., Klose, J., Kolbeck, R., Korte, M., and Barde, Y. A. (2008) Biosynthesis and processing of endogenous BDNF: CNS neurons store and secrete BDNF, not pro-BDNF. *Nat. Neurosci.* **11**, 131–133
- Yang, J., Siao, C. J., Nagappan, G., Marinic, T., Jing, D., McGrath, K., Chen, Z. Y., Mark, W., Tessarollo, L., Lee, F. S., Lu, B., and Hempstead, B. L. (2009) Neuronal release of proBDNF. *Nat. Neurosci.* **12**, 113–115
- Yamamoto, H., and Gurney, M. E. (1990) Human platelets contain brain-derived neurotrophic factor. *J. Neurosci.* **10**, 3469–3478
- Barde, Y. A., Edgar, D., and Thoenen, H. (1982) Purification of a new neurotrophic factor from mammalian brain. *EMBO J.* **1**, 549–553
- Burnouf, T., Kuo, Y. P., Blum, D., Burnouf, S., and Su, C. Y. (2012) Human platelet concentrates: a source of solvent/detergent-treated highly enriched brain-derived neurotrophic factor. *Transfusion* **52**, 1721–1728
- Munkholm, K., Vinberg, M., and Kessing, L. V. (2016) Peripheral blood brain-derived neurotrophic factor in bipolar disorder: a comprehensive systematic review and meta-analysis. *Mol. Psychiatry* **21**, 216–228
- Szuhany, K. L., Bugatti, M., and Otto, M. W. (2015) A meta-analytic review of the effects of exercise on brain-derived neurotrophic factor. *J. Psychiatr. Res.* **60**, 56–64
- Dieni, S., Matsumoto, T., Dekkers, M., Rauskolb, S., Ionescu, M. S., Deogracias, R., Gundelfinger, E. D., Kojima, M., Nestel, S., Frotscher, M., and Barde, Y. A. (2012) BDNF and its pro-peptide are stored in presynaptic dense core vesicles in brain neurons. *J. Cell Biol.* **196**, 775–788
- Kolbeck, R., Bartke, I., Eberle, W., and Barde, Y. A. (1999) Brain-derived neurotrophic factor levels in the nervous system of wild-type and neurotrophin gene mutant mice. *J. Neurochem.* **72**, 1930–1938
- Kwok, S. C., Chakraborty, D., Soares, M. J., and Dai, G. (2013) Relative expression of proprotein convertases in rat ovaries during pregnancy. *J. Ovarian Res.* **6**, 91
- Radka, S. F., Holst, P. A., Fritsche, M., and Altar, C. A. (1996) Presence of brain-derived neurotrophic factor in brain and human and rat but not mouse serum detected by a sensitive and specific immunoassay. *Brain Res.* **709**, 122–301
- Zeiler, M., Moser, M., and Mann, M. (2014) Copy number analysis of the murine platelet proteome spanning the complete abundance range. *Mol. Cell. Proteomics* **13**, 3435–3445
- Koppel, I., Tuvikene, J., Lekk, I., and Timmusk, T. (2015) Efficient use of a translation start codon in BDNF exon I. *J. Neurochem.* **134**, 1015–1025
- West, A. E., Pruunsild, P., and Timmusk, T. (2014) Neurotrophins: transcription and translation. *Handb. Exp. Pharmacol.* **220**, 67–100
- Cheng, K. T., Ong, H. L., Liu, X., and Ambudkar, I. S. (2011) Contribution of TRPC1 and Orai1 to Ca^{2+} entry activated by store depletion. *Adv. Exp. Med. Biol.* **704**, 435–449
- Kolbeck, R., Jungbluth, S., and Barde, Y. A. (1994) Characterisation of neurotrophin dimers and monomers. *Eur. J. Biochem.* **225**, 995–1003
- Wetsel, W. C., Rodriguiz, R. M., Guillemot, J., Rousselet, E., Essalmani, R., Kim, I. H., Bryant, J. C., Marcinkiewicz, J., Desjardins, R., Day, R., Constam, D. B., Prat, A., and Seidah, N. G. (2013) Disruption of the expression of the proprotein convertase PC7 reduces BDNF production and affects learning and memory in mice. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17362–17367
- Maynard, D. M., Heijnen, H. F., Horne, M. K., White, J. G., and Gahl, W. A. (2007) Proteomic analysis of platelet α -granules using mass spectrometry. *J. Thromb. Haemost.* **5**, 1945–1955
- Kerschensteiner, M., Gallmeier, E., Behrens, L., Leal, V. V., Misgeld, T., Klinkert, W. E., Kolbeck, R., Hoppe, E., Oropeza-Wekerle, R. L., Bartke, I., Stadelmann, C., Lassmann, H., Wekerle, H., and Hohlfield, R. (1999) Activated human T cells, B cells, and monocytes produce brain-derived neurotrophic factor *in vitro* and in inflammatory brain lesions: a neuroprotective role of inflammation? *J. Exp. Med.* **189**, 865–870
- Snappan, M., Lemasson, M., Brill, M. S., Blais, M., Massouh, M., Ninkovic, J., Gravel, C., Berthod, F., Götz, M., Barker, P. A., Parent, A., and Saghatelian, A. (2009) Vasculature guides migrating neuronal precursors in the adult mammalian forebrain via brain-derived neurotrophic factor signaling. *J. Neurosci.* **29**, 4172–4188
- Heijnen, H. F., Schiel, A. E., Fijnheer, R., Geuze, H. J., and Sixma, J. J. (1999) Activated platelets release two types of membrane vesicles: microvesicles

- by surface shedding and exosomes derived from exocytosis of multivesicular bodies and α -granules. *Blood* **94**, 3791–3799
29. Hillman, E. M. (2014) Coupling mechanism and significance of the BOLD signal: a status report. *Annu. Rev. Neurosci.* **37**, 161–181
30. Aatonen, M. T., Ohman, T., Nyman, T. A., Laitinen, S., Gronholm, M., and Siljander, P. R. (2014) Isolation and characterization of platelet-derived extracellular vesicles. *J. Extracell. Vesicles* **3**, 10.3402/jev.v3.24692
31. Cohen, S. (2008) Origins of growth factors: NGF and EGF. *J. Biol. Chem.* **283**, 33793–33797
32. Polyakova, M., Stuke, K., Schuemberg, K., Mueller, K., Schoenknecht, P., and Schroeter, M. L. (2015) BDNF as a biomarker for successful treatment of mood disorders: a systematic & quantitative meta-analysis. *J. Affect. Disord.* **174**, 432–440
33. Pardridge, W. M., Kang, Y. S., and Buciak, J. L. (1994) Transport of human recombinant brain-derived neurotrophic factor (BDNF) through the rat blood-brain barrier *in vivo* using vector-mediated peptide drug delivery. *Pharm. Res.* **11**, 738–746
34. Kestin, A. S., Ellis, P. A., Barnard, M. R., Errichetti, A., Rosner, B. A., and Michelson, A. D. (1993) Effect of strenuous exercise on platelet activation state and reactivity. *Circulation* **88**, 1502–1511
35. Day, R. B., and Link, D. C. (2014) Megakaryocytes in the hematopoietic stem cell niche. *Nat. Med.* **20**, 1233–1234
36. Méndez-Ferrer, S., Michurina, T. V., Ferraro, F., Mazloom, A. R., Macarthur, B. D., Lira, S. A., Scadden, D. T., Ma'ayan, A., Enikolopov, G. N., and Frenette, P. S. (2010) Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* **466**, 829–834

Structural Basis of Host Autophagy-related Protein 8 (ATG8) Binding by the Irish Potato Famine Pathogen Effector Protein PexRD54^{*[S]♦}

Received for publication, June 24, 2016, and in revised form, July 18, 2016. Published, JBC Papers in Press, July 25, 2016, DOI 10.1074/jbc.M116.744995

 Abbas Maqbool^{‡1},  Richard K. Hughes^{‡1},  Yasin F. Dagdas[§],  Nicholas Tregidgo[‡],  Erin Zess[§],  Khaoula Belhaj[§],
 Adam Round^{¶||},  Tolga O. Bozkurt^{**},  Sophien Kamoun[§], and  Mark J. Banfield^{‡2}

From the [‡]Department of Biological Chemistry, John Innes Centre, and the [§]Sainsbury Laboratory, Norwich Research Park, Norwich NR4 7UH, United Kingdom, the ^{**}Department of Life Sciences, Imperial College, London SW7 2AZ, United Kingdom, [¶]EMBL Grenoble, 71 Avenue des Martyrs, CS 90181, 38042 Grenoble Cedex 9, France, and ^{||}EPSAM, Keele University, Keele ST5 5GB, United Kingdom

Filamentous plant pathogens deliver effector proteins to host cells to promote infection. The *Phytophthora infestans* RXLR-type effector PexRD54 binds potato ATG8 via its ATG8 family-interacting motif (AIM) and perturbs host-selective autophagy. However, the structural basis of this interaction remains unknown. Here, we define the crystal structure of PexRD54, which includes a modular architecture, including five tandem repeat domains, with the AIM sequence presented at the disordered C terminus. To determine the interface between PexRD54 and ATG8, we solved the crystal structure of potato ATG8CL in complex with a peptide comprising the effector's AIM sequence, and we established a model of the full-length PexRD54-ATG8CL complex using small angle x-ray scattering. Structure-informed deletion of the PexRD54 tandem domains reveals retention of ATG8CL binding *in vitro* and *in planta*. This study offers new insights into structure/function relationships of oomycete RXLR effectors and how these proteins engage with host cell targets to promote disease.

During selective autophagy, specific cellular constituents can be targeted to autophagic pathways for subcellular trafficking or degradation (1–3). The autophagy toolkit includes around 40 ATG (autophagy-related) proteins. Together, they help initiate, regulate, and form the constituents of autophagic pathways. The role of selective autophagy in the response to pathogen challenge in animal cells is increasingly being appreciated and includes direct elimination of microorganisms and control of immunity-related signaling (4, 5). In turn, microorganisms have developed mechanisms to perturb host-selective autophagy to either shut it down and promote infection (4, 5) or activate it and re-direct

nutrients to the parasite (6). There is also evidence that membrane formation and trafficking, as controlled by ATG proteins, are exploited by numerous viruses (7). To date, the role of host-selective autophagy in host-microbe interactions has mostly been studied in mammals. The role of host-selective autophagy in plant-microbe interactions, and how it is manipulated by plant pathogens, remains poorly understood.

ATG8 is a ubiquitin-like protein that performs multiple functions in autophagy. It is cycled, via conjugation and deconjugation reactions, to the membrane lipid phosphatidylethanolamine, and this localization is important for autophagosome biogenesis (8). The intracellular animal pathogen *Legionella pneumophila* targets this process by delivering type IV secreted effector protein RavZ, which irreversibly deconjugates ATG8 from membranes and restricts autophagy (9). ATG8 also functions as an adaptor to interact with proteins containing an ATG8-interacting motif (AIM).³ AIM-containing proteins can serve as receptors for cargo destined for autophagosomes. The core AIM sequence is defined as $\Omega XX\Psi$, where Ω is an aromatic amino acid (Trp, Tyr, or Phe); X is any residue, and Ψ is an aliphatic amino acid (Leu, Ile, and Val) (10–12). Frequently, residues just to the N terminus of the $\Omega XX\Psi$ motif are acidic in nature. Structural studies have elucidated how the AIM sequence binds ATG8, with key features including the Ω and Ψ residues binding within hydrophobic pockets, and the motif adopting a β -strand structure that extends the β -sheet of ATG8 (1, 13–15). It is generally thought that AIMS adopt a disordered or flexible conformation in the absence of a binding partner (11, 16). Mechanisms for pathogens to perturb host-selective autophagy include delivery of factors that interfere with recruitment of endogenous AIM-containing proteins to ATG8 or that re-direct additional cellular components to autophagosomes.

Filamentous plant pathogens cause devastating diseases of crops that are of both historical significance (17) and relevant to global agriculture today (18). *Phytophthora infestans*, the Irish potato famine pathogen, facilitates disease on its hosts by delivering effector proteins that modulate host cell processes to the

^{*} This work was supported by European Research Council Proposal “NGRB,” Biotechnology and Biological Sciences Research Council Grants BB/J00453 and BB/M002462, the John Innes Foundation, and the Gatsby Charitable Foundation. The authors declare that they have no conflicts of interest with the contents of this article.

[✂] Author's Choice—Final version free via Creative Commons CC-BY license.

[♦] This article was selected as a Paper of the Week.

^[S] This article contains supplemental Video 1.

The atomic coordinates and structure factors (codes 5L7S and 5L83) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

¹ Both authors contributed equally to this work.

² To whom correspondence should be addressed. Tel.: 44-1603-450742; Fax: 44-1603-450018; E-mail: mark.banfield@jic.ac.uk.

³ The abbreviations used are: AIM, ATG8-family interacting motif; SPR, surface plasmon resonance; NTA, nitrilotriacetic acid; RFP, red fluorescent protein; ITC, isothermal titration calorimetry; SAXS, small angle x-ray scattering; NSD, normalized spatial discrepancy.

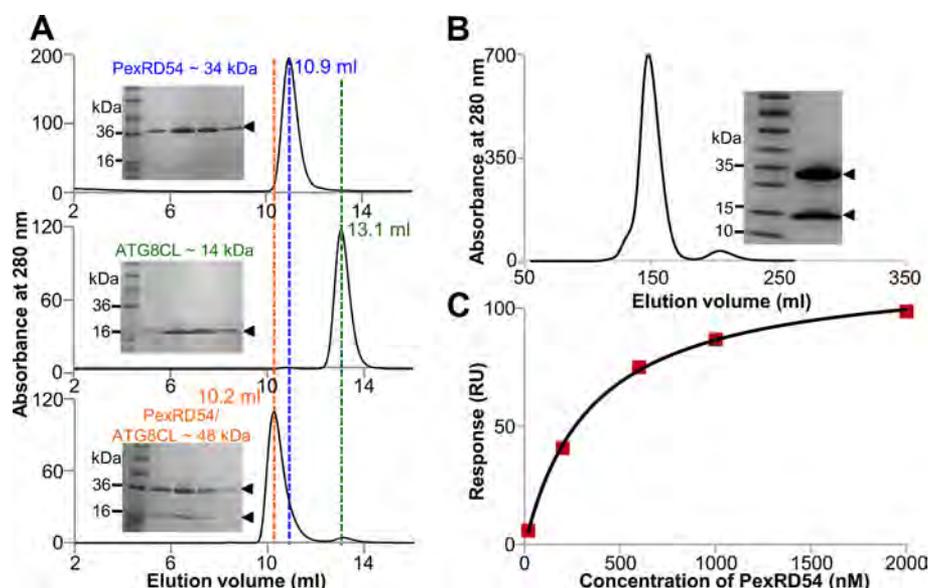


FIGURE 1. **Interaction of PexRD54 and ATG8CL proteins *in vitro*.** *A*, analytical gel filtration traces obtained for PexRD54 (top), ATG8CL (middle), and a 1:1 mixture of the complex (bottom). Insets show SDS-polyacrylamide gels of the fractions collected across the elution peaks. *B*, gel filtration trace derived from preparative purification of the PexRD54-ATG8CL complex following co-expression in *E. coli*. Inset, SDS-polyacrylamide gel containing purified complex. *C*, binding curve derived from SPR single cycle kinetics data for PexRD54 binding to ATG8CL.

benefit of the parasite (19), a strategy used by many biotrophic plant pathogens (20–22). Many putative *P. infestans* effectors contain a conserved N-terminal RXLR (Arg-Xaa-Leu-Arg) motif for host translocation (23). Furthermore, about half of these effectors are predicted to adopt the conserved WY domain fold in their C-terminal regions, which encodes their biochemical activity (24–26). Although recent studies have begun to elucidate the virulence-associated targets and functions of *P. infestans* RXLR effectors (27–34), these have yet to be identified for the vast majority of these proteins.

Recently, a *P. infestans* RXLR effector, PexRD54, which contains an AIM sequence Trp-Glu-Ile-Val “WEIV” positioned at the C terminus (residues 378–381), was identified (35). It was shown that PexRD54 specifically interacts with a member of the ATG8 family of proteins from potato, ATG8CL, *in vitro* and *in planta*. In plant cells, PexRD54 activates selective autophagy by increasing the number of ATG8CL-containing autophagosomes and stabilizing ATG8CL. Furthermore, PexRD54 was shown to antagonize the function of the host autophagy cargo receptor Joka2 by competing for binding with ATG8CL. As Joka2 contributed toward immunity against *P. infestans*, which was counteracted by PexRD54, it was concluded that this effector acts as an inhibitor of Joka2 function.

To better understand how PexRD54 interacts with potato ATG8CL to perturb host-selective autophagy, we have investigated the structural basis of effector-host target interaction. We determined the crystal structures of PexRD54 and ATG8CL in complex with the C-terminal AIM peptide of this effector. We also obtained a structure of the PexRD54-ATG8CL complex by docking the crystal structures into an envelope derived from solution scattering data. Site-directed mutagenesis of the PexRD54 C-terminal AIM region, and ATG8CL binding to a PexRD54 AIM-based peptide array, mapped the key residues that define the PexRD54-ATG8CL interface. Finally, we used structure-informed deletions to show that the WY domains of

PexRD54 are dispensable for ATG8CL binding suggesting an alternative function for these domains. Together, these data provide a mechanistic understanding of how translocated effectors engage with their host targets and offer new methods for engineering control of plant diseases.

Results

PexRD54 Forms a Stable Complex with ATG8CL in Vitro—To investigate complex formation between PexRD54 and ATG8CL, we expressed both proteins separately in *Escherichia coli* and purified them to homogeneity (Fig. 1A). To determine whether the two proteins form a stable complex in solution, we mixed them in an equimolar ratio prior to injection on a Superdex S75 10/300 analytical gel filtration column and compared the resulting elution volume to the elution volumes of the individual proteins. As shown in Fig. 1A, PexRD54 elutes at 10.9 ml and ATG8CL at 13.1 ml when these proteins are run independently. After mixing, a new peak at an earlier elution volume (10.2 ml) is apparent, and SDS-PAGE analysis shows this peak contains both proteins. This shift in the elution peak is indicative of complex formation and that this complex is stable over the time course of the experiment. Based on a calibration curve, elution volumes from this column of 10.9, 13.1, and 10.2 ml correspond to ~44, ~18, and ~58 kDa. All these represent overestimates of the predicted molecular masses of the proteins on their own or in complex (PexRD54 ~34 kDa, ATG8CL ~15 kDa, and PexRD54-ATG8CL complex ~49 kDa) but indicate monomeric forms of each state exist in solution.

Next, we determined whether the PexRD54-ATG8CL complex could be formed on purification following co-expression in *E. coli*. We cloned PexRD54 and ATG8CL into different expression vectors, with only the ATG8CL containing a His₆ tag (see under “Experimental Procedures”). Following expression and preparative tandem immobilized metal affinity chromatography/gel filtration chromatography of the clarified cell lysate, a

Structure/Function of PexRD54

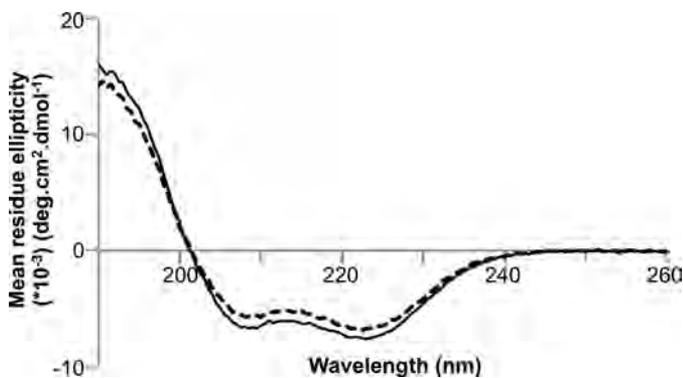


FIGURE 2. **CD spectra of PexRD54.** Far-UV CD spectra of wild-type PexRD54 (solid line) and its variant PexRD54^{378-AEIA-381} (dashed line) confirming similar secondary structure content (predominantly α -helical).

single peak was obtained at an elution volume consistent with a complex between PexRD54 and ATG8CL (Fig. 1B, an elution volume of 151 ml on this column corresponds to ~ 50 kDa, predicted molecular mass of the complex is ~ 49 kDa). SDS-PAGE analysis of the fractions confirmed the presence of both proteins (Fig. 1B). This shows that a complex between PexRD54 and ATG8CL is likely formed in cells and can be purified from cell culture directly.

Finally, we used surface plasmon resonance (SPR) to investigate the affinities of complex formation between PexRD54 and ATG8CL (Fig. 1C). Using this technique, we determined that PexRD54 binds to ATG8CL with a K_d of 388 ± 47 nM. The AIM motif disrupting PexRD54^{378-AEIA-381} variant (where the Trp and Val of the “WEIV” AIM motif are replaced by alanine) did not bind to ATG8CL using SPR, consistent with previous results (35). The overall fold of the PexRD54^{378-AEIA-381} variant was equivalent to wild-type protein as assessed by circular dichroism (CD) spectroscopy (Fig. 2).

PexRD54 Is a Tandem Repeat WY Domain Effector with a Disordered C-terminal AIM—To discover the molecular architecture of PexRD54, we determined the crystal structure of the effector domain of this protein (residues Val-92 to Val-381) at 2.90 Å resolution. Although PexRD54 could be crystallized alone, the crystal that gave rise to the best x-ray dataset was obtained from a sample including both PexRD54 and ATG8CL after co-expression in *E. coli* (see under “Experimental Procedures”). Although SDS-polyacrylamide gel analysis of dissolved crystals showed that both proteins were present in these crystals, no electron density for ATG8CL was observed. The structure of PexRD54 was solved using single wavelength anomalous diffraction, and the final model was refined to final R_{work} and R_{free} values of 23.1 and 25.6%, respectively (Table 1). Inspection of the packing of PexRD54 revealed that ATG8CL could be accommodated in the crystal, within a region of unaccounted for space near the C terminus of the effector. The structure of PexRD54 includes 16 α -helices (Fig. 3A and supplemental video 1). Five N-terminal residues (92–96), the residues in two loops (248–250 and 331–334), and 11 C-terminal residues (371–381), which include the AIM motif, were not included in the final model due to poor electron density in these regions.

Previous bioinformatics analysis predicted the presence of multiple WY domains in PexRD54 (24). Our structural analysis

revealed that PexRD54 includes five tandem WY domains that pack to form an elongated molecule (Fig. 3A). This is a conformation not yet observed for RXLR effectors with multiple WY domains. The WY domain is a conserved structural unit consisting of three α -helices and two characteristic hydrophobic amino acids, frequently W (Trp) and Y (Tyr), which contribute to a stable hydrophobic core (24, 25). Structural superposition of the archetypal WY domain of the *Phytophthora capsici* RXLR-WY effector AVR3a11 on each of the WY domains of PexRD54 is shown in Fig. 3B, with root mean square deviations derived from each superposition given in Table 2. As more structures are determined, it is increasingly clear that WY domains can tolerate variations at the Trp and Tyr positions, while maintaining the hydrophobic core and overall fold. This is in addition to the remarkable overall structural conservation among WY domains despite a lack of pairwise sequence identity, which is as low as 13% between PexRD54 and AVR3a11 (Table 2).

Host Protein ATG8CL Binds the PexRD54 AIM Sequence via Two Hydrophobic Pockets—In the PexRD54 structure, we did not observe the last 10 amino acids that contain the AIM motif, or the ATG8CL protein itself, in the electron density. Therefore, to visualize the interaction between PexRD54 and ATG8CL, we determined the crystal structure of ATG8CL in complex with a PexRD54 C-terminal pentapeptide. This pentapeptide includes the AIM motif, with residues Asp-377–Trp-378–Glu-379–Ile-380–Val-381. To produce crystals of ATG8CL + pentapeptide, we used an ATG8CL construct lacking four N-terminal residues and five C-terminal residues.

The structure of the complex was solved by molecular replacement and refined to 1.90 Å with final R_{work} and R_{free} values of 17.6 and 19.9%, respectively (Table 1). Positive difference electron density within the likely AIM binding region of ATG8CL indicated the presence of bound pentapeptide. The final model contains two molecules of ATG8CL + pentapeptide in the asymmetric unit. The electron density maps for both complexes were of equivalent quality, and subsequent analysis focuses on one representative monomer.

The structure of ATG8CL contains two domains, an N-terminal helical domain ($\alpha 1$ and $\alpha 2$) and a C-terminal domain that adopts a β -grasp (ubiquitin-like) fold of four β -strands ($\beta 1$ – $\beta 4$) flanked by two helices ($\alpha 3$ and $\alpha 4$) (Fig. 4A). ATG8CL adopts a very similar structure to that observed for ATG8s from other organisms. For example, ATG8CL overlays on the structures of GATE-16 (Protein Data Bank code 1EO6, 60% sequence identity with ATG8CL) and GABARAP (Protein Data Bank code 4XC2, 57% sequence identity with ATG8CL) with a root mean square deviation of 0.8 and 0.9 Å, respectively, for 115 α -carbons.

In the complex, the pentapeptide adopts an extended conformation forming a parallel β -sheet with $\beta 2$ of ATG8CL. The peptide binds within a narrow channel at the surface of ATG8CL via hydrophobic and hydrogen bond interactions (Fig. 4A). The side chain of PexRD54 Trp-378 is contained within a hydrophobic pocket formed at the interface between the β -grasp and N-terminal helical domains of ATG8CL, whereas the side chain of PexRD54 Val-381 binds a distinct hydrophobic pocket between $\beta 2$ and an adjacent helix on the

TABLE 1
PexRD54/ATG8CL x-ray data collection and refinement statistics

	PexRD54		
	Native	Iodide	ATG8CL native
Data collection statistics			
Wavelength (Å)	0.9795	2.0	0.9795
Space group	$P3_121$	$P3_121$	$I4_132$
Cell dimensions			
a, b, c (Å)	89.16, 89.16, 144.32	91.67, 91.67, 144.66	172.80, 172.80, 172.80
Resolution (Å) ^a	77.21–2.90 (2.90–2.98)	79.39–3.50 (3.50–3.59)	86.09–1.90 (1.90–1.95)
R_{merge} (%)	7.0 (134.9)	13.9 (116.8)	13.0 (132.5)
$I/\sigma I$	24.9 (2.9)	22.5 (3.9)	27.4 (3.4)
Completeness (%)			
Overall	99.8 (99.7)	99.9 (100)	100 (100)
Anomalous		99.9 (99.8)	
Unique reflections	15,256 (1132)	9319 (676)	34,386 (2623)
Redundancy			
Overall	12.1 (12.3)	31.6 (29.2)	32.8 (31.9)
Anomalous		16.8 (15.1)	
CC(1/2) (%) ^a	99.9 (79.6)	99.9 (91.3)	100 (86.4)
Refinement and model statistics			
Resolution (Å)	77.21–2.90 (2.98–2.90)		86.09–1.90 (1.95–1.90)
$R_{\text{work}}, R_{\text{free}}$ (%)	23.1/25.6 (40.5/32.5)		17.6/19.9 (24.2/25.3)
No. of atoms			
Protein	2224		235
B -Factors			
Protein	98.9		24.00
Root mean square deviations			
Bond lengths (Å)	0.007		0.011
Bond angles (°)	1.047		1.50
Ramachandran plot (%) ^b			
Favored	94.25		98.71
Allowed	5.75		1.29
Outliers	0		0
MolProbity Score	1.45 (100th percentile)		1.14 (100th percentile)

^a The highest resolution shell is shown in parentheses.^b Data are as calculated by MolProbity.

C-terminal domain of ATG8CL (Fig. 4A). In addition to hydrophobic interactions, the indole nitrogen of Trp-378 forms a hydrogen bond with the side chain of ATG8CL Glu-18 (Fig. 4A). The side chain of PexRD54 Glu-379 makes hydrogen bonds and ionic interactions with the side chains of ATG8CL Lys-47 and ATG8CL Arg-68 (Fig. 4A). Another prominent ionic interaction is formed between the side chain of PexRD54 Asp-377 and ATG8CL Lys-47 (Fig. 4A).

Molecular Envelope of the Full-length PexRD54 and ATG8CL Complex—Despite having determined the crystal structures of PexRD54 and of ATG8CL bound to the PexRD54 AIM motif pentapeptide, structural information on how the full-length proteins interact was still lacking. To gain insight into this, we collected solution x-ray scattering data (small angle x-ray scattering (SAXS)) of both PexRD54 alone and the PexRD54-ATG8CL complex following co-expression and purification as described previously.

Analysis of the solution scattering data (“Experimental Procedures”) revealed that the PexRD54 particle has a radius of gyration of 26.1 Å (from Guinier analysis) or 26.7 Å (from $P(r)$ function (Fig. 5A, left)), with a maximal dimension (D_{max}) of 92 Å. This compares well with the maximal dimension in the crystal structure of ~87 Å. The predicted molecular mass from the Porod-Debye analysis is 26–34 kDa, which is close to the mass determined by LC-MS (34.023 kDa). The PexRD54-ATG8CL complex particle has a radius of gyration of 32.6 Å (from Guinier analysis) or 34.1 Å (from $P(r)$ function (Fig. 5A, right)) with a D_{max} of 120 Å. The predicted molecular mass from the Porod-Debye analysis is 41–54 kDa, and the mass of the proteins in the complex as deter-

mined by LC-MS (48.694 kDa) fits well within this range. *Ab initio* shape reconstructions of the particles were generated, and the crystal structure of PexRD54 (for the PexRD54 data) was docked into its envelope (Figs. 5B, left, and 6, A and B). A complex between PexRD54 and ATG8CL + pentapeptide consistent with the scattering data was generated using CORAL (36) and subsequently docked into the appropriate envelope (Figs. 5B, right, and 6, A and C). The latter model provides a molecular snapshot of a *P. infestans* translocated effector protein bound to a host target.

Characterization of the PexRD54 AIM Region Binding to ATG8CL—To build on the structural studies above, we used two complementary biochemical approaches to investigate the role of individual residues in the AIM region of PexRD54 in binding to ATG8CL.

First, we used alanine-scanning mutagenesis to substitute Ala at six positions in the PexRD54 AIM region, Pro-373, Asp-377, Trp-378, Glu-379, Ile-380, and Val-381. Each of these proteins was expressed and purified as described for wild type. We then used analytical gel filtration to qualitatively assay whether these variants support complex formation with ATG8CL. As predicted, we did not observe interaction of PexRD54 W378A with ATG8CL (Fig. 7). For each of the other mutations, we still observed an interaction with ATG8CL, including PexRD54 V381A. Second, we designed a nitrocellulose-anchored peptide array of 200 variant AIM peptides, based on the final 10 amino acids of PexRD54, where each amino acid was changed to all other possible amino acids. The peptides were anchored at the N terminus to best mimic the presentation of the PexRD54 AIM region to ATG8CL. We visualized

Structure/Function of PexRD54

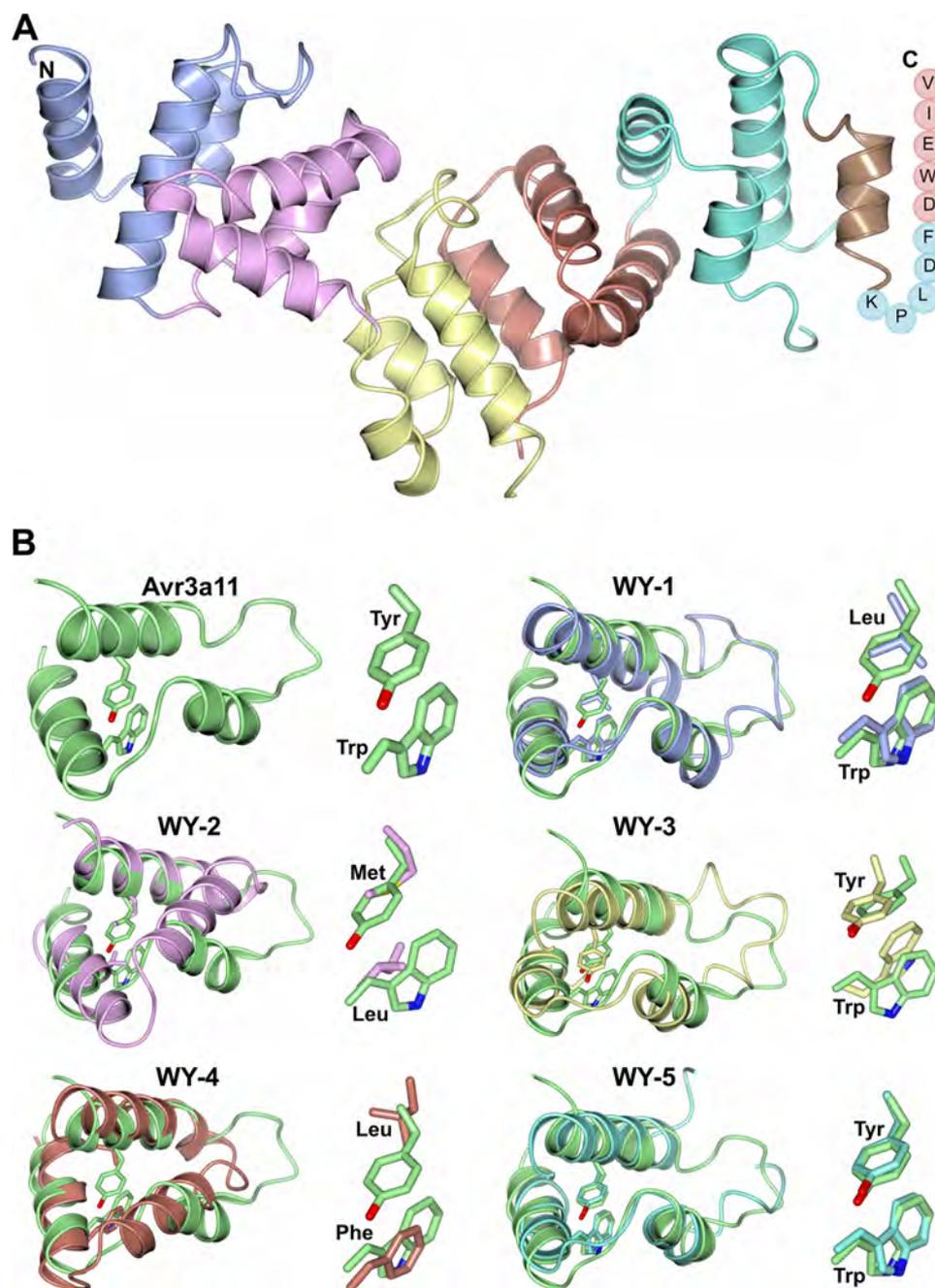


FIGURE 3. Crystal structure of PexRD54. *A*, schematic representation of the crystal structure of PexRD54 showing the five tandem WY domains (blue, magenta, yellow, coral, and cyan) and the disordered AIM motif at the C terminus (circles with single letter amino acid codes shown). The N and C termini are labeled. *B*, superimposition of the WY domains of AVR3a11 (top left, green) on the WY domains from PexRD54. The characteristic hydrophobic residues of each WY domain are also shown in stick representation. The PexRD54 WY domains are colored as in *A*.

TABLE 2

Root mean square deviations (r.m.s.d.) derived from the overlays shown in Fig. 3*B*, including the number of carbon atoms in the overlay, the identity of the “WY” amino acids, and percentage sequence identity to AVR3a11

	r.m.s.d.	Residue range	WY amino acids	Sequence identity to AVR3a11
	Å			%
WY-1	1.81/37	Ser-97–Gly-150	WL	13
WY-2	2.35/32	Asn-151–Gly-198	LM	18
WY-3	2.89/39	Asn-199–Asn-247	WY	16
WY-4	2.80/41	Phe-251–Ser-299	FL	14
WY-5	1.73/41	Ser-302–Ile-354	WY	20

ATG8CL binding to the peptide array using an ATG8CL fusion with glutathione *S*-transferase (GST) and a His tag (see “Experimental Procedures”), followed by incubation with an anti-GST-HRP antibody (Amersham Biosciences) and detection of chemiluminescence (Fig. 4*B*). The results of the peptide array clearly highlight the importance of the hydrophobic residues 378 and 381 of the PexRD54 AIM motif (Trp and Val) in binding ATG8CL. For position 378, the strongest binding was seen for Trp and Phe, with limited binding of Tyr and the aliphatic amino acids. Position 381

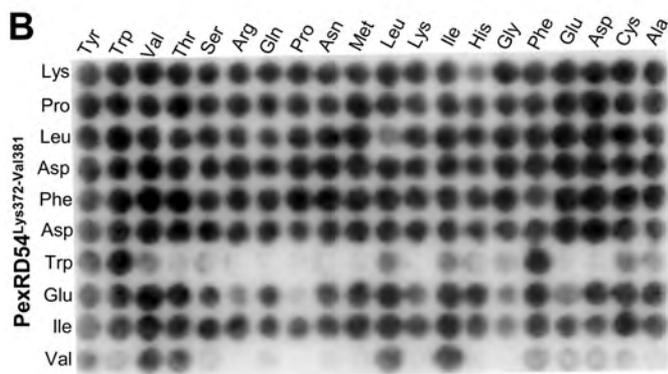
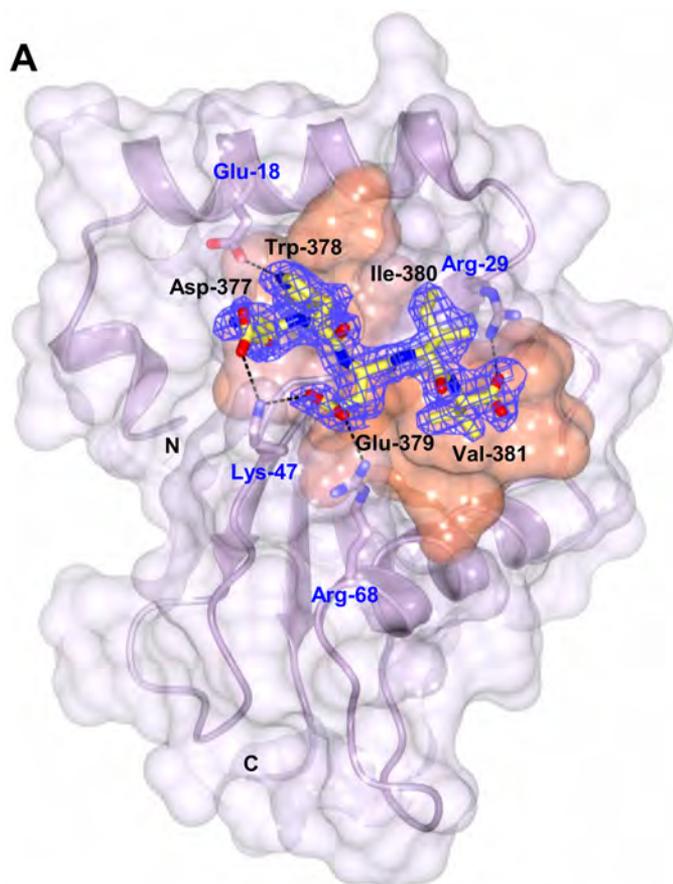


FIGURE 4. Crystal structure of ATG8CL bound to the PexRD54(377–381)-peptide and specificity of peptide binding. *A*, schematic representation of ATG8CL/PexRD54(377–381)-peptide complex highlighting key interactions. ATG8CL is shown in *magenta* schematic representation with the molecular surface that contacts the PexRD54(377–381)-peptide shown in *orange*. The PexRD54(377–381)-peptide is shown as *sticks* with *yellow* carbon atoms. The electron density omit map of the peptide ligand ($F_{\text{obs}} - F_{\text{calc}}$ map) is shown in *blue mesh* and contoured at 2σ . Electrostatic interactions are indicated with *black dashed lines*. *B*, results of the peptide array analyzing the effect of single amino acid substitutions (*top*) at all positions of 10-mer peptide of PexRD54 (Lys-372–Val-381, *side*). GST-tagged ATG8CL was visualized using an anti-GST-HRP antibody.

favors the bulky aliphatic amino acids, with limited binding also observed by bulky hydrophobic residues. Interestingly, with the exception of bulky hydrophobic residues. Interestingly, with the exception of bulky hydrophobic residues, any amino acid can be accommodated at positions 379 and 380, and binding is still observed. Furthermore, any amino acid can be accommodated at positions 372–377 without a significant reduction in binding, suggesting that these residues may only act

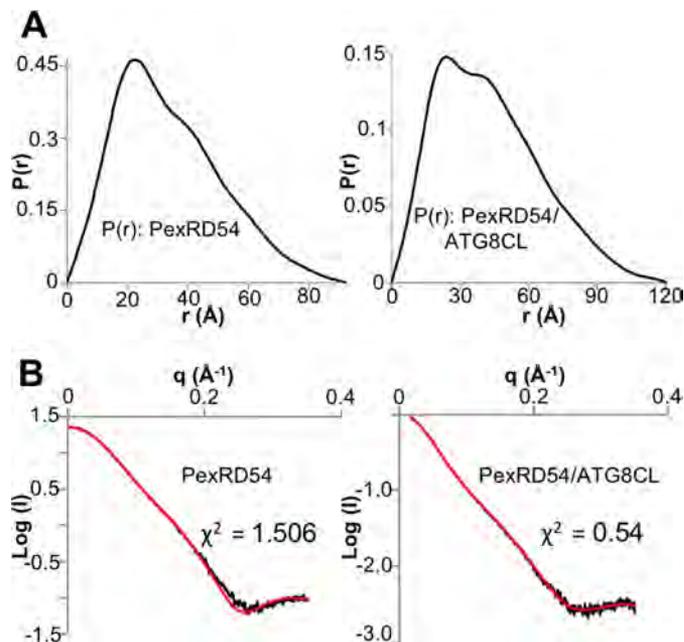


FIGURE 5. Analysis of SAXS data. *A*, $P(r)$ distribution curves used for *ab initio* modeling. *Left*, PexRD54; *right*, PexRD54-ATG8CL complex. D_{max} was set at 92 nm (PexRD54) and 120 nm (PexRD54/ATG8CL complex). Data were cropped at 0.35 \AA^{-1} for analysis. *B*, *left*, fit of the theoretical scattering curve of PexRD54 from CRYSOLOG to the PexRD54 scattering data (*black*). *Right*, fit of the theoretical scattering curve of the PexRD54-ATG8CL complex from CORAL (*red*) to the PexRD54-ATG8CL scattering data (*black*).

as a linker between the WY domain region of PexRD54 and the C-terminal AIM motif.

WY Domains of PexRD54 Are Dispensable for the Interaction with ATG8CL *In Vitro* and *In Planta*—Although the AIM region of PexRD54 appears necessary and sufficient for the interaction with ATG8CL, we explored whether the WY domains of PexRD54, which include $\sim 96\%$ of the protein expressed here, impact the binding of the effector to ATG8CL. For this, we produced two structure-informed deletions of PexRD54, removing either the first three WY domains (but leaving the C-terminal helix of WY-3, which forms an N-terminal extension of WY-4), generating PexRD54 $^{\Delta 218}$, or the first four WY domains (leaving only WY-5), producing PexRD54 $^{\Delta 298}$ (Fig. 8, *A* and *B*). These proteins were expressed and purified as for wild-type PexRD54 and confirmed to be predominantly α -helical by CD spectroscopy (Fig. 9). We used ITC to calculate the affinity of interaction for these constructs with ATG8CL, which gave a K_d of 69 nM for PexRD54 $^{\Delta 218}$ and a K_d of 39 nM for PexRD54 $^{\Delta 298}$ (Fig. 8, *A* and *B*). These values are broadly in line with the K_d of 383 nM obtained for the ATG8CL interaction with wild-type PexRD54 (35). The AIM motif disrupting PexRD54 $^{\Delta 218\Delta E1A}$ and PexRD54 $^{\Delta 298\Delta E1A}$ variants showed no binding to ATG8CL but retained a similar fold to PexRD54 $^{\Delta 218}$ and PexRD54 $^{\Delta 298}$ as judged by CD spectroscopy (Fig. 9).

We also tested whether the PexRD54 $^{\Delta 218}$ and PexRD54 $^{\Delta 298}$ deletions retained the ability to bind ATG8CL *in planta* by co-immunoprecipitation (co-IP) from *Nicotiana benthamiana* leaves transiently expressing these proteins following delivery of the genes by infiltration with *Agrobacterium tumefaciens* (agroinfiltration). In these assays both RFP-PexRD54 deletion mutants still interacted with ATG8CL (Fig. 8C). Full-length

Structure/Function of PexRD54

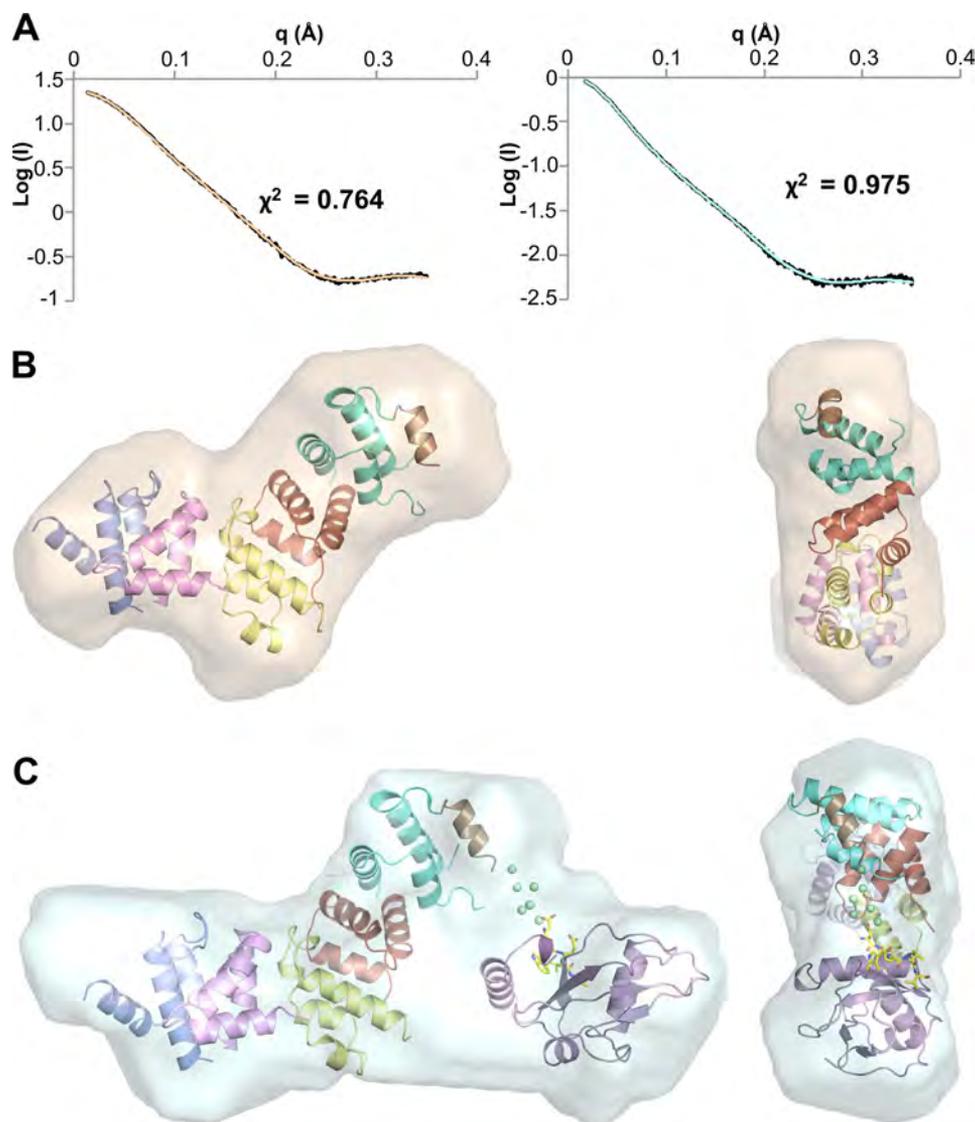


FIGURE 6. **PexRD54 and PexRD54-ATG8CL complex analyzed by small angle x-ray scattering.** *A*, fits of the most probable (lowest NSD) dummy atom models from DAMMIN for PexRD54 (*left*) and PexRD54/ATG8CL (*right*). The fit to the experimental data (in *black*) is shown in *wheat* and *cyan*, respectively, with χ^2 shown as an *inset*. *B*, superposition of the crystal structure of PexRD54 with the most probable *ab initio* envelope of PexRD54 (*wheat surface*). *C*, superposition of the CORAL rigid body model of PexRD54/ATG8CL + pentapeptide with the most probable *ab initio* envelope of the complex (*cyan surface*). For *B* and *C*, two views are shown, face-on (*left*) and end-on (*right*). The fits shown in *A* and the envelopes shown in *B* and *C* are from the same run of DAMMIN.

RFP-PexRD54 and the AIM motif disrupting variant RFP-PexRD54^{378-AEIA-381} were used as controls.

Discussion

Understanding the mechanistic basis of translocated effector protein function in support of pathogen infection and colonization is a major focus of research in plant-microbe interactions. Such studies reveal how manipulation of host cell processes by pathogen-derived molecules can promote virulence and also identify plant systems, such as autophagy, whose importance in disease or general host cell physiology may be underappreciated. In a few cases, the structural basis for bacterial plant pathogen effector interaction with a host protein or peptide has been described (37–40). However, such studies of filamentous plant pathogen effectors are lacking. The *P. infestans* RXLR-type effector PexRD54 (PITG_09316) perturbs host-selective autophagy for the benefit of the pathogen via

interaction with ATG8CL (35). Here, we focused on the biochemical and structural basis of PexRD54's interaction with ATG8CL to understand how the pathogen co-opts autophagic pathways.

Structural conservation in RXLR-type effectors from the oomycetes, in the absence of confidently assignable sequence similarity, has previously been established (24, 25). Although each of the five structurally conserved three-helical bundle (WY domain) repeats in PexRD54 adopts the same overall fold, they pack together to form a unique structure different from that of the two WY domain repeat effector ATR1 from *Hyaloperonospora arabidopsidis* (41). Detailed analysis of the PexRD54 structure suggests trajectories for the evolution of WY domain proteins through gain or loss of functional units presented on the N or C terminus of the core three-helical bundle. First, the minimal three helix WY domain fold seen in

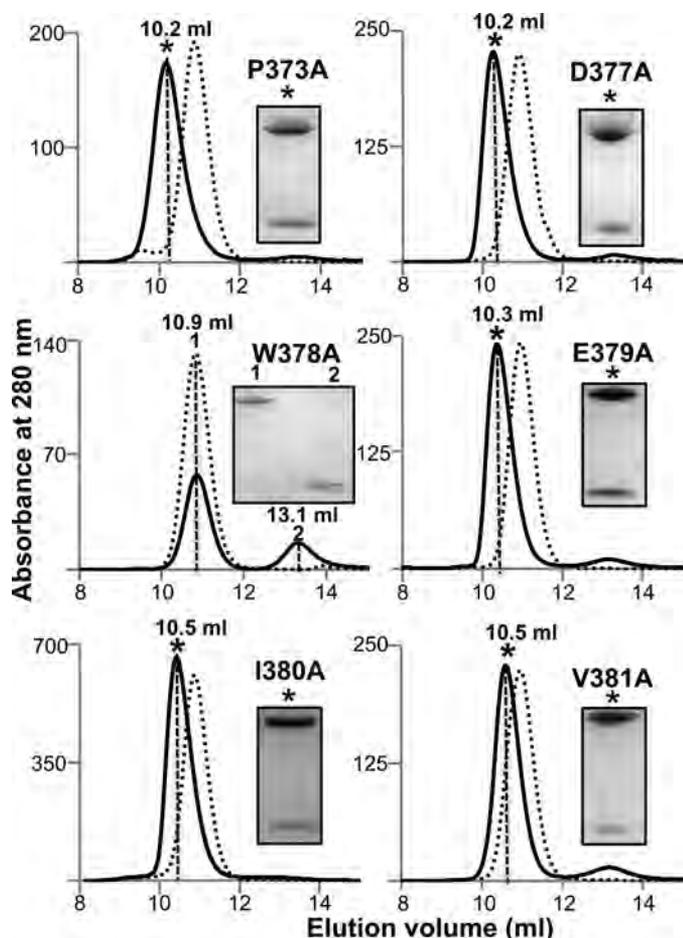


FIGURE 7. Analysis of the interaction between PexRD54 variants and ATG8CL by gel filtration. Analytical gel filtration traces were obtained for PexRD54 variants mutated in the AIM region and incubated with ATG8CL (1:1 mixture). Insets show SDS-polyacrylamide gels of the fractions at the elution peaks as marked by the dashed lines.

PexRD54 is found in *P. infestans* effector PexRD2 (24), but in other RXLR-type effectors of known structure an N-terminal helix is present resulting in a four-helical bundle. Interestingly, in PexRD54, the C-terminal helices of WY-1, WY-3, and WY-4 are positioned such that they also serve as N-terminal helical extensions to WY-2, WY-4, and WY-5 to build four-helical bundles as observed in AVR3a4 (42), AVR3a11, and ATR1. Second, in ATR1 the tandem repeats of the four helix bundle are separated by a fifth “linker” helix. When the first WY domain of ATR1 is overlaid on WY-5 of PexRD54, the fifth linker helix is positioned on the final helix of PexRD54 (brown in Fig. 3A). In both protein structures, this helix then serves to present the proximal regions, either a second WY domain as seen in ATR1 or the AIM region as seen in PexRD54. Finally, PexRD54:WY-3 does not have an N-terminal helix and does not form a four helical bundle. This correlates with a significant kink in the PexRD54 structure between WY-2 and WY-3. Each of these observations serves to highlight the plasticity of the WY-fold and how it can be utilized to deliver new template structures with the potential for functional diversification. It is interesting to note that conserved structure in the absence of confidently assignable sequence similarity is emerging as a recurring theme for filamentous plant pathogen effectors (43, 44).

Little is known about how plant autophagic pathways are controlled and manipulated by pathogens. The structure of ATG8CL bound to the PexRD54 AIM peptide revealed the fundamental mechanisms of AIM recognition by plant ATG8s are similar to those seen in other organisms. The two critical hydrophobic residues of the Ω XX Ψ motif, Trp and Val in PexRD54, are bound in two hydrophobic pockets on the surface of ATG8CL (Fig. 4A). Furthermore, our mutagenesis and peptide-binding studies confirm the important roles for these residues in the interaction. The identity of the residues to the N terminus of the AIM, which in other systems comprise acidic residues (11), do not seem to be important in this case. Previously, it was shown that the binding of PexRD54 to another ATG8 family member, ATG8IL, was weaker *in planta* and *in vitro*. These two proteins share 50% sequence identity. Interestingly, three amino acids are changed between ATG8CL and ATG8IL at the ATG8CL/PexRD54 AIM peptide interface: I33V, L56M, and V164I. ATG8CL Ile-33 is located at the base of the pocket that binds PexRD54 Trp-378, whereas ATG8CL Leu-56 and ATG8CL Val-64 are both located in the second hydrophobic pocket that faces PexRD54 Val-381. The interactions between ATG8s and AIM peptides are dominated by hydrophobic interactions, and the subtle changes delivered by these mutations may be responsible for the weaker binding affinity of ATG8IL over ATG8CL, although this remains to be tested *in vitro* and will be the subject of future work.

The previous study (35) and the work described here reveal the importance of the interaction between PexRD54 and ATG8CL, as mediated by the effector’s C-terminal AIM region. This region includes only ~3% of the amino acids downstream of the RXLR-dEER motif, but deletion of WY domains 1–4 does not significantly affect ATG8CL binding *in vitro* or *in planta*. This raises the following question. How do the five WY domains contribute to PexRD54 function? This effector has been shown to stimulate host autophagosome formation, and it was hypothesized that the pathogen exploits this for its own benefit in either promoting nutrient recycling or counteracting defense. Future work will address how the PexRD54 WY domains may contribute to autophagosome formation and/or act as a receptor to localize specific cellular cargo to autophagic pathways.

Experimental Procedures

Gene Cloning

All constructs were verified by DNA sequencing.

PexRD54—For protein expression in *E. coli*, DNA encoding PexRD54 residues Val-92 to Val-381 was amplified from RFP-PexRD54 (35) and cloned into pOPINA or pOPINS3C (45) by In-Fusion cloning (Clontech). The resultant vectors expressed PexRD54 protein without a fusion tag (pOPINA) or with the N-terminal His₆-SUMO tag (pOPINS3C), respectively. DNA encoding PexRD54 residues Arg-219 to Val-381 was amplified from pOPINA-PexRD54 and cloned into pOPINS3C. DNA encoding PexRD54 residues Ser-299 to Val-381 was amplified from pOPINA-PexRD54 (and cloned into pOPINS3C) or from pOPINS3C-PexRD54 (and cloned into pOPINA). Single point mutants within the AIM region of PexRD54 were encoded

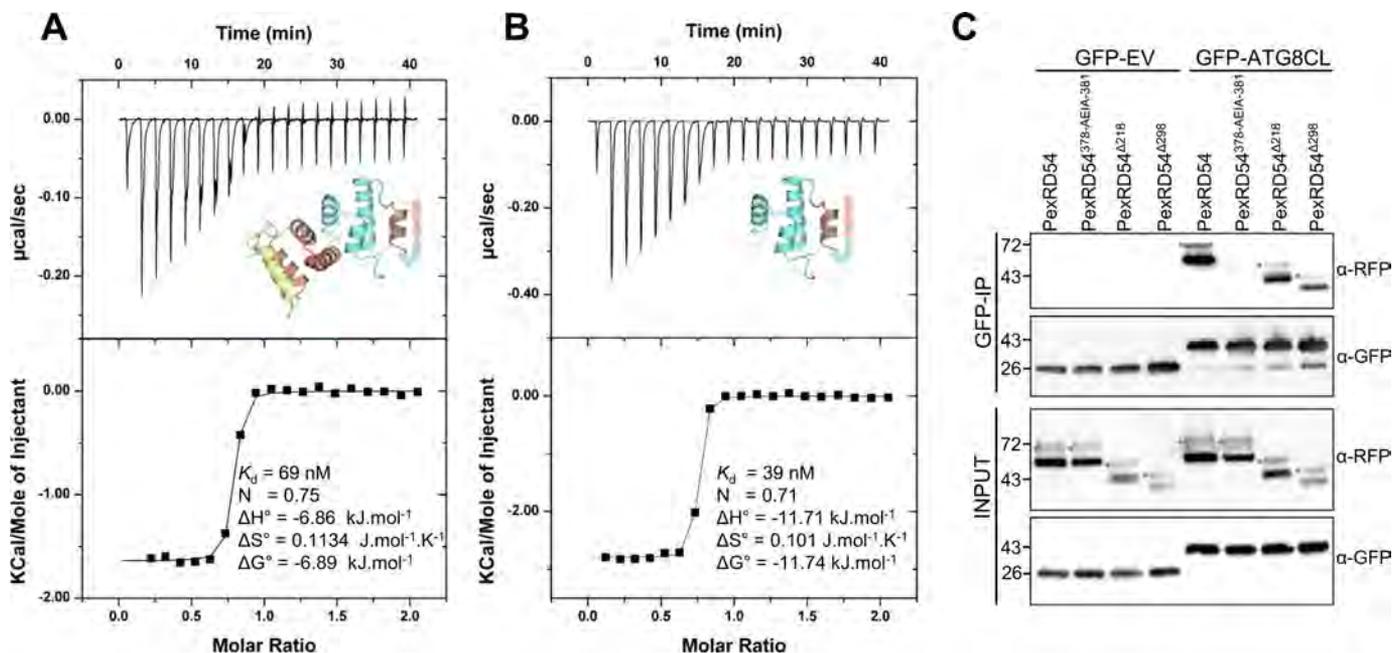


FIGURE 8. Interaction of PexRD54 Δ^{218} and PexRD54 Δ^{298} with ATG8CL *in vitro* and *in planta*. The binding affinities of PexRD54 Δ^{218} (A) and PexRD54 Δ^{298} (B) to ATG8CL were determined by ITC. Following a heats-of-dilution correction, a single-site binding model was used to fit the data using the MicroCal Origin software (data are shown on the top, with the fit on the bottom). The insets in the top panel depict the PexRD54 truncation used in the experiment, colored as in Fig. 3A. C, validation of PexRD54 Δ^{218} and PexRD54 Δ^{298} interaction with ATG8CL in plant cells by co-immunoprecipitation. Red asterisks indicate expected band sizes of the PexRD54 constructs. Degradation is due to autophagy, as seen previously (35).

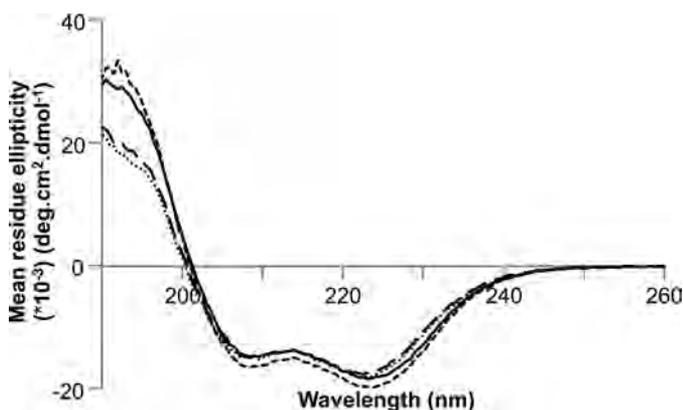


FIGURE 9. CD spectra of truncated PexRD54 constructs. Far-UV CD spectra of PexRD54 Δ^{218} (solid line), PexRD54 Δ^{298} (long dash line), PexRD54 Δ^{218} AEIA (short dashed line), and PexRD54 Δ^{298} AEIA (dotted line) variants confirming a similar secondary structure composition (predominantly α -helical).

within primers that were then used to amplify the full-length construct from pOPINS3C-PexRD54 followed by ligation into pOPINS3C. For protein expression *in planta*, DNA encoding PexRD54 residues Arg-219 to Val-381 or Ser-299 to Val-381 were amplified from RFP-PexRD54 and cloned into pENTR (ThermoFisher, UK). The expression constructs RFP-PexRD54 Δ^{218} and RFP-PexRD54 Δ^{298} were generated by Gateway LR reaction (Invitrogen) using the destination vector pH7WGR2 (N-terminal RFP fusion).

ATG8CL—For protein expression in *E. coli*, DNA encoding Met-1 to Phe-119 of ATG8CL was amplified from pOPINF-ATG8CL (35) and cloned into pOPINE (45), producing ATG8CL with a non-cleavable C-terminal His₆ tag. DNA encoding Ser-5 to Asn-114 of ATG8CL was amplified from pOPINF-ATG8CL and cloned into pOPINF, expressing

ATG8CL with a cleavable N-terminal His₆ tag (called ATG8CL* hereafter). For probing the peptide array, DNA encoding ATG8CL residues Met-1 to Phe-119 was amplified from pOPINE-ATG8CL and cloned into pOG3182 (Oxford Genetics). DNA encoding the ATG8CL-GST fusion was amplified from ATG8CL-pOG3182 and cloned into pOPINE. The resultant pOPINE-ATG8CL-GST vector expressed ATG8CL protein with a non-cleavable C-terminal GST-His₆ tag. For protein expression *in planta*, GFP-EV and GFP-ATG8CL constructs were described previously (35).

Heterologous Protein Production and Purification

Purified proteins were concentrated and stored in 20 mM HEPES buffer, pH 7.5, containing 150 mM NaCl, except where stated.

PexRD54 and Its Variants—For analytical gel filtration and ITC, all PexRD54 proteins were produced using *E. coli* BL21-arabinose-inducible cells and purified as described previously (35). For SPR, the same purification protocol was followed, with the exception of the final gel filtration step, which used 20 mM HEPES, pH 7.5, 500 mM NaCl.

ATG8CL—ATG8CL, expressed from pOPINF, was produced in *E. coli* BL21(DE3) and purified as described previously (35). When produced from pOPINE, a single Ni²⁺-NTA capture step followed by gel filtration produced soluble protein. The same strategy was used for purifying pOPINE-ATG8CL-GST-His. For SPR, ATG8CL was purified using 20 mM HEPES, pH 7.5, 500 mM NaCl in the gel filtration step. For crystallization, pOPINF-ATG8CL* was expressed and purified as for pOPINF-ATG8CL, except auto-induction media were used to culture the *E. coli*.

PexRD54-ATG8CL Complex—For crystallization and SAXS analysis of the complex, pOPINA-PexRD54 and pOPINE-

ATG8CL were co-transformed and expressed in BL21(DE3). Purification used the same protocol as for ATG8CL produced from pOPINE.

Protein-Protein Interaction Studies

Analytical Gel Filtration—Analytical gel filtration chromatography was performed at 4 °C using a Superdex 75 10/300 column (GE Healthcare) pre-equilibrated in 20 mM HEPES, pH 7.5, 150 mM NaCl. 100 μ l of sample was injected at a flow rate of 0.8 ml/min, and 0.5-ml fractions were collected for analysis. To study complex formation, proteins were mixed and incubated on ice for at least 1 h prior to loading.

Surface Plasmon Resonance—SPR experiments were performed at 18 °C using a BIAcore T200 system (GE Healthcare) and an NTA sensor chip (GE Healthcare). Protein samples were prepared in 20 mM HEPES, pH 7.5, 500 mM NaCl, and all the measurements were recorded in the same buffer at a flow rate of 30 μ l/min. A single cycle kinetics approach was used to study the interaction between PexRD54 and ATG8CL. The NTA chip was activated by injecting 10 μ l of 0.5 mM NiCl₂ over flow cell 2, which was also used to immobilize His-tagged ATG8CL to a response level of 85 ± 2 . Increasing concentrations of PexRD54 (20, 200, 600, 1000, and 2000 nM) were injected over flow cell 1 and 2 for 90 s. After the final injection, the dissociation was recorded for 300 s. Two startup cycles were run where the chip was activated and ATG8CL immobilized in the same manner, but buffer only was injected instead of PexRD54. This was subtracted to account for any dissociation of ATG8CL from the sensor chip. The sensor chip was regenerated by injecting 10 μ l of 350 mM EDTA. The data were analyzed using BIAcore T200 BIAevaluation software (GE Healthcare) and then plotted with Microsoft Excel.

Isothermal Titration Calorimetry—Calorimetry experiments were recorded at 15 °C in 20 mM HEPES, pH 7.5, 150 mM NaCl, using an iTC200 instrument (MicroCal Inc.). The calorimetric cell was filled with 80 μ M PexRD54 truncation (PexRD54 ^{Δ 218} or PexRD54 ^{Δ 298}) and titrated with 0.8 mM ATG8CL from the syringe. A single injection of 0.5 μ l of ATG8CL was followed by 19 injections of 2 μ l each. Injections were made at 120-s intervals with a stirring speed of 750 rpm. The raw titration data were integrated and fitted to a one-site binding model using the MicroCal Origin software.

In Planta Co-immunoprecipitation—3–4-week-old *N. benthamiana* plants were used for transient expression experiments. T-DNA expression vectors encoding PexRD54 constructs, ATG8CL constructs, or empty vector were transformed into the *A. tumefaciens* GV3101 strain. Transformed agrobacteria were diluted in 5 mM MES, 10 mM MgCl₂, pH 5.6, and mixed in 1:1 ratio to a final A₆₀₀ of 0.2 prior to leaf infiltration.

N. benthamiana leaves transiently expressing proteins were harvested 2 days post-infiltration. Protein extraction, immunoprecipitation, and Western blotting analyses were performed as described previously (35). For blots shown in Fig. 8, mouse monoclonal single step GFP-HRP antibody (Santa Cruz Biotechnology) was used for GFP immunoblot experiments. For RFP blots, polyclonal RFP antibody (Invitrogen) was used as primary antibody and anti-rat HRP antibody (Sigma, UK) was used as secondary antibody.

Crystallization, Data Collection, and Structure Solution

PexRD54 (in the Presence of ATG8CL)—For crystallization, the PexRD54-ATG8CL complex produced by co-expression was concentrated to 10 mg/ml in 20 mM HEPES, 150 mM NaCl, pH 7.5. Crystallization experiments used 4- μ l hanging drops with a 2:1 protein/precipitant ratio. For data collection, crystals were grown in 18% PEG 10K, 0.1 M sodium acetate, pH 5.0, 0.18 M tri-ammonium citrate and transferred to a cryoprotectant solution consisting of 22% PEG 10K, 0.1 M sodium acetate, pH 5.0, 0.18 M tri-ammonium citrate and 10% ethylene glycol. To enable structure solution, crystals were soaked for ~45 s in well solution supplemented with 500 mM potassium iodide and then cryoprotected as above.

Native and single wavelength anomalous diffraction x-ray data sets were collected at the Diamond Light Source, United Kingdom, beamline I02. The datasets were processed using the Xia2 pipeline (46), see Table 1. The structure was solved using the single wavelength anomalous diffraction approach with the data collected from the crystal soaked in potassium iodide solution. Iodide sites were identified with Phenix (47). These positions were used to estimate initial phases using PHASER EP from the CCP4 suite (48), followed by density improvement with PARROT (49). An initial model was built using BUCCANEER (50) followed by manual rebuilding and refinement using COOT (51) and REFMAC5 (52). Next, molecular replacement with Phaser, followed by the Phenix AutoBuild wizard, was used to produce an initial model of PexRD54 using the native x-ray data. The final model was produced through iterative rounds of refinement using REFMAC5 and manual rebuilding with COOT. Structure validation used the tools provided in COOT and MOLPROBITY (53).

ATG8CL—ATG8CL* mixed with a 3-fold molar excess of pentapeptide (Asp-Trp-Glu-Ile-Val) was incubated at 4 °C for 24 h and concentrated to 80 mg/ml in 20 mM HEPES, 150 mM NaCl, pH 7.5. Crystallization experiments used 2- μ l sitting drops with a 1:1 protein/precipitant ratio. Crystals were produced in 0.2 M ammonium sulfate, 0.1 M Tris buffer, pH 8.0, and 36% PEG3350 and transferred to the precipitant solution with the addition of 10% ethylene glycol as a cryoprotectant. X-ray diffraction data were collected at the Diamond Light Source, UK, beamline I04, and the data were processed as above (Table 1). The structure was solved by molecular replacement using PHASER, as implemented in Phenix. The molecular replacement search model was generated by submitting the complete sequence of ATG8CL to the Phyre web server (54). Based on the solution, an initial model was produced using the AutoBuild wizard in Phenix. At this stage, clear electron density was apparent for the Asp-Trp-Glu-Ile-Val pentapeptide in both molecules of ATG8CL*. The final model was completed and validated as described for PexRD54. Data collection and refinement statistics for PexRD54 and ATG8CL are given in Table 1.

SAXS Measurements, Data Processing, and Analysis

SAXS data were collected at the ESRF beamline BM29 (Grenoble, France (55, 56)) and at the Diamond Light Source, UK, beamline B21. For BM29, measurements were made at an energy of 12.5 keV, camera length of 2.81 m, and q range

Structure/Function of PexRD54

0.003–5 nm⁻¹. For B21, measurements were made at an energy of 12.4 keV, camera length of 4.018 m, and q range 0.004–3.8 nm⁻¹. Measurements of 40 μ l of protein solution at three different concentrations (0.5, 1.0, and 2.0 mg/ml European Synchrotron Radiation Facility (ESRF); 2.5, 5.0, and 10.0 mg/ml Diamond Light Source) were made for each sample (and buffer). Matched buffer measurements taken before and after every sample were averaged and used for background subtraction. Merging of separate concentrations and further analysis steps were performed manually using the ATSAS package (57, 58). DATCMP was used to exclude any individual frames showing signs of radiation damage using standard thresholds for the beamlines. For uncomplexed PexRD54, data collected at the ESRF were used for further analysis. Inspection of the SAXS data for the PexRD54-ATG8CL complex suggested the optimum dataset incorporated both the ESRF (low angles and wide angles) and DLS (mid-range angles) data, and these were merged manually. The forward scattering $I(0)$ and radius of gyration (R_g) for each particle were calculated from the Guinier approximation. The molecular mass of the samples was estimated using the Porod invariant (59) and the maximum particle sizes (D_{max}) were determined from the pair distribution function computed by GNOM (60) using PRIMUS (61). For both PexRD54 and the PexRD54-ATG8CL complex, 40 *ab initio* models were calculated using DAMMIN (62). DAMSEL compared these models and calculated a mean normalized spatial discrepancy (NSD) of 0.545 ± 0.02 for PexRD54 (discarding only one model with $NSD > \text{mean} \pm 2 \times \text{S.D.}$), and a mean NSD of 0.635 ± 0.03 for PexRD54-ATG8CL complex (no models discarded). DAMSEL also identified the most probable (lowest NSD) model. All non-discarded models were aligned, averaged, and compared using DAMSUP, DAMAVER, and DAMFILT in ATSAS for analysis. Rigid body modeling of the PexRD54-ATG8CL complex was achieved with CORAL (36), with the inclusion of the missing residues and linker region that were not visible in the electron density maps of PexRD54 or ATG8CL. The fits of the most probable *ab initio* models to the experimental data were calculated by DAMMIN, the theoretical scattering of PexRD54 was calculated with CRY SOL (63), and the fit of the PexRD54-ATG8CL complex was as calculated by CORAL. Rigid body models of PexRD54 and the PexRD54-ATG8CL complex were overlaid with the *ab initio* models using SUPCOMB (64) and viewed in PyMOL.

Peptide Library

The PexRD54-AIM peptide library was synthesized by Kinexus (Vancouver, Canada) and included 200 peptides where each amino acid in the last 10 amino acids of PexRD54 was changed to every other amino acid. The peptides were spotted on cellulose membrane (Invatis, Germany) with free C termini. Peptide interactions with the ATG8CL-GST-His fusion protein were determined as described previously. The membrane was blocked with 5% (w/v) nonfat dried milk in TBS-T, washed with TBS-T, and overlaid with 1 μ g/ml purified ATG8CL-GST-His fusion protein for 2 h at room temperature. The membrane was washed in TBS-T, and bound proteins were detected with HRP-conjugated anti-GST antibody (1:5000) (RPN1236; GE Healthcare, UK).

Circular Dichroism Spectroscopy

CD spectroscopy experiments were performed using a Chirascan-Plus CD spectrophotometer (Applied Photophysics). Purified proteins in 20 mM HEPES, pH 7.5, 150 mM NaCl at a concentration of at least 10 mg/ml were diluted to 0.2 mg/ml in 20 mM di-potassium phosphate, pH 7.2. CD measurements were carried out in a quartz glass cell with a 0.5-mm path length. To obtain overall CD spectra, wavelength scans between 190 and 260 nm were collected at 15 °C using a 2.0-nm bandwidth, 0.5-nm step size, and time per point of 1 s. The data were collected over four accumulations and averaged. The raw data in millidegree units were corrected for background and converted to mean residue molar ellipticity.

Author Contributions—A. M., R. K. H., Y. F. D., T. O. B., S. K., and M. J. B. designed the research; A. M., R. K. H., Y. F. D., N. T., and E. Z. performed the experiments; K. B. provided reagents and analytic tools; A. M., R. K. H., Y. F. D., A. R., T. O. B., and M. J. B. analyzed the data; A. M., R. K. H., and M. J. B. wrote the paper with editorial input from all authors.

Acknowledgments—We thank the Diamond Light Source (beamlines I02, I04, and B21 under proposals MX7641 and MX9475) and the ESRF (beamline BM29) for access to x-ray data collection facilities and Clare Stevenson (JIC Surface Plasmon Resonance facility) for help with SPR.

References

- Hurley, J. H., and Schulman, B. A. (2014) Atomistic autophagy: the structures of cellular self-digestion. *Cell* **157**, 300–311
- Shaid, S., Brandts, C. H., Serve, H., and Dikic, I. (2013) Ubiquitination and selective autophagy. *Cell Death Differ.* **20**, 21–30
- Reggiori, F., and Klionsky, D. J. (2013) Autophagic processes in yeast: mechanism, machinery and regulation. *Genetics* **194**, 341–361
- Deretic, V., Saitoh, T., and Akira, S. (2013) Autophagy in infection, inflammation and immunity. *Nat. Rev. Immunol.* **13**, 722–737
- Levine, B., Mizushima, N., and Virgin, H. W. (2011) Autophagy in immunity and inflammation. *Nature* **469**, 323–335
- Niu, H., Xiong, Q., Yamamoto, A., Hayashi-Nishino, M., and Rikihisa, Y. (2012) Autophagosomes induced by a bacterial Beclin 1 binding protein facilitate obligatory intracellular infection. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 20800–20807
- Dreux, M., and Chisari, F. V. (2010) Viruses and the autophagy machinery. *Cell Cycle* **9**, 1295–1307
- Nakatogawa, H., Ichimura, Y., and Ohsumi, Y. (2007) Atg8, a ubiquitin-like protein required for autophagosome formation, mediates membrane tethering and hemifusion. *Cell* **130**, 165–178
- Choy, A., Dancourt, J., Mugo, B., O'Connor, T. J., Isberg, R. R., Melia, T. J., and Roy, C. R. (2012) The Legionella effector RavZ inhibits host autophagy through irreversible Atg8 deconjugation. *Science* **338**, 1072–1076
- Alemu, E. A., Lamark, T., Torgersen, K. M., Birgisdottir, A. B., Larsen, K. B., Jain, A., Olsvik, H., Øvervatn, A., Kirkin, V., and Johansen, T. (2012) ATG8 family proteins act as scaffolds for assembly of the ULK complex: sequence requirements for LC3-interacting region (LIR) motifs. *J. Biol. Chem.* **287**, 39275–39290
- Noda, N. N., Ohsumi, Y., and Inagaki, F. (2010) Atg8-family interacting motif crucial for selective autophagy. *FEBS Lett.* **584**, 1379–1385
- Rogov, V., Dötsch, V., Johansen, T., and Kirkin, V. (2014) Interactions between autophagy receptors and ubiquitin-like proteins form the molecular basis for selective autophagy. *Mol. Cell* **53**, 167–178
- Ichimura, Y., Kumanomidou, T., Sou, Y. S., Mizushima, T., Ezaki, J., Ueno, T., Kominami, E., Yamane, T., Tanaka, K., and Komatsu, M. (2008) Structural basis for sorting mechanism of p62 in selective autophagy. *J. Biol. Chem.* **283**, 22847–22857

14. Noda, N. N., Kumeta, H., Nakatogawa, H., Satoo, K., Adachi, W., Ishii, J., Fujioka, Y., Ohsumi, Y., and Inagaki, F. (2008) Structural basis of target recognition by Atg8/LC3 during selective autophagy. *Genes Cells* **13**, 1211–1218
15. Klionsky, D. J., and Schulman, B. A. (2014) Dynamic regulation of macroautophagy by distinctive ubiquitin-like proteins. *Nat. Struct. Mol. Biol.* **21**, 336–345
16. Noda, N. N., Satoo, K., Fujioka, Y., Kumeta, H., Ogura, K., Nakatogawa, H., Ohsumi, Y., and Inagaki, F. (2011) Structural basis of Atg8 activation by a homodimeric E1, Atg7. *Mol. Cell* **44**, 462–475
17. Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F. N., Kamoun, S., Krause, J., Thines, M., Weigel, D., and Burbano, H. A. (2013) The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *Elife* **2**, e00731
18. Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., and Gurr, S. J. (2012) Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**, 186–194
19. Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M., Alvarado, L., Anderson, V. L., Armstrong, M. R., et al. (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461**, 393–398
20. Dodds, P. N., and Rathjen, J. P. (2010) Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat. Rev. Genet.* **11**, 539–548
21. Win, J., Chaparro-Garcia, A., Belhaj, K., Saunders, D. G., Yoshida, K., Dong, S., Schornack, S., Zipfel, C., Robatzek, S., Hogenhout, S. A., and Kamoun, S. (2012) Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harb. Symp. Quant. Biol.* **77**, 235–247
22. Wirthmueller, L., Maqbool, A., and Banfield, M. J. (2013) On the front line: structural insights into plant–pathogen interactions. *Nat. Rev. Microbiol.* **11**, 761–776
23. Whisson, S. C., Boevink, P. C., Moleleki, L., Avrova, A. O., Morales, J. G., Gilroy, E. M., Armstrong, M. R., Grouffaud, S., van West, P., Chapman, S., Hein, I., Toth, I. K., Pritchard, L., and Birch, P. R. (2007) A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* **450**, 115–118
24. Boutemy, L. S., King, S. R., Win, J., Hughes, R. K., Clarke, T. A., Blumenschein, T. M., Kamoun, S., and Banfield, M. J. (2011) Structures of *Phytophthora* RXLR effector proteins: a conserved but adaptable fold underpins functional diversity. *J. Biol. Chem.* **286**, 35834–35842
25. Win, J., Krasileva, K. V., Kamoun, S., Shirasu, K., Staskawicz, B. J., and Banfield, M. J. (2012) Sequence divergent RXLR effectors share a structural fold conserved across plant pathogenic oomycete species. *PLoS Pathog.* **8**, e1002400
26. Win, J., Morgan, W., Bos, J., Krasileva, K. V., Cano, L. M., Chaparro-Garcia, A., Ammar, R., Staskawicz, B. J., and Kamoun, S. (2007) Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Cell* **19**, 2349–2369
27. Bos, J. I., Armstrong, M. R., Gilroy, E. M., Boevink, P. C., Hein, I., Taylor, R. M., Zhendong, T., Engelhardt, S., Vetukuri, R. R., Harrower, B., Dixelius, C., Bryan, G., Sadanandom, A., Whisson, S. C., Kamoun, S., and Birch, P. R. (2010) *Phytophthora infestans* effector AVR3a is essential for virulence and manipulates plant immunity by stabilizing host E3 ligase CMPG1. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9909–9914
28. McLellan, H., Boevink, P. C., Armstrong, M. R., Pritchard, L., Gomez, S., Morales, J., Whisson, S. C., Beynon, J. L., and Birch, P. R. (2013) An RxLR effector from *Phytophthora infestans* prevents re-localisation of two plant NAC transcription factors from the endoplasmic reticulum to the nucleus. *PLoS Pathog.* **9**, e1003670
29. Saunders, D. G., Breen, S., Win, J., Schornack, S., Hein, I., Bozkurt, T. O., Champouret, N., Vleeshouwers, V. G., Birch, P. R., Gilroy, E. M., and Kamoun, S. (2012) Host protein BSL1 associates with *Phytophthora infestans* RXLR effector AVR2 and the *Solanum demissum* immune receptor R2 to mediate disease resistance. *Plant Cell* **24**, 3420–3434
30. King, S. R., McLellan, H., Boevink, P. C., Armstrong, M. R., Bukharova, T., Sukarta, O., Win, J., Kamoun, S., Birch, P. R., and Banfield, M. J. (2014) *Phytophthora infestans* RXLR effector PexRD2 interacts with host MAPKKKε to suppress plant immune signaling. *Plant Cell* **26**, 1345–1359
31. Bozkurt, T. O., Schornack, S., Win, J., Shindo, T., Ilyas, M., Oliva, R., Cano, L. M., Jones, A. M., Huitema, E., van der Hoorn, R. A., and Kamoun, S. (2011) *Phytophthora infestans* effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20832–20837
32. Gilroy, E. M., Taylor, R. M., Hein, I., Boevink, P., Sadanandom, A., and Birch, P. R. (2011) CMPG1-dependent cell death follows perception of diverse pathogen elicitors at the host plasma membrane and is suppressed by *Phytophthora infestans* RXLR effector AVR3a. *New Phytol.* **190**, 653–666
33. Wang, X., Boevink, P., McLellan, H., Armstrong, M., Bukharova, T., Qin, Z., and Birch, P. R. (2015) A host KH RNA-binding protein is a susceptibility factor targeted by an RXLR effector to promote late blight disease. *Mol. Plant* **8**, 1385–1395
34. Boevink, P. C., Wang, X., McLellan, H., He, Q., Naqvi, S., Armstrong, M. R., Zhang, W., Hein, I., Gilroy, E. M., Tian, Z., and Birch, P. R. (2016) A *Phytophthora infestans* RXLR effector targets plant PP1c isoforms that promote late blight disease. *Nat. Commun.* **7**, 10311
35. Dagdas, Y. F., Belhaj, K., Maqbool, A., Chaparro-Garcia, A., Pandey, P., Petre, B., Tabassum, N., Cruz-Mireles, N., Hughes, R. K., Sklenar, J., Win, J., Menke, F., Findlay, K., Banfield, M. J., Kamoun, S., and Bozkurt, T. O. (2016) An effector of the Irish potato famine pathogen antagonizes a host autophagy cargo receptor. *Elife* **5**, e10856
36. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V., and Svergun, D. I. (2012) New developments in the program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350
37. Cheng, W., Munkvold, K. R., Gao, H., Mathieu, J., Schwizer, S., Wang, S., Yan, Y. B., Wang, J., Martin, G. B., and Chai, J. (2011) Structural analysis of *Pseudomonas syringae* AvrPtoB bound to host BAK1 reveals two similar kinase-interacting domains in a type III effector. *Cell Host Microbe* **10**, 616–626
38. Dong, J., Xiao, F., Fan, F., Gu, L., Cang, H., Martin, G. B., and Chai, J. (2009) Crystal structure of the complex between *Pseudomonas* effector AvrPtoB and the tomato Pto kinase reveals both a shared and a unique interface compared with AvrPto-Pto. *Plant Cell* **21**, 1846–1859
39. Xing, W., Zou, Y., Liu, Q., Liu, J., Luo, X., Huang, Q., Chen, S., Zhu, L., Bi, R., Hao, Q., Wu, J. W., Zhou, J. M., and Chai, J. (2007) The structural basis for activation of plant immunity by bacterial effector protein AvrPto. *Nature* **449**, 243–247
40. Desveaux, D., Singer, A. U., Wu, A. J., McNulty, B. C., Musselwhite, L., Nimchuk, Z., Sondek, J., and Dangel, J. L. (2007) Type III effector activation via nucleotide binding, phosphorylation, and host target interaction. *PLoS Pathog.* **3**, e48
41. Chou, S., Krasileva, K. V., Holton, J. M., Steinbrenner, A. D., Alber, T., and Staskawicz, B. J. (2011) *Hyaloperonospora arabidopsidis* ATR1 effector is a repeat protein with distributed recognition surfaces. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13323–13328
42. Yaeno, T., Li, H., Chaparro-Garcia, A., Schornack, S., Koshiba, S., Watanabe, S., Kigawa, T., Kamoun, S., and Shirasu, K. (2011) Phosphatidylinositol monophosphate-binding interface in the oomycete RXLR effector AVR3a is required for its stability in host cells to modulate plant immunity. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14682–14687
43. De Guillen, K., Ortiz-Vallejo, D., Gracy, J., Fournier, E., Kroj, T., and Padilla, A. (2015) Structure analysis uncovers a highly diverse but structurally conserved effector family in phytopathogenic fungi. *PLoS Pathog.* **11**, e1005228
44. Pedersen, C., Ver Loren van Themaat, E., McGuffin, L. J., Abbott, J. C., Burgis, T. A., Barton, G., Bindschedler, L. V., Lu, X., Maekawa, T., Wessling, R., Cramer, R., Thordal-Christensen, H., Panstruga, R., and Spanu, P. D. (2012) Structure and evolution of barley powdery mildew effector candidates. *BMC Genomics* **13**, 694
45. Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I., and Owens, R. J. (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35**, e45
46. Winter, G. (2010) xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190

47. Adams, P. D., Afonine, P. V., Bunkóczy, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
48. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242
49. Cowtan, K. (2010) Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 470–478
50. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011
51. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
52. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
53. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21
54. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858
55. Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., Morawe, C., Mueller-Dieckmann, C., Ohlsson, S., Schmid, W., Surr, J., Theveneau, P., *et al.* (2013) Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J. Synchrotron Radiat.* **20**, 660–664
56. Round, A., Felisaz, F., Fodinger, L., Gobbo, A., Huet, J., Villard, C., Blanchet, C. E., Pernot, P., McSweeney, S., Roessle, M., Svergun, D. I., and Cipriani, F. (2015) BioSAXS sample changer: a robotic sample changer for rapid and reliable high-throughput x-ray solution scattering experiments. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 67–75
57. Konarev, P. V., Petoukhov, M. V., Volkov, V. V., and Svergun, D. I. (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **39**, 277–286
58. Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285
59. Ciccariello, S., Goodisman, J., and Brumberger, H. (1988) On the Porod Law. *J. Appl. Crystallogr.* **21**, 117–128
60. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503
61. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282
62. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886
63. Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRY SOL—A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773
64. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41

Structural Insights into the Polyphyletic Origins of Glycyl tRNA Synthetases^{*S}♦

Received for publication, March 31, 2016, and in revised form, May 9, 2016. Published, JBC Papers in Press, May 23, 2016, DOI 10.1074/jbc.M116.730382

Marco Igor Valencia-Sánchez^{†1}, Annia Rodríguez-Hernández^{‡S2}, Ruben Ferreira^{¶1}, Hugo Aníbal Santamaría-Suárez[‡], Marcelino Arciniega[‡], Anne-Catherine Dock-Bregeon^{||}, Dino Moras^{**}, Brice Beinstainer^{**},  Haydyn Mertens^{‡‡},  Dmitri Svergun^{‡‡}, Luis G. Briebe[§], Morten Grøtli[¶], and  Alfredo Torres-Larios^{‡3}

From the [†]Departamento de Bioquímica y Biología Estructural, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria, Apartado Postal 70-243, Mexico City 04510, México, the [§]Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Guanajuato 04510, México, the [¶]Department of Chemistry and Molecular Biology, University of Gothenburg, SE-412 96 Gothenburg, Sweden, the ^{||}Laboratoire de Biologie Intégrative des Milieux Marins, CNRS UMR8227, 29680 Roscoff, France, the ^{**}Centre for Integrative Biology, Department of Integrated Structural Biology, Institute of Genetics and of Molecular and Cellular Biology, CNRS UMR 7104, 1 Rue Laurent Fries, Illkirch, France, and the ^{‡‡}European Molecular Biology Laboratory, Hamburg Outstation, c/o DESY, Notkestrasse 85, Hamburg 22603, Germany

Glycyl tRNA synthetase (GlyRS) provides a unique case among class II aminoacyl tRNA synthetases, with two clearly widespread types of enzymes: a dimeric (α_2) species present in some bacteria, archaea, and eukaryotes; and a heterotetrameric form ($\alpha_2\beta_2$) present in most bacteria. Although the differences between both types of GlyRS at the anticodon binding domain level are evident, the extent and implications of the variations in the catalytic domain have not been described, and it is unclear whether the mechanism of amino acid recognition is also dissimilar. Here, we show that the α -subunit of the $\alpha_2\beta_2$ GlyRS from the bacterium *Aquifex aeolicus* is able to perform the first step of the aminoacylation reaction, which involves the activation of the amino acid with ATP. The crystal structure of the α -subunit in the complex with an analog of glycyl adenylate at 2.8 Å resolution presents a conformational arrangement that properly positions the cognate amino acid. This work shows that glycine is recognized by a subset of different residues in the two types of GlyRS. A structural and sequence analysis of class II catalytic domains shows that bacterial GlyRS is closely related to alanyl tRNA synthetase, which led us to define a new subclassification of these ancient enzymes and to propose an evolution-

ary path of $\alpha_2\beta_2$ GlyRS, convergent with α_2 GlyRS and divergent from AlaRS, thus providing a possible explanation for the puzzling existence of two proteins sharing the same fold and function but not a common ancestor.

Aminoacyl tRNA synthetases are ancient enzymes that attach cognate amino acids to their corresponding tRNAs (1–3). This task is performed in two steps as follows: amino acid activation with ATP, followed by attachment of the resulting aminoacyl adenylate to the cognate tRNA (4–6). According to their sequence and structural features, there are two main classes of non-related aaRS,⁴ which evolved independently from two different domains corresponding to two modes of ATP binding (7–10). These classes are further divided into subclasses according to protein sequence, structural features of the catalytic domain, the presence of accessory domains, and similarity of amino acids (11–13). Most aaRSs are descended from a single ancestor (monophyletic); however, there are two clearly recognized exceptions, lysyl tRNA synthetase and GlyRS (2, 12, 14–16). GlyRS, a class II aaRS, can be found as an α_2 homodimer in eukarya, archaea, and some bacteria (14, 17–20) and as an $\alpha_2\beta_2$ heterotetramer in most bacteria and chloroplasts (21–23). Notably, the limited extent of the sequence conservation between the two forms does not allow inference to a common ancestor between them. Nonetheless, the catalytic domains contained in the α -subunits of both GlyRSs share the same fold (11, 14, 16, 22–24). The eukaryotic GlyRS belongs to the subclass IIa, specific for hydrophobic and small polar amino acids, with similarities with HisRS, ThrRS, ProRS, and SerRS.

* The authors declare that they have no conflicts of interest with the contents of this article.

M. I. V. S. and A. T. L. dedicate this work to Prof. Enrique Merino Pérez, IBT-Universidad Nacional Autónoma de México, for his constant support and guidance.

✂ Author's Choice—Final version free via Creative Commons CC-BY license.

♦ This article was selected as a Paper of the Week.

 This article contains supplemental Figs. S1 and S2.

The atomic coordinates and structure factors (code 5F5W) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

¹ Recipient of a grant from CONACyT to perform graduate studies. Part of this work in was performed to fulfill the requirements of a doctoral dissertation in the Posgrado en Ciencias Bioquímicas, Universidad Nacional Autónoma de México.

² Supported by a CONACyT postdoctoral fellowship.

³ Supported by Consejo Nacional de Ciencia y Tecnología (CONACyT) Grants CB2015-253323 and PDCPN2014-47543 and Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica, Universidad Nacional Autónoma de México (PAPIIT-DGAPA) Grant IN202416. To whom correspondence should be addressed: Circuito Exterior s/n, Ciudad Universitaria, Apartado Postal 70-243, México D.F. 04510, México. Fax: 52-55-5622-5630; E-mail: torres@ifc.unam.mx.

⁴ The abbreviations used are: aaRS, aminoacyl tRNA synthetase; Gly-SA, G5A, or GSAd, 5'-O-[N-(L-glycyl)sulfamoyl] adenosine; SAXS, small angle x-ray scattering; TCS, transitive consistency score; SEC-MALS, size exclusion chromatography coupled to multiangle light scattering; PCA, principal component analysis; r.m.s.d., root mean square deviation; PDB, Protein Data Bank; GlyRS, glycyl tRNA synthetase; HisRS, histidyl tRNA synthetase; ThrRS, threonyl tRNA synthetase; ProRS, prolyl tRNA synthetase; SerRS, seryl tRNA synthetase; PheRS, phenylalanyl tRNA synthetase; AlaRS, alanyl tRNA synthetase; SepRS, phosphoserine tRNA synthetase; PylRS, pyrrolysyl tRNA synthetase.

Bacterial GlyRS, however, belongs to the subclass IIc, together with PheRS, AlaRS, SepRS, and PylRS, the most heterogeneous group of the three subclasses (25–27). However, there are some studies that place AlaRS or bacterial GlyRS within subclass IIa and/or do not consider the existence of two types of GlyRS belonging to different subclasses (28–32).

The ~35-kDa α -subunit of bacterial GlyRS contains the aminoacylation site (22, 24, 33, 34). However, no activity has been demonstrated for this domain alone, suggesting that the β -subunit is required for catalysis (21, 33, 35–37). The ~65-kDa β -subunit has the most important tRNA recognition elements but is not homologous to the anticodon binding domain of eukaryotic GlyRS (23, 37). Two crystal structures are available for the bacterial GlyRS α -subunit in the apo-conformation (PDB entries 3rf1 and 1j5w), leaving unclear whether the recognition details of the small substrates, glycine and ATP, are shared by the two types of enzymes.

Here, we show that the bacterial GlyRS α -subunit, which has all of the molecular determinants needed for the first step of the reaction, is indeed able to perform this catalysis. The crystal structure of this subunit in complex with an analog of glycyl adenylate (5'-O-[N-(L-glycyl)sulfamoyl] adenosine, also known as Gly-SA, G5A, or GSAd) allows the conformational changes correlated with glycine recognition to be defined. A comparison of the α -subunit of bacterial GlyRS with the activation domain of archaeal and eukaryotic GlyRS establishes that the two classes of GlyRS employ different chemical strategies to recognize glycine as a substrate. Moreover, a structural and sequence analysis performed on the activation domain of class II aaRS defines a new subclass IIc, comprising AlaRS and bacterial $\alpha_2\beta_2$ GlyRS. This consequently modifies subclass IIc to contain $(\alpha\beta)_2$ PheRS, SepRS, and possibly PylRS. This classification standing on common structural motifs and active site residues in class IIc enzymes allows us to propose a divergence of bacterial GlyRS from AlaRS, which resolves the puzzling existence of two proteins sharing the same fold and function but not a common ancestor.

Results

α -Subunit of Bacterial GlyRS Is Able to Perform the Amino Acid Activation—We were able to obtain highly purified α -subunit GlyRS of the hyperthermophilic bacterium *Aquifex aeolicus* (α -AaGlyRS) by means of a heat treatment and an astringent His tag affinity chromatography step. A final purification step using size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) indicated a homogeneous dimeric population of 69.1 kDa (theoretical mass = 67.4 kDa), in agreement with previous reports (21, 36). Small angle x-ray scattering (SAXS) further confirmed the dimeric nature of the ensemble. Because of the evolutionary conservation of all amino acids involved in glycine activation (see below), we speculated that the α -subunit alone would be able to catalyze the first step of the reaction, the attachment of glycine to ATP. With the use of an alternative method based on thin layer chromatography to monitor the activity, we found that α -AaGlyRS was indeed able to perform the first step of aminoacylation (Fig. 1). In contradiction with previous reports, the α -subunit showed weak activity at pH values ranging from 6.0 to 8.0 and

glycine concentrations from 80 μ M to 10 mM. Under the best possible reaction conditions, the observed K_m for glycine was 0.11 ± 0.016 mM, similar to a previously reported value for the full-length *Escherichia coli* enzyme (21). Nevertheless, the k_{cat} reaction was extremely slow, only 2.3×10^{-4} s $^{-1}$. Previous studies have performed an ATP-PP $_i$ exchange assay, which follows the reverse reaction of incorporation of 32 P from PP $_i$ into ATP (21, 33, 35–37), whereas in our study the Gly-AMP synthesis was measured directly. These findings provide biological relevance to our reported crystal structure (see below) and support the functional existence of an $\alpha_2\beta_2$ quaternary structure, rather than an $(\alpha\beta)_2$ organization.

Binding to a Transition State Analog Promotes Conformational Changes in the α -Subunit of Bacterial GlyRS—To understand amino acid and nucleotide recognition in bacterial GlyRS, we solved the crystal structure of α -AaGlyRS in complex with GSAd at 2.81 Å resolution (Table 1). The electron density map unambiguously showed all features of the bound GSAd and its molecular surroundings in all five molecules in the asymmetric unit (Fig. 2B).

As shown previously (PDB code 1j5w (24)), the α -subunit forms a homodimer (Fig. 2A), with 58% of its 2492 Å 2 interface formed by a 97-residue helical region located at the C-terminal part of the subunit and situated on top of the signature antiparallel β -strand of class II synthetases. Further details on the description of the general architecture of this subunit have been given previously (24).

Comparison of α -AaGlyRS with previously solved structures of this subunit in the apo-form reveals conformational changes in the region formed by residues 112–123 (Fig. 2, C and D). This segment is the topological equivalent of the so-called amino acid loop found in HisRS, ProRS, ThrRS, SerRS, and eukaryotic GlyRS (38). Here, the movement of this region causes an ~5 Å displacement of Trp-115 when compared with the apo structure, allowing it to form a cation- π interaction with the glycine moiety of GSAd (Fig. 2D, see below) and to form a hydrogen bond with the O1S atom of GSAd (equivalent to the O1P atom of the cognate Gly-AMP (Fig. 3D)). As a result, the size of the active site pocket is reduced from an area and volume of 706 Å 2 and 960 Å 3 to 451 Å 2 and 636 Å 3 .

In correlation with this movement, thermal shift assays showed a displacement of the T_m values of α -AaGlyRS in the presence of GSAd, from 83.2 °C (no ligand) to 92.3 °C (8 mM GSAd). We could not detect significant changes in the presence of glycine, ATP, or ATP + glycine. The reported crystal structure of the α -subunit of bacterial GlyRS shows a productive, biologically meaningful complex that allows the recognition mechanisms in the two types of GlyRS to be compared.

Mechanism of Glycine Recognition Differs in the Two Types of Known GlyRS—Although nucleotide recognition shows common features in both types of GlyRSs (Fig. 3), critical differences are found at the amino acid recognition level, which is centered on the amino group. In eukaryotic GlyRS, the amino group is recognized by three conserved and negatively charged glutamic acid residues (Glu-522, Glu-296, and Glu-245 in PDB code 2zt8) (Fig. 4, A and C) (39). In contrast, in the bacterial type GlyRS, the amino group is recognized through five conserved different residues (Trp-115, Gln-76, Gln-78, Thr-33, and Glu-

Insights into the Polyphyletic Origins of GlyRSs

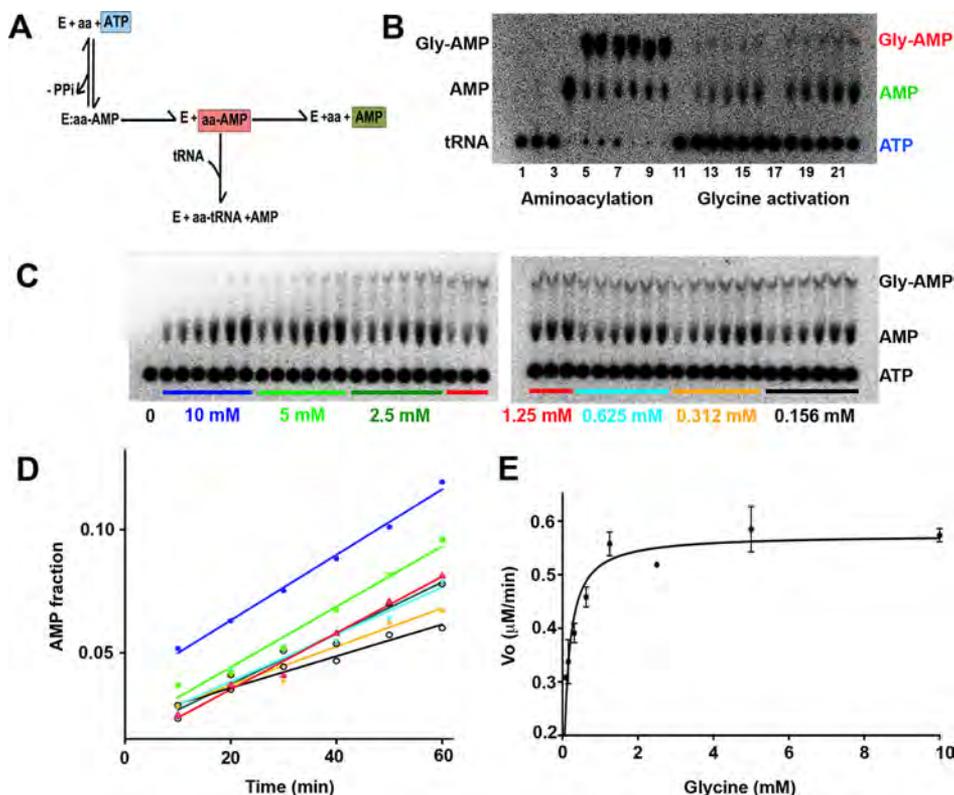


FIGURE 1. α -Subunit of the tetrameric $\alpha_2\beta_2$ GlyRS from *A. aeolicus* (α -AaGlyRS) is able to activate the amino acid. *A*, schematic diagram of the first step of aminoacylation. *B*, control experiments. Comparison of full aminoacylation reaction versus amino acid activation. Lanes 1–10, aminoacylation reaction performed as described previously (57–59). Lane 1, no enzyme added, and no P1 nuclease added. Lane 2, *Anaerolinea thermophila*, a protein that bears in the same sequence both subunits ($\alpha\beta$ -GlyRS) added, and no P1 nuclease added. Lane 3, *A. aeolicus* ($\alpha + \beta$ -GlyRS) added, and no P1 nuclease added. Lane 4, no enzyme added, and P1 nuclease added. Lanes 5–7, 5, 10, and 15 min of aminoacylation reaction using $\alpha + \beta$ -GlyRS (with P1 nuclease added). Lanes 8–10, 5, 10, and 15 min of aminoacylation reaction using $\alpha\beta$ -GlyRS (with P1 nuclease added). Lanes 11–22, glycine activation reaction. Lane 11, zero time point using α -AaGlyRS. Lanes 12–16, 10, 20, 30, 40, and 50 min of the glycine activation reaction using α -AaGlyRS. Lane 17, zero time point using $\alpha\beta$ -GlyRS. Lanes 18–22, 10, 20, 30, 40, and 50 min of the glycine activation reaction using the $\alpha\beta$ -GlyRS. *C*, amino acid activation. α -AaGlyRS at 40 μ M, in the presence of decreasing glycine concentrations, 0.5 mM ATP, 50 mM Tris, pH 8.0, 50 mM KCl, 10 mM MgCl₂. Time points were taken every 10 min for 60 min for each concentration, and the formation of AMP was monitored for each point. *D*, initial velocities (kinetics of AMP formation from the experiment in *C*). Steady state time courses using different glycine concentrations. *E*, Michaelis-Menten plot. Initial velocities were plotted against substrate concentration; error bars indicate the standard deviation for each point.

TABLE 1
Data collection and refinement statistics

One crystal was used.

α -AaGlyRS-GSAd	
Data collection	
Space group	P22 ₁ 2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	101.8, 130.0, 145.5
α , β , γ (°)	90.0, 90.0, 90.0
Resolution (Å)	83.43–2.81 (2.91–2.81) ^a
<i>R</i> _{sym} or <i>R</i> _{merge}	0.084 (0.584)
<i>I</i> / σ <i>I</i>	9.7 (2.2)
Completeness (%)	100 (100)
Redundancy	4.4 (4.5)
Mean (<i>I</i>) half-set correlation <i>CC</i> _{1/2}	0.990 (0.770)
Wilson <i>B</i> -factor (Å ²)	58.4
Refinement	
Resolution (Å)	80.19–2.81 (2.86–2.81)
No. of reflections	47,741 (2607)
<i>R</i> _{work} / <i>R</i> _{free}	0.239/0.252 (0.362/0.363)
No. of atoms	
Protein	11,740
Ligand/ion	135
Water	2
<i>B</i>-Factors	
Protein (by chain)	45.1, 53.6, 56.9, 69.6, 73.1
Ligand/ion	40.8, 47.5, 52.6, 68.0, 74.6
Water	45.8
r.m.s.d.	
Bond lengths (Å)	0.011
Bond angles (°)	1.462

^a Highest resolution shell is shown in parentheses.

156) (Figs. 2*D* and 4, *B*, *D*, and *E*). Interestingly, Gln-76 interacts directly with the glycine carbonyl group and also with the amino moiety through hydrogen bonding, with Gln-78, of a solvent molecule (Fig. 4*D*). The amino group of GSAd also interacts with the side chains of Thr-33 and Glu-156 (Fig. 4, *B* and *D*).

The differences with eukaryotic GlyRS are also extended to the region of the amino acid loop. In eukaryotic GlyRS, the hydroxyl group of a non-conserved Tyr-386 contacts the carbonyl group of GSAd (Fig. 4*E*). In contrast, in bacterial GlyRS, an absolutely conserved Trp-115 makes a clear cation- π interaction with the amino moiety of GSAd and also contacts the O1S atom of GSAd (Figs. 2*D* and 4*E*).

In addition, in eukaryotic GlyRS, the binding of the noncognate alanine is prevented in part by a non-conserved Ser (or Ala) residue (524) and the highly negatively charged recognition cavity (Fig. 4, *C* and *F*). In contrast, in bacterial GlyRS, this steric hindrance is made by a pair of highly conserved Thr residues (158 and 140) together with a nearby solvent molecule (Fig. 4*F*).

The structural sequence alignment combined with multiple sequence alignments indicate seven highly conserved residues that are shared by the two types of GlyRS (Fig. 3*B*). Five of these

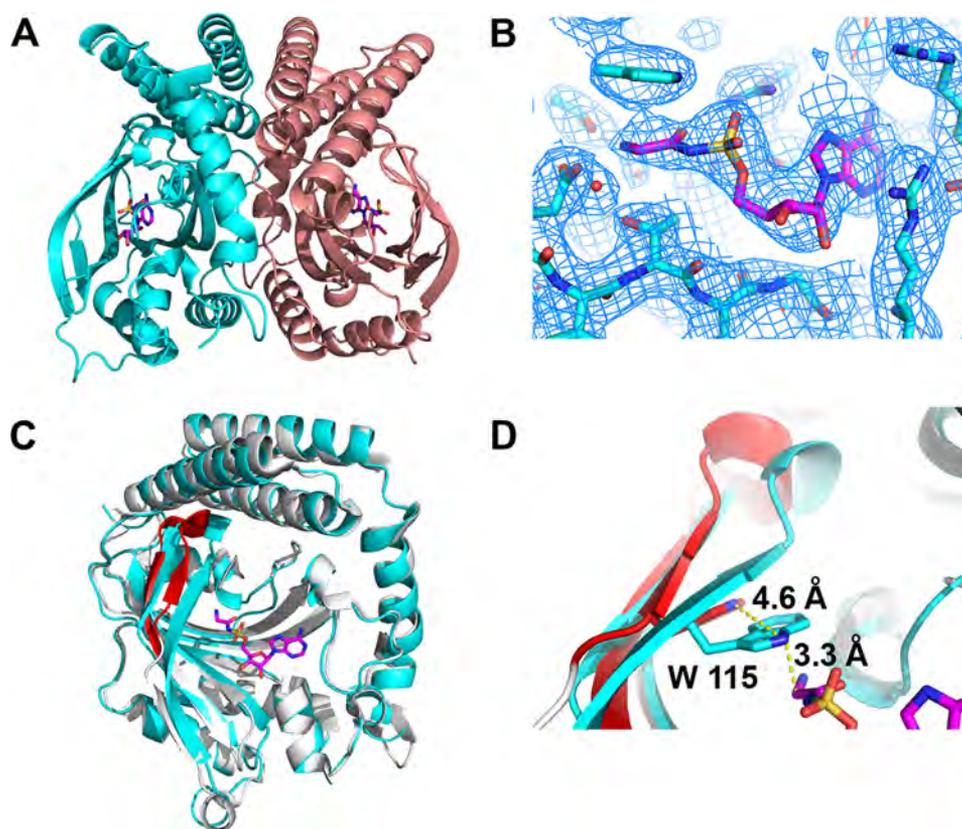


FIGURE 2. **Binding of a glycyl adenylate analog promotes a conformational change in α -AaGlyRS.** *A*, overall dimeric structure of α -AaGlyRS with GSAd shown in magenta. There are five monomers, each with a ligand bound, in the asymmetric unit. *B*, simulated annealed, $2F_o - F_c$ omit electron density map contoured at 1σ around the GSAd-binding site. Two solvent molecules were located near the glycine recognition site that could be well fitted and observed on double difference and omit maps for monomer B and could be seen as peaks bigger than 3.5σ on difference, $F_o - F_c$ electron density maps on the four other monomers. No other solvent molecules were added to the model. *C*, superposition of α -AaGlyRS-GSAd with the apo structure of *Campylobacter jejuni* (PDB code 3rgl). The sequences of the subunits have an identity of 60% and a similarity of 77%. The overall r.m.s.d. is 0.959 Å. The binding of GSAd causes conformational changes in the region of residues 112–123, shown in red in *C* and *D*. *D*, close-up of the superposition showing the conformational change upon binding of GSAd, which involves a 4.6 Å movement of Trp-115 to make a cation- π interaction with GSAd. The superposition with another α -subunit structure from *T. maritima* (PDB code 1j5w) does not show this displacement, suggesting that this conformational change is due to GSAd binding.

are involved in ATP recognition, and as such, these residues are highly conserved and shared by many aaRSs. There is only one residue that is shared by the two types of GlyRS for the recognition of glycine (Glu-156 in bacterial GlyRS and Glu-522 in eukaryotic GlyRS). Not another single residue is shared in any other region of the catalytic domain. Comparison of the electrostatic potentials within the glycine binding pockets shows a highly negatively charged region in eukaryotic GlyRS and a less polar environment in bacterial GlyRS. The comparison of the two types of GlyRSs and further analysis as shown below indicate that their differences are so profound, up to the level of amino acid recognition, that it is extremely unlikely for them to share a common ancestor, even when their catalytic domains have the same overall fold.

Bacterial GlyRS Presents Key Differences Compared with Other Class II Synthetases—A hallmark of class II synthetases is the presence of key residues located in each of their three characteristic motifs (Fig. 5A) (7). In bacterial GlyRS, the signature sequence of motif 2, formed by the first two strands of the antiparallel β -sheet that forms the floor of the active site, presents several changes compared with the consensus sequence (Fig. 5B). In particular, there are three key differences. 1) Unique to bacterial GlyRS, an arginine residue (Arg57) is located where a hydrophobic residue is otherwise always pres-

ent (Fig. 5, A–C). Additionally, an Arg residue makes a salt bridge with two highly conserved Asp residues (Fig. 5C). In the rest of class II synthetases, a Phe, Tyr, Val, or His residue helps to stabilize a hydrophobic region located at the dimeric interface (Fig. 5C). 2) Similar to AlaRS, there is an insertion (Pro-59 in bacterial GlyRS and two to six residues in AlaRS) between an absolutely conserved Arg and an acidic residue (Fig. 5, A and B). 3) In many aaRSs, a fully conserved acidic residue recognizes the amino moiety of the cognate amino acid (Fig. 5, A, B and D). In bacterial GlyRS, AlaRS, HisRS, SepRS, and PylRS, another residue occupies this position (Fig. 5B). In bacterial GlyRS, a glutamine residue is substituted in place of this acidic residue (Fig. 5, A, B and D). In α -AaGlyRS, Gln-78 makes a hydrogen bond with a solvent molecule that interacts with the amino moiety of the cognate glycine (Fig. 5D). These sequence-structure observations confirm that bacterial GlyRS has non-canonical features that are somewhat shared with other atypical aaRSs, like AlaRS, and not with eukaryal GlyRS.

Bacterial GlyRS and AlaRS Form a Structural Subclass of aaRS—To gain further insight into the evolutionary origin of bacterial GlyRS, we performed pairwise structural alignments of the catalytic core (proposed to reflect the primordial synthetase (40)), among the available crystal structures of all types of class II aaRS using two different algorithms (Dali (41) and

Insights into the Polyphyletic Origins of GlyRSs

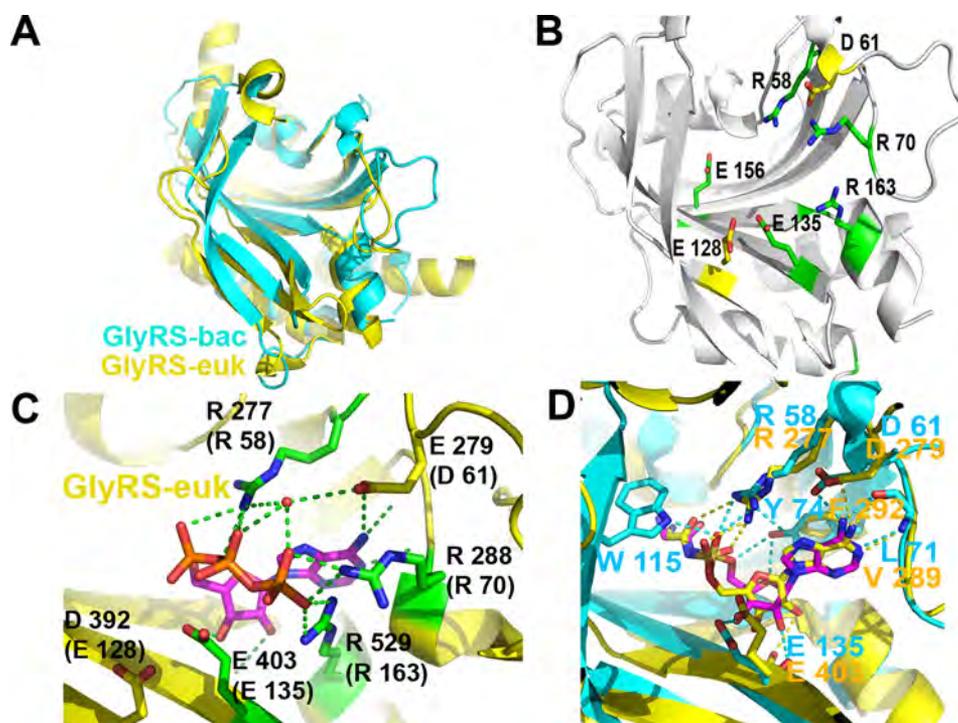


FIGURE 3. Features shared by the two types of GlyRSs. *A*, superposition of the catalytic core, according to Dali, of bacterial GlyRS (residues 1–169) and eukaryotic GlyRS (residues 97–536, with the exception of extensions comprising residues 129–137, 145–242, and 301–349, PDB code 2zt8). The Z-score is 14.9 for the 152 residues aligned, with an r.m.s.d. of 2.7 and 17% identity. *B*, residues that are highly (yellow) or absolutely conserved (green) and also shared by bacterial GlyRS and eukaryotic GlyRS according to the structural and sequence alignment of all available sequences are shown on the structure of bacterial GlyRS. The following panels show the role and/or location of some of these residues in more detail. *C*, five of the seven residues depicted in *B* are highly conserved class II residues that recognize ATP (PDB code 2zt7). *D*, in comparison with eukaryotic GlyRS, α -AaGlyRS Tyr-74 makes additional interactions through its OH group, which contacts the O4' and O5' atoms of GSAd and the NE1 atom of Trp-115 that contacts the O1S atom of GSAd.

STAMP (42)). We also performed a multiple sequence alignment using the structural information through T-Coffee Expresso (43). Finally, we performed a principal component analysis (PCA) on the C α atoms of 80 core residues common to the structure of class II aaRSs, according to the algorithm implemented in Bio3D (44). The information derived from each of the four approaches is summarized in Figs. 6–10, Tables 2 and 3, and supplemental Figs. S1 and S2. We were able to clearly identify four subclasses of class II aaRSs. These subclasses could be defined either using the whole catalytic domain just devoid of big insertions (Fig. 6, Table 2, and supplemental Figs. S1 and S2) or a small C α core of 80 atoms (Figs. 7–10). The analysis of the C α core of 80 atoms showed that the main differences between the groups are found in two helical regions and one β -strand (PC1 and PC2, residues 16–22, 46–58, and 61–64 of the C α core, Figs. 7 and 9A). Notably, the PCA analysis (Fig. 9) shows that only these two components, which together account for almost 50% of the structural variance of the core between the class II structures (Fig. 9B), allow us to define a clear separation between the subgroups, most importantly, the group formed by bacterial GlyRS and AlaRS and the rest of class II aaRSs (Figs. 9C and 10). The four subgroups are maintained even if bacterial GlyRS or AlaRS is removed from the PCA. At the whole catalytic domain level, the superposition of representatives of the four subclasses shows a general fold agreement (Fig. 6, average Dali Z-score of 13.4). A closer approach allows us to visualize the differences among the four established divisions (groups with Dali Z-scores equal to or higher than 18.5, which turned out to be equal to the branches

of the STAMP analysis based on r.m.s.d.) and the agreement between them (Fig. 6, Table 2, and supplemental Fig. S1). Each subclass is mostly defined by the relative angles of two helices located at the back of the active site (subclass IIa *versus* IIb), by the length of these helices (subclass IIa and IIb *versus* IIc and IId), or by the length and relative orientation of three strands of the active site β -sheet (subclass IIc *versus* IId). Notably, each subclass is in full agreement with previously proposed divisions. However, subclass IIc is now divided in two distinct subclasses: IIc, including PheRS, SepRS, and possibly PylRS, which groups with subclass IIc according to the structural classification (Figs. 8–10 and supplemental Fig. S1) but with subclass IIb according to the sequence analysis (supplemental Fig. S2). The other subclass, IId, includes AlaRS and bacterial GlyRS, which match not only in helical length but also the length and orientation of three strands of the active site β -sheet (Fig. 11).

Bacterial GlyRS and AlaRS Share Key Residues Important for Both Protein Structure and Substrate Recognition—The structural sequence alignment coupled with multiple sequence alignments (Fig. 11, A and B) indicate that there are 15 residues either absolutely or highly conserved by bacterial GlyRS and AlaRS. Five of these are involved in ATP recognition (Fig. 11, B and C). Outside of this, the amino acid recognition region surprisingly revealed that the equivalent of Trp-115 in bacterial GlyRS is also found in AlaRS (Trp-192, PDB code 2zzg) (Fig. 11, B, C and E) (45). This residue is absolutely conserved in AlaRSs from all domains of life and bacterial GlyRS and only in these synthetases. Notably, there is a conformational change in the region of Trp-192 in AlaRS in response to ligand binding (45),

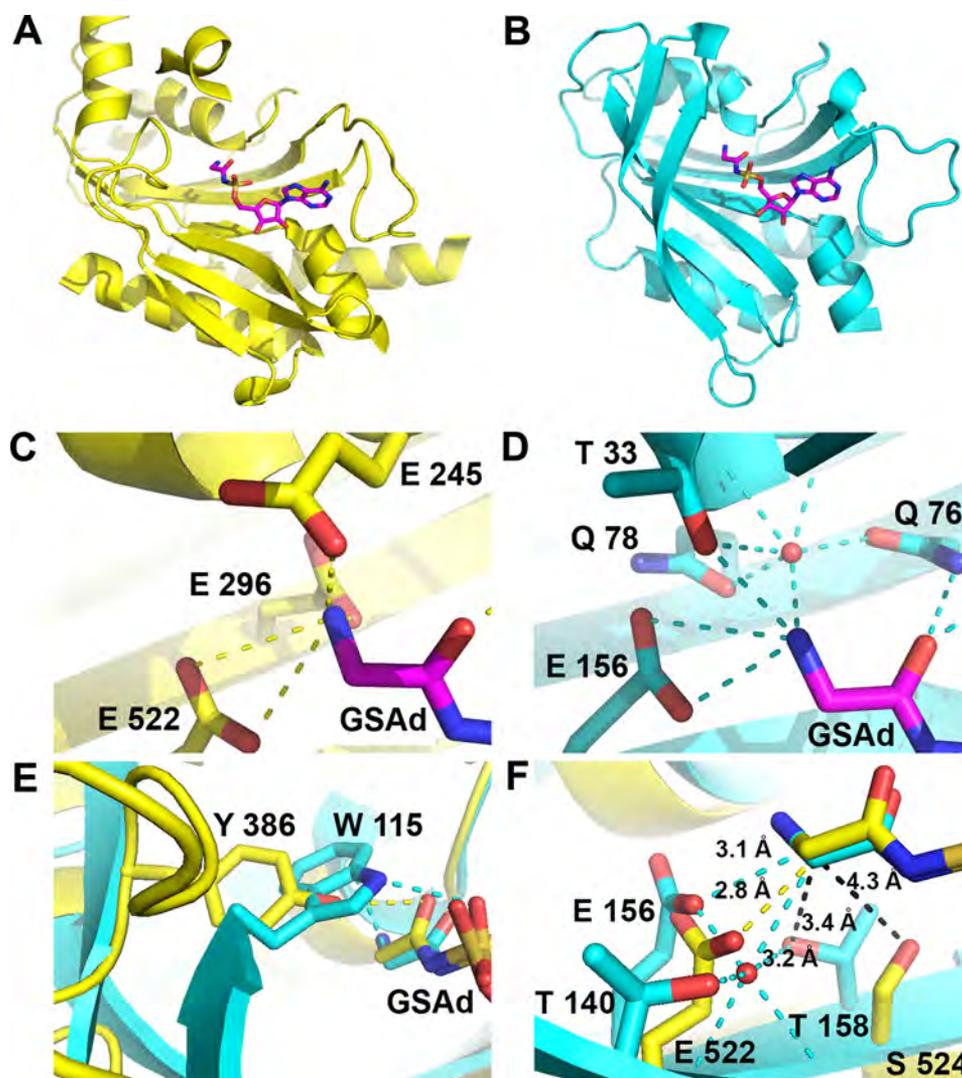


FIGURE 4. **Distinct modes of amino acid recognition in polyphyletic GlyRSs.** *A*, ribbon diagram of the catalytic core of eukaryotic GlyRS (PDB code 2zt8). *B*, ribbon diagram of the catalytic core of bacterial GlyRS. *C*, stick diagram of residues involved in the amino group of glycine recognition in eukaryotic GlyRS. *D*, stick diagram of residues involved in the amino group of glycine recognition in bacterial GlyRS. Thr-33 and Gln-76 have no topological equivalent in eukaryotic GlyRS. All residues involved in the amino group recognition are absolutely conserved in each type of GlyRS. *E*, non-conserved Tyr-386 of the amino acid loop in eukaryotic GlyRS (yellow) compared with the fully conserved Trp-115 of bacterial GlyRS (cyan). *F*, discrimination of non-cognate amino acids by steric hindrance. Non-conserved Ser-524 in eukaryotic GlyRS (yellow) compared with highly conserved Thr-158 and 140 in bacterial GlyRS (cyan).

similar to the one seen in bacterial GlyRS. There is also a highly conserved Ala-31 (Ala-99 in AlaRS), which is 4 Å from Trp-115 (Fig. 11, *B* and *E*), besides an equivalent of Glu-156 (Asp-248 in AlaRS) that recognizes the amino group of the cognate amino acid (Fig. 11, *B*, *C* and *D*). In addition, the location of a water molecule that interacts with the amino moiety of the cognate amino acid (Fig. 11*D*) complements the remarkable similarities in the active site cavity. Interestingly, a partial solvent-mediated recognition of the amino group is also found, for example, in ThrRS (46).

Furthermore, there are four residues, outside the active site cavity, that are highly conserved in AlaRS and bacterial GlyRS (Fig. 11, *B* and *F*). These amino acids are Phe-13, Gly-124, Tyr-141, and Gln-143 in bacterial GlyRS, which correspond to Phe-69, Gly-199, Phe-216, and Gln-218 in AlaRS (Fig. 11, *B* and *F*). As mentioned previously, there are no such residues, indicative of a much higher sequence-structure relatedness, in the comparison of both types of GlyRS.

Taken together, the structural and sequence similarities between the catalytic domains of AlaRS and bacterial GlyRS demonstrate a greater degree of conservation than that observed between the two types of GlyRSs. These similarities and differences allow for a new subclass of aaRSs to be defined that is clearly different from any other subgroup.

Discussion

In this work, we explored three long-standing questions regarding bacterial GlyRS. The first was based on previous experimental evidence suggesting that the full $\alpha_2\beta_2$ tetramer was needed for any reaction step. The second concerned the comparison of substrate recognition and the active sites of the two types of GlyRSs at the molecular level. Finally, the third question deals with the subclassification of class IIc aaRSs and the possible origin of bacterial GlyRS.

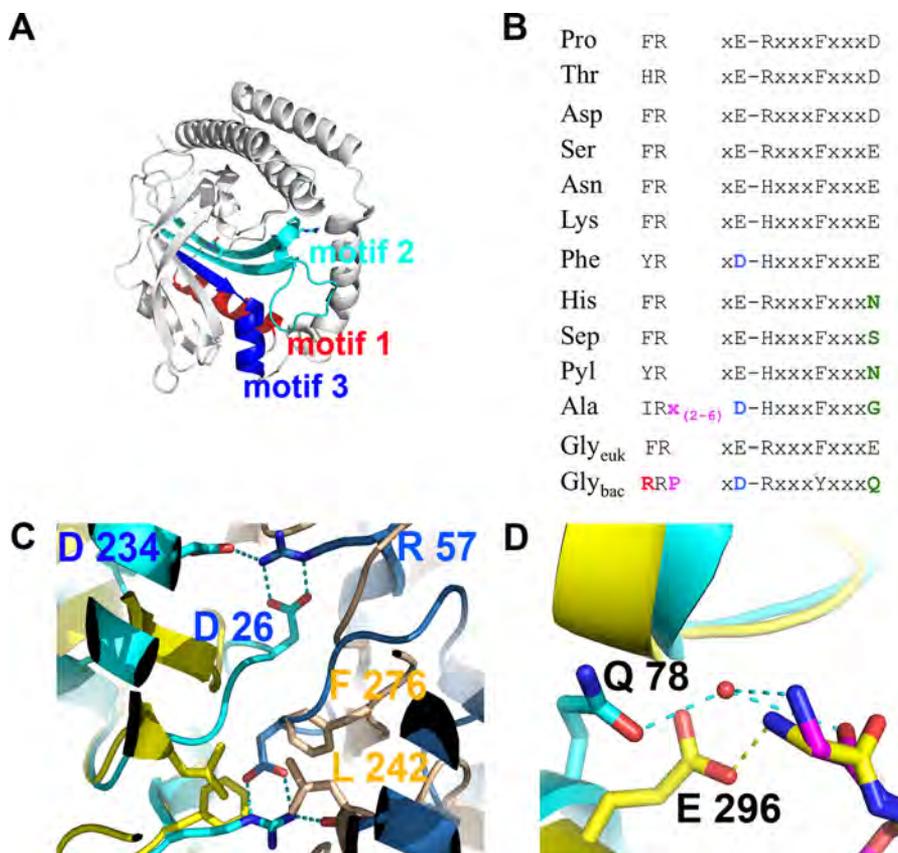


FIGURE 5. Bacterial GlyRS shows key differences with class II aaRSs. *A*, three class II motifs are color-coded and labeled on the bacterial α -GlyRS monomer. Motif 1, residues 3–17; motif 2, residues 48–82; and motif 3, residues 155–169. *B*, signature sequences of motif 2. The consensus sequences were obtained from the analysis of the seed alignments from the Pfam families PF00587 (subclass IIa), PF00152 (subclass IIb), PF01409 (PheRS), PF01411 (AlaRS), and PF02091 (bacterial GlyRS), as well as from our sequence analysis based on structure using Multiseq (see “Experimental Procedures”). The four main differences of bacterial GlyRS with most of class II synthetases are marked in color. Arg-57 and Gln-78 are further analyzed in *C* and *D*, respectively. *C*, differences at the dimerization interface (Arg-57). In bacterial GlyRS (cyan and blue), Arg-57 mediates a salt bridge with Asp-26 and Asp-234. In eukaryotic GlyRS (yellow and light orange), a hydrophobic patch made by Phe-276 and Leu-242 is found in the equivalent region. In some sequences, an Ile residue substitutes Arg-57. This change is correlated with a substitution of Asp-26 by a Thr or Asn residue. *D*, differences in the cognate amino acid binding pocket. Gln-78 in bacterial GlyRS (cyan) helps to bind a solvent molecule that recognizes in part the amino group of the cognate amino acid. In the equivalent position in all class II aaRS, exemplified with eukaryotic GlyRS (yellow), Glu-296 helps to recognize the amino group of glycine.

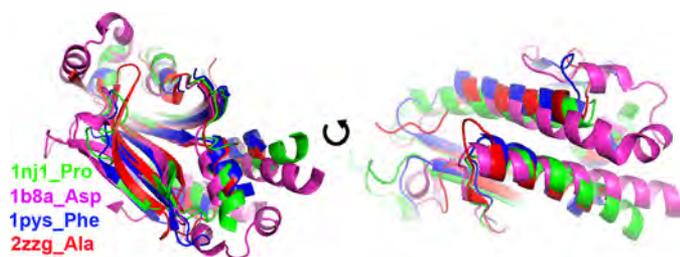


FIGURE 6. Initial selection of a catalytic core of class II aaRSs. A collection of 111 crystal structures from class II aaRSs was chosen according to their diversity in specificity, species, and ligands bound in the active site. The catalytic cores of these structures was selected by including the residues located between motifs 1 and 3 and excluding additional insertions or domains not including common structural motifs for most aaRSs. An overall multiple structural alignment was made using the STAMP algorithm (60) as implemented in MultiSeq in VMD (61, 62). This panel shows the superposition of structures of the catalytic core described and used to obtain the Dali Z-score matrix shown in Table 2, as well as the dendrograms shown in supplemental Fig. S1 (based on structure according to STAMP) and supplemental Fig. S2 (based on sequence according to T-Coffee and built with PhyML). It represents a structure from each obtained subgroup.

α -Subunit of Bacterial GlyRS Is Dimeric and Active for the First Step of Aminoacylation—The α -subunit of bacterial GlyRS contains all the determinants for aminoacyl-adenylate synthesis, and here we show that the activation of glycine can indeed

take place with this chain alone. Given the high sequence conservation of this subunit in bacteria (Pfam family PF2091), it is likely that the same observation would apply to all available enzymes.

The rigorous purification of the α -subunit, which included a heating step to 75 °C for 30 min (T_m of the α -subunit = 83.2 °C, according to thermal shift assays), nickel-nitrilotriacetic acid affinity purification in the presence of relatively high amounts of salts (1 M KCl and 0.5 M urea), and SEC-MALS purification, allowed us to obtain a homogeneous dimeric population of 69.1-kDa protein, free of even trace amounts of higher molecular weight complexes. Previous works have reported that the α -subunit of bacterial GlyRS contains the aminoacyl adenylate synthesis site, but, unlike this study, no activity had been seen. These findings also support the functional existence of an $\alpha_2\beta_2$ quaternary structure for bacterial GlyRS, rather than an $(\alpha\beta)_2$ organization, as seen in most PheRSs (47).

Two Types of GlyRS Recognize Glycine in Distinct Ways—The crystal structure of a bacterial GlyRS α -subunit in a complex with GSA_d allows comparison between both types of GlyRS. To function in a meaningful biological context, these synthetases must exclude any amino acid other than glycine, including the

TABLE 2

DALI Z-score of pairwise comparisons of representative structures of class II aaRSs

Z-scores higher than 18.5 are indicated in bold and grouped by colors that indicate class II subclasses: a, (green); b, (magenta); c, (blue); d, (red).

	PDB	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1ati Gly		18.8	17.0	19.0	19.6	18.2	13.4	14.1	14.0	18.1	14.2	16.1	14.5	14.7
2	1qe0 His	18.7		16.6	16.7	18.5	16.7	12.6	12.5	12.7	18.3	14.7	17.1	12.6	13.7
3	1evk Thr	17.0	16.6		23.2	19.2	17.4	10.0	13.0	10.6	13.7	12.8	14.5	11.9	12.0
4	1ser Ser	19.0	16.7	23.2		21.0	19.1	10.5	14.2	11.7	15.5	14.5	14.9	13.6	12.6
5	1nj1 Pro	19.6	18.5	19.2	21.0		22.4	12.7	12.6	13.0	17.5	14.2	16.1	13.4	13.1
6	2cj9 Ser	18.2	16.7	17.4	19.1	22.4		10.6	11.3	11.6	15.3	12.9	13.7	12.8	13.1
7	1e1o Lys	13.4	12.6	10.1	10.5	12.7	10.6		19.2	23.3	15.2	13.4	14.3	11.5	13.5
8	1b8a Asp	14.1	12.5	13.0	14.2	12.6	11.3	19.2		23.4	15.0	11.8	13.4	12.9	14.2
9	1x54 Asn	13.8	12.7	10.6	11.7	13.0	11.6	23.3	23.4		15.0	11.9	13.3	11.6	12.9
10	2zim Pyl	18.1	18.3	13.8	15.4	17.5	15.3	15.2	15.0	15.0		18.5	19.4	16.1	17.3
11	1pys Phe	11.9	14.7	12.8	12.9	14.2	12.9	13.4	11.8	11.9	18.5		20.9	13.2	14.6
12	2du3 Sep	16.1	17.1	14.5	14.9	16.1	13.7	14.3	13.4	13.3	19.4	20.9		14.0	14.8
13	2zzg Ala	14.5	12.6	11.9	13.6	13.4	12.8	11.5	12.9	11.6	16.1	12.0	14.0		19.5
14	5f5w Gly	14.7	13.7	11.8	12.7	13.1	13.1	13.5	14.2	12.8	17.3	12.2	14.8	19.5	

near-cognate alanine. This recognition problem was solved in different ways by the two types of GlyRSs, which further highlight their distinct origin.

Both types of GlyRSs share 1) the same basic architecture of class II synthetases, a six-stranded antiparallel β -sheet. However, the active site domains are not similar enough to share the same subclass, 2) the same ATP-binding site, which is a common feature of all class II synthetases, built using residues of the characteristic motifs 2 and 3. However, the motif 2 signature is different in the two types of GlyRSs, 3) a glutamic acid residue (Glu-156) that helps to recognize the amino moiety of the cognate amino acid glycine.

Beyond these general similarities, there are key differences at the molecular level between the two types of GlyRSs. First, other than the class II antiparallel β -sheet, there are no other structural elements or residues shared by the two types of GlyRS in this domain. Second, the character of the eukaryotic GlyRS dimerization interface is mainly hydrophobic like most class II synthetases, whereas that of the bacterial GlyRS is stabilized by a number of salt bridges. Third, the most impressive difference between the two classes of GlyRSs is their highly distinct strategies for amino acid recognition. Whereas bacterial GlyRS interacts with the amino group of glycine through five absolutely conserved residues, eukaryotic GlyRS uses three absolutely conserved glutamic acids, creating an electronegative pocket, to perform this task. To further illustrate the different character of this region, an absolutely conserved residue in

eukaryotic GlyRS (Arg-410) that interacts with two highly conserved residues, Glu-245 and Glu-522, is substituted by a highly conserved Phe (142) in bacterial GlyRS.

The dynamics of the reaction, exemplified by the amino acid loop and its topological analog in bacterial GlyRS, also differs in both types of GlyRS. In eukaryotic GlyRS, the hydroxyl group of a non-conserved Tyr residue contacts the carboxyl moiety of the cognate glycine. In bacterial GlyRS, an absolutely conserved Trp (115) residue closes the amino acid binding cavity, forming a cation- π interaction with the amino group of the cognate glycine. Finally, the discrimination of the near-cognate alanine is made in bacterial GlyRS by two nearly absolutely conserved Thr residues (140 and 158) that also bind a solvent molecule, although in eukaryotic GlyRS this task is done mainly through a non-conserved Ser or Ala residue (524), equivalent of Thr-158 (Fig. 4B). The equivalent of Thr-140 is a non-conserved position (Ala, His, or Ser-408) in eukaryotic GlyRS.

Such disparate differences have only been partially observed until now in the case of SerRSs, where a subset of methanogenic archaea possesses a SerRS with different amino acid recognition elements than those found in the majority of SerRSs (48). However, in this case, both types of SerRS appear to belong to subclass IIa of aaRSs (Figs. 8, 9C, and 10 and PDB entries 2dq0 and 2cj9). Overall, the features of the homologous catalytic core of the two types of GlyRSs are so different that it is extremely unlikely for them to share a common origin.

Insights into the Polyphyletic Origins of GlyRSs

TABLE 3

Proposed subgroups of class II aaRSs

Subclass	Type of amino acid	Amino acid	Quaternary structure	aa site ¹	m2 (#outliers) ²	Particular Features ³	Post transfer editing ⁴	
IIa	small hydrophobic, small polar	Gly	α_2	3'		ACBD ⁵ shared with Pro, Thr and HisRSs		
		Pro	α_2	3'			Ala, Cys	
		Thr	α_2	3'			Bacterial editing domain shared with AlaRS	Ser
		His	α_2	3'	1			
IIb	charged and large	Ser	α_2	3'				
		Asp	α_2	3'		ACBD shared with Asn and LysRSs		
		Asn	α_2	3'				
		Lys	α_2	3'				
IIc	large	Pyl ⁶	α_2	3'	1			
		Phe	$(\alpha\beta)_2$	2'	1		Tyr	
		Sep	α_4	2'	1			
IId	hydrophobic and small	Ala	α_2	3'	3	Helicoidal domain on the C-ter ⁷	Gly, Ser	
		Gly	$\alpha_2\beta_2$	3'	4	Helicoidal domain on the C-ter ⁷		

¹ Aminoacylation site (2' or 3' OH of the terminal ribose of the cognate tRNA).

² Number of residues in motif two that do not agree with the canonical sequence.

³ Although this proposed classification is based solely on the structural and sequence features of the catalytic core of class IIaaRSs, there are some characteristics that are shared among its members.

⁴ Editing carried out after the first step of the reaction against standard amino acids.

⁵ ABCD is anticodon binding domain.

⁶ PylRS can be either subclass IIc according to the structural analysis or subclass IIb according to the sequence analysis. It is placed in the frontier in this table because of the aminoacylation site properties of PheRS and SepRS, which are different from PylRS.

⁷ These domains have limited similarity between each other (Dali Z-score 2.6, r.m.s.d. 4.5 Å, % identity = 12).

There Are Four Structural Subclasses of Class II aaRSs—We used the standard accepted structure and sequence comparison methods to detect structural relationships, including the following: an all-against-all similarity matrix using Dali Z-scores (41, 49, 50); a dendrogram built according to the overall C α r.m.s.d. values derived from the STAMP superposition algorithm (42) as implemented in MultiSeq (51, 52); a PCA and r.m.s.d. dendrogram based on a conserved core of 80 C α atoms using bio3D (44); and a dendrogram built with PhyML (53) according to a weighted multiple sequence alignment made using structural information, as implemented in T-Coffee Expresso (43) and validated using Transitive Consistency Score (TCS) (54). Using these tools and the crystal structures from several species and in complex with different ligands, we were able to define four different class II divisions that are in general agreement with previous proposals. We made this classification proposal based mainly on the structural features of a monomeric catalytic core. Remarkably, even if this core is reduced to

a minimal structure of 80 C α atoms, it is still possible to define the four subclasses. It is noteworthy that the new adjustment of class IIc and IId takes into account the physicochemical character of the cognate amino acids, the aminoacylation site, and correct quaternary structures among the four proposed subclasses (Table 3). PheRS and SepRS are clear functional outsiders from the subclass IId (AlaRS and bacterial GlyRS); the corresponding amino acids are much bigger than Ala and Gly; the aminoacylation sites are different (3'-OH for AlaRS and GlyRS and 2'-OH for Phe and SepRS (27)), and the quaternary structures are also different, most notably between bacterial GlyRS ($\alpha_2\beta_2$) and PheRS ($(\alpha\beta)_2$) (Table 3). In this way, subclasses IIc and IId encompass much more homogeneous and coherent members, as opposed to the previous subclass IIc, which included a wide variety of aaRSs (25–27).

These analyses define a clear subgroup for AlaRS and bacterial GlyRS. According to the sequence analysis, however, it is less clear how to classify PylRS, which has been suggested to

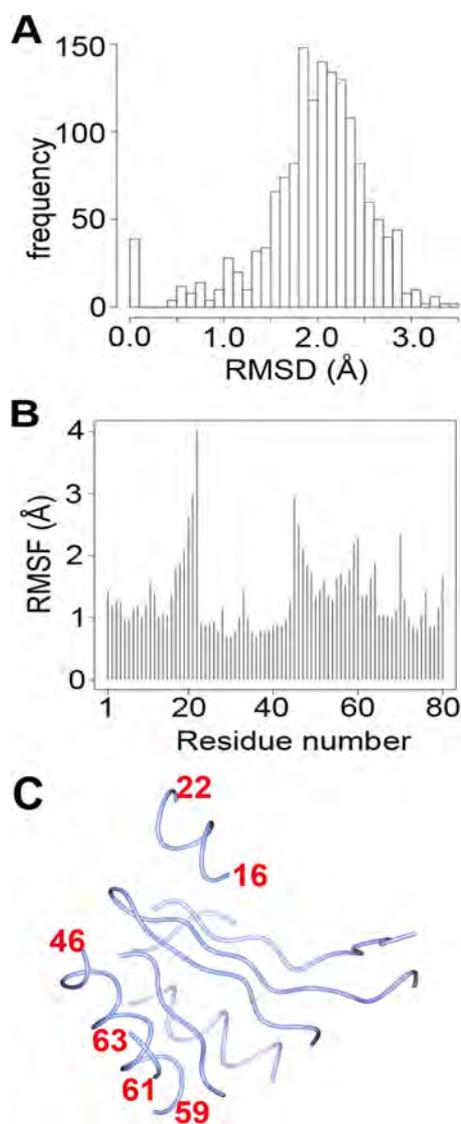


FIGURE 7. Structural features of the conserved core of 80 C α atoms of class II aaRSs. 39 structures from the alignment described in Fig. 6 were further trimmed, and the largest positional differences were excluded to determine the invariant core of class II aaRSs using the Bio3D package in R (68). An r.m.s.d. and a PCA, shown in Figs. 8 and 9, were performed using the 80 residues core depicted in C. A, histogram of pairwise r.m.s.d. values B, root mean square fluctuations (RMSF) plot. C, location on the 80-residue core of the main differences shown in B.

belong to subclass IIc (55) (consistent with our structure-based classification) or IIb (56) (as in our sequence analysis). There also appears to be a discrepancy in the case of HisRS at the sequence level. However, HisRS can be confidently placed in class IIa due to the presence of an anticodon recognition domain located on the C-terminal region, which is another signature of this subclass.

Recent phylogenetic trees have not commented on the proposed subdivision (27, 55). However, a relationship between AlaRS and bacterial GlyRS has even been suggested based on the insertions in the catalytic core and the C-terminal helical domain (57, 58). To address this, in this work we presented a complete analysis of all class II synthetases, defined a new subclass, and explored the similarities between AlaRS and bacterial GlyRS.

Bacterial GlyRS and AlaRS Are Highly Related—In contrast to its comparison with eukaryal GlyRS, bacterial GlyRS shares several key features with AlaRS. First, these two enzymes display similar catalytic cores whose architectures are distinct from other subclasses. Second, they present a similarly modified motif 2 signature: an insertion between an Arg and an acidic residue and a variation in this position of Glu for Asp, where the overwhelming consensus is a Glu residue in most class II aaRSs. Third, in the active site, they share the same strategy for the recognition of the amino group. Fourth, they share an absolutely conserved Trp involved in the recognition of the cognate amino acids and a highly conserved Ala in the vicinity of this site. An interesting case of a structural/functional convergence is seen in this particular feature in methanogenic archaea SerRS, where a different motif approaches a Trp (396) in a different conformation but with the same apparent role (PDB code 2cj9). Fifth, outside of the active site, AlaRS and bacterial GlyRS share four other highly conserved residues.

Proposed Models for the Origin of the Two Types of GlyRS—GlyRSs exhibit an unusual scenario for two enzymes with the same activity and the same overall fold of the catalytic domain. In most cases, these similarities would indicate that they share a common origin. Accordingly, we may consider different scenarios for the structural source of the catalytic core of GlyRSs as follows. 1) A single common ancestor, where an initial $\alpha_2\beta_2$ GlyRS gave rise to another GlyRS α_2 or vice versa. We consider this hypothesis unlikely, according to the analysis presented in this work. 2) Two different ancestors with an emergence of a dimeric eukaryotic type α_2 from an ancient ancestor for all class II synthetases and another emergence of $\alpha_2\beta_2$ GlyRS from AlaRS or a pre-AlaRS.

We believe there are several points that support this second hypothesis. First, there is high structural and sequence similarity between the catalytic cores of AlaRS and bacterial GlyRS. Second, the amino acids Ala and Gly are related in nature, only differing by one methyl group and thus may be recognized in a similar manner. In fact, it is possible to delineate changes in bacterial GlyRS in the positions of Thr-158, Thr-140, and/or Glu-256 or in the equivalent and absolutely conserved Gly-250, Val-215, and Asp-248 positions of AlaRS to partially interconvert the amino acid specificity (45, 59). Third, AlaRS presents editing activity against Gly and Ser to cope with the misincorporation of these amino acids (1/300 for Ser and 1/170 for Gly) in the active site (60). It is possible that a pre-AlaRS, able to aminoacylate both Ala and Gly, may have given rise to GlyRS. Taken together, we favor that the second scenario would more easily solve the paradox of the presence of two proteins with the same function and sharing the same primordial fold but not a single common ancestor. The two types of GlyRS represent a beautiful case of isofunctional paralogs between species (61).

Experimental Procedures

Protein Purification

The cDNA coding for α -AaGlyRS was synthesized by GenScript, optimized according to *E. coli* codon frequency, and subcloned into the pET-28 vector (Novagen). Recombinant α -

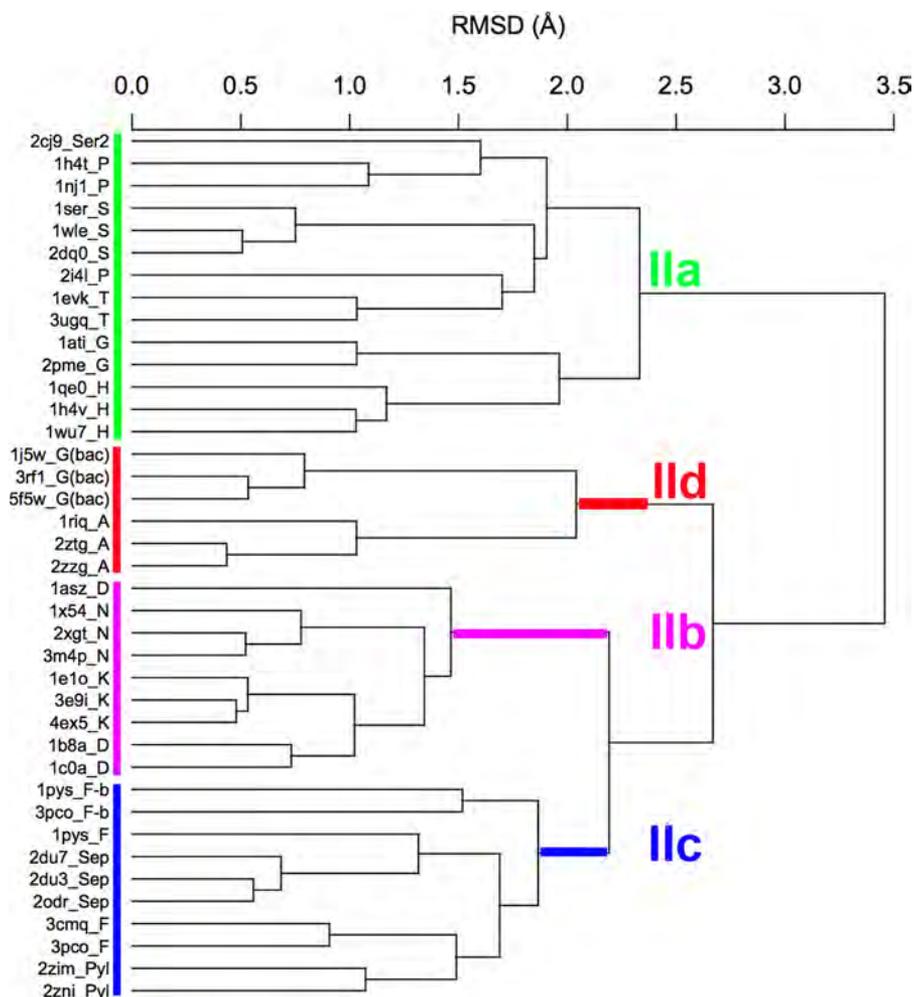


FIGURE 8. **Bacterial GlyRS and AlaRS form a subclass of aaRSs.** r.m.s.d. cluster dendrogram of class II aaRSs. The dendrogram was calculated according to the C α core of 80 residues (see Fig. 7). This tree is in full agreement with previous subclass definitions but with an additional subgroup formed by bacterial GlyRS and AlaRS.

AaGlyRS was expressed with an N-terminal His₆ affinity purification tag in *E. coli* BL21(DE3) cells. Cells were grown at 37 °C until an A₆₀₀ of 1.0 was reached. At that time, cells were induced for 10 h with a final concentration of 1 mM isopropyl β -D-thiogalactopyranoside at 37 °C and were harvested by centrifugation (5515 \times g for 20 min). The resulting cell pellet from 2 liters of culture was frozen at -80 °C until needed. For α -AaGlyRS purification, the cell pellet was thawed and then resuspended in 20 ml of buffer A (50 mM sodium phosphate, pH 6.3, 100 mM potassium chloride, 1 mM MgCl₂). Cells were lysed by sonication on ice and then centrifuged at 20,410 \times g for 30 min. The resulting supernatant was heated at 75 °C for 30 min and then centrifuged again at 20,410 \times g for 30 min. The supernatant was supplemented with the following additives: 1 M KCl, 0.5 M urea, 50 mM imidazole, and 10% glycerol (final concentrations) and applied to a 5-ml Ni²⁺-Sepharose FF column (GE Healthcare) connected to an Äkta FPLC system (GE Healthcare). The column was then washed with buffer A plus additives to remove nonspecifically bound proteins. The target protein was eluted with a 100-ml gradient of buffer A plus 500 mM imidazole. Fractions containing α -AaGlyRS were pooled and concentrated (Amicon Ultra-filter Millipore, 30 kDa) to 15 mg/ml. The protein was >99% pure as judged by denaturing gel electrophore-

sis. The protein was loaded onto a size exclusion chromatography Superdex S-75 10/300 analytical column (GE Healthcare) connected to DAWN Heleos-II and Optilab RI detectors (Wyatt Technologies). The column was run at a flow rate of 0.5 ml/min. One peak was eluted that included 99.9% of the injected mass and corresponded to the dimeric form of α -AaGlyRS, according to the light scattering measurements. Prior to crystallization, the protein was desalted into 20 mM HEPES, pH 7.2, 50 mM NaCl, and 1 mM DTT.

Crystallization and Structure Determination

Crystals of α -AaGlyRS were obtained by sitting-drop vapor diffusion. Drops containing 1 μ l of α -AaGlyRS at 15 mg/ml yielded crystals at 18 °C when mixed with 1 μ l of a reservoir solution containing 30% polyethylene glycol monomethyl ether 2000 and 100 mM potassium thiocyanate. Crystals were soaked for 2 days in a solution containing a final concentration of 28 mM GSAd by mixing 1.3 μ l of the drop containing crystals, 2.5 μ l of reservoir solution, and 1.5 μ l of GSAd at 100 mM. The synthesis of GSAd was performed as described previously (62). Crystals were cryoprotected in a solution prepared with the mother liquor supplemented with 15% glycerol and then flash-frozen in liquid nitrogen. Diffraction data were collected at 100

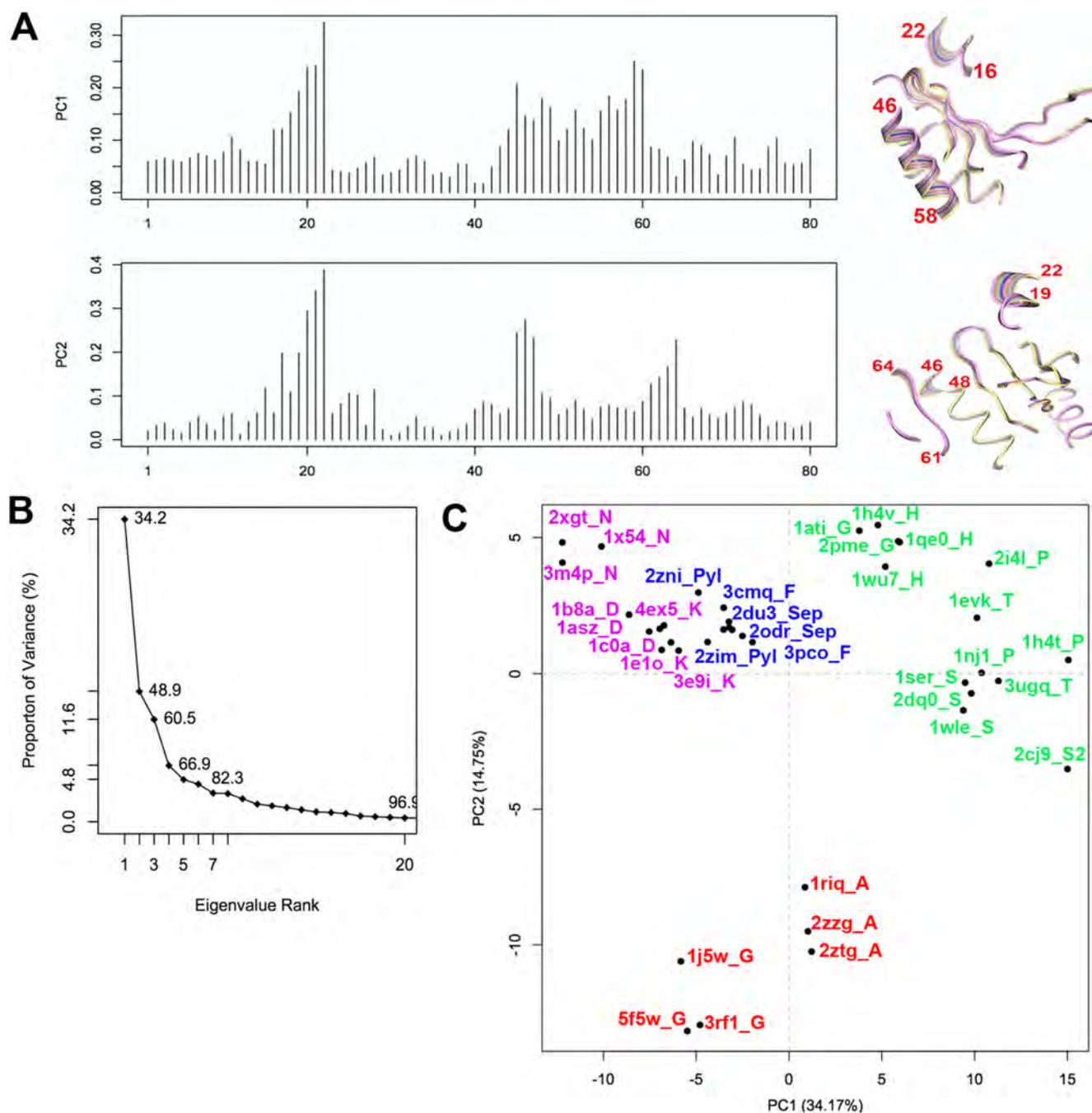


FIGURE 9. **Bacterial GlyRS and AlaRS form a subclass of aaRSs.** PCA of class II aaRSs. *A*, contribution of each of the 80 residues C α core, as defined in the analysis described in Fig. 7, to the first two principal components found. *B*, contribution of each component to the total variance of the distribution of structures. *C*, plot along PC1 and PC2, which describe the change that covers 49% of the structural variance in class II aaRSs.

K using a wavelength of 0.9785 Å at the Life Sciences Collaborative Access Team (LS-CAT) 21-ID-G beamline at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL). Data were indexed and processed with iMOS-FLM (63) and reduced with Aimless (64). The structure was solved by molecular replacement using Balbes (65), with a final search model based on the structure of the α -subunit GlyRS from *Thermotoga maritima* (PDB code 1j5w). There are five monomers in the asymmetric unit, with a crystal solvent content of 57%. The initial R_{free} was 0.35. The ligand GSAd was fitted by means of the LigandFit program of the PHENIX suite (66). Refinement was alternated with manual building/refine-

ment in COOT (67), PHENIX, and the PDB_REDO server (68). Non-crystallographic symmetry restraints were used for the refinement. The model presents no Ramachandran outliers, with 1355 residues (97%) on the favored and 36 residues (3%) on the allowed regions. *B*-factor refinement was limited to one value per amino acid. Monomers A and B present the best adjustment to the electron density according to the wwPDB validation report (0 residues with RSR-Z scores >2), and all figures were prepared using these monomers as templates. Two solvent molecules were located near the glycine recognition site. They could be well fitted and seen on double difference and omit maps for monomer B, and they can be seen as peaks bigger

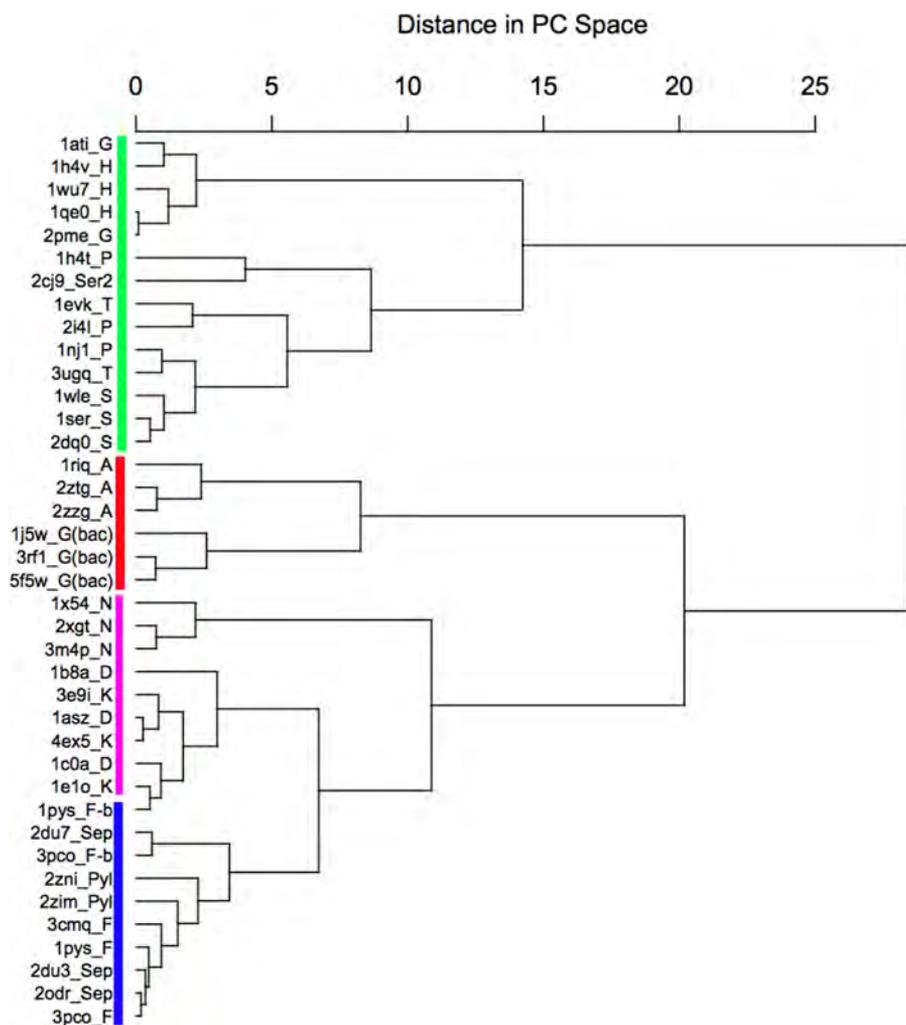


FIGURE 10. **PCA cluster dendrogram of classII aaRSs.** Clustering dendrogram along PC1 and PC2 as described in Fig. 9.

than 3.5σ on difference, $F_o - F_c$ electron density maps on the four other monomers. No other solvent molecules were added to the model. σA -weighted $2F_o - F_c$ -simulated annealing omit maps were used to validate further the quality of the model. Data collection and refinement statistics are summarized in Table 1. Analysis of the interface was performed with the PISA server (69). Because of the absence of ideal geometric parameters from MSDchem, the geometry of the GSAd ligand was adjusted to that observed in PDB code 3hy0 (structure of catalytic fragment of *E.coli* AlaRS in the complex with GlySA). Validation of the GSAd ligand was performed with the ValLigURL server (70). Figures were prepared with PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4, Schrödinger, LLC).

Thermal Shift Assays

According to the protocol of Ref. 71, purified α -AaGlyRS in 25 mM NaH_2PO_4 , 25 mM KCl, and 10 mM MgCl_2 , at a final concentration of 1 mg/ml, was mixed with different concentrations of ligands and a 1:100 dilution of SYPRO Orange dye (Invitrogen) in a final volume of 10 μl . The dye was excited at 490 nm and the emitted light intensity was recorded at 575 nm. Data were collected at 1 $^\circ\text{C}$ intervals from 25 to 99 $^\circ\text{C}$ on

a StepOnePlus real time PCR system and analyzed using the Protein Thermal Shift software from Applied Biosystems. No significant differences were detected using three different values of pH (5, 7, and 8). Assays were performed in triplicate.

Measurements of SAXS Data

Data collection for SAXS of α -AaGlyRS was performed following standard procedures in the European Molecular Biology Laboratory (EMBL) on the storage ring Petra III (DESY, Hamburg, Germany) on the P12 beamline (72). The scattered intensity was recorded as a function of the scattering vector s using a wavelength of 0.124 nm at 23 $^\circ\text{C}$. The measurements were carried out in 30 mM HEPES, pH 7.4, 25 mM NaCl, and 1 mM DTT, with exposure times of 20×0.05 s. The average of the data was normalized and subtracted the scattering attributed to the solvent using automatic procedures (73). The SAXS data were processed using PRIMUS (73), where the values of the forward scattering intensity I_0 and radii of gyration R_g were evaluated from the experimental SAXS patterns using the Guinier approximation; these parameters were also computed from the entire scattering curve using Porod's law by the calculation of the distance distribution function $P(r)$ using the program

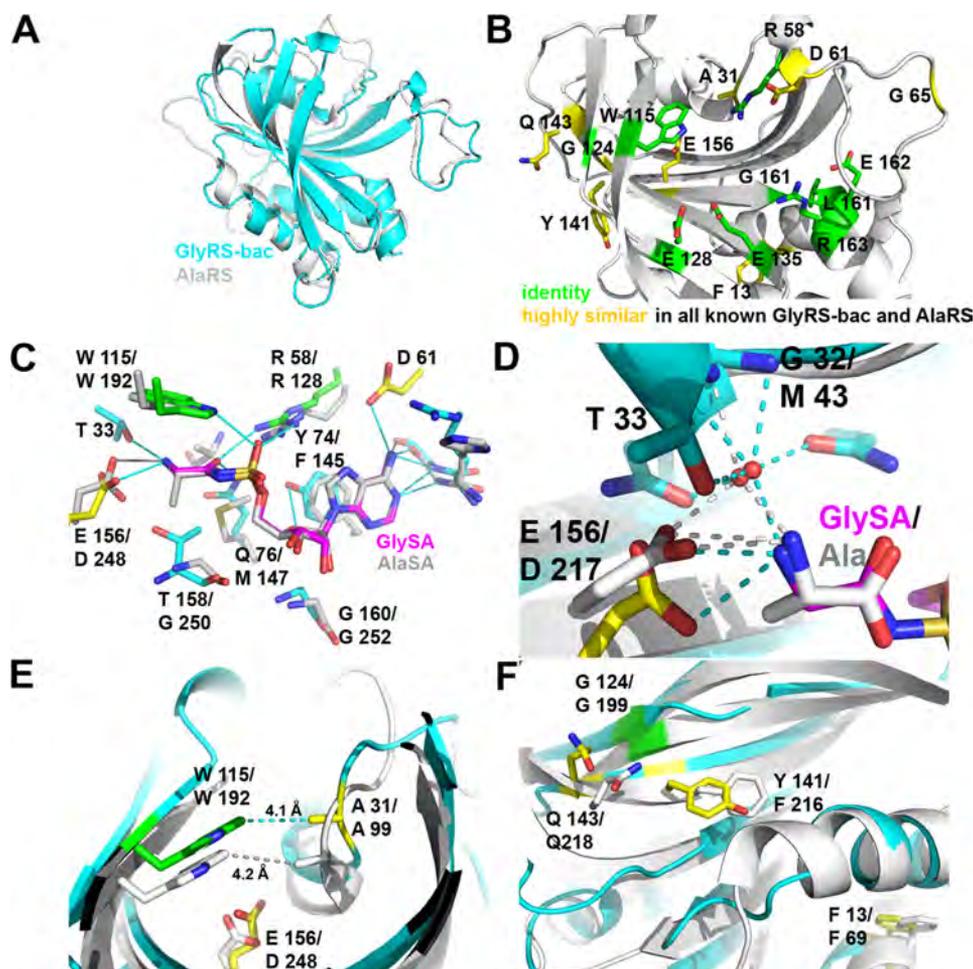


FIGURE 11. Bacterial GlyRS and AlaRS share several features. *A*, superposition of the catalytic core, according to Dali, of bacterial GlyRS (residues 1–169) and AlaRS (PDB code 2zzg, residues 58–260). The Z-score is 19.5 for the 160 residues aligned, with r.m.s.d. of 2.3 and 22% identity. *B*, residues that are highly (yellow) or absolutely conserved (green) and also shared by bacterial GlyRS and AlaRS, according to the structural and sequence alignment of all available sequences, are shown on the structure of bacterial GlyRS. The following panels show the role and/or location of some of these residues in more detail. *C*, superposition of the active site region. Some shared residues (Trp-115/Trp-192) and Glu-156/Asp-248 are related with the cognate amino acid recognition. Thr-158/Gly-250 are proposed to be related with amino acid discrimination. Some other residues (Arg-58/Arg-128), for example, are common class II residues aimed to recognize the ATP molecule. *D*, amino group of the cognate amino acid is in part recognized by a solvent molecule and an acidic residue (Glu-156/Asp-217) in both bacterial GlyRS and AlaRS (PDB code 1yfs, in complex with Ala). A positive electron density peak is also seen in the same position of the solvent molecule for PDB code 2zzg, the model that was used for all the figures of this enzyme shown in this work. *E*, Trp-115/Trp-192 makes a cation- π interaction on the cognate amino acid that helps to close the active site cavity, together with Ala-31/Ala-99 located in the vicinity. *F*, four other residues are absolutely or highly conserved in both enzymes and are not related to substrate recognition.

GNOM (74). Three different approaches were used to determine the molecular weight as follows: the I_0 value, the Porod volume (from the *Pr* function), and the excluded volume (from DAMMIF calculations). The low resolution shape was reconstructed *ab initio* by the DAMMIF method (75). The theoretical SAXS patterns of dimer and monomer from the crystallographic structure were predicted using CRY SOL (76) and were compared graphically with the experimental data using Sasplot (73). The missing flexible residues of the His tag/tobacco etch virus were modeled using Coral (77). The crystallographic structure with the flexible residues was superposed with the *ab initio* envelope using Supcomb (78).

Activity Assays

Glycine Activation—The first step of the aminoacylation reaction was measured using a previously described method to quantify the synthesis of aminoacyl-adenylate (79). Reaction conditions were initially screened for optimization of

pH, enzyme, ATP, and glycine concentrations. The final reaction mix contained the α -subunit of Aa-GlyRS at 40 μ M in 50 mM Tris, pH 8.0, 50 mM KCl, 10 mM MgCl₂, 0.5 mM ATP, 0.25 μ Ci of [α -³²P]ATP and varying concentrations of glycine. Reactions mixes were pre-incubated for 5 min at 45 °C and started by the addition of enzyme. 1- μ l time points were quenched in 4 μ l of buffer containing 0.1% SDS and 400 mM sodium acetate, pH 5.2. 2- μ l aliquots were spotted on pre-washed PEI-cellulose thin layer chromatography (TLC) plates and developed in 0.1 M ammonium acetate, 5% acetic acid solution. The TLC plates were dried and exposed to a phosphor-screen overnight to monitor Gly-[³²P]AMP, [³²P]AMP, and [³²P]ATP.

Images were obtained on an Amersham Biosciences Storm 820 Imager and quantified with ImageQuant TL version 2005. The total fraction of [³²P]AMP was plotted against time to obtain initial velocities; the slopes from the linear data were then plotted against glycine concentration and fit to a hyper-

Insights into the Polyphyletic Origins of GlyRSs

bolic Michaelis-Menten equation using Prism6 software. All experiments were performed in triplicate.

Aminoacylation Reactions—According to Refs. 80–82, the aminoacylation of *A. aeolicus* tRNA^{Gly} was assayed using identical buffer conditions as for the glycine activation reaction, with the exception of using 5 mM ATP and the omission of labeled ATP. Instead *in vitro* transcribed tRNA^{Gly} was radiolabeled at the 3'-end using [α -³²P]ATP and the *E. coli* tRNA nucleotidyltransferase, and labeled tRNA was purified using a QIAquick nucleotide removal kit (Qiagen), followed by buffer exchange using Amicon Ultra 3-kDa filter (Millipore). All reactions were initiated by the addition of enzyme, and 1- μ l aliquots were quenched in 10 μ l of buffer containing 0.1% SDS and 400 mM sodium acetate, pH 5.2, and 0.1 mM PI nuclease SIGMA (N8630-1VL) or 0.1% SDS, 400 mM sodium acetate, pH 5.2. 0.5 μ l of the quenched mix were spotted on pre-washed PEI-cellulose TLC plates and developed and treated like the ones for the glycine activation reaction.

Structural and Sequence Analysis

STAMP—A collection of 111 crystal structures from class II aaRSs was chosen according to their diversity in specificity, species, and ligands bound in the active site to take into account conformational changes. The catalytic cores of these structures was selected by including the residues located between motifs 1 and 3 and excluding additional insertions or domains not comprising common structural motifs for most aaRSs. An example of the trimmed catalytic cores is found in Fig. 6. The PDB entries of the structures used and a partial description of them are found in [supplemental Fig. S1](#). The approximate length of each structure of the working set was 200 residues. An overall multiple structural alignment was made using the STAMP algorithm (42) as implemented in MultiSeq in VMD (51, 52). The program generates trees according to four different criteria as follows: Qres (structure similarity per residue); Qh (structural homology); percent identity, and r.m.s.d. Bacterial GlyRS and AlaRS appeared in a separate group in the trees calculated according to all four parameters. Although all the trees were highly similar to each other, the percent identity and r.m.s.d. trees were in full agreement with previous subclass definitions. The same results were obtained when the algorithm was applied to just a sample of representative aaRSs.

Structure-based multiple sequence alignments were made for the cases of the two types of GlyRSs and AlaRS. All the sequences contained in Pfam families PF01411 (AlaRS), PF02091 (bacterial GlyRS) and PF00587 (subclass IIa, but trimmed for eukaryal GlyRS) were analyzed. The catalytic cores were selected and aligned with the ClustalW plug-in in MultiSeq. Non-redundant sets were generated using the sequence QR tool in MultiSeq (83). A profile alignment was calculated taking into account the structures aligned for each type of aaRS.

DALI—A sample of structural cores as defined above was used to perform an all-against-all three-dimensional structural comparison with the Dali pairwise comparison server. A matrix was generated based on the resulting Z-score values. Although no tree was derived from this matrix, four clear groups are defined, taking into account a cutoff of 18.5. The Z-score value for structures belonging to other subgroups was between 11.9

and 16.5. In most cases other than PylRS, a clear gap is seen between members and non-members of the subclass.

T-Coffee Espresso—A non-redundant set of sequences of the catalytic core of 63 structures from different species used for the alignment with STAMP were aligned using structural information with the T-Coffee Espresso server (43). The obtained alignment was further validated by means of the suggested TCS method, which identifies the most correct positions on the alignment (54). An estimated phylogenetic tree was built with PhyML (53) according to the tcs_weighted obtained file, where the sequences are ordered according to their TCS score. The tree was visualized with Archaeopteryx (84).

Root Mean Square Deviation and PCA with Bio3D—The analysis was performed using the Bio3D package in R (44). 39 structures derived from the STAMP analysis were selected, three of different species and in complex with different ligands for each type of aaRS. The aligned structures were used in Bio3D to define an invariant core with the minimal structural variance among all the protein structures. This core of 80 C α atoms was the basis for the r.m.s.d. dendrogram calculation and PCA. The structures were superimposed onto this core, and the variances shown by PC1 and PC2 were used to define the four different subgroups.

Author Contributions—M. I. V. S., A. T. L., A. C. D. B., and D. M. conceived the project. M. I. V. S. and H. A. S. S. purified α -AaGlyRS and performed thermal shift assays. A. R. H. and L. B. C. performed the activity experiments. M. I. V. S. purified, crystallized, and solved the structure of the α -AaGlyRS-GlyAMS complex. M. I. V. S., H. M., and D. S. performed and analyzed the SAXS experiments. R. F. and M. G. synthesized the Gly-SA compound. D. M., B. B., M. A., M. I. V. S., and A. T. L. performed the structural and sequence analysis. All authors analyzed the data and contributed to manuscript preparation. M. I. V. S. and A. T. L. wrote the manuscript.

Acknowledgments—We thank Elizabeth Nallely Cabrera González and Nevena Cvetesic for scientific and technical guidance, Prof. John J. Perona for the gift of gene coding for the tRNA nucleotidyltransferase, and Martha Contreras and Beatriz Aguirre López for technical support. The critical reading by Jonathan Ipsaro and Rajan Sankaranarayanan is greatly acknowledged. Use of the Advanced Photon Source, an Office of Science User Facility operated for the United States Department of Energy Office of Science at Argonne National Laboratory, was supported by the United States Department of Energy under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corp. and Michigan Technology Tri-Corridor Grant 085P1000817.

References

1. Berg, P., and Offengand, E. J. (1958) An enzymatic mechanism for linking amino acids to RNA. *Proc. Natl. Acad. Sci. U.S.A.* **44**, 78–86
2. Ibba, M., and Soll, D. (2000) Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **69**, 617–650
3. Pang, Y. L., Poruri, K., and Martinis, S. A. (2014) tRNA synthetase: tRNA aminoacylation and beyond. *Wiley Interdiscip. Rev. RNA* **5**, 461–480
4. Berg, P. (1956) Acyl adenylate—synthesis and properties of adenyl acetate. *J. Biol. Chem.* **222**, 1015–1023
5. Preiss, J., Berg, P., Ofengand, E. J., Bergmann, F. H., and Dieckmann, M. (1959) The chemical nature of the RNA-amino acid compound formed by amino acid-activating enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **45**, 319–328

6. Berg, P., Bergmann, F. H., Ofengand, E. J., and Dieckmann, M. (1961) Enzymic synthesis of amino acyl derivatives of ribonucleic acid. *J. Biol. Chem.* **236**, 1735–1740
7. Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347**, 203–206
8. Cusack, S., Berthet-Colominas, C., Härtle, M., Nassar, N., and Leberman, R. (1990) A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. *Nature* **347**, 249–255
9. Burbaum, J. J., Starzyk, R. M., and Schimmel, P. (1990) Understanding structural relationships in proteins of unsolved three-dimensional structure. *Proteins* **7**, 99–111
10. Burbaum, J. J., and Schimmel, P. (1991) Structural relationships and the classification of aminoacyl-transfer RNA-synthetases. *J. Biol. Chem.* **266**, 16965–16968
11. Wolf, Y. I., Aravind, L., Grishin, N. V., and Koonin, E. V. (1999) Evolution of aminoacyl-tRNA synthetases—analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res.* **9**, 689–710
12. O'Donoghue, P., and Luthey-Schulten, Z. (2003) On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol. Mol. Biol. Rev.* **67**, 550–573
13. Schimmel, P. (2008) Development of tRNA synthetases and connection to genetic code and disease. *Protein Sci.* **17**, 1643–1652
14. Shiba, K., Schimmel, P., Motegi, H., and Noda, T. (1994) Human glycyl-tRNA synthetase. Wide divergence of primary structure from bacterial counterpart and species-specific aminoacylation. *J. Biol. Chem.* **269**, 30049–30055
15. Fournier, G. P., Andam, C. P., and Gogarten, J. P. (2015) Ancient horizontal gene transfer and the last common ancestors. *BMC Evol. Biol.* **15**, 70
16. Chien, C. I., Chen, Y. W., Wu, Y. H., Chang, C. Y., Wang, T. L., and Wang, C. C. (2014) Functional substitution of a eukaryotic glycyl-tRNA synthetase with an evolutionarily unrelated bacterial cognate enzyme. *PLoS ONE* **9**, e94659
17. Logan, D. T., Mazauric, M. H., Kern, D., and Moras, D. (1995) Crystal structure of glycyl-transfer-RNA synthetase from *Thermus thermophilus*. *EMBO J.* **14**, 4156–4167
18. Mazauric, M. H., Reinbolt, J., Lorber, B., Ebel, C., Keith, G., Giegé, R., and Kern, D. (1996) An example of non-conservation of oligomeric structure in prokaryotic aminoacyl-tRNA synthetases. Biochemical and structural properties of glycyl-tRNA synthetase from *Thermus thermophilus*. *Eur. J. Biochem.* **241**, 814–826
19. Freist, W., Logan, D. T., and Gauss, D. H. (1996) Glycyl-tRNA synthetase. *Biol. Chem. Hoppe-Seyler* **377**, 343–356
20. Okamoto, K., Kuno, A., and Hasegawa, T. (2005) Recognition sites of glycine tRNA for glycyl-tRNA synthetase from hyperthermophilic archaeon, *Aeropyrum pernix* K1. *Nucleic Acids Symposium Series* **49**, 299–300
21. Ostrem, D. L., and Berg, P. (1974) Glycyl transfer ribonucleic acid synthetase from *Escherichia coli*: purification, properties, and substrate binding. *Biochemistry* **13**, 1338–1348
22. Mazauric, M. H., Keith, G., Logan, D., Kreutzer, R., Giegé, R., and Kern, D. (1998) Glycyl-tRNA synthetase from *Thermus thermophilus*—wide structure divergence with other prokaryotic glycyl-tRNA synthetases and functional inter-relation with prokaryotic and eukaryotic glycylation systems. *Eur. J. Biochem.* **251**, 744–757
23. Tang, S. N., and Huang, J. F. (2005) Evolution of different oligomeric glycyl-tRNA synthetases. *FEBS Lett.* **579**, 1441–1445
24. Tan, K., Zhou, M., Zhang, R., Anderson, W. F., and Joachimiak, A. (2012) The crystal structures of the α -subunit of the $\alpha(2)\beta(2)$ tetrameric glycyl-tRNA synthetase. *J. Struct. Funct. Genomics* **13**, 233–239
25. Woese, C. R., Olsen, G. J., Ibba, M., and Söll, D. (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **64**, 202–236
26. Perona, J. J., and Hadd, A. (2012) Structural diversity and protein engineering of the aminoacyl-tRNA synthetases. *Biochemistry* **51**, 8705–8729
27. Englert, M., Moses, S., Hohn, M., Ling, J., O'Donoghue, P., and Söll, D. (2013) Aminoacylation of tRNA 2'- or 3'-hydroxyl by phosphoseryl- and pyrrolysyl-tRNA synthetases. *FEBS Lett.* **587**, 3360–3364
28. Ribas de Pouplana, L., and Schimmel, P. (2001) Two classes of tRNA synthetases suggested by sterically compatible dockings on tRNA acceptor stem. *Cell* **104**, 191–193
29. Nawaz, M., and Martinis, S. A. (2008) *Wiley Encyclopedia of Chemical Biology*. 10.1002/9780470048672.webc008. John Wiley & Sons, Inc., New York
30. Giege, R., and Springer, M. (2013) Aminoacyl tRNA synthetases in the bacterial world. *EcoSal. Plus.* 10.1128/ecosalplus.ESP-0002–201
31. Klipcan, L., Saftro, M., and Schimmel, P. (2013) Anticodon G recognition by tRNA synthetases mimics the tRNA core. *Trends Biochem. Sci.* **38**, 229–232
32. Caetano-Anollés, G., Wang, M., and Caetano-Anollés, D. (2013) Structural phylogenomics retrodicts the origin of the genetic code and uncovers the evolutionary impact of protein flexibility. *PLoS ONE* **8**, e72225
33. Toth, M. J., and Schimmel, P. (1990) A mutation in the small (α) subunit of glycyl-transfer RNA-synthetase affects amino-acid activation and subunit association parameters. *J. Biol. Chem.* **265**, 1005–1009
34. Toth, M. J., and Schimmel, P. (1990) Deletions in the large (β) subunit of a hetero-oligomeric aminoacyl-transfer RNA-synthetase. *J. Biol. Chem.* **265**, 1000–1004
35. Ostrem, D. L., and Berg, P. (1970) Glycyl-transfer RNA synthetase—an oligomeric protein containing dissimilar subunits. *Proc. Natl. Acad. Sci. U.S.A.* **67**, 1967–1974
36. McDonald, T., Breite, L., Pangburn, K. L., Hom, S., Manser, J., and Nagel, G. M. (1980) Overproduction, purification, and subunit structure of *Escherichia coli* glycyl transfer ribonucleic-acid synthetase. *Biochemistry* **19**, 1402–1409
37. Nagel, G. M., Cumberledge, S., Johnson, M. S., Petrella, E., and Weber, B. H. (1984) The β subunit of *E. coli* glycyl-tRNA synthetase plays a major role in tRNA recognition. *Nucleic Acids Res.* **12**, 4377–4384
38. Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008) Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. *J. Mol. Biol.* **378**, 634–652
39. Arnez, J. G., Dock-Bregeon, A. C., and Moras, D. (1999) Glycyl-tRNA synthetase uses a negatively charged pit for specific recognition and activation of glycine. *J. Mol. Biol.* **286**, 1449–1459
40. Schimmel, P., Giegé, R., Moras, D., and Yokoyama, S. (1993) An operational RNA code for amino acids and possible relationship to genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8763–8768
41. Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549
42. Russell, R. B., and Barton, G. J. (1992) Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins* **14**, 309–323
43. Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., and Notredame, C. (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.* **34**, W604–W608
44. Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A., and Caves, L. S. (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696
45. Guo, M., Chong, Y. E., Shapiro, R., Beebe, K., Yang, X. L., and Schimmel, P. (2009) Paradox of mistranslation of serine for alanine caused by AlaRS recognition dilemma. *Nature* **462**, 808–812
46. Sankaranarayanan, R., Dock-Bregeon, A. C., Rees, B., Bovee, M., Caillet, J., Romby, P., Francklyn, C. S., and Moras, D. (2000) Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase. *Nat. Struct. Biol.* **7**, 461–465
47. Finarov, I., Moor, N., Kessler, N., Klipcan, L., and Saftro, M. G. (2010) Structure of human cytosolic phenylalanyl-tRNA synthetase: evidence for kingdom-specific design of the active sites and tRNA binding patterns. *Structure* **18**, 343–353
48. Bilokapic, S., Maier, T., Ahel, D., Gruic-Sovulj, I., Söll, D., Weygand-Durasevic, I., and Ban, N. (2006) Structure of the unusual seryl-tRNA

- synthetase reveals a distinct zinc-dependent mode of substrate recognition. *EMBO J.* **25**, 2498–2509
49. Lichtarge, O. (2001) Getting past appearances: the many-fold consequences of remote homology. *Nat. Struct. Biol.* **8**, 918–920
 50. Majorek, K. A., Dunin-Horkawicz, S., Steczkiewicz, K., Muszewska, A., Nowotny, M., Ginalski, K., and Bujnicki, J. M. (2014) The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification. *Nucleic Acids Res.* **42**, 4160–4179
 51. Roberts, E., Eargle, J., Wright, D., and Luthey-Schulten, Z. (2006) MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* **7**, 382
 52. Eargle, J., Wright, D., and Luthey-Schulten, Z. (2006) Multiple alignment of protein structures and sequences for VMD. *Bioinformatics* **22**, 504–506
 53. Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321
 54. Chang, J. M., Di Tommaso, P., and Notredame, C. (2014) TCS: A new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. *Mol. Biol. Evol.* **31**, 1625–1637
 55. Kavran, J. M., Gundllapalli, S., O'Donoghue, P., Englert, M., Söll, D., and Steitz, T. A. (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11268–11273
 56. Fournier, G. P., Huang, J., and Gogarten, J. P. (2009) Horizontal gene transfer from extinct and extant lineages: biological innovation and the coral of life. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 2229–2239
 57. Swairjo, M. A., Otero, F. J., Yang, X. L., Lovato, M. A., Skene, R. J., McRee, D. E., Ribas de Pouplana, L., and Schimmel, P. (2004) Alanyl-tRNA synthetase crystal structure and design for acceptor-stem recognition. *Mol. Cell* **13**, 829–841
 58. Smith, T. F., and Hartman, H. (2015) The evolution of class II aminoacyl-tRNA synthetases and the first code. *FEBS Lett.* **589**, 3499–3507
 59. Beebe, K., Ribas De Pouplana, L., and Schimmel, P. (2003) Elucidation of tRNA-dependent editing by a class II tRNA synthetase and significance for cell viability. *EMBO J.* **22**, 668–675
 60. Schimmel, P. (2011) Mistranslation and its control by tRNA synthetases. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2965–2971
 61. Gabaldón, T., and Koonin, E. V. (2013) Functional and evolutionary implications of gene orthology. *Nat. Rev. Genet.* **14**, 360–366
 62. Redwan, I. N., Ingemyr, H. J., Ljungdahl, T., Lawson, C. P., and Grotli, M. (2012) Solid-Phase synthesis of 5'-O-[N-(acyl)sulfamoyl]adenosine derivatives. *Eur. J. Org. Chem.* **2012**, 3665–3669
 63. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281
 64. Evans, P. R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282–292
 65. Long, F., Vagin, A. A., Young, P., and Murshudov, G. N. (2008) BALBES: a molecular-replacement pipeline. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 125–132
 66. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
 67. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
 68. Joosten, R. P., Long, F., Murshudov, G. N., and Perrakis, A. (2014) The PDB_REDO server for macromolecular structure model optimization. *IUCr. J.* **1**, 213–220
 69. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
 70. Kleywegt, G. J., and Harris, M. R. (2007) ValLigURL: a server for ligand-structure comparison and validation. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 935–938
 71. Niesen, F. H., Berglund, H., and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–2221
 72. Blanchet, C. E., Hermes, C., Svergun, D. I., and Fiedler, S. (2015) A small and robust active beamstop for scattering experiments on high-brilliance undulator beamlines. *J. Synchrotron Radiat.* **22**, 461–464
 73. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282
 74. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503
 75. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346
 76. Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRYSOLE—A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773
 77. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350
 78. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41
 79. Gruic-Sovulj, I., Uter, N., Bullock, T., and Perona, J. J. (2005) tRNA-dependent aminoacyl-adenylate hydrolysis by a nonediting class I aminoacyl-tRNA synthetase. *J. Biol. Chem.* **280**, 23978–23986
 80. Ledoux, S., and Uhlenbeck, O. C. (2008) [³²P]-labeling tRNA with nucleotidyltransferase for assaying aminoacylation and peptide bond formation. *Methods* **44**, 74–80
 81. Rodríguez-Hernández, A., Bhaskaran, H., Hadd, A., and Perona, J. J. (2010) Synthesis of Glu-tRNA(Gln) by engineered and natural aminoacyl-tRNA synthetases. *Biochemistry* **49**, 6727–6736
 82. Rodríguez-Hernández, A., and Perona, J. J. (2011) Heat maps for intramolecular communication in an RNP enzyme encoding glutamine. *Structure* **19**, 386–396
 83. O'Donoghue, P., and Luthey-Schulten, Z. (2005) Evolutionary profiles derived from the QR factorization of multiple structural alignments gives an economy of information. *J. Mol. Biol.* **346**, 875–894
 84. Han, M. V., and Zmasek, C. M. (2009) phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* **10**, 356

Conserved Sequence Preferences Contribute to Substrate Recognition by the Proteasome*^[S]

Received for publication, March 23, 2016, and in revised form, May 16, 2016. Published, JBC Papers in Press, May 17, 2016, DOI 10.1074/jbc.M116.727578

Houqing Yu^{†§}, Amit K. Singh Gautam[‡], Shameika R. Wilmington^{†§}, Dennis Wylie[¶], Kirby Martinez-Fonts^{‡§}, Grace Kago[‡], Marie Warburton[‡], Sreenivas Chavali^{||}, Tomonao Inobe^{**}, Ilya J. Finkelstein[‡], M. Madan Babu^{||1}, and Andreas Matouschek^{†§2}

From the [†]Department of Molecular Biosciences and the [¶]Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, Texas 78712, the [‡]Department of Molecular Biosciences, Northwestern University, Evanston, Illinois 60208, the ^{||}Medical Research Council, Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom, and ^{**}Frontier Research Core for Life Sciences, University of Toyama, 3190 Gofuku, Toyama-shi, Toyama 930-8555, Japan

The proteasome has pronounced preferences for the amino acid sequence of its substrates at the site where it initiates degradation. Here, we report that modulating these sequences can tune the steady-state abundance of proteins over 2 orders of magnitude in cells. This is the same dynamic range as seen for inducing ubiquitination through a classic N-end rule degron. The stability and abundance of His3 constructs dictated by the initiation site affect survival of yeast cells and show that variation in proteasomal initiation can affect fitness. The proteasome's sequence preferences are linked directly to the affinity of the initiation sites to their receptor on the proteasome and are conserved between *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and human cells. These findings establish that the sequence composition of unstructured initiation sites influences protein abundance *in vivo* in an evolutionarily conserved manner and can affect phenotype and fitness.

The proteasome controls the concentrations of thousands of regulatory proteins, removes misfolded and damaged proteins in cells, and digests foreign proteins to produce peptides that are displayed by the major histocompatibility complex (MHC) at the cell surface (1, 2). Proteins are targeted to the proteasome by ubiquitin chains, but these chains also have other biological functions (3, 4). The pattern through which ubiquitin moieties are linked and their number in the chains convey some targeting specificity. For example, chains of four or more ubiquitin

moieties linked through Lys-48 of ubiquitin are thought to be the canonical proteasome targeting signal (5), whereas short tags of one ubiquitin or chains linked through Lys-63 are associated with membrane trafficking (6–8) and DNA repair (9).

Recent research shows that a much broader spectrum of ubiquitin linkages is associated with proteasome degradation (10, 11). In other cases, the same ubiquitin linkage can target proteins to different cellular processes (2, 3, 12–18). Even ubiquitin-tagged proteins that are recognized by the proteasome are not always degraded. It has been proposed that competition between different ubiquitin receptors (19) can protect some ubiquitinated proteins from proteasomal degradation. Additionally, disassembly of some polyubiquitin chains by specialized deubiquitinating enzymes on the proteasome can inhibit degradation (20).

Targeting information may also be encoded directly in the substrate protein itself. Efficient degradation requires the presence of an unstructured or disordered region in the substrate protein. The proteasome engages the substrate's disordered region to initiate unfolding and translocation to the proteolytic sites (21–24). The selection of the initiation site by the proteasome is part of the mechanism that confers specificity to degradation (25–28). Indeed, the absence of proteasomal initiation sites explains the unexpected stability of several natural proteins in yeast, such as the ubiquitin-conjugating enzyme Cdc34 and the proteasomal shuttle receptor Rad23 (27, 28). The requirement of initiation sites for degradation is also reflected in the half-lives of proteins measured in large scale proteomics experiments on yeast and mammalian cells (23, 28, 29). Proteins that contain predicted proteasome initiation regions have shorter half-lives than proteins that lack these regions (23, 28).

We recently investigated the proteasome's preferences for the amino acid sequence of its initiation sites *in vitro* by comparing the rates by which purified yeast proteasomes degraded a series of model proteins (28). However, it is possible that the proteasome's intrinsic preferences are overridden *in vivo*. For example, regulatory proteins such as p97/Cdc48/VCP may deliver already unfolded proteins to the proteasome (30–36). These considerations raise several questions. Are the initiation sequence preferences identified *in vitro* operational *in vivo*? If so, do the preferences affect protein abundance substantially and influence phenotype and cell fitness? How does the magnitude of the effect compare with that achieved by the regulation

* This work was supported in part by National Institutes of Health Grants U54GM105816, R21CA196456, and R21CA191664 (to A. M.), Welch Foundation Grants F-1817 (to A. M.) and F-1808 (to I. J. F.), Cancer Prevention and Research Institute of Texas Grant RP140328 (to A. M.), the Program to Disseminate Tenure Track System from the Ministry of Education, Culture, Sports, Science and Technology, Japan (to T. I.), the United Kingdom Medical Research Council Grant MC_U105185859 (to M. M. B.), and the European Molecular Biology Organization Long-Term Fellowship (to S. C.) and Young Investigator Program (to M. M. B.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

✂ Author's Choice—Final version free via Creative Commons CC-BY license.

[S] This article contains supplemental Tables S1–S3.

¹ Lister Institute Research Prize Fellow.

² To whom correspondence should be addressed: Dept. of Molecular Biosciences, The University of Texas at Austin, 2506 Speedway Stop A5000, Austin, TX 78712. Tel.: 512-232-4045; Fax: 512-471-1218; E-mail: matouschek@austin.utexas.edu.

of ubiquitination? Are the sequence preferences for the initiation site conserved in different organisms?

Here, we investigate whether the proteasome has sequence preferences in cells using a scalable assay to monitor protein stability. We find that changes in the initiation sequence in model proteasome substrates tune protein degradation rates and adjust protein steady-state abundance over 2 orders of magnitude. These differences in abundance correspond to the dynamic range that is achieved by controlling ubiquitination. Modulating proteasomal initiation can change protein abundance sufficiently to affect cellular fitness by targeting His3 protein to degradation. We observe that the proteasome's sequence preferences are conserved between *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and cultured human cells (HEK293 cells). The sequence preferences reflect the binding affinity of the initiation regions to their recognition site on the proteasome and are related to specific physical and chemical properties of the amino acid sequences.

Results

Monitoring Proteasomal Degradation in Vivo—To investigate protein targeting to the proteasome in *S. cerevisiae*, we assayed the abundance of fluorescent proteasome substrates by measuring total cell fluorescence. We constructed YFP variants with or without different degradation signals or degrons and expressed them from the constitutive *tpi1* promoter (37) on a CEN plasmid in *S. cerevisiae*. To correct for differences in plasmid copy number, transcription and translation levels, and cell size, we also expressed the RFP³ dsRed-Express2 (38) from a constitutive *pgk1* promoter (37) on the same plasmid (Fig. 1, A and B). The ratio of YFP over RFP fluorescence of individual cells served as a measure of the steady-state concentration of the YFP variants (39, 40). A high numerical value of the ratio of yellow fluorescence intensity to red fluorescence intensity (high YFP/RFP ratio) reports high YFP protein abundance and thus inefficient proteasomal degradation, and vice versa.

We targeted YFP to the proteasome by attaching the UbL domain of yeast Rad23 to its N terminus. The UbL domain is recognized by receptors on the proteasome (41–43), but the UbL domain and YFP lack disordered regions at which the proteasome can initiate degradation so that the UbL-YFP protein accumulated in cells and was easily detected by flow cytometry (Fig. 1B). Treating the cells with bortezomib, which partially inhibits the proteasome in *S. cerevisiae* (44, 45), did not increase UbL-YFP levels noticeably (Fig. 1B). Attaching a 51-residue C-terminal tail derived from subunit 9 of the F_0 component of the *Neurospora crassa* ATP synthase (Su9, sequence P in supplemental Table S1) to UbL-YFP reduced the yellow cell fluorescence to low levels slightly above the background fluorescence of cells not expressing YFP (Fig. 1B), suggesting that the UbL-YFP-Su9 protein was degraded efficiently. The red fluorescence of the RFP reference protein was not affected significantly (Fig. 1B). In the presence of bortezomib, UbL-YFP-Su9 fluorescence increased, showing that degradation depends on the proteasome, but to levels lower than UbL-YFP, as expected

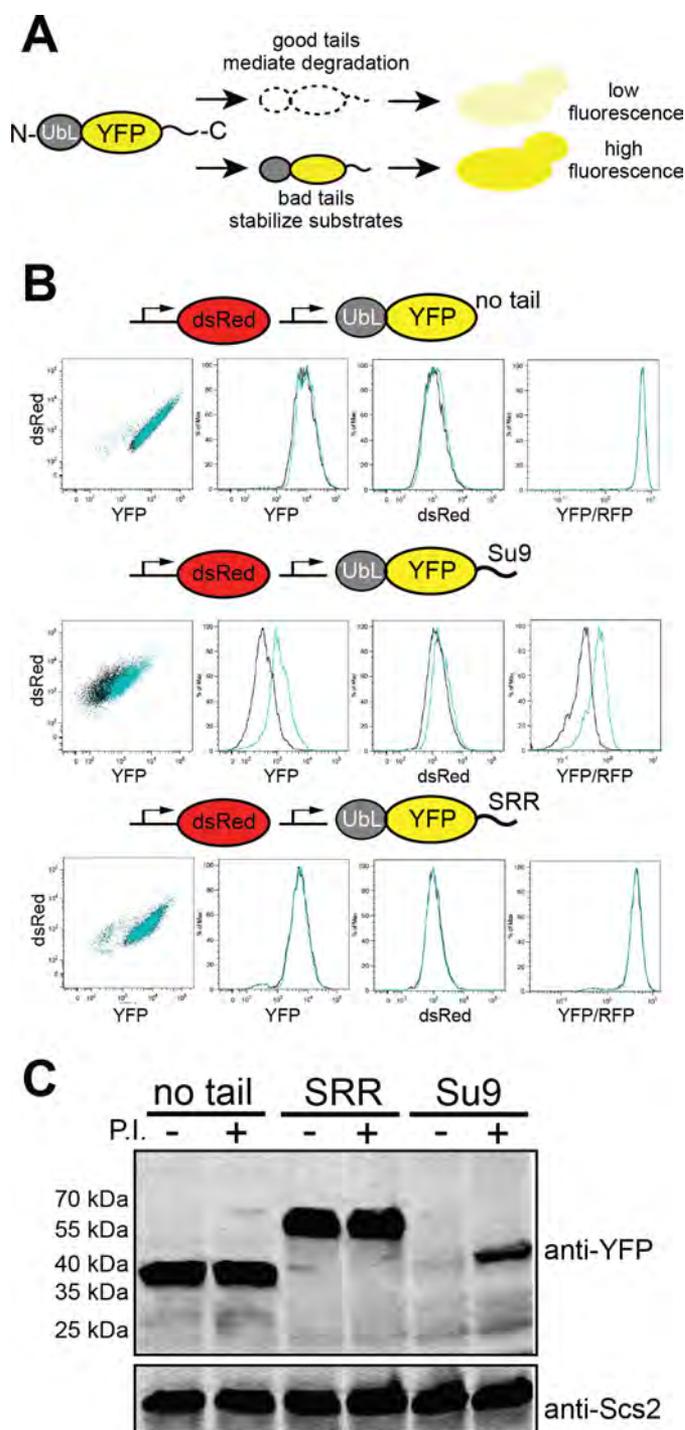


FIGURE 1. Assessing proteasomal initiation in *S. cerevisiae*. A, outline of the fluorescence-based degradation assay in *S. cerevisiae*. Proteasome substrates consisted of an N-terminal UbL domain derived from *S. cerevisiae* Rad23, followed by a YFP domain, and finally a disordered tail at the C terminus. Tails that allow the proteasome to initiate degradation resulted in a low cellular YFP signal, whereas tails that are not recognized by the proteasome led to a high YFP signal. B, cell fluorescence profiles of *S. cerevisiae* cultures expressing proteasome substrates with different initiation sequences monitored by flow cytometry. Cells were treated with proteasome inhibitor (100 μ M bortezomib, cyan population) or DMSO (black population). 10,000 cells were collected in each flow cytometry run. C, cellular abundance of YFP substrates (UbL-YFP-tail) without an initiation sequence (no tail, Q) or with poor (SRR tail, E in supplemental Table S1) or effective (Su9 tail, P in supplemental Table S1) initiation sequences was assayed by Western blotting. The proteasome was inhibited with 100 μ M bortezomib where indicated; the integral endoplasmic reticulum membrane protein Scs2, detected with a specific Scs2 antibody, served as the loading control.

³ The abbreviations used are: RFP, red fluorescent protein; DHFR, dihydrofolate reductase; 3-AT, 3-amino-1,2,4-triazole; IQR, interquartile range.

Sequence Preferences in Proteasome Degradation

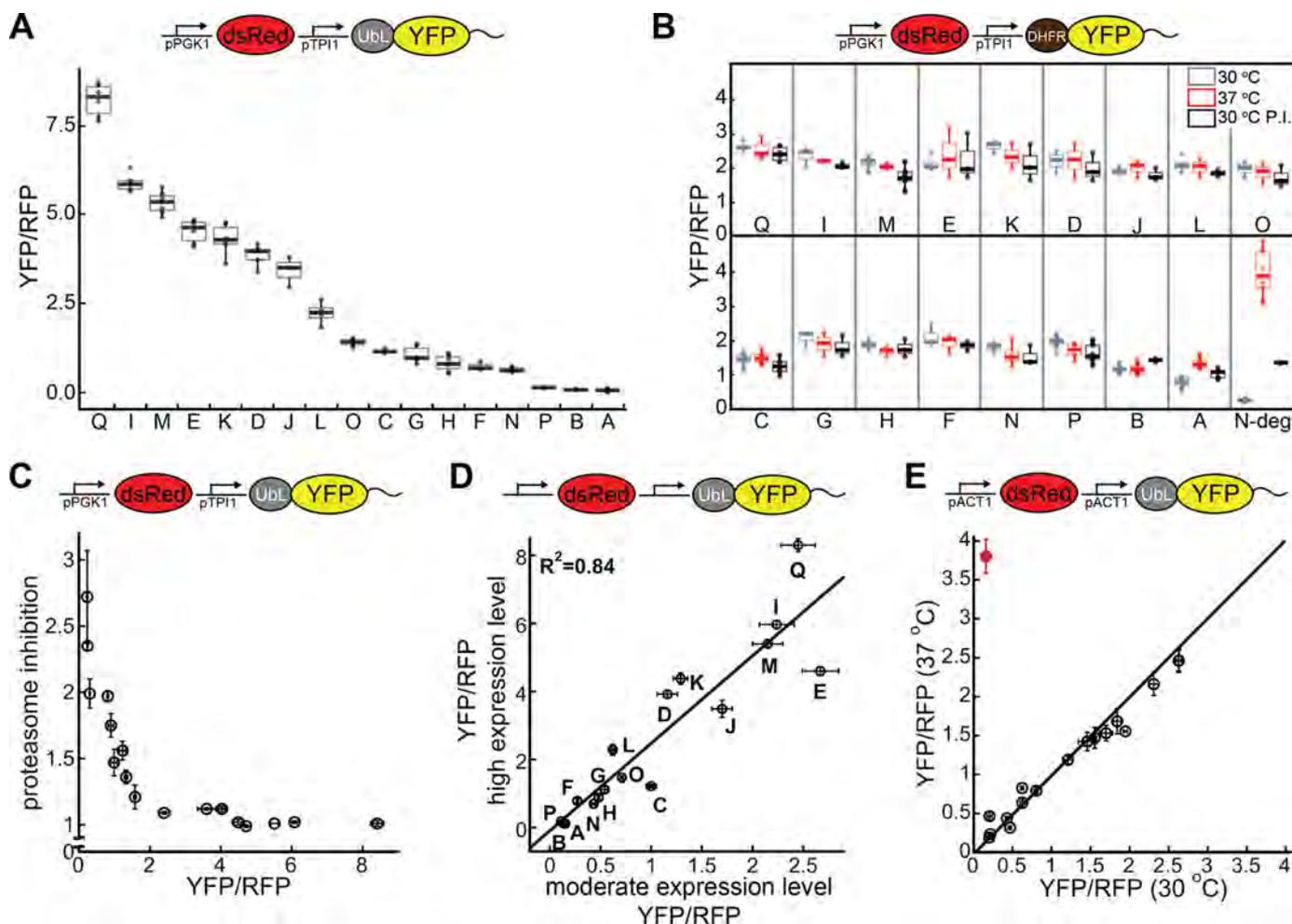


FIGURE 2. Proteasomal preferences for initiation sequences in *S. cerevisiae*. *A*, boxplots of corrected median cellular YFP fluorescence (median YFP/RFP values) for cultures expressing fluorescent proteasome substrates with different tails at their C termini (pPGK1 dsRed, pTPI1 UbL-YFP-tail; tail sequences shown in supplemental Table S1). Whiskers contain data within 1.5 interquartile range (IQR) of box. The IQR of the data is the difference between the 3rd quartile (75th percentile) and 1st quartile (25th percentile) and thus corresponds to the height of the box. *B*, boxplots (1.5 IQR whiskers) of YFP/RFP values for proteasome substrates in which the UbL domain was replaced with a DHFR domain (DHFR-YFP-tail) in the E1 temperature-sensitive strain *uba1-204* at the permissive (30 °C, gray) and restrictive (37 °C, red) temperatures and after proteasome inhibition (30 °C + 100 μ M bortezomib, black). *N-deg*, YFP substrate with N-end rule degen. *C*, graph plots median YFP/RFP values upon proteasome inhibition against the YFP/RFP values of each proteasome substrate. The recovery was calculated as the ratio of the median YFP/RFP values for cultures grown after addition of 100 μ M bortezomib and DMSO. For stable YFP proteins, the YFP/RFP value does not change upon proteasome inhibition (recovery \approx 1), and for well degraded proteins, the YFP/RFP value recovers as the proteasome is inhibited (recovery $>$ 1). *D*, graph plots median YFP/RFP values for cell cultures expressing proteasome substrates expressed at high levels (pPGK1 dsRed, pTPI1 UbL-YFP-tail) against the median YFP/RFP values for the same protein expressed at moderate levels (pACT1 dsRed, pACT1 UbL-YFP-tail). The correlation coefficient is calculated for a fit to a straight line. *E*, median YFP/RFP values of cell cultures expressing proteasome substrates with different tails at their C termini at moderate expression levels (pACT1 dsRed, pACT1 UbL-YFP-tail, black circles) in the E1 temperature-sensitive strain *uba1-204* at the permissive (30 °C) or the restrictive (37 °C) temperature. A ubiquitination-dependent N-end rule substrate is also shown (Ub-R-KK-YFP-Su9, red circle). The median YFP/RFP values for each construct were calculated from 10,000 cells collected in one flow cytometry run. Data points in panels *C*, *D*, and *E* represent mean values determined from at least three repeat experiments; error bars indicate S.E.

if proteasome inhibition is incomplete (Fig. 1B). Analysis of cell extracts by SDS-PAGE and Western blotting confirmed that the UbL-YFP-Su9 protein was depleted from cells in a proteasome-dependent manner and that the protein was degraded completely as no partially degraded protein fragments could be detected (Fig. 1C). Replacing the Su9 tail with a sequence consisting almost entirely of Ser residues (serine-rich region or SRR, sequence E in supplemental Table S1) stabilized the protein and restored UbL-YFP-SRR levels almost to those seen for UbL-YFP without an initiation region (Fig. 1B). Proteasome inhibition by bortezomib did not cause further increase in UbL-YFP-SRR levels (Fig. 1, B and C). These results show that only some disordered tails allow the proteasome to initiate degradation in cells.

Initiation Sequence Preferences in Vivo—Next, we fused UbL-YFP to 14 additional disordered C-terminal tails (supplemental Table S1) and used the fluorescence assay described above to investigate intracellular degradation. The steady-state abundance of these proteins, as judged by YFP/RFP ratios, varied \sim 70-fold between the most and least stable proteins (Fig. 2A and Table 1).

Degradation of the YFP proteins depended on the UbL domain and was not due to ubiquitination of the disordered tails. *S. cerevisiae* encodes only one ubiquitin-activating enzyme (46), Uba1, and the temperature-sensitive *uba1-204* allele makes it possible to reduce protein ubiquitination substantially by shifting cells to the restrictive temperature (47). We replaced the UbL domain with a DHFR domain and fused

TABLE 1

YFP/RFP ratios of constructs in *S. cerevisiae*, *S. pombe*, and mammalian cells (HEK293)

The median of YFP/RFP ratio for each construct was calculated from 10,000 cells collected in one flow cytometry run and used to indicate the abundance of the YFP substrate in yeast cells as described under "Experimental Procedures." Data represent mean values and standard errors determined from at least three repeat experiments. ND is not determined; NS indicates derived from influenza A virus non-structural protein 1 (NS1); GRR indicates glycine-rich region (derived from human p105); NB indicates derived from influenza B virus glycoprotein NB; SNS indicates tandem repeat of SP2-NB-SP2; NBS indicates tandem repeat of NB-NB-SP2; DRR indicates aspartic acid (D)-rich region (derived from *S. cerevisiae* Cdc34); SP1 indicates peptide region 1 in influenza A virus M2 protein used to produce antisera; SP2 indicates peptide region 2 in influenza A virus M2 protein used to produce antisera; SPmix indicates tandem repeat of SP1 and SP2 (SP2-SP1-SP2-SP1-SP2); PEST indicates sequence from human IκBα; SRR indicates serine-rich region (derived from herpes virus 1 ICP4); Su9, derived from subunit 9 of *N. crassa* ATP synthase component F_0 ; eRR, derived from *E. coli* lacI; ODC indicates derived from ornithine decarboxylase; 35 indicates derived from *S. cerevisiae* cytochrome b_2 .

Tail	Name	<i>S. cerevisiae</i>			<i>S. pombe</i>	HEK293
		pACT1 UbL-YFP-tail pACT1 dsRed	pTPI1 UbL-YFP-tail pPGK1 dsRed	pTPI1 UbL-GFP-tail pPGK1 dsRed		
A	35	0.15 ± 0.01	0.12 ± 0.02	0.13 ± 0.01	0.73 ± 0.06	0.20 ± 0.02
B	ODC	0.13 ± 0.01	0.13 ± 0.01	0.034 ± 0.004	1.0 ± 0.3	0.13 ± 0.02
C	Poly(G)	1.00 ± 0.05	1.22 ± 0.05	1.07 ± 0.08	ND	0.51 ± 0.01
D	GRR	1.2 ± 0.1	3.9 ± 0.1	1.9 ± 0.1	4.8 ± 0.2	0.83 ± 0.05
E	SRR	2.7 ± 0.2	4.6 ± 0.1	3.81 ± 0.09	6.54 ± 0.07	0.87 ± 0.02
F	NB	0.27 ± 0.01	0.79 ± 0.07	0.27 ± 0.02	2.8 ± 0.2	0.31 ± 0.02
G	NS	0.54 ± 0.03	1.12 ± 0.09	0.80 ± 0.03	2.3 ± 0.1	0.27 ± 0.02
H	SP1	0.48 ± 0.01	0.88 ± 0.08	0.63 ± 0.01	3.0 ± 0.1	0.47 ± 0.03
I	SP2	2.2 ± 0.2	6.0 ± 0.1	4.30 ± 0.06	5.65 ± 0.07	0.86 ± 0.04
J	SPmix	1.7 ± 0.1	3.5 ± 0.3	2.40 ± 0.07	5.94 ± 0.09	0.834 ± 0.004
K	SNS	1.29 ± 0.07	4.4 ± 0.2	2.7 ± 0.2	4.77 ± 0.07	0.70 ± 0.02
L	NBS	0.62 ± 0.02	2.3 ± 0.2	1.01 ± 0.07	5.2 ± 0.1	0.54 ± 0.02
M	DRR	2.2 ± 0.2	5.4 ± 0.1	4.6 ± 0.3	4.9 ± 0.2	0.76 ± 0.03
N	eRR	0.43 ± 0.02	0.70 ± 0.02	0.40 ± 0.01	1.10 ± 0.05	0.28 ± 0.02
O	PEST	0.71 ± 0.01	1.47 ± 0.04	0.98 ± 0.08	2.85 ± 0.08	0.21 ± 0.01
P	Su9	0.11 ± 0.03	0.19 ± 0.01	0.14 ± 0.02	1.7 ± 0.2	0.16 ± 0.01
Q	No tail	2.5 ± 0.2	8.3 ± 0.2	5.1 ± 0.3	6.8 ± 0.1	0.83 ± 0.06

the DHFR-YFP variants to the same 16 tails in a *uba1-204* strain. The steady-state levels of 14 of these proteins were similar both at the restrictive temperature and at the permissive temperature in the absence or presence of bortezomib (Fig. 2C). Shifting the *uba1-204* strain to the restrictive temperature does inhibit ubiquitin-dependent degradation of YFP substrate with a classic N-end rule degron (see below) more than 20-fold (Fig. 2B). These results indicate that most DHFR-YFP-tail variants are neither ubiquitinated nor degraded by the proteasome. The steady-state level of DHFR-YFP-35, which has a tail derived from the pre-sequence of *S. cerevisiae* cytochrome b_2 (sequence A in supplemental Table S1), increased at the restrictive temperature and at the permissive temperature in the presence of bortezomib (Fig. 2C), suggesting that this sequence did become ubiquitinated to some extent. The steady-state level of DHFR-YFP-ODC, which has a tail derived from the 37 C-terminal amino acids of ornithine decarboxylase (sequence B in supplemental Table S1), increased by a small amount at the permissive temperature in the presence of bortezomib but was not ubiquitin-dependent (Fig. 2B). Proteasomal degradation of ornithine decarboxylase is known to be ubiquitin-independent (48, 49), suggesting that DHFR-YFP-ODC is directly targeted to the proteasome by the ODC tail. However, UbL-YFP-ODC was degraded substantially more efficiently than DHFR-YFP-ODC (Fig. 2, A and B), showing that tethering to the proteasome is important for robust degradation.

Inhibiting the proteasome with bortezomib increased accumulation of the least stable proteins by ~3-fold, had no effect on the most stable proteins, and affected the proteins in-between proportionally to their abundance (Fig. 2C). Thus, degradation of unstable UbL-YFP-tail proteins was by the proteasome.

In principle, promoter strength might influence the proteasomal degradation of different proteins via aggregation or sat-

uration of the folding or degradation machinery. To test this possibility, we expressed the set of UbL-YFP variants as well as RFP reference protein from two *act1* promoters (50) on the same CEN plasmid, which reduced cellular levels of a non-degraded UbL-YFP protein ~5-fold compared with expression from the *tpi1* promoter (data not shown). At these lower expression levels, there was a 25-fold difference in abundance between proteins that degraded effectively (e.g. UbL-YFP-Su9; sequence P) and the proteins that degraded poorly (e.g. UbL-YFP-SRR; sequence E) (Table 1 and Fig. 2, D and E). Importantly, the steady-state levels of the UbL-YFP-tail proteins expressed from stronger and weaker promoters were highly correlated ($R^2 = 0.84$; Fig. 2D). Thus, different expression levels did not cause a significant difference in the contribution of the tails to the initiation of proteasomal degradation. Again, degradation depended on the UbL domain and the proteasome (data not shown) and was not affected by ubiquitination, except for UbL-YFP-35 (sequence A in supplemental Table S1; Fig. 2E). UbL-YFP-35 was stabilized ~2-fold when ubiquitination was inhibited, compared with a 25-fold stabilization when the tail was removed. Thus, the 35 tail was ubiquitinated to some extent, but its ubiquitination made a relatively small contribution to proteasome targeting.

Proteasome Initiation Tunes Degradation over a Similar Range as Ubiquitination—Next, we asked whether initiation regions could modulate proteasomal degradation over the same dynamic range as achieved through the regulation of ubiquitination of a classic degron. We replaced the UbL domain of UbL-YFP-Su9 with an N-end rule degron consisting of a ubiquitin domain followed by a destabilizing (Arg) or stabilizing (Val) residue and a linker derived from *Escherichia coli* lacI, which contains two Lys residues (51, 52). In the cell, the ubiquitin domain is cleaved off by ubiquitin hydrolases, and an Arg residue leads to ubiquitination of the degron but a Val does not.

Sequence Preferences in Proteasome Degradation

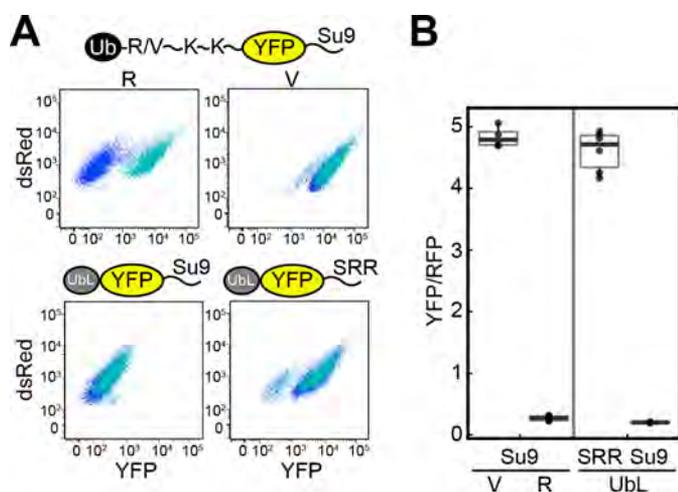


FIGURE 3. Proteasome initiation tunes protein abundance over a similar range as ubiquitination. *A*, cell fluorescence profiles of *S. cerevisiae* cultures expressing UbL-YFP-tail and N-end rule substrates. *Top row*, cells expressing N-end rule degron substrates with a destabilizing residue as the N degnon (Ub-R-KK-YFP-Su9) or a stabilizing residue as the N degnon (Ub-V-KK-YFP-Su9); *bottom row*, cells expressing UbL-YFP proteins with a tail that serves as an effective proteasome initiation site (Su9; P in supplemental Table S1) or a poor proteasome initiation site (SRR; E in supplemental Table S1). The proteins were expressed in the E1 temperature-sensitive strain *uba1-204* at the permissive temperature (30 °C, blue) or the restrictive temperature (37 °C, cyan). *B*, boxplots (1.5 IQR whiskers) of corrected YFP fluorescence (median YFP/RFP values) for cells expressing the N-end rule substrates and UbL substrates analyzed in *A*. The median YFP/RFP value for each construct was calculated from 10,000 cells collected in one run in flow cytometry.

Steady-state levels of the R-KK-YFP-Su9 protein were low and similar to those of UbL-YFP-Su9 (Fig. 3*A*, blue populations), whereas levels of the V-KK-YFP-Su9 protein were high and similar to those of UbL-YFP-SRR (Fig. 3*A*, blue populations). Inhibiting ubiquitination by shifting *uba1-204* cells to the restrictive temperature increased protein levels for R-KK-YFP-Su9 but did not affect V-KK-YFP-Su9, UbL-YFP-Su9 or UbL-YFP-SRR levels (Fig. 3*A*, cyan populations). Altering the N-end rule degron from Arg to Val changed YFP levels ~26-fold (Fig. 3*B*), whereas modulating initiation by replacing the Su9 tail of UbL-YFP-Su9 with an SRR tail changed YFP fluorescence ~24-fold (Fig. 3*B*). Thus, the identity of the initiation region can be as important as the regulation of ubiquitination in targeting proteins to proteasomal degradation.

Steady-state Levels Correlate with Degradation Rates—The steady-state abundance of UbL-YFP-tail variants depended on the rates at which they were degraded by the proteasome in the cell. We measured degradation rates by inhibiting protein synthesis with cycloheximide and measured the amount of YFP substrate remaining over time (Fig. 4*A*). The half-lives of YFP substrates with different tails correlated well ($R^2 = 0.76$) with their steady-state levels (Fig. 4*B*). The half-lives also varied over a similar dynamic range as the steady-state levels, with the least stable protein being degraded 68-fold faster than the most stable protein (Table 2). Thus, altering proteasomal initiation by changing the amino acid sequence of the disordered tails can tune degradation rates in the cell over a wide dynamic range. We conclude that the steady-state accumulation of the different proteins reflected the rates with which they were degraded by the proteasome.

We also measured degradation rates for a subset of UbL-YFP-tail proteins by expressing them from a *gal1* promoter (37) and then shutting off expression by adding glucose. The rate constants determined in these experiments were very similar to the rate constants measured in the cycloheximide shut-off experiments for the same proteins expressed from the strong *tpi1* promoter (Fig. 4*C*).

Proteasomal Sequence Preferences Are Consistent for Different Proteins and Can Affect Fitness—Next, we tested whether altering proteasomal initiation of degradation could regulate the abundance of proteins other than YFP. We first replaced the YFP domain of UbL-YFP-tail proteins with jellyfish green fluorescent protein (GFP) (53), and found that the cellular levels of the UbL-GFP-tail and UbL-YFP-tail proteins were affected by the tail sequences in similar ways ($R^2 = 0.94$ for a linear fit, Table 1).

Fluorescent proteins do not occur naturally in *S. cerevisiae*, and their overexpression can have pleiotropic effects (50). To test a different protein, we chose *S. cerevisiae* His3, which is required for growth in medium lacking histidine (54). We constructed UbL-His3-tail proteins (Fig. 5) with the same 16 initiation regions analyzed in the YFP constructs, as well as a UbL-His3 protein without a tail and two His3 variants in which the UbL domain was replaced with a DHFR domain (DHFR-His3 and DHFR-His3-Su9). We tested whether expressing these proteins could complement growth of a *his3* mutant strain, using a competitive inhibitor of His3 (3-AT) to enhance assay sensitivity (55). Yeast with the parental control vector did not grow in the absence of histidine, but strains expressing the His3 variants that are not expected to be degraded, namely UbL-His3, DHFR-His3, and DHFR-His3-Su9, restored growth (Fig. 5*B*).

The growth phenotypes of the 16 different strains expressing UbL-His3-tail proteins fell into three broad groups, robust complementation, modest complementation, and poor or no complementation (Fig. 5*B*). The initiation sequences in the tails that prevented UbL-His3-tail proteins to complement in this assay also resulted in rapid degradation of the corresponding UbL-YFP-tail proteins. The tails that gave modest complementation resulted in intermediate UbL-YFP-tail degradation rates, and the tails that fully complemented resulted in slow degradation of UbL-YFP-tail proteins (*cf.* Figs. 2*A* and 4*A* with 5*B*). Expression of the His3 fusion proteins was not deleterious to yeast cells because supplementing the medium with histidine restored wild-type growth (Fig. 5*B*). In summary, the proteasome shows distinct preferences for the sequence of the disordered region in its substrate where it initiates degradation. These preferences are not dependent on the nature of the protein that is degraded. Furthermore, regulation of protein degradation by modulation of proteasome initiation can affect cell fitness.

Initiation Sequence Preferences Are Similar in Different Organisms—The proteasome is evolutionarily conserved (56), although its processivity can vary substantially between different organisms (57). This raises the question whether the initiation rules for proteasomal degradation are the same for different organisms. To address this question, we tested the degradation of UbL-YFP-tail proteins in *S. pombe* and in cul-

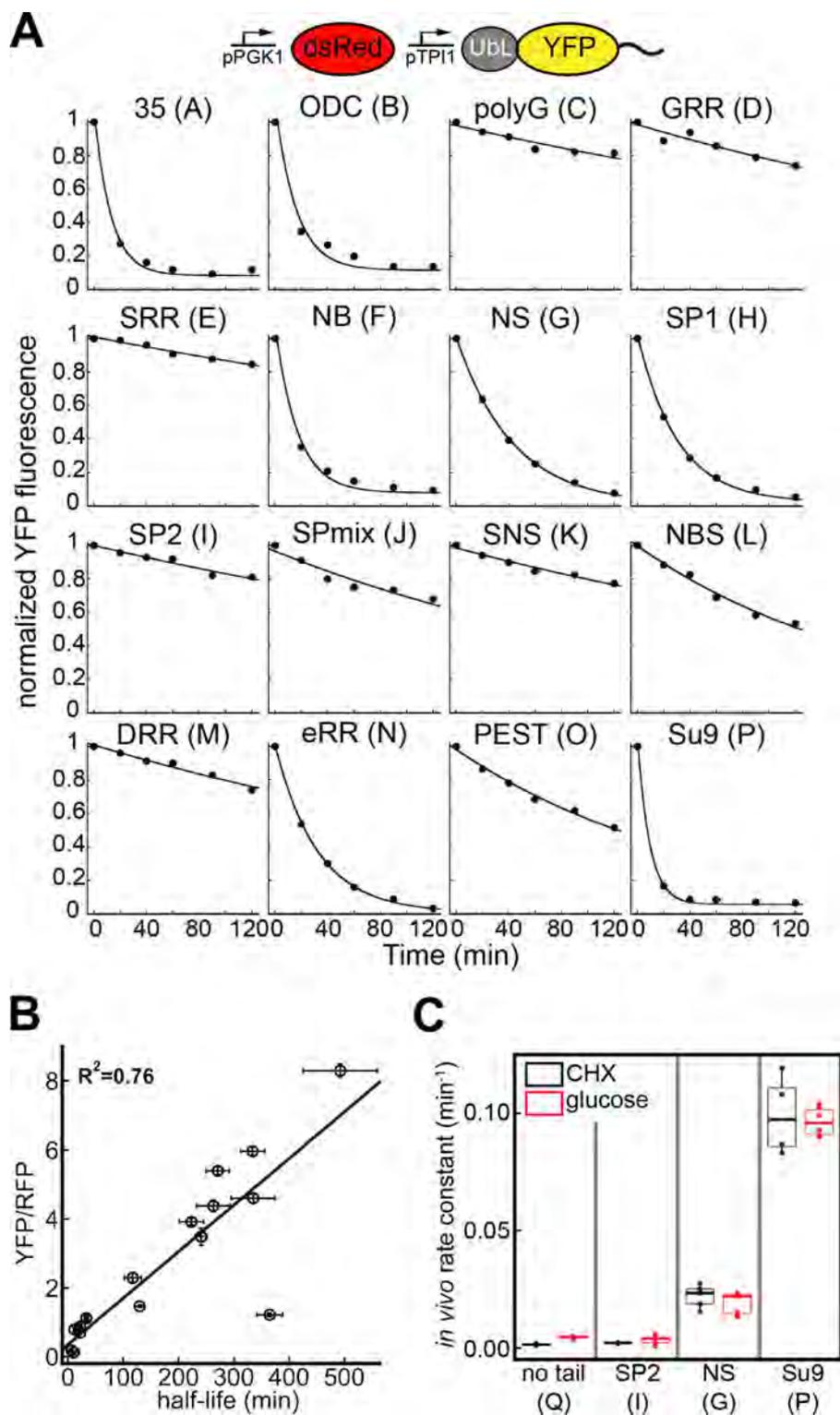


FIGURE 4. Steady-state protein abundances correlate with degradation rates. *A*, normalized time courses of YFP fluorescence illustrating *in vivo* degradation of UbL-YFP-tail constructs for 16 different tails after inhibition of protein synthesis by the addition of 125 μM cycloheximide. *Graphs* show one representative dataset of at least three independent experiments. *B*, relationship between protein abundance and half-life in yeast. The steady-state YFP/RFP ratios of fluorescent substrates with 16 different tails are plotted against the half-lives determined by nonlinear fitting of the data shown in *A* to a single exponential decay. The correlation coefficient is calculated for a fit to a straight line. Data points represent mean values determined from at least three repeat experiments; *error bars* indicate S.E. *C*, degradation rate constants of YFP substrates in yeast obtained from the cycloheximide chase experiment in *A* (*black*) and from a glucose shutdown assay (*red*) are shown. The median cellular YFP fluorescence for each construct at each time point was calculated from 10,000 cells collected in one run in flow cytometry; repeat experiments yielded the median values indicated in the *boxplots* shown (1.5 IQR whiskers).

Sequence Preferences in Proteasome Degradation

TABLE 2

Binding affinities of initiation sequences to the proteasome and *in vivo* degradation rate constants of fluorescent substrates (Ubl-YFP-tail) in yeast

See Table 1 for definitions of names.

Name	K_i	<i>In vivo</i> degradation rate constant
	μM	min^{-1}
35	36 ± 6	0.071 ± 0.004
ODC	28 ± 3	0.068 ± 0.007
Poly(G)		0.0019 ± 0.0001
GRR	130 ± 20	0.0033 ± 0.003
SRR		0.0022 ± 0.0003
NB	63 ± 12	0.07 ± 0.01
NS	158 ± 3	0.022 ± 0.002
SP1	86 ± 4	0.035 ± 0.004
SP2	322 ± 15	0.0021 ± 0.0001
SPmix		0.0029 ± 0.0001
SNS	133 ± 12	0.0028 ± 0.0003
NBS		0.0063 ± 0.0008
DRR		0.0026 ± 0.0002
eRR	36 ± 5	0.032 ± 0.001
PEST		0.0054 ± 0.0003
Su9		0.099 ± 0.008
No tail		0.0015 ± 0.0002

tured human HEK293 cells. To ensure proteasome targeting, we used the Ubl domain of *S. pombe* Rhp23 (58) for the *S. pombe* experiments and the Ubl domain of human HR23B (59) for the HEK293 experiments, as well as appropriate vectors, promoters, and red fluorescent proteins (Fig. 6A). The steady-state levels of the Ubl-YFP-tail proteins, as assayed by YFP/RFP values, in the different organisms were highly correlated (Fig. 6, B–D). We conclude that the proteasomes of *S. cerevisiae*, *S. pombe*, and *Homo sapiens* share similar preferences for the amino acid sequence of their initiation sites.

Initiation Sequence Preferences Reflect Proteasome Affinity—Degradation rates of Ubl-YFP-tail proteins in *S. cerevisiae* correlated with degradation rates measured *in vitro* for a set of proteins with the same tails as tested here (Fig. 7A). In the *in vitro* experiments, the tails were attached to a DHFR domain that was targeted to the proteasome by a tetra-ubiquitin chain, and these Ub₄-DHFR-tail proteins were then degraded by purified *S. cerevisiae* proteasome (28).

To test whether initiation sequence preferences reflect a direct interaction with the proteasome, we developed an assay in which the different initiation regions compete with a substrate for degradation by purified *S. cerevisiae* proteasome. The substrate consisted of superfolder GFP with the Ubl domain from *S. cerevisiae* Rad23 fused to its N terminus and a disordered region of 95 amino acids derived from *S. cerevisiae* cytochrome *b*₂ fused to its C terminus. Degradation of this substrate by the proteasome in the presence of ATP, as assayed by loss of GFP fluorescence, followed Michaelis-Menten kinetics, with the V_{max} scaling linearly with proteasome concentration and the K_m remaining constant (Fig. 7B). We fused the different disordered tails described above to the C terminus of *E. coli* DHFR and purified the proteins from *E. coli* by affinity chromatography. We first characterized DHFR-35 and DHFR-eRR (where eRR is derived from *E. coli* lacI) (see supplemental Table S1 for sequences; 35, A; eRR, N). Both tails supported robust degradation of Ubl-YFP proteins in the cell (Figs. 2, 4–6). Increasing the concentrations of the DHFR-tail proteins progressively inhibited degradation of Ubl-GFP-95 (Fig. 7, C and

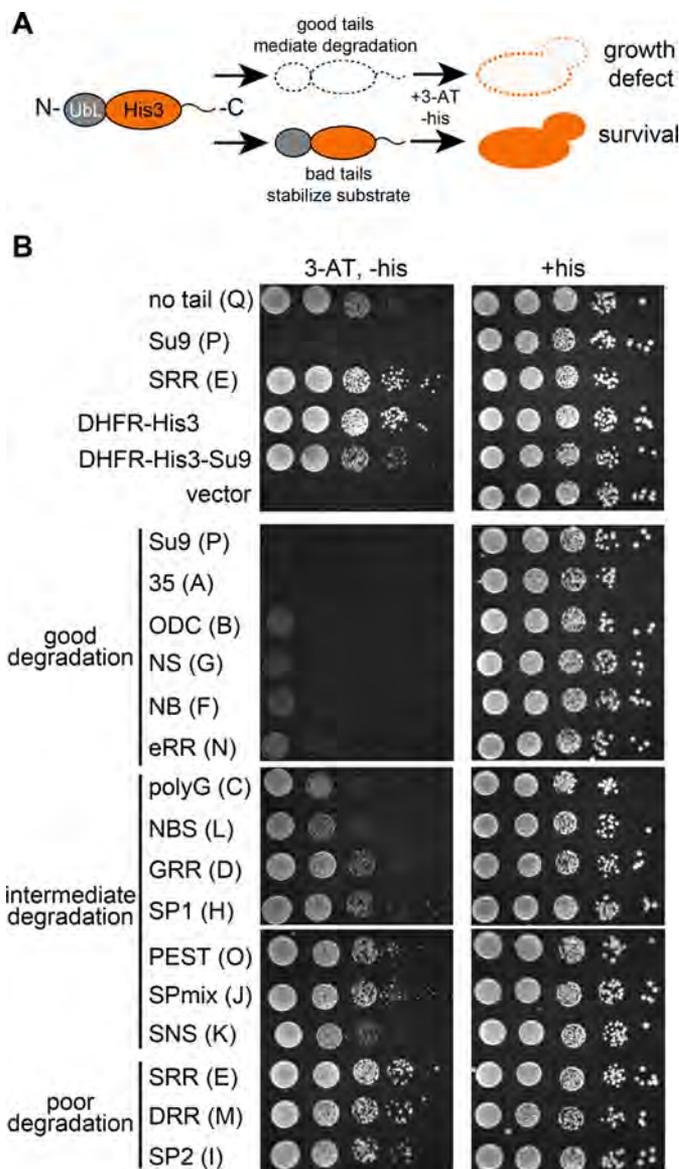


FIGURE 5. Analysis of proteasome initiation using His3 protein degradation in yeast. A, outline of the yeast growth assay. Imidazoleglycerol-phosphate dehydratase (His3) was targeted to the proteasome by fusing the Ubl domain of *S. cerevisiae* Rad23 to its N terminus and different tails to its C terminus. Only tails that provide effective proteasome initiation regions mediate His3 protein degradation. In *his3* mutant cells, the absence of His3 protein causes growth defects in medium lacking histidine. B, cells expressing the indicated His3 fusion proteins in late log phase were serially diluted and stamped onto selective (+3-AT, -his) or non-selective (+his) synthetic medium. Plates were incubated at 30 °C for 3 days before imaging.

D). Inhibition was overcome by increasing the Ubl-GFP-95 concentration, suggesting that Ubl-GFP-95 and the DHFR-tail proteins compete for binding to the proteasome's receptor for the initiation region (Fig. 7D). The apparent inhibition constants (K_i), with which the DHFR-tail proteins inhibited Ubl-GFP-95 degradation, therefore reflected the affinity of the tails for the proteasome.

We then selected a subset of initiation regions and measured their ability to inhibit Ubl-GFP-95 degradation (Table 2). The K_i values ranged from $\sim 30 \mu\text{M}$ for the highest affinity interaction to $\sim 300 \mu\text{M}$ for the weakest interaction. The inhibition constants for different tails correlated well ($R^2 = 0.73$) with

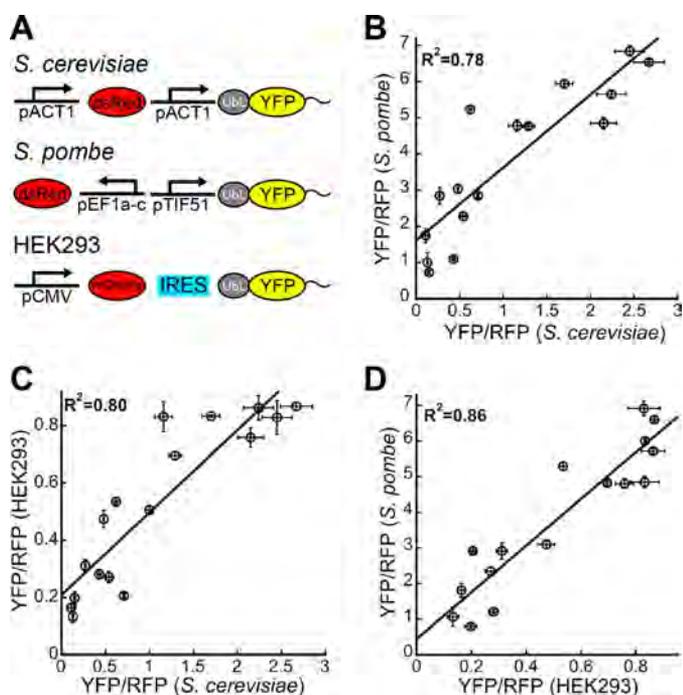


FIGURE 6. Conserved sequence preferences in different species. *A*, schematic representation of constructs used to assess proteasome targeting in *S. cerevisiae*, *S. pombe*, and cultured human cells (HEK293). In *S. cerevisiae*, YFP substrates and dsRed were both expressed from constitutive *act1* promoters on a CEN plasmid. In *S. pombe*, YFP substrates and dsRed were expressed from the constitutive promoters *tif51* and *ef1a-c*, respectively, after integration into genomic DNA. In HEK293 cells, mCherry and the YFP substrates were expressed from a single CMV promoter with their coding sequences separated by an internal ribosome entry site (*IRES*) on the mammalian expression vector pCDNA5. *B–D*, graphs plot corrected median cellular YFP fluorescence (median YFP/RFP values) for each construct expressed in two organisms against each other: *S. cerevisiae* and *S. pombe* (*B*), *S. cerevisiae* and HEK293 cells (*C*), and *S. pombe* and HEK293 cells (*D*). Correlation coefficients are calculated for fits to a straight line. The median YFP/RFP ratio for each construct was calculated from 10,000 cells collected in one flow cytometry run. Data points represent mean values determined from at least three repeat experiments; error bars indicate S.E.

their ability to support degradation, with tighter-binding tails leading to lower protein abundance in cells (Fig. 7E). Thus, the ability of disordered tails to initiate proteasomal degradation appears to be determined by their affinity for the proteasome.

Sequence Features of Initiation Regions—Which sequence characteristics of disordered tails dictate the proteasome's initiation preferences? To study initiation preferences in greater detail, we used the 16 sequences characterized above (supplemental Table S1) as well as 99 additional sequences derived primarily from human and yeast proteins, and we tested their effects on steady-state levels of UbL-YFP-tail proteins *in vivo* (supplemental Table S2). We calculated a set of parameters of these sequences that report on their chemical or physical properties, such as hydrophobicity, charge, sequence complexity, flexibility, etc. (Fig. 8A). The sequence characteristics of the complete set of 115 sequences tested represent the properties of the human proteome well (Fig. 8B).

The scales for which the parameters reporting sequence properties are assessed were not developed with protein degradation in mind, and there is no *a priori* reason to assume a linear relationship between these parameters and degradation rates. Therefore, we took a nonparametric approach when relating

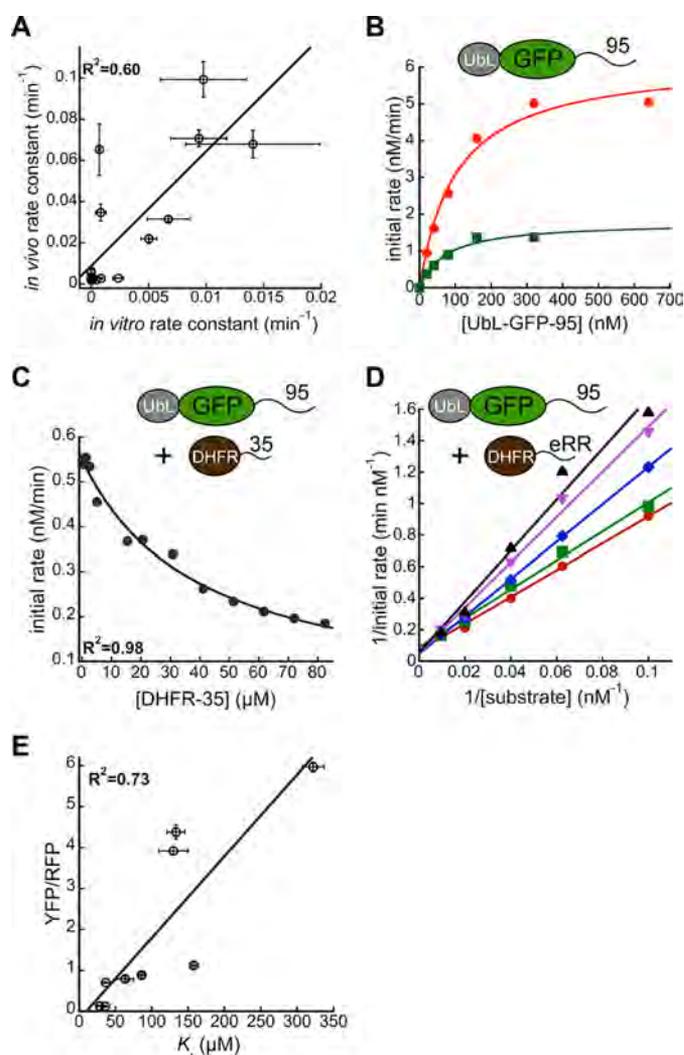


FIGURE 7. Binding of initiation sequences to the proteasome. *A*, correlation between substrate degradation rate constants *in vitro* and *in vivo*. Degradation rate constants *in vivo* for UbL-YFP-tail constructs as quoted in Table 2 are plotted against rate constants for degradation of Ub₄-DHFR-tail substrates by purified yeast proteasome (adopted from Ref. 28). The correlation coefficient is calculated for a fit to a straight line. *B*, Michaelis-Menten plot for UbL-GFP-95 degradation by 10 nM (green) or 40 nM (red) purified *S. cerevisiae* proteasome. The substrate consisted of an N-terminal UbL domain derived from *S. cerevisiae* Rad23, followed by superfolder GFP and a 95-amino acid-long tail derived from *S. cerevisiae* cytochrome *b₂*. *C*, initial degradation rates of UbL-GFP-95 in the presence of different concentrations of purified DHFR-35 by purified yeast proteasome are plotted and fitted to an equation describing competitive inhibition to calculate the inhibition constant K_i for DHFR-35. *D*, initial degradation rates for different concentrations of UbL-GFP-95 in the presence of 0 μ M (red), 2.6 μ M (green), 10 μ M (blue), 31 μ M (pink), or 95 μ M (black) DHFR-eRR. *E*, corrected median cellular YFP fluorescence (median YFP/RFP values) for cultures expressing fluorescent proteasome substrates with different tails at their C termini (pPGK1 dsRed, pTP11 UbL-YFP-tail; Table 1) are plotted against the K_i values for DHFR-tail constructs with the same tail (Table 2). The correlation coefficient is calculated for a fit to a straight line. UbL-GFP-95 degradation was followed by monitoring fluorescence intensity over time using a Tecan plate reader at room temperature as described under the "Experimental Procedures." Tail sequences are shown in supplemental Table S1 (35, A; eRR, N).

the sequence parameters to protein degradation, using Spearman correlation coefficients to compare only the rank ordering of the different sequences by the various parameters with the rank ordering by YFP/RFP ratios. The correlations were tested for statistical significance (no correlation is the null hypothe-

Sequence Preferences in Proteasome Degradation

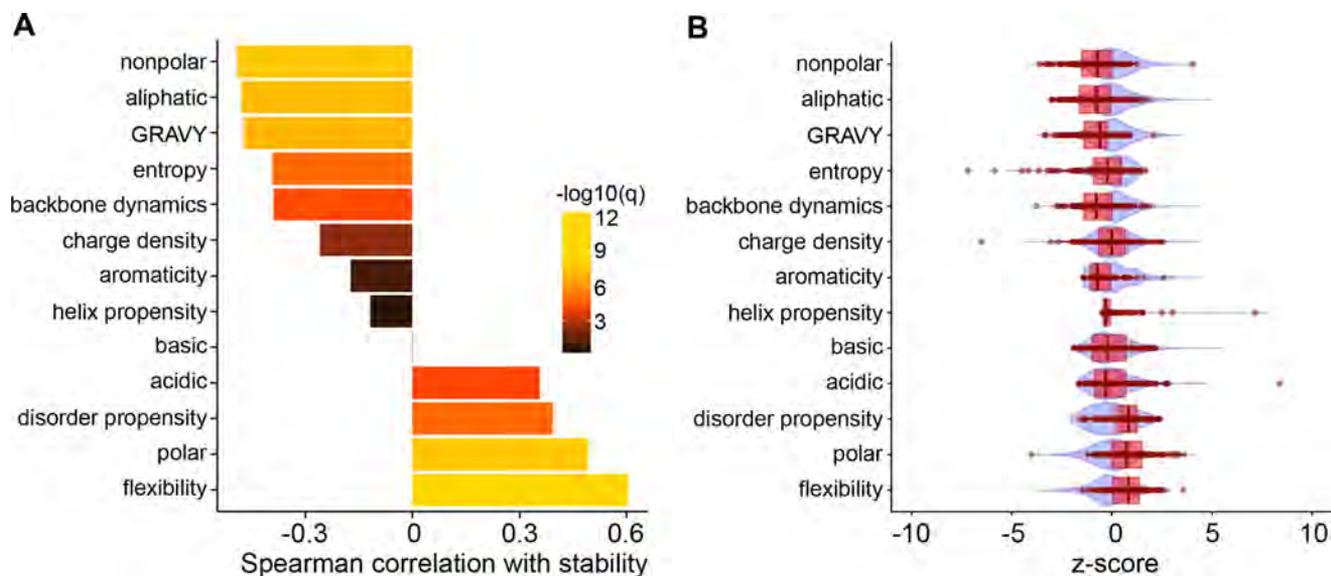


FIGURE 8. **Bioinformatics analysis of initiation sequences.** A, Spearman correlations between experimentally determined stabilities (measured by steady-state YFP/RFP ratios in Table 1 and supplemental Table S2 (pPGK1 dsRed, pTPI1 UbL-YFP-tail)) and physicochemical parameters of peptide sequences. The parameters for 115 sequences were calculated as described under the “Experimental Procedures.” Spearman correlation coefficients between the physicochemical parameters of the different sequences and YFP/RFP ratios were calculated. Statistical significance of these associations was accessed by pairwise *t* tests. The *p* values from *t* tests were then used to compute false discovery rate *q*-values by the Benjamini-Hochberg method (60), which are indicated in color ($-\log_{10}(q)$). B, comparison of the distribution of the parameters calculated for the 115 sequences analyzed in this study (box plots in red) and the same values calculated for the human proteome (violin plot in blue). A designed polypeptide library that covers the human proteome (T7-pep) (104) was used to represent the sequence characteristics of human proteome. The parameters were calculated as described under the “Experimental Procedures.”

sis), with *p* values adjusted for false discovery by the Benjamini-Hochberg method (Fig. 8A) (60).

Consistent with our previous study (28), we found sequence complexity (measured as sequence entropy) to be correlated negatively with protein stability, suggesting that initiation sequences with biased amino acid compositions support proteasome initiation inefficiently (Fig. 8A). In addition to sequence complexity, hydrophobicity of the initiation region as measured by the fraction of aliphatic residues, the fraction of nonpolar residues, or the GRAVY algorithm (61) correlated negatively with stability (Fig. 8A). In contrast, sequence polarity (fraction of Asp, Glu, His, Lys, Asn, Arg, Ser, and Thr) and sequence acidity (fraction of Asp and Glu) correlated positively with stability (Fig. 8A). Thus, proteasome initiation seems more efficient at less hydrophilic and more hydrophobic regions.

Sequence dynamics also affect proteasome initiation and stiffer initiation regions appeared to promote better recognition and degradation. This relationship was found using two independent algorithms to predict sequence flexibility. The FLEXPLOT algorithm (62) predicts sequence flexibility by an algorithm calibrated using crystallographic B-factors, whereas the DynaMine tool predicts backbone dynamics through an algorithm calibrated by residue dynamics measured by nuclear magnetic resonance spectroscopy (63, 64). The two algorithms define their scales in opposite directions; the FLEXPLOT algorithm assigns numerically lower scores to stiffer sequences, and the DynaMine algorithm assigns numerically higher scores. Analysis based on the DynaMine algorithm resulted in a negative correlation between its score and stability, whereas the FLEXPLOT algorithm results in a positive correlation (Fig. 8A).

Discussion

We find that the proteasome has distinct preferences for the amino acid sequence where it initiates degradation *in vivo*. Initiation regions tune protein abundance over 2 orders of magnitude, which is the same range as achieved by controlling ubiquitination of a classical N-end rule degron. Proteasomal initiation may regulate protein abundance over an even larger range, but our experimental system is limited by the level of protein expression and by signal over noise detection. Therefore, the proteasome’s sequence preferences at the initiation step of degradation can contribute substantially to the regulation of protein abundance in the cell.

The ubiquitin code is ambiguous in that the proteasome can recognize polyubiquitin chains that typically target proteins to other processes. For example, polyubiquitin chains linked through Lys-63 are usually part of membrane trafficking processes, but they can target a protein to the proteasome both *in vitro* and *in vivo* (18). Proteins that are tagged with ubiquitin chains for fates other than degradation may have evolved to lack effective proteasome initiation sites to reduce the likelihood of degradation if they bind the proteasome. Conversely, if the proteasome is able to engage its target efficiently at the initiation region, even a weakly binding ubiquitin chain may induce degradation. Indeed, the presence of predicted proteasome initiation regions correlates with shorter protein half-lives *in vivo* (23, 27–29). Thus, natural variation in the unstructured regions of protein paralogs could tune protein abundance by modulating proteasomal initiation. Similarly, mechanisms that alter disordered regions such as alternative splicing (65, 66) could be used to tune protein abundance.

At the same time, the different polyubiquitin chains that target proteins to the proteasome physiologically are not necessarily synonymous. Cell cycle progression requires the degradation of regulatory proteins in the correct order. Some of these regulatory proteins are ubiquitinated by the same master ubiquitin ligase and ubiquitin-conjugating enzyme, but the polyubiquitin chains are synthesized with different efficiencies (67–69). The initiation step could contribute to substrate ordering if the late substrates, which are ubiquitinated less efficiently, also had less effective initiation regions and vice versa. Thus, the proteasome's initiation sequence preferences could contribute to the temporal control of degradation.

The proteasome's sequence preferences may also contribute to the accumulation of disease-related proteins. Several neurodegenerative diseases are associated with the buildup of proteins that appear to be targeted for degradation but escape destruction (70). Several of these proteins contain highly biased amino acid sequences. For example, the protein linked to Huntington disease (HTT exon1) is rich in proline and glutamine residues. It is in a disordered conformation but escapes proteasomal degradation even when ubiquitinated, apparently because of its biased sequence (28, 71, 72). Similar biased sequences also exist in the PRNP protein, which is associated with prion disease, in α -synuclein, which is associated with Parkinson disease and other disease-related proteins. The worst initiation regions analyzed here all share a biased amino acid composition with some amino acids strongly over-represented and others missing entirely (28). It is possible that the amino acid composition of the proteins associated with neurodegenerative diseases also makes them more difficult to be recognized by the proteasome and contributes to their accumulation in cells.

The proteasome's sequence preferences are conserved in *S. cerevisiae*, *S. pombe*, and cultured human cells, and they reflect a direct physical interaction between the initiation region in the substrate protein and the proteasome. The receptor of the initiation region on the proteasome is not known, but a likely site is the degradation channel that leads through the ATPase ring of the activator cap to the proteolytic chamber in the core of the proteasome.

The degradation channel is lined by two sets of loops, the P1 and P2 loops, and their sequences are conserved between eukaryotic proteasomes and bacterial AAA⁺ proteases, which fulfill similar functions as the proteasome (2, 73–79). The biochemical mechanism of the *E. coli* AAA⁺ protease ClpXP is particularly well understood (79). The P1 loops undergo conformational changes during ATP-dependent proteolysis and are thought to act as paddles that move the polypeptide chain through the pore (77, 78, 80–82) toward a second binding site formed by the P2 loops (80). Binding to the second site may prevent backsliding of the polypeptide chains between strokes of the P1 paddles (80).

By analyzing ~100 different initiation regions, we were able to determine some sequence properties that govern how well the proteasome is able to initiate degradation. For example, stiffer initiation regions lead to more rapid degradation, perhaps because they increase the distance through space that an initiation region can explore. The P1 loops are located 1–3 nm

from the entrance to the degradation channel as judged by the proteasome structure (83–86) so that stiffer sequences may be able to reach the P1 loops and engage the translocation motor more efficiently than more flexible sequences. The preference for hydrophobic initiation sequences may reflect their ability to interact with the P1 paddles, which have the consensus sequence aromatic-hydrophobic-Gly. The stabilizing effect of acidic sequences may be a reflection of weaker interaction between these sequences and the P2 loop region, which contains Glu and Asp in the proteasome. The sequence preferences could also be a reflection of consensus sequences in initiation regions that are recognized by their receptor on the proteasome. The relationship between sequence complexity and proteasome degradation could then be explained by the fact that biased sequences would be less likely to contain sequences that resemble the consensus motif than diverse sequences (28).

Bacterial AAA⁺ proteases can recognize their substrates and initiate degradation at the same sequences. These targeting sequences fulfill both the recognition and initiation functions, although recognition can be enhanced by targeting adaptors (87). Some 20 ClpXP degrons have been described in *E. coli* (88). The degrons are 10–12 amino acids long (88) and thus appear to be shorter than the proteasome initiation regions discussed here. The shorter length is consistent with the fact that the P1 loops in ClpX are also closer to the entrance to the degradation channel than in the proteasome. One of the bacterial targeting sequences, the SsrA tag, is particularly well characterized and binds ClpXP with ~1 μ M affinity (79). The sequence of the SsrA degron is precisely defined in the sense that single amino acid substitutions can reduce degradation drastically (88). Some eukaryotic proteins are targeted to the proteasome in the absence of ubiquitin, with ODC as the best characterized example (89). However, even the ODC initiation region binds the proteasome with considerably lower affinity than the SsrA tag. Efficient ODC degradation requires the protein Antizyme to serve as an adaptor that increases ODC's affinity to the proteasome (90, 91) in analogy to the bacterial adaptor proteins. So far the proteasome initiation regions seem to be less precisely defined than the ClpXP degrons (88), and the results summarized here did not reveal specific consensus motifs that are recognized by the proteasome. Larger sequence libraries will have to be analyzed to reveal such consensus sequences, if they exist.

In summary, we show that *in vivo* the proteasome has pronounced preferences for the amino acid sequence of its targets at the site where it initiates degradation. The selection of initiation sites represents a degradation code that is embedded within the target proteins. This initiation code is as important and operates in parallel to the ubiquitin code. Its evolutionary conservation suggests that the mechanism may alter the abundance distribution of the proteome in a wide variety of cells and organisms, affecting diverse genetic and regulatory processes.

Experimental Procedures

Substrate Proteins—Proteasome substrate proteins were derived from *E. coli* DHFR, *S. cerevisiae* His3, superfolder GFP (53), and a rapidly maturing derivative of YFP (a gift from B. S. Glick (University of Chicago)). N-terminal UbL domains from

Sequence Preferences in Proteasome Degradation

S. cerevisiae Rad23 or its homologs in *S. pombe* (Rhp23) and human cells (human HR23B) were connected to His3, GFP, or YFP by the linker sequence (VDGGSGGGS). C-terminal tails were attached through a 2-amino acid linker (Pro-Arg), and the amino acid sequences of the tails are shown in supplemental Tables S1 and S2.

Protein Expression and Purification—Yeast proteasome was purified from *S. cerevisiae* strain YYS40 by immunoaffinity chromatography using FLAG antibodies (M2-agarose affinity beads, Sigma), as described previously (92). Proteasome preparations were analyzed by SDS-PAGE and compared with published compositions (83). Each proteasome preparation was tested for activity by measuring degradation of the proteasome substrate UbL-DHFR-95 and for contamination by proteases by testing for stability of proteins that lack a proteasome-binding tag (DHFR-95) as described previously (57, 93).

The substrate (UbL-GFP-95) and competitor proteins (His₁₀-DHFR-tail) used in *in vitro* inhibition assays were overexpressed in *E. coli* and purified using standard methods. Constructs were cloned into a pET3a vector and expressed from the T7 promoter in *E. coli* strains BL21(DE3)pLysS or Rosetta(DE3)pLysS (Novagen). Proteins were purified by TALON metal affinity beads (catalog no. 635502, Clontech) following the manufacturer's instructions. Purified proteins were dialyzed into buffer containing 50 mM Tris, pH 7.4, 300 mM NaCl, and 1% glycerol for storage. Protein concentrations were determined by measuring light absorbance at 280 nm and using extinction coefficients predicted from the proteins' sequence (ExpASY's ProtParam). The integrity and purity of proteins were evaluated by SDS-PAGE.

Competition Assays—The degradation of a fluorescent substrate protein (UbL-GFP-95) *in vitro* was monitored by measuring GFP fluorescence intensity over time in 384-well plates using a plate reader (Infinite M1000 PRO, Tecan) as described previously (94). Assays were carried out at 30 °C by adding fluorescent substrates at the indicated concentrations to 40 nM purified yeast proteasome in a reaction buffer (50 mM Tris, 5 mM MgCl₂, 2.5% glycerol, 1 mM ATP, 4 mM DTT, 0.2 mg ml⁻¹ bovine serum albumin, 10 mM creatine phosphate, 0.1 mg ml⁻¹ creatine kinase, pH 7.5). Fluorescence intensity was read every 30 s for 1 h (excitation, 485 nm/5 nm bandwidth; emission, 535 nm/10 nm bandwidth). Protein amounts were confirmed by comparing the fluorescence intensity of the reaction in each well with calibration curves relating fluorescence intensity to protein concentration. Each assay was repeated at least three times. Initial degradation rates representing the slope of the decay curves at time 0 were calculated as the product of the amplitude, and the rate constant was determined by nonlinear fitting of the time-dependent fluorescence change to the equation describing single exponential decay to a constant offset using the software package KaleidaGraph (version 4.1, Synergy Software).

Yeast Expression—In *S. cerevisiae*, fluorescent proteins were expressed from a CEN plasmid (YCplac33) and His3 proteins from a 2-micron plasmid (pYES2), both with a URA3 selection marker. The plasmids were transformed into *S. cerevisiae* strain BY4741, which carries a deletion of the efflux pump Pdr5 (see

supplemental Table S3 for genotype) (95) using Frozen-EZ Yeast Transformation II kit (Zymo Research).

In *S. pombe*, fluorescent proteins were expressed after integration into the genome using pDUAL-derived plasmids (96). 2 μg of plasmid were digested with NotI and purified (Qiagen). Frozen competent cells of *S. pombe* (see supplemental Table S3 for genotype) were thawed in a 30 °C water bath for 2 min, mixed with purified plasmid DNA, and 30% PEG 3350. The mixture was vortexed, incubated at 30 °C for 1 h, heat shocked at 43 °C for 15 min, and placed at room temperature for 10 min. The cells were then pelleted at 1600 × *g* for 3 min and resuspended in 500 μl of ½YE (0.25% yeast extract, 1.5% glucose). Cells were shaken for 1 h at 30 °C and spread onto synthetic medium lacking leucine. Clones with a single cassette integrated were validated by PCR.

The proteasome substrates and the RFP dsRed-Express2 (38) were expressed from separate promoters on the *S. cerevisiae* and *S. pombe* plasmids as specified. Proteasome activity was inhibited by 100 μM bortezomib. Protein synthesis was inhibited by 125 μM cycloheximide. 2% glucose was used to turn off protein expression from the promoter *gal1*.

Cell Culture Expression—Fluorescent proteins were expressed from the vector pCDNA5 (Life Technologies, Inc.) transiently transfected into HEK293 cells. The proteasome substrates and the RFP mCherry (97) were expressed from a single CMV promoter with the coding regions separated by an internal ribosome entry site derived from encephalomyocarditis virus (98).

HEK293 cells were cultured at 37 °C and 5% CO₂ in DMEM supplemented with 10% FBS, 100 units/ml penicillin, and 100 units/ml streptomycin (Life Technologies, Inc.). Cells were seeded in a 6-well plate at 0.5 × 10⁶ cells/ml 24 h prior to transfection. 1 μg of DNA was transfected into cells with Lipofectamine 2000 (Life Technologies, Inc.) for 24 h. Cells were washed with PBS and recovered in complete DMEM for 24 h before analysis.

Yeast Growth Assay—His3 proteins were expressed from the inducible *gal1* promoter on a 2-micron plasmid (pYES2) with a URA3 selection marker in *S. cerevisiae* strain *pdr5Δ* (see supplemental Table S3 for genotype). Cells were grown with galactose as the carbon source at 30 °C to late log phase, serially diluted (*A*₆₀₀ from 10⁻¹ to 10⁻⁶) and stamped on plates with synthetic defined drop-out medium. Where indicated, 10 mM 3-AT was added to the medium as a competitive inhibitor of His3 to enhance assay sensitivity (55). Plates were incubated at 30 °C for 3 days for imaging.

Flow Cytometry—Yeast cells were grown at 30 °C to early log phase. HEK293 cells were grown for 24 h after transfection, trypsinized with TrypLE (Life Technologies, Inc.), washed, suspended in 0.5 ml of phosphate-buffered saline (PBS), and fixed by the addition of 0.5 ml of 3.7% formaldehyde in PBS. The fluorescence signals in the dsRed (or mCherry) and YFP (or GFP) channels were measured in a flow cytometer (LSR Fortessa, BD Biosciences) and analyzed by FlowJo software to calculate the medians of the YFP over RFP fluorescence ratios for each population. Each assay was repeated at least three times.

Western Blot—Yeast cells were grown to mid-log phase and lysed by vortexing with glass beads (BioSpec Products). Protein extracts were prepared and analyzed by Western blotting using standard protocols as described (28). YFP fusion proteins were detected with a mouse monoclonal anti-enhanced GFP antibody (1:1000, Clontech, catalog no. 632569) and an Alexa-800-labeled goat anti-mouse secondary antibody (1:20,000, Rockland Immunochemicals, catalog no. 610-132-121). Scs2 was detected by an anti-Scs2 rabbit polyclonal antibody (28) (1:1000, gift from J. Brickner, Northwestern University) and an Alexa-680 goat anti-rabbit secondary antibody (1:20,000, Invitrogen, catalog no. A21109). Protein amounts were estimated by direct infrared fluorescence imaging (Odyssey LI-COR Biosciences).

Bioinformatics and Statistical Analysis—The physicochemical properties of 115 sequences were calculated using the following tools: helix propensity, Agadir (99); backbone dynamics, DynaMine (63, 64); disorder propensity, IUPred (100); fraction of aliphatic, acidic, basic, nonpolar, aromatic or polar residues: EMBOSS pepstats (101); entropy, SEG (102, 103); flexibility, FLEXPLOT (62); and hydrophobicity, GRAVY (61). The charge density was calculated as the fraction of basic amino acids minus the fraction of acidic amino acids. Note that the flexibility prediction method of Vihenen *et al.* (62) was optimized for short windows of nine amino acid residues, although we calculated an aggregate flexibility score (mean flex (62)) by averaging the flexibility values for all consecutive nine-residue windows. To avoid making assumptions about the relationships between degradation rates and the numerical values of the sequence properties on their varied scales, we quantified the association of each metric with protein stability using the nonparametric Spearman rank correlation coefficients. Statistical significance of these associations was assessed by testing the hypothesis that a particular parameter does not correlate with degradation using pairwise *t* tests. The *p* values resulting from these tests were then used to compute false discovery rate *q*-values using the Benjamini-Hochberg method (60).

Author Contributions—H. Y., A. K. S. G., S. R. W., D. W., G. K., K. M., M. W., S. C., and T. I. performed the experiments, analyzed the data, and co-wrote the paper. I. J. F., M. M. B., and A. M. directed the experiments, analyzed the data, and co-wrote the paper.

Acknowledgments—We thank Drs. Ben S. Glick (University of Chicago), Jason Brickner (Northwestern University), Raymond Deshaies (California Institute of Technology), Robert T. Sauer (Massachusetts Institute of Technology), Akihisa Matsuyama (RIKEN), as well as Christopher M. Yellman, Blerta Xhemalce, and Richard Salinas (The University of Texas at Austin) for advice and reagents. We are grateful to Kimberly Bowen and Rachael Rossi for constructing *S. pombe* strains. We acknowledge the use of the instruments of the Microscopy and Imaging Facility of the Institute for Cell and Molecular Biology at The University of Texas at Austin.

References

1. Finley, D. (2009) Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **78**, 477–513
2. Finley, D., Chen, X., and Walters, K. J. (2016) Gates, channels, and switches: elements of the proteasome machine. *Trends Biochem. Sci.* **41**, 77–93
3. Komander, D., and Rape, M. (2012) The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–229
4. Husnjak, K., and Dikic, I. (2012) Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. *Annu. Rev. Biochem.* **81**, 291–322
5. Thrower, J. S., Hoffman, L., Rechsteiner, M., and Pickart, C. M. (2000) Recognition of the polyubiquitin proteolytic signal. *EMBO J.* **19**, 94–102
6. Terrell, J., Shih, S., Dunn, R., and Hicke, L. (1998) A function for monoubiquitination in the internalization of a G protein-coupled receptor. *Mol. Cell* **1**, 193–202
7. Hicke, L., and Riezman, H. (1996) Ubiquitination of a yeast plasma membrane receptor signals its ligand-stimulated endocytosis. *Cell* **84**, 277–287
8. Galan, J. M., and Haguenaer-Tsapis, R. (1997) Ubiquitin lys63 is involved in ubiquitination of a yeast plasma membrane protein. *EMBO J.* **16**, 5847–5854
9. Jackson, S. P., and Durocher, D. (2013) Regulation of DNA damage responses by ubiquitin and SUMO. *Mol. Cell* **49**, 795–807
10. Kirkpatrick, D. S., Hathaway, N. A., Hanna, J., Elsasser, S., Rush, J., Finley, D., King, R. W., and Gygi, S. P. (2006) Quantitative analysis of *in vitro* ubiquitinated cyclin B1 reveals complex chain topology. *Nat. Cell Biol.* **8**, 700–710
11. Xu, P., Duong, D. M., Seyfried, N. T., Cheng, D., Xie, Y., Robert, J., Rush, J., Hochstrasser, M., Finley, D., and Peng, J. (2009) Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell* **137**, 133–145
12. Dimova, N. V., Hathaway, N. A., Lee, B.-H., Kirkpatrick, D. S., Berkowitz, M. L., Gygi, S. P., Finley, D., and King, R. W. (2012) APC/C-mediated multiple monoubiquitylation provides an alternative degradation signal for cyclin B1. *Nat. Cell Biol.* **14**, 168–176
13. Kravtsova-Ivantsiv, Y., Cohen, S., and Ciechanover, A. (2009) Modification by single ubiquitin moieties rather than polyubiquitination is sufficient for proteasomal processing of the p105 NF- κ B precursor. *Mol. Cell* **33**, 496–504
14. Lu, Y., Lee, B.-H., King, R. W., Finley, D., and Kirschner, M. W. (2015) Substrate degradation by the proteasome: a single-molecule kinetic analysis. *Science* **348**, 1250834–1250834
15. Shabek, N., Herman-Bachinsky, Y., and Ciechanover, A. (2009) Ubiquitin degradation with its substrate, or as a monomer in a ubiquitination-independent mode, provides clues to proteasome regulation. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11907–11912
16. Baboshina, O. V., and Haas, A. L. (1996) Novel multiubiquitin chain linkages catalyzed by the conjugating enzymes E2EPF and RAD6 are recognized by 26 S proteasome subunit 5. *J. Biol. Chem.* **271**, 2823–2831
17. Hofmann, R. M., and Pickart, C. M. (2001) *In vitro* assembly and recognition of Lys-63 polyubiquitin chains. *J. Biol. Chem.* **276**, 27936–27943
18. Saeki, Y., Kudo, T., Sone, T., Kikuchi, Y., Yokosawa, H., Toh-e, A., and Tanaka, K. (2009) Lysine 63-linked polyubiquitin chain may serve as a targeting signal for the 26S proteasome. *EMBO J.* **28**, 359–371
19. Nathan, J. A., Kim, H. T., Ting, L., Gygi, S. P., and Goldberg, A. L. (2013) Why do cellular proteins linked to K63-polyubiquitin chains not associate with proteasomes? *EMBO J.* **32**, 552–565
20. Jacobson, A. D., Zhang, N.-Y., Xu, P., Han, K.-J., Noone, S., Peng, J., and Liu, C.-W. (2009) The lysine 48 and lysine 63 ubiquitin conjugates are processed differently by the 26 S proteasome. *J. Biol. Chem.* **284**, 35485–35494
21. Prakash, S., Tian, L., Ratliff, K. S., Lehotzky, R. E., and Matouschek, A. (2004) An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nat. Struct. Mol. Biol.* **11**, 830–837
22. Takeuchi, J., Chen, H., and Coffino, P. (2007) Proteasome substrate degradation requires association plus extended peptide. *EMBO J.* **26**, 123–131
23. van der Lee, R., Lang, B., Kruse, K., Gsponer, J., Sánchez de Groot, N., Huynen, M. A., Matouschek, A., Fuxreiter, M., and Babu, M. M. (2014) Intrinsically disordered segments affect protein half-life in the cell and during evolution. *Cell Rep.* **8**, 1832–1844
24. Zhao, M., Zhang, N.-Y., Zurawel, A., Hansen, K. C., and Liu, C.-W. (2010) Degradation of some polyubiquitinated proteins requires an intrinsic proteasomal binding element in the substrates. *J. Biol. Chem.* **285**, 4771–4780

Sequence Preferences in Proteasome Degradation

25. Prakash, S., Inobe, T., Hatch, A. J., and Matouschek, A. (2009) Substrate selection by the proteasome during degradation of protein complexes. *Nat. Chem. Biol.* **5**, 29–36
26. Inobe, T., Fishbain, S., Prakash, S., and Matouschek, A. (2011) Defining the geometry of the two-component proteasome degron. *Nat. Chem. Biol.* **7**, 161–167
27. Fishbain, S., Prakash, S., Herrig, A., Elsasser, S., and Matouschek, A. (2011) Rad23 escapes degradation because it lacks a proteasome initiation region. *Nat. Commun.* **2**, 192
28. Fishbain, S., Inobe, T., Israeli, E., Chavali, S., Yu, H., Kago, G., Babu, M. M., and Matouschek, A. (2015) Sequence composition of disordered regions fine-tunes protein half-life. *Nat. Struct. Mol. Biol.* **22**, 214–221
29. Guharoy, M., Bhowmick, P., Sallam, M., and Tompa, P. (2016) Tripartite degrons confer diversity and specificity on regulated protein degradation in the ubiquitin-proteasome system. *Nat. Commun.* **7**, 10239
30. Rape, M., Hoppe, T., Gorr, I., Kalocay, M., Richly, H., and Jentsch, S. (2001) Mobilization of processed, membrane-tethered SPT23 transcription factor by CDC48(UFD1/NPL4), a ubiquitin-selective chaperone. *Cell* **107**, 667–677
31. Richly, H., Rape, M., Braun, S., Rumpf, S., Hoege, C., and Jentsch, S. (2005) A series of ubiquitin binding factors connects CDC48/p97 to substrate multiubiquitylation and proteasomal targeting. *Cell* **120**, 73–84
32. Wójcik, C., Rowicka, M., Kudlicki, A., Nowis, D., McConnell, E., Kujawa, M., and DeMartino, G. N. (2006) Valosin-containing protein (p97) is a regulator of endoplasmic reticulum stress and of the degradation of N-end rule and ubiquitin-fusion degradation pathway substrates in mammalian cells. *Mol. Biol. Cell* **17**, 4606–4618
33. Beskow, A., Grimberg, K. B., Bott, L. C., Salomons, F. A., Dantuma, N. P., and Young, P. (2009) A conserved unfoldase activity for the p97 AAA-ATPase in proteasomal degradation. *J. Mol. Biol.* **394**, 732–746
34. Raman, M., Havens, C. G., Walter, J. C., and Harper, J. W. (2011) A genome-wide screen identifies p97 as an essential regulator of DNA damage-dependent CDT1 destruction. *Mol. Cell* **44**, 72–84
35. Verma, R., Oania, R., Fang, R., Smith, G. T., and Deshaies, R. J. (2011) Cdc48/p97 mediates UV-dependent turnover of RNA Pol II. *Mol. Cell* **41**, 82–92
36. Barthelme, D., and Sauer, R. T. (2012) Identification of the Cdc48-20S proteasome as an ancient AAA⁺ proteolytic machine. *Science* **337**, 843–846
37. Partow, S., Siewers, V., Bjørn, S., Nielsen, J., and Maury, J. (2010) Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast* **27**, 955–964
38. Strack, R. L., Strongin, D. E., Bhattacharyya, D., Tao, W., Berman, A., Broxmeyer, H. E., Keenan, R. J., and Glick, B. S. (2008) A noncytotoxic DsRed variant for whole-cell labeling. *Nat. Methods* **5**, 955–957
39. Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., and Segal, E. (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* **30**, 521–530
40. Yen, H.-C., Xu, Q., Chou, D. M., Zhao, Z., and Elledge, S. J. (2008) Global protein stability profiling in mammalian cells. *Science* **322**, 918–923
41. Elsasser, S., Gali, R. R., Schwickart, M., Larsen, C. N., Leggett, D. S., Müller, B., Feng, M. T., Tübing, F., Dittmar, G. A., and Finley, D. (2002) Proteasome subunit Rpn1 binds ubiquitin-like protein domains. *Nat. Cell Biol.* **4**, 725–730
42. Husnjak, K., Elsasser, S., Zhang, N., Chen, X., Randles, L., Shi, Y., Hofmann, K., Walters, K. J., Finley, D., and Dikic, I. (2008) Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* **453**, 481–488
43. Shi, Y., Chen, X., Elsasser, S., Stocks, B. B., Tian, G., Lee, B.-H., Shi, Y., Zhang, N., de Poot, S. A., Tübing, F., Sun, S., Vannoy, J., Tarasov, S. G., Engen, J. R., Finley, D., and Walters, K. J. (2016) Rpn1 provides adjacent receptor sites for substrate binding and deubiquitination by the proteasome. *Science* **351**, aad9421
44. Collins, G. A., Gomez, T. A., Deshaies, R. J., and Tansey, W. P. (2010) Combined chemical and genetic approach to inhibit proteolysis by the proteasome. *Yeast* **27**, 965–974
45. Crawford, L. J., Walker, B., Ovaa, H., Chauhan, D., Anderson, K. C., Morris, T. C., and Irvine, A. E. (2006) Comparative selectivity and specificity of the proteasome inhibitors BzLLCCHO, PS-341, and MG-132. *Cancer Res.* **66**, 6379–6386
46. McGrath, J. P., Jentsch, S., and Varshavsky, A. (1991) UBA 1: an essential yeast gene encoding ubiquitin-activating enzyme. *EMBO J.* **10**, 227–236
47. Ghaboosi, N., and Deshaies, R. J. (2007) A conditional yeast E1 mutant blocks the ubiquitin-proteasome pathway and reveals a role for ubiquitin conjugates in targeting Rad23 to the proteasome. *Mol. Biol. Cell* **18**, 1953–1963
48. Bercovich, Z., Rosenberg-Hasson, Y., Ciechanover, A., and Kahana, C. (1989) Degradation of ornithine decarboxylase in reticulocyte lysate is ATP-dependent but ubiquitin-independent. *J. Biol. Chem.* **264**, 15949–15952
49. Murakami, Y., Matsufuji, S., Kameji, T., Hayashi, S., Igarashi, K., Tamura, T., Tanaka, K., and Ichihara, A. (1992) Ornithine decarboxylase is degraded by the 26S proteasome without ubiquitination. *Nature* **360**, 597–599
50. Strack, R. L., Keenan, R. J., and Glick, B. S. (2011) Noncytotoxic DsRed derivatives for whole-cell labeling. *Methods Mol. Biol.* **699**, 355–370
51. Bachmair, A., Finley, D., and Varshavsky, A. (1986) *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179–186
52. Bachmair, A., and Varshavsky, A. (1989) The degradation signal in a short-lived protein. *Cell* **56**, 1019–1032
53. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88
54. Alifano, P., Fani, R., Liò, P., Lazcano, A., Bazzicalupo, M., Carlomagno, M. S., and Bruni, C. B. (1996) Histidine biosynthetic pathway and genes: structure, regulation, and evolution. *Microbiol. Rev.* **60**, 44–69
55. Hawkes, T. R., Thomas, P. G., Edwards, L. S., Rayner, S. J., Wilkinson, K. W., and Rice, D. W. (1995) Purification and characterization of the imidazoleglycerol-phosphate dehydratase of *Saccharomyces cerevisiae* from recombinant *Escherichia coli*. *Biochem. J.* **306**, 385–397
56. Fort, P., Kajava, A. V., Delsuc, F., and Coux, O. (2015) Evolution of proteasome regulators in eukaryotes. *Genome Biol. Evol.* **7**, 1363–1379
57. Kraut, D. A., Israeli, E., Schrader, E. K., Patil, A., Nakai, K., Nanavati, D., Inobe, T., and Matouschek, A. (2012) Sequence- and species-dependence of proteasomal processivity. *ACS Chem. Biol.* **7**, 1444–1453
58. Wilkinson, C. R., Seeger, M., Hartmann-Petersen, R., Stone, M., Wallace, M., Semple, C., and Gordon, C. (2001) Proteins containing the UBA domain are able to bind to multi-ubiquitin chains. *Nat. Cell Biol.* **3**, 939–943
59. Schaubert, C., Chen, L., Tongaonkar, P., Vega, I., Lambertson, D., Potts, W., and Madura, K. (1998) Rad23 links DNA repair to the ubiquitin/proteasome pathway. *Nature* **391**, 715–718
60. Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300
61. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **157**, 105–132
62. Vihinen, M., Torkkila, E., and Riikonen, P. (1994) Accuracy of protein flexibility predictions. *Proteins* **19**, 141–149
63. Cilia, E., Pancsa, R., Tompa, P., Lenaerts, T., and Vranken, W. F. (2014) The DynaMine webserver: predicting protein dynamics from sequence. *Nucleic Acids Res.* **42**, W264–W270
64. Cilia, E., Pancsa, R., Tompa, P., Lenaerts, T., and Vranken, W. F. (2013) From protein sequence to dynamics and disorder with DynaMine. *Nat. Commun.* **4**, 2741
65. Buljan, M., Chalancon, G., Eustermann, S., Wagner, G. P., Fuxreiter, M., Bateman, A., and Babu, M. M. (2012) Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell* **46**, 871–883
66. Weatheritt, R. J., Davey, N. E., and Gibson, T. J. (2012) Linear motifs confer functional diversity onto splice variants. *Nucleic Acids Res.* **40**, 7123–7131
67. Rape, M., Reddy, S. K., and Kirschner, M. W. (2006) The processivity of multiubiquitination by the APC determines the order of substrate degradation. *Cell* **124**, 89–103
68. Williamson, A., Banerjee, S., Zhu, X., Philipp, I., Iavarone, A. T., and

- Rape, M. (2011) Regulation of ubiquitin chain initiation to control the timing of substrate degradation. *Mol. Cell* **42**, 744–757
69. Lu, Y., Wang, W., and Kirschner, M. W. (2015) Specificity of the anaphase-promoting complex: a single-molecule study. *Science* **348**, 1248737
70. Popovic, D., Vucic, D., and Dikic, I. (2014) Ubiquitination in disease pathogenesis and treatment. *Nat. Med.* **20**, 1242–1253
71. Holmberg, C. I., Staniszewski, K. E., Mensah, K. N., Matouschek, A., and Morimoto, R. I. (2004) Inefficient degradation of truncated polyglutamine proteins by the proteasome. *EMBO J.* **23**, 4307–4318
72. Juenemann, K., Schipper-Krom, S., Wiemhoefer, A., Kloss, A., Sanz Sanz, A., and Reits, E. A. (2013) Expanded polyglutamine-containing N-terminal huntingtin fragments are entirely degraded by mammalian proteasomes. *J. Biol. Chem.* **288**, 27068–27084
73. Wang, J., Song, J. J., Franklin, M. C., Kamtekar, S., Im, Y. J., Rho, S. H., Seong, I. S., Lee, C. S., Chung, C. H., and Eom, S. H. (2001) Crystal structures of the HslVU peptidase-ATPase complex reveal an ATP-dependent proteolysis mechanism. *Structure* **9**, 177–184
74. Yamada-Inagawa, T., Okuno, T., Karata, K., Yamanaka, K., and Ogura, T. (2003) Conserved pore residues in the AAA protease FtsH are important for proteolysis and its coupling to ATP hydrolysis. *J. Biol. Chem.* **278**, 50182–50187
75. Hinnerwisch, J., Fenton, W. A., Furtak, K. J., Farr, G. W., and Horwich, A. L. (2005) Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation. *Cell* **121**, 1029–1041
76. Zhang, F., Wu, Z., Zhang, P., Tian, G., Finley, D., and Shi, Y. (2009) Mechanism of substrate unfolding and translocation by the regulatory particle of the proteasome from *Methanocaldococcus jannaschii*. *Mol. Cell* **34**, 485–496
77. Martin, A., Baker, T. A., and Sauer, R. T. (2008) Pore loops of the AAA⁺ ClpX machine grip substrates to drive translocation and unfolding. *Nat. Struct. Mol. Biol.* **15**, 1147–1151
78. Beckwith, R., Estrin, E., Worden, E. J., and Martin, A. (2013) Reconstitution of the 26S proteasome reveals functional asymmetries in its AAA⁺ unfoldase. *Nat. Struct. Mol. Biol.* **20**, 1164–1172
79. Baker, T. A., and Sauer, R. T. (2012) ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim. Biophys. Acta* **1823**, 15–28
80. Iosefson, O., Olivares, A. O., Baker, T. A., and Sauer, R. T. (2015) Dissection of axial-pore loop function during unfolding and translocation by a AAA⁺ proteolytic machine. *Cell Rep.* **12**, 1032–1041
81. Martin, A., Baker, T. A., and Sauer, R. T. (2008) Diverse pore loops of the AAA⁺ ClpX machine mediate unassisted and adaptor-dependent recognition of ssrA-tagged substrates. *Mol. Cell* **29**, 441–450
82. Koga, N., Kameda, T., Okazaki, K.-I., and Takada, S. (2009) Paddling mechanism for the substrate translocation by AAA⁺ motor revealed by multiscale molecular simulations. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18237–18242
83. Lander, G. C., Estrin, E., Matyskiela, M. E., Bashore, C., Nogales, E., and Martin, A. (2012) Complete subunit architecture of the proteasome regulatory particle. *Nature* **482**, 186–191
84. Matyskiela, M. E., Lander, G. C., and Martin, A. (2013) Conformational switching of the 26S proteasome enables substrate degradation. *Nat. Struct. Mol. Biol.* **20**, 781–788
85. Ślędz, P., Unverdorben, P., Beck, F., Pfeifer, G., Schweitzer, A., Förster, F., and Baumeister, W. (2013) Structure of the 26S proteasome with ATP-γS bound provides insights into the mechanism of nucleotide-dependent substrate translocation. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7264–7269
86. Unverdorben, P., Beck, F., Ślędz, P., Schweitzer, A., Pfeifer, G., Plitzko, J. M., Baumeister, W., and Förster, F. (2014) Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5544–5549
87. Kirstein, J., Molière, N., Dougan, D. A., and Turgay, K. (2009) Adapting the machine: adaptor proteins for Hsp100/Clp and AAA⁺ proteases. *Nat. Rev. Microbiol.* **7**, 589–599
88. Flynn, J. M., Neher, S. B., Kim, Y. L., Sauer, R. T., and Baker, T. A. (2003) Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol. Cell* **11**, 671–683
89. Erales, J., and Coffino, P. (2014) Ubiquitin-independent proteasomal degradation. *Biochim. Biophys. Acta* **1843**, 216–221
90. Wu, H.-Y., Chen, S.-F., Hsieh, J.-Y., Chou, F., Wang, Y.-H., Lin, W.-T., Lee, P.-Y., Yu, Y.-J., Lin, L.-Y., Lin, T.-S., Lin, C.-L., Liu, G.-Y., Tzeng, S.-R., Hung, H.-C., and Chan, N.-L. (2015) Structural basis of antizyme-mediated regulation of polyamine homeostasis. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11229–11234
91. Chen, H., MacDonald, A., and Coffino, P. (2002) Structural elements of antizymes 1 and 2 are required for proteasomal degradation of ornithine decarboxylase. *J. Biol. Chem.* **277**, 45957–45961
92. Saeki, Y., Isono, E., and Toh-E, A. (2005) Preparation of ubiquitinated substrates by the PY motif-insertion method for monitoring 26S proteasome activity. *Methods Enzymol.* **399**, 215–227
93. Kraut, D. A., and Matouschek, A. (2011) Proteasomal degradation from internal sites favors partial proteolysis via remote domain stabilization. *ACS Chem. Biol.* **6**, 1087–1095
94. Martinez-Fonts, K., and Matouschek, A. (2016) A rapid and versatile method for generating proteins with defined ubiquitin chains. *Biochemistry* **55**, 1898–1908
95. Fleming, J. A., Lightcap, E. S., Sadis, S., Thoroddsen, V., Bulawa, C. E., and Blackman, R. K. (2002) Complementary whole-genome technologies reveal the cellular response to proteasome inhibition by PS-341. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1461–1466
96. Matsuyama, A., Shirai, A., Yashiroda, Y., Kamata, A., Horinouchi, S., and Yoshida, M. (2004) pDUAL, a multipurpose, multicopy vector capable of chromosomal integration in fission yeast. *Yeast* **21**, 1289–1305
97. Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N., Palmer, A. E., and Tsien, R. Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572
98. Wilmington, S. R., and Matouschek, A. (2016) An inducible system for rapid degradation of specific cellular proteins using proteasome adaptors. *PLoS ONE* **11**, e0152679
99. Muñoz, V., and Serrano, L. (1994) Elucidating the folding problem of helical peptides using empirical parameters. *Nat. Struct. Biol.* **1**, 399–409
100. Dosztányi, Z., Csizmek, V., Tompa, P., and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433–3434
101. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277
102. Wootton, J. C. (1994) Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput. Chem.* **18**, 269–285
103. Wootton, J. C., and Federhen, S. (1996) Analysis of compositionally biased regions in sequence databases. *Methods Enzymol.* **266**, 554–571
104. Larman, H. B., Zhao, Z., Laserson, U., Li, M. Z., Ciccia, A., Gakidis, M. A., Church, G. M., Kesari, S., Leproust, E. M., Solimini, N. L., and Elledge, S. J. (2011) Autoantigen discovery with a synthetic human peptidome. *Nat. Biotechnol.* **29**, 535–541

Autoinhibitory Interdomain Interactions and Subfamily-specific Extensions Redefine the Catalytic Core of the Human DEAD-box Protein DDX3*

Received for publication, October 26, 2015, and in revised form, November 18, 2015. Published, JBC Papers in Press, November 23, 2015, DOI 10.1074/jbc.M115.700625

Stephen N. Floor^{‡§1}, Kendall J. Condon[‡], Deepak Sharma^{¶||}, Eckhard Jankowsky^{¶||}, and Jennifer A. Doudna^{‡§***##\$2}

From the [‡]Department of Molecular and Cell Biology, [§]Howard Hughes Medical Institute, ^{**}Department of Chemistry, and ^{##}Innovative Genomics Initiative, University of California, Berkeley, California 94720, the ^{§§}Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, and the [¶]Department of Biochemistry and ^{||}Center for RNA Molecular Biology, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106

DEAD-box proteins utilize ATP to bind and remodel RNA and RNA-protein complexes. All DEAD-box proteins share a conserved core that consists of two RecA-like domains. The core is flanked by subfamily-specific extensions of idiosyncratic function. The Ded1/DDX3 subfamily of DEAD-box proteins is of particular interest as members function during protein translation, are essential for viability, and are frequently altered in human malignancies. Here, we define the function of the subfamily-specific extensions of the human DEAD-box protein DDX3. We describe the crystal structure of the subfamily-specific core of wild-type DDX3 at 2.2 Å resolution, alone and in the presence of AMP or nonhydrolyzable ATP. These structures illustrate a unique interdomain interaction between the two ATPase domains in which the C-terminal domain clashes with the RNA-binding surface. Destabilizing this interaction accelerates RNA duplex unwinding, suggesting that it is present in solution and inhibitory for catalysis. We use this core fragment of DDX3 to test the function of two recurrent medulloblastoma variants of DDX3 and find that both inactivate the protein *in vitro* and *in vivo*. Taken together, these results redefine the structural and functional core of the DDX3 subfamily of DEAD-box proteins.

DEAD-box proteins are ATP-dependent RNA-binding proteins that remodel RNA structures and RNA-protein complexes, stably clamp RNA, and promote fluidity within RNA granules (1–3). The human DEAD-box protein DDX3 (encoded by *DDX3X*) and its yeast ortholog Ded1p have been implicated in numerous functions including translation initiation (4–12). Messenger RNA molecules containing especially long or structured 5' leader sequences are particularly sensitive

to DDX3 activity (6, 7, 12, 13). *DDX3X* is frequently mutated in numerous cancer types (5), such as chronic lymphocytic leukemia (14–16), natural killer/T-cell lymphoma (17), head and neck squamous cell carcinoma (18, 19), and lung cancer (20). *DDX3X* is also one of the most frequently mutated genes in the highly malignant brain tumor medulloblastoma (21–24). In medulloblastoma, many mutations are predicted to inactivate DDX3, and some have been demonstrated to diminish activities *in vitro* (17, 25).

DEAD-box proteins are defined by 12 different motifs that function in ATP binding or hydrolysis and RNA binding, or couple ATP and RNA binding (1). Outside of these conserved motifs, each DEAD-box protein subfamily has unique tails that lie N- or C-terminal to the helicase core and contain elements that define the unique properties of that subfamily. For example, DDX21 has a GUCT domain in its C-terminal extension (26, 27), DDX5 has tandem P68HR domains in its C-terminal extension, and DDX43 has a KH1 domain in its N-terminal extension. However, as the tails of each DEAD-box protein subfamily are idiosyncratic, whereas the cores are very similar (28, 29), it is essential to study individual subfamilies of DEAD-box proteins in detail to understand the role of subfamily-specific tails.

DDX3 is a member of the Ded1/DDX3 subfamily, along with the *Saccharomyces cerevisiae* ortholog *DED1*, and Vasa/DDX4 (5, 30). The tails of Ded1/DDX3 subfamily members are thought to be largely unstructured and contain diverse motifs with different functions. For example, the N-terminal tail of DDX3 contains a Crm1-dependent nuclear export sequence (31) and an eIF4E-binding motif (10, 11), whereas the C-terminal tail contains conserved sequences of unknown function that are essential for oligomerization (5, 32). The tails of *DED1* additionally contain assembly domains that modulate translation and alleviate lethality associated with protein overexpression when deleted (10). The minimal functional core of DEAD-box proteins has been defined as the isolated phenylalanine upstream of the Q-motif through roughly 35 residues beyond Motif VI (29). However, this analysis does not consider subfamily-specific extensions. Similarly, prior structural work has truncated one (25) or both (33) of the Ded1/DDX3 subfamily-specific regions down to the boundary of the helicase core, but it is unknown how active these truncations are when compared with full-length DDX3. Moreover, previous structures of DDX3

* This work was supported by National Institutes of Health NIGMS Grants GM0067700 and GM099720 (to E. J.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

The atomic coordinates and structure factors (codes 5E7I, 5E7J, and 5E7M) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

¹ A Howard Hughes Medical Institute fellow of the Helen Hay Whitney Foundation.

² An investigator of the Howard Hughes Medical Institute. To whom correspondence should be addressed: Dept. of Molecular and Cell Biology, University of California, 708A Stanley Hall, Berkeley, CA 94720. Tel.: 510-643-0225; Fax: 510-643-0080; E-mail: doudna@berkeley.edu.

are of either inactive constructs (33) or point mutants (25), so the relevance of the crystallized conformations is unclear. Therefore, crucial details of the regions uniquely conserved within the Ded1/DDX3 subfamily and the conformational landscape of active, wild-type DDX3 remain incompletely understood.

Here, we assay the function of the tails of human DDX3 by generating a series of truncations and exploring their activity *in vitro* and *in vivo*. We define an active but truncated construct of DDX3 and solve its crystal structure at 2.2 Å resolution in a partially closed state, representing the highest resolution structure of active, wild-type DDX3 to date. We find that this partially closed state is autoinhibited and demonstrate that mutations predicted to destabilize this conformation accelerate RNA duplex unwinding by DDX3. Using molecular dynamics simulations, we show that the ATP-binding loop of DDX3 samples transient interactions with ATP and that the partially closed state is stable in solution. Lastly, we test two recurrent medulloblastoma variants of DDX3 and find that they inactivate duplex unwinding by up to 3 log units. Our work defines a functional truncation of DDX3 that purifies to high yield, elucidates the function of the signature tails found in the Ded1/DDX3 subfamily of DEAD-box proteins, presents high-resolution structural information of active DDX3 of utility for molecular modeling and drug design, and demonstrates the consequences of two medulloblastoma-associated DDX3 variants.

Experimental Procedures

Recombinant Protein Purification—The Ded1/DDX3 subfamily core of DDX3 was expressed using a construct containing *Escherichia coli* codon-optimized, human DDX3X amino acids 132–607 fused to a His₆-MBP (maltose-binding protein) tag and expressed in *E. coli* BL21 Star by induction with isopropyl-1-thio- β -D-galactopyranoside at 16 °C for 18 h. Cell pellets were lysed by sonication, clarified by centrifugation at $\sim 30,000 \times g$, and purified by nickel chromatography including a 1 M NaCl wash to remove bound nucleic acids. The His₆-MBP tag was cleaved using tobacco etch virus protease during dialysis into 200 mM NaCl, 10% (v/v) glycerol, 20 mM HEPES, pH 7, and 0.5 mM TCEP.³ The sample was then purified using heparin affinity chromatography, eluted at 400 mM NaCl, 10% glycerol, 20 mM HEPES, pH 7, and 0.5 mM TCEP, and applied to a Superdex 75 gel filtration column equilibrated in 500 mM NaCl, 10% glycerol, 20 mM HEPES, pH 7.5, and 0.5 mM TCEP. Fractions were then concentrated and supplemented with 20% (v/v) glycerol and flash-frozen for kinetics, or used directly for crystallization. Typical yield was ~ 10 mg of purified protein per liter. Point mutants were generated by site-directed mutagenesis.

Full-length DDX3X with a His₆ tag was cloned into a pET-29a vector and expressed in *E. coli* BL21 (37 °C). Cells were processed as described previously for the purification of Ded1p (34). Lysates were passed through pre-equilibrated nickel-aga-

rose beads and washed with increasing imidazole concentrations (5–60 mM) (34). DDX3X was eluted in 250 mM imidazole. The His₆ tag was cleaved using tobacco etch virus protease in 50 mM Tris-Cl (pH 8.0), 0.5 mM EDTA, 1 mM DTT, and 40% (v/v) glycerol. DDX3X was further purified by adsorption to phosphocellulose resin (P11, Whatman) and elution with NaCl, as described for Ded1p (34). Eluted fractions were analyzed by SDS-PAGE and Western blotting using anti-His antibody to confirm removal of the His₆ tag. DDX3X fractions were supplemented to 40% (v/v) glycerol, flash-frozen in liquid nitrogen, and stored at -80 °C.

X-ray Crystallography—Purified protein was concentrated to ~ 5 mg ml⁻¹ and mixed 1:1 with precipitant solution containing from 6 to 12% PEG 3000 and 100 mM sodium citrate, pH 5.0, and crystallization was achieved by hanging drop vapor diffusion at 18 °C within 24 h. Nucleotide-bound crystals were grown identically but supplemented with 10 mM of the nucleotide in solution. Homogeneous pieces of branched crystals were harvested for data collection, which was conducted at Beamline 8.3.1 of the Advanced Light Source. Data were indexed, integrated, and scaled using XDS (35), phased using molecular replacement with PHASER (36) with both domains from Protein Data Bank (PDB) 2I4I (33) as independent search models, and refined and built using PHENIX (37) and Coot (38). High resolution was determined by $CC_{1/2} \geq 10\%$ (39). Structures were visualized with PyMOL (40).

RNA Duplex Unwinding Assays—Assays were performed as described (41) with minor modifications. Briefly, duplex RNAs containing a 3' overhang were formed by radiolabeling a single-stranded RNA, annealing, and using gel purification. The two RNA sequences are 5'-AGCACCGUAAAGACGC-3' and 5'-GCGUCUUUACGGUGCUUAAAACAAAACAAAACAA-AACAAA-3'. Reactions contained trace duplex RNA and 1 μ M protein and were initiated by the addition of 2 mM MgATP.

Yeast Complementation Assays—A yeast strain containing the genomic region surrounding *DED1* on a centromeric plasmid was used for all experiments (10). Mutations were made in a *HIS3*-marked plasmid and exchanged for the wild-type allele by plasmid shuffling using counterselection with 5-fluoroorotic acid. Growth assays show 10-fold dilution from an optical density of 1 and were conducted using rich medium (YPD) plates at the temperature indicated.

Multiple Sequence Alignments—Sequences for *Drosophila melanogaster* Vasa, *Homo sapiens* DDX4, *Danio rerio* PL10, *Mus musculus* PL10, *M. musculus* DDX3Y, *H. sapiens* DDX3Y, *M. musculus* DDX3X, *H. sapiens* DDX3X, *D. melanogaster* belle, *S. cerevisiae* DED1, and *Schizosaccharomyces pombe* sum3 were retrieved from the National Center for Biotechnology Information (NCBI) and aligned using MUSCLE (42). Sequence alignments were visualized using Jalview (43). The DEAD-box protein motifs shown in Fig. 6A were generated using WebLogo with input from all human DEAD-box protein sequences aligned with MAFFT (44).

Molecular Dynamics Simulations—Molecular dynamics simulations were performed using Gromacs version 4.6.5 (45). The apo structure was used as the starting point for all simulations (PDB 5E7I) with missing loops built using UCSF Chimera version 1.8.1 (46) and Modeller version 9.12 (47). A rhombic

³ The abbreviations used are: TCEP, tris(2-carboxyethyl)phosphine; AMPPNP, 5'-adenylyl- β , γ -imidodiphosphate; CTE, C-terminal extension; NTE, N-terminal extension; ABL, ATP-binding loop; CC, correlation coefficient.

Redefining the DDX3 Catalytic Core

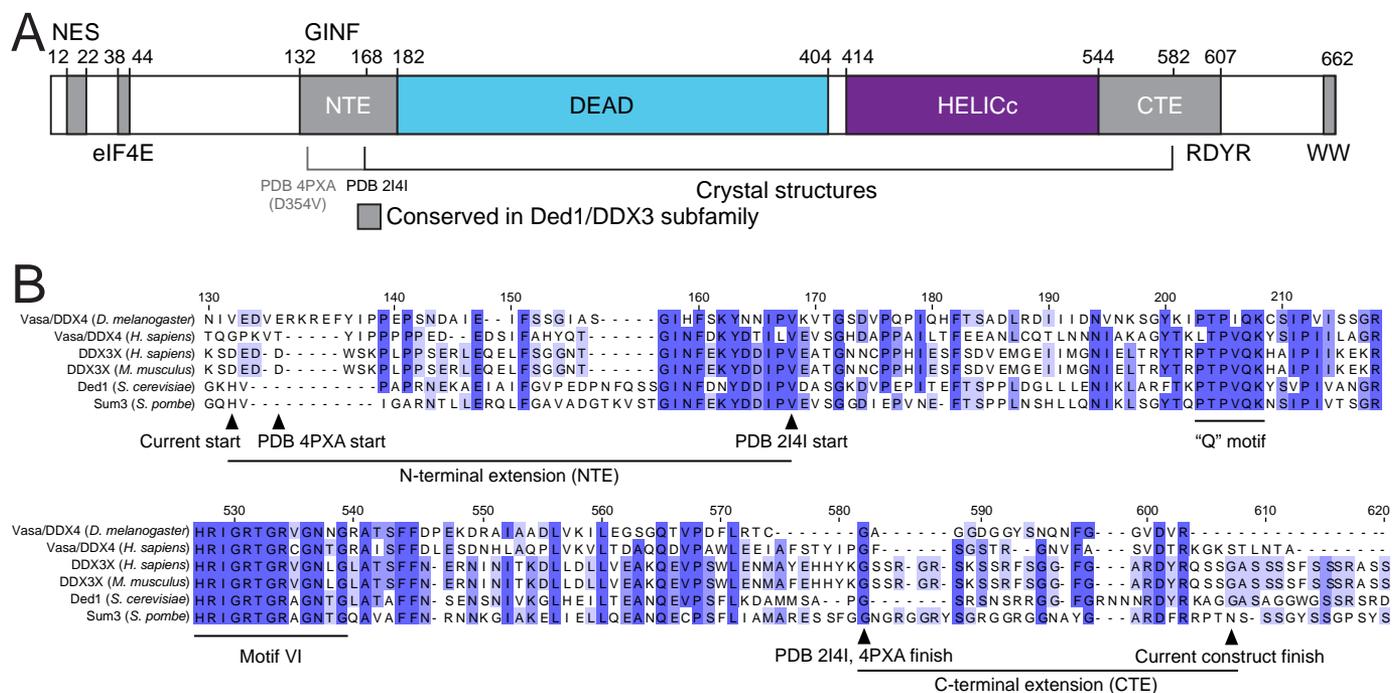


FIGURE 1. Conserved regions specific to the Ded1/DDX3 subfamily of DEAD-box proteins. *A*, a linear diagram of the features of the Ded1/DDX3 family, including the core helicase DEAD and HELICc RecA-like domains, the Crm1-dependent nuclear export sequence (NES), and the eIF4E-binding site. Conserved regions specific to the Ded1/DDX3 subfamily are indicated in gray; GINF, RDYR, and WW refer to amino acid motifs. The NTE and CTE are indicated. *B*, sequence alignments showing the N- and C-terminal extensions found in the Ded1/DDX3 family and indicating construct boundaries used in this study and two previous crystal structures. Numbers correspond to human DDX3X.

dodecahedral, periodic simulation box was used with a buffer of 12 Å between the solute and the boundary. Na⁺ and Cl⁻ ions were added for charge neutralization followed by vacuum and solvent equilibration using transferable intermolecular potential 3 point (TIP3P) water (48). The production simulation length was 100 ns, and the step size was 2 fs. For modeling, the ATP-bound, Vasa closed-state domains of the AMPPNP structure (PDB 5E7M) were superposed onto the corresponding domains of the Vasa structure using PyMOL. Then, missing loops were added using UCSF Chimera and Modeller. Initial energies were high, likely reflecting a clash due to imperfect alignment, but quickly equilibrated in vacuum. Interatomic distances along trajectories were calculated using VMD version 1.9.1 (49).

Results

The Ded1/DDX3 Subfamily Contains Conserved Regions outside the Helicase Core—To define the minimal active construct of human DDX3X, we aligned Ded1/DDX3 family members from diverse species. In addition to the RecA-like core domains, five regions of sequence conservation unique to this family are apparent (Fig. 1A). The N-terminal conserved sequences correspond to a Crm1-dependent nuclear export sequence (31) and an eIF4E-binding site (10, 11), and the C terminus contains an RDYR motif and an invariant WW dipeptide motif (Fig. 1) (5). In addition, there are regions adjacent to the helicase core conserved between DDX3, Ded1, and Vasa/DDX4 (Fig. 1A) (5, 25). The N-terminal extension (NTE; residues 132–168; Fig. 1B) is predicted by PSIPRED (50) to form a short α -helix from residues 145 to 151, whereas the C-terminal extension (CTE; residues 582–607; Fig. 1B) has no predicted structure

but is highly positively charged (pI ~12). Recent structural and biochemical work demonstrated that DDX3 constructs containing the NTE are competent for ATP hydrolysis (25), but it is unknown how this activity compares with full-length human DDX3.

The CTE Is Essential for RNA Duplex Unwinding and Affects Yeast Growth—To compare the activity of full-length DDX3 with truncated variants, we expressed and purified full-length DDX3 and truncations of the NTE, CTE, or both and then measured RNA duplex unwinding activities. Truncation of 131 residues from the N terminus and 55 residues from the C terminus yields a functional core of DDX3 that robustly unwinds RNA duplexes, although this construct has a roughly 5-fold lower functional affinity for RNA when compared with the full-length protein (Fig. 2A). Both full-length and truncated protein show sigmoidal functional binding isotherms, suggesting that DDX3 functions as an oligomer, as yeast DED1 (32).

Removal of the CTE containing the RDYR motif severely diminishes duplex unwinding, independent of the presence of the NTE (Fig. 2A). It is possible that removal of this positively charged region in the CTE negatively impacts RNA binding, as suggested by weaker binding to heparin resin (data not shown) and by the RNA binding defect caused by deletion of the entire C-terminal tail of Ded1p up to the helicase core (29). Alternatively, or in addition, this region might be critical for oligomerization (32).

To assess the biological function of the truncated proteins, we tested the ability of truncated versions of DED1 to support yeast growth. We generated truncations of DED1 in a plasmid and shuffled these into a yeast strain containing the sole copy of

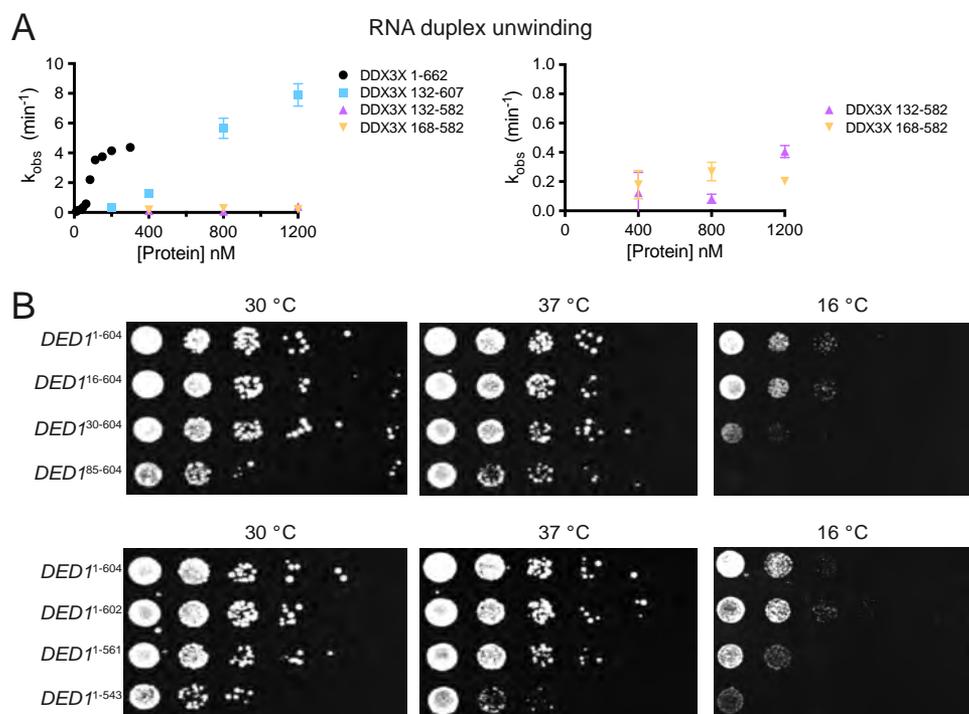


FIGURE 2. **Subfamily-specific extensions to the DEAD-box core are essential for DDX3 function.** *A*, RNA duplex unwinding rates at the indicated concentrations for four truncations of DDX3. Error bars indicate S.D. *B*, yeast growth assays for strains containing truncated copies of the essential gene *DED1* demonstrate a requirement for the region conserved in the Ded1/DDX3 family. *DED1* residues 16, 30, 85, 543, 561, and 604 correspond to DDX3 residues 23, 47, 122, 592, 607, and 661, respectively. Strains containing truncations of both tails were nonviable (data not shown).

DED1 on a plasmid under control of its endogenous promoter (10). Under normal growth conditions, *DED1* truncations lacking both the nuclear export sequence and the eIF4E-binding site complement yeast growth fully. Truncation up to the boundary of the NTE is viable but confers a slow growth phenotype (Fig. 2*B*). It is intriguing that this strain fully complements *DED1* because Ded1p is thought to be primarily nuclear localized when the nuclear export sequence is deleted (51). We were unable to generate strains lacking the ATP-binding loop (ABL; *DED1*^{130–604}; data not shown). The *DED1*^{30–604} and *DED1*^{85–604} strains are cold-sensitive, indicating a special requirement for the N-terminal tail at cold temperatures. Similarly, truncation up to the boundary of the CTE is tolerated, but deletion of the CTE results in slowed growth (Fig. 2*B*), whereas further truncation into the CTE exhibits cold sensitivity (10) or very weak complementation (29). These data demonstrate that the activity of the Ded1/DDX3 subfamily of DEAD-box proteins is surprisingly resilient to truncation of the subfamily-specific tails, but only up to conserved regions adjacent to the helicase core. Furthermore, inclusion of the CTE is essential for activity *in vitro* and *in vivo*.

The 2.2 Å Crystal Structure of Wild-type DDX3 132–607—To better understand the role of the NTE and CTE, we solved the crystal structure of AMP-bound wild-type DDX3 132–607 to ~2.2 Å resolution (Table 1). The DEAD-domain is oriented uniquely with respect to the HELICc domain when compared with two other structures of DDX3 (25, 33) (Fig. 3*A*). Interestingly, the orientation of the present structure and DDX3 135–582 (Δ CTE; PDB 4PXA) is more similar than DDX3 168–582 (Δ NTE and Δ CTE; PDB 2I4I), suggesting that inclusion of the NTE biases crystallization toward this conformation. However,

the interdomain orientation between the present structure and 4PXA is different, likely because PDB 4PXA contains a D354V mutation, which is located at the interdomain interface in both structures. Comparison of the present crystal structure with the structure of Vasa bound to RNA shows that the crystallized conformation of DDX3 is refractory to RNA binding, as the HELICc domain overlaps with the bound RNA in the Vasa structure (Fig. 3*B*; PDB 2DB3) (52). The CTE is predicted to be disordered by PSIPRED, and we observe no density past residue 584 in any of our structures, despite the obvious requirement for this region for protein function *in vitro* and *in vivo* (Fig. 2). Therefore, DDX3 preferentially crystallizes in a partially closed conformation with interdomain interactions between the DEAD and HELICc domains that are refractory to catalysis.

We observe a short α -helix from residues 146–151 (the 150's helix) in the NTE as predicted by PSIPRED and seen in the DDX3 135–582 structure (25). The NTE additionally contains the ABL, which is disordered in our structure and 4PXA (25) but forms a short α -helix in the structure of Vasa bound to RNA (52). Therefore, we crystallized DDX3 132–607 in the presence of no ligand, ADP, or AMPPNP to see whether the conformation of the ABL was altered (Table 1). In all cases, the ABL was disordered and difficult to model robustly. We elected to not build residues 155–165 of the NTE in any of these structures as repeated model building and refinement indicated insufficient density in this region to specify a unique structure. Thus, the ABL is dynamic in the presence of adenosine phosphates, and may fold into an α -helix cooperatively with interdomain closure and RNA binding, as seen in the Vasa structure bound to ssRNA (52).

TABLE 1
Data collection and refinement statistics

Statistics for the highest-resolution shell are shown in parentheses. RMS, root mean square.

	DDX3 132–607 AMPNP (5E7M)	DDX3 132–607 AMP (5E7J)	DDX3 132–607 apo (5E7I)
Wavelength (Å)	1.115869	1.115869	1.115869
Resolution range	47.29–2.304 (2.386–2.304)	46.18–2.229 (2.309–2.229)	84.33–2.223 (2.303–2.223)
Space group	P 21 21 21	P 21 21 21	P 1 21 1
Unit cell	53.93 100.28 107.26 90 90 90	51.35 89.91 107.65 90 90 90	92.71 105.38 94.61 90 114.544 90
Total reflections	198,184 (19,421)	96,467 (5247)	311,845 (18,681)
Unique reflections	26,335 (1930)	24,246 (1660)	79,732 (6607)
Multiplicity	7.5 (7.6)	4.0 (2.8)	3.9 (2.8)
Completeness (%)	0.97 (0.98)	0.96 (0.76)	0.98 (0.82)
Mean $I/\sigma(I)$	10.80 (0.12)	16.59 (2.42)	8.24 (0.95)
Wilson B-factor	75.58	31.24	33
R_{merge}	0.1274 (15.21)	0.06914 (0.5794)	0.1617 (1.306)
R_{meas}	0.1372 (16.32)	0.07978 (0.7047)	0.1873 (1.593)
$CC_{1/2}$	0.999 (0.111)	0.998 (0.931)	0.991 (0.362)
CC^*	1 (0.447)	1 (0.982)	0.998 (0.729)
Reflections used in refinement	25,653 (1928)	24,004 (1644)	79,682 (6593)
Reflections used for R_{free}	1281 (95)	1200 (81)	3985 (332)
R_{work}	0.2279 (0.5683)	0.2328 (0.4521)	0.2270 (0.3966)
R_{free}	0.2856 (0.5753)	0.2649 (0.4761)	0.2599 (0.4120)
CC_{work}	0.953 (0.377)	0.940 (0.556)	0.950 (0.446)
CC_{free}	0.942 (0.283)	0.950 (0.563)	0.940 (0.477)
Number of non-hydrogen atoms	3390	3446	10,060
Macromolecules	3354	3446	10,060
Ligands	31	23	0
Protein residues	424	433	1272
RMS (bonds)	0.01	0.003	0.004
RMS (angles)	1.43	0.82	0.84
Ramachandran favored (%)	93	98	98
Ramachandran allowed (%)	6.5	2.1	2.1
Ramachandran outliers (%)	0.48	0	0.24
Rotamer outliers (%)	2.7	0.27	1
Clashscore	16.42	4.22	5.37
Average B-factor	100.35	37.4	46.72
Macromolecules	100.18	37.4	46.72
Ligands	118.05	42.53	
Solvent	102.92		

Mutation of an Interdomain Interacting Residue Accelerates Duplex Unwinding—If the crystallized conformation of DDX3 is present in solution, then it should inhibit duplex unwinding by DDX3 because it is refractory to RNA binding (Fig. 3). The interdomain interface buries $\sim 560 \text{ \AA}^2$, suggesting that it may be stable in solution. Therefore, we hypothesized that mutation of the interdomain interface should accelerate duplex unwinding. Four residues make apparent interdomain contacts: Asp-354 and Glu-388 form a salt bridge with His-527 (separated by 4.0 and 2.7 Å, respectively), and Asp-506 hydrogen-bonds to the backbone of Arg-276 and caps an α -helix (Fig. 4A; 2.8 Å distance). We targeted residues Asp-354 and Glu-388 for mutation because Asp-506 and His-527 are members of conserved motifs Va and VI, respectively (1). Mutation of residue Asp-354 to either alanine or tryptophan (to sterically block formation of the partially closed state) results in decreased activity (Fig. 4B). Although Asp-354 is not part of a pan-DEAD-box motif, it is conserved within the Ded1/DDX3 subfamily, including Vasa. Interestingly, the structure of DDX3 with a D354V mutation shows a different closed state (25), supporting the role of this residue in interdomain interactions. In contrast, mutation of Glu-388 to alanine, arginine, or tryptophan accelerates duplex unwinding by a factor of two (Fig. 4B). In the closed, RNA-bound structure of *D. melanogaster* Vasa, residue Glu-438 (DDX3 Glu-388) is solvent-exposed and distal to the RNA-binding site, making it unlikely that these mutations increase activity by altered RNA or interdomain interactions. As a comparison, we tested mutation of the conserved GINF motif in the NTE (Fig. 1A) and found that it decreases duplex unwinding by

a factor of two, supporting the conclusion that the ABL is necessary for catalysis (Fig. 4B) (25). In sum, mutations predicted to destabilize the crystallized interdomain interface accelerate duplex unwinding by DDX3, consistent with the presence of the partially closed structure in solution and its inhibitory nature.

The Partially Closed Structure Is Stable over 100 ns of Molecular Dynamics Simulation—As the crystal structure was solved at pH 5 and crystals can trap transient structures, we tested the stability of the partially closed structure and the interactions of the ABL with ATP at neutral pH by performing 100-ns molecular dynamics simulations of apo DDX3 132–607 apo and ATP-bound DDX3 132–607. The ATP-bound state was modeled off the closed form of Vasa bound to RNA to attempt to induce structure formation in the ABL, but instead quickly equilibrated toward an unobserved, alternative closed state, which was not pursued further. In contrast, the partially closed interdomain interface remained stable for the full 100-ns simulation when started from the apo crystal structure (PDB 5E7I) (Fig. 5A). Similarly, in both the partially closed and the ATP-bound states, the 150's helix remained stably docked to the side of the DEAD domain (Fig. 5B). In contrast, the ABL is dynamic. In the ATP-bound simulation, the ABL makes transient interactions with the adenine moiety of ATP (Fig. 5C) involving interactions between Lys-162 and ATP and Phe-160 cation- π stacking with Lys-162 (Fig. 5D). In this state, Lys-162 forms a bipartite hydrogen bond with $\sim 3 \text{ \AA}$ separation from both a ribose oxygen and the adenine N3 of ATP and is separated from the Phe-160 aromatic ring by $\sim 4 \text{ \AA}$. These simulations show

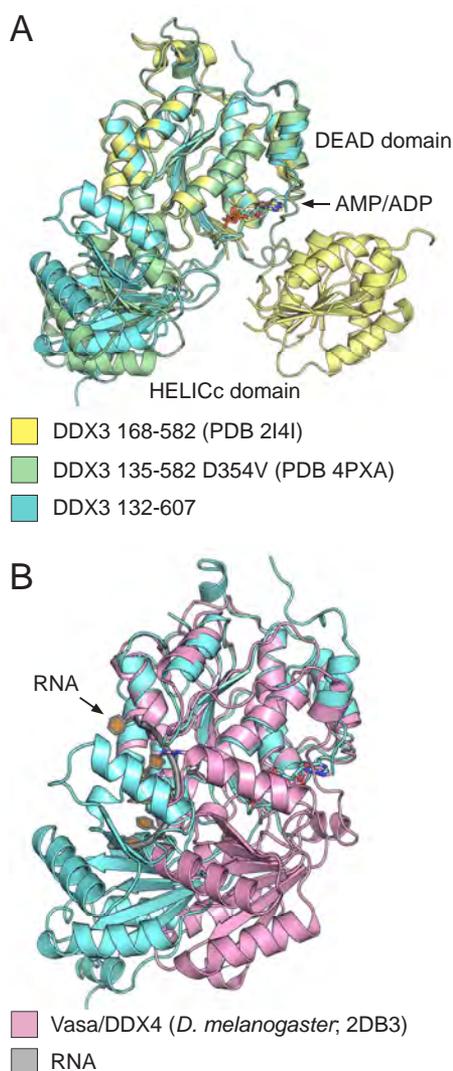


FIGURE 3. The 2.2 Å crystal structure of the conserved core of wild-type DDX3. *A*, the structure of DDX3 132–607 bound to AMP (blue) is shown along with the structure of DDX3 135–582 D354V (green; PDB 4PXA) and 168–582 (yellow; PDB 2I4I). Structures are aligned by the DEAD domain, highlighting the rotation of the C-terminal HELICc domain between the three structures. *B*, the partially closed state of DDX3 (blue) clashes with the RNA-binding site based on a comparison with the DEAD-box protein Vasa bound to RNA (Vasa: pink; RNA: gray).

that the partially closed state of DDX3 (Fig. 3) is stable *in silico* and suggest that the ABL becomes structured in a cooperative manner with RNA binding.

Medulloblastoma Variants of DDX3 Inactivate the Protein—DDX3 is among the most frequently mutated genes in the highly malignant brain tumor medulloblastoma (21–24). Most variants are predicted to inactivate the catalytic activity of DDX3, and some have been shown to decrease the ATPase activity (25). The truncated construct of DDX3 containing the CTE and NTE is highly active and easy to purify, facilitating analysis of disease-associated variants of DDX3. We therefore selected three recurrent variants found in the DEAD-box motifs Ia and VI, R276K, R276Y, and R534H (Fig. 6A) (21–23), and additionally tested alanine substitutions and made these mutations in the DDX3 132–607 construct. The corresponding residue of Arg-276 in Vasa is Arg-328, which binds to the RNA

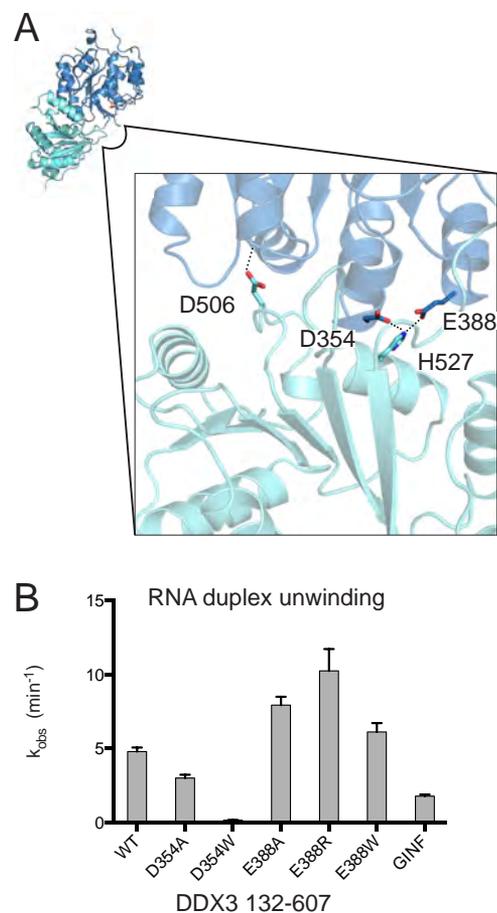


FIGURE 4. Mutations that destabilize the partially closed state accelerate duplex unwinding. *A*, contacts between the DEAD and HELICc domains oriented on the view as in Fig. 3 (upper left). Asp-354 and Glu-388 are separated from His-527 by 4.0 and 2.7 Å, respectively, and Asp-506 is separated from the backbone of Arg-276 by 2.8 Å. *B*, RNA duplex unwinding conducted at 1 μM protein with DDX3 132–607 and mutations indicated. GINF is a mutation of the GINF motif to AAAA (residues 157–160). Error bars indicate standard error of the fit parameter.

backbone, and Arg-534 is Vasa Arg-582, which interacts with the γ -phosphate of ATP. All mutant proteins purified to high yield (>5 mg l⁻¹) and decreased the rate of duplex unwinding, from 1-fold to 1000-fold (Fig. 6B). In concert with the *in vitro* results, mutations R276A and R276K support yeast growth (Fig. 6C), whereas the other three mutations could not complement *DED1* (data not shown). Similarly, mutation of the residue corresponding to Arg-534 in *S. cerevisiae* *PRP28* causes dominant negative lethality *in vivo* (53). In sum, medulloblastoma variants of DDX3 at Arg-276 or Arg-534 are inactivating, further confirming that full-length, inactive DDX3 is selected for by this tumor (21–23, 25), as well as demonstrating the utility of DDX3 132–607 in determining the functional consequences of disease-associated variants of DDX3.

Discussion

DEAD-box proteins consist of two RecA-like domains that comprise the “helicase” core surrounded by variable regions that are unique to individual subfamilies (1, 29). Here, we studied the role of the N- and C-terminal extensions that are essential for function in the Ded1/DDX3 subfamily of DEAD-box proteins (5). We found that removal of the N-terminal 131 and

Redefining the DDX3 Catalytic Core

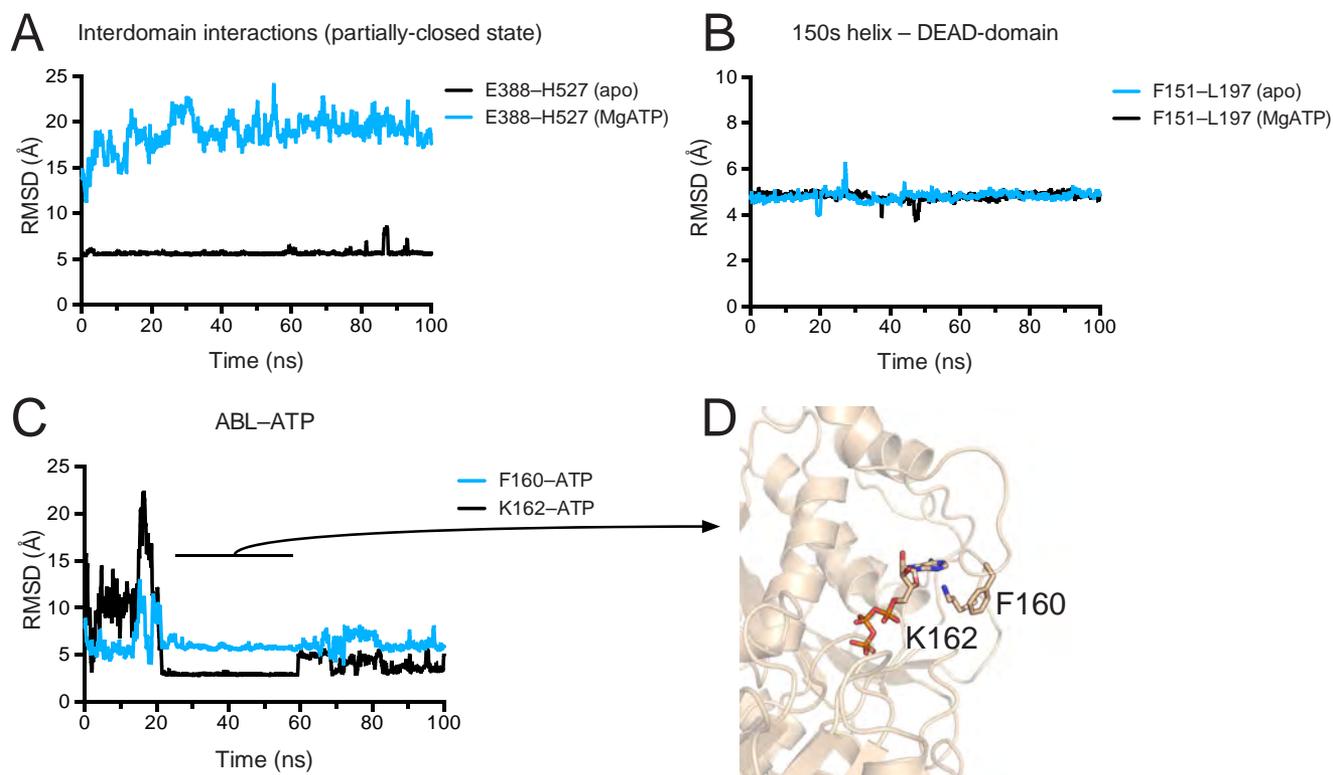


FIGURE 5. Molecular dynamics simulations suggest a stable interdomain interface and transient ABL-ATP interactions. *A*, the partially closed structure is stable over a 100-ns trajectory. Separation between the side chains Glu-388 (C_{β}) and His-527 (C_{γ}) is shown for apo DDX3 132–607 (*black*) and a closed-state model based off the Vasa structure (*cyan*; PDB 2DB3; see “Experimental Procedures”). *RMSD*, root mean square deviation. *B*, the 150's α -helix forms stable interactions with the DEAD domain (colors as in *A*). Distance is measured between Phe-151 (C_{γ}) and Leu-197 (C_{γ}). *C*, the ABL makes transient, stable interactions with the adenine group of ATP. *Black*: distance between the C_{γ} of Phe-160 and the H2 of ATP; *cyan*: distance between the N_{γ} of Lys-162 and the N3 of ATP. In *A–C*, distances are smoothed over a sliding window of 20 frames. *D*, structural model of the ABL interacting with ATP at 50 ns of simulation.

C-terminal 55 residues yields an active construct of DDX3, but further truncation is deleterious (Fig. 2). We then solved the highest resolution crystal structure of an active construct of DDX3 to date (Fig. 3), illustrating a unique autoinhibited state of the protein (Fig. 4) and providing an excellent starting structure for molecular dynamics simulations (Fig. 5). We find that two untested variants found in the highly malignant brain tumor medulloblastoma inactivate the protein *in vitro* and are lethal in yeast (Fig. 6). In sum, our data characterize the essential, conserved core of the Ded1/DDX3 subfamily, which will prove useful when interpreting variants found in human malignancies.

DDX3 132–607 crystallized with a unique interdomain interface not seen in previous crystal structures (Fig. 4*A*). The interface buries $\sim 560 \text{ \AA}^2$ of surface area and overlaps with the RNA-binding surface, suggesting that it is inhibitory to RNA duplex unwinding (Fig. 3*B*). Indeed, introduction of point mutations predicted to destabilize the interdomain interface accelerated the rate of duplex unwinding, despite being distal to the RNA, ATP, or closed-state interdomain interface (Fig. 4). In support of the presence of this state in solution at neutral pH, it was stable during our molecular dynamics simulation (Fig. 5). Therefore, DDX3 contains a cryptic second binding site for the HELICc domain on the DEAD-domain, which is present in solution and is inhibitory for function. *S. cerevisiae* Prp5p was also crystallized in an inhibitory conformation, and destabilization of this “twisted” state accelerated catalysis (54). It is possi-

ble that specific proteins may bind and stabilize inhibitory conformations of DEAD-box proteins to negatively regulate catalysis, as opposed to the many MIF4G domains that bind and activate catalysis (55–59).

Other DEAD-box subfamilies have different subfamily-specific NTE and CTEs. For example, the structure of RNA-bound DDX19 (PDB 3GOH) (60) contains an NTE helix similar to DDX3 and Vasa. However, this region forms part of a β -sheet in the ADP-bound structure of DDX19 (PDB 3EWS) (60). In the crystal structure of *S. cerevisiae* Prp5p, an NTE helix stabilizes the twisted conformation by interacting with the DEAD and HELICc domains (54). Therefore, structural plasticity at the N-terminal boundary of the DEAD domain may be a feature common to many DEAD-box proteins. Perhaps the most extreme example characterized to date of a CTE is in the mitochondrial DEAD-box protein Mss116p, where the CTE forms an entire domain essential for RNA binding and is a fundamental piece of the helicase core (61, 62), unlike most DEAD-box proteins. Thus, to understand the function of an individual DEAD-box subfamily, it is essential to characterize sequences conserved within the subfamily and beyond the DEAD-box helicase core.

Interestingly, the Ded1/DDX3 subfamily of DEAD-box proteins multimerizes both *in vitro* and *in vivo* (32). Oligomerization depends on the C-terminal tail of Ded1, and truncation of the C-terminal 69 residues blocks multimerization and hinders duplex unwinding (32). Our data show that removal of this region strongly

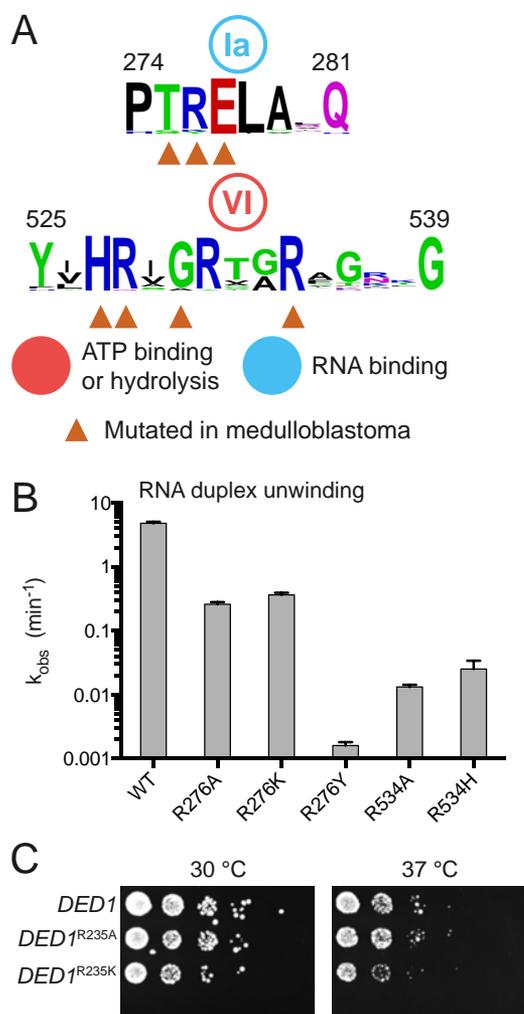


FIGURE 6. Medulloblastoma-associated variants inactivate DDX3. *A*, motifs Ia and VI of human DEAD-box proteins with residues altered in the medulloblastoma indicated. *B*, RNA duplex unwinding conducted with DDX3 132–607 and recurrent medulloblastoma variants indicated. *Error bars* indicate standard error of the fit parameter. *C*, yeast with mutations at *DED1* Arg-235 to alanine or lysine are tolerated, whereas R235Y, R492A, and R492H were nonviable (not shown). *DED1* Arg-235 and Arg-492 correspond to DDX3 Arg-276 and Arg-534, respectively.

suppresses duplex unwinding *in vitro*, but only if the conserved RDYR motif is deleted (Figs. 1A and 2A). Future experiments will test the role of the RDYR motif and C-terminal sequences in oligomerization of the Ded1/DDX3 subfamily.

Human *DDX3X* is altered in numerous malignancies, and different disorders have unique mutation spectrums indicative of their distinct requirements for DDX3 function. For example, nearly all variants found in patients afflicted by the malignant brain tumor medulloblastoma are nonsynonymous single nucleotide variants yielding point mutants predicted to inactivate the protein, yet there are no premature stop codons, frameshifts, or splice variants (21–24). Thus, full-length but inactive protein is selected by this tumor type. In contrast, in blood cancers such as natural-killer/T-cell lymphoma (17), Burkitt lymphoma (63), or chronic lymphocytic leukemia (15, 16), *DDX3X* variants include nonsynonymous single nucleotide variants but also many premature stop codons, frameshifts, and splice variants. The elucidation of the minimal conserved func-

tional core and the new, high-resolution structures of DDX3 presented here is of broad utility for molecular modeling and when predicting the function of truncating variants of DDX3 present in patient samples.

Author Contributions—S. N. F. and J. A. D. conceptualized the study; S. N. F., D. S., and K. J. C. designed the methodology; S. N. F., D. S., and K. J. C. investigated the study; S. N. F. and J. A. D. wrote the original draft; S. N. F., J. A. D., and E. J. performed funding acquisition; and J. A. D. and E. J. supervised the study. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments—We thank Angie Hilliker and the lab of Jasper Rine for help with the yeast experiments and for sharing yeast strains, Kaihong Zhou and Alison Smith for early protein characterization, Yoon-Jae Cho for advice regarding the medulloblastoma variants, and Yun Bai for critical reading of the manuscript. Structural data were collected at Beamline 8.3.1 of the Advanced Light Source at the Lawrence Berkeley National Laboratory, which acknowledges support from the University of California Office of the President, Multi-campus Research Programs and Initiatives Grant MR-15-328599, and the Program for Breakthrough Biomedical Research, which is partially funded by the Sandler Foundation. Molecular dynamics starting models were preprocessed with the UCSF Chimera package, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS National Institutes of Health Grant P41-GM103311).

References

- Linder, P., and Jankowsky, E. (2011) From unwinding to clamping: the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* **12**, 505–516
- Hubstenberger, A., Noble, S. L., Cameron, C., and Evans, T. C. (2013) Translation repressors, an RNA helicase, and developmental cues control RNP phase transitions during early development. *Dev. Cell* **27**, 161–173
- Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C. C., Eckmann, C. R., Myong, S., and Brangwynne, C. P. (2015) The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7189–7194
- Chuang, R. Y., Weaver, P. L., Liu, Z., and Chang, T. H. (1997) Requirement of the DEAD-Box protein Ded1p for messenger RNA translation. *Science* **275**, 1468–1471
- Sharma, D., and Jankowsky, E. (2014) The Ded1/DDX3 subfamily of DEAD-box RNA helicases. *Crit. Rev. Biochem. Mol. Biol.* **49**, 343–360
- Sen, N. D., Zhou, F., Ingolia, N. T., and Hinnebusch, A. G. (2015) Genome-wide analysis of translational efficiency reveals distinct but overlapping functions of yeast DEAD-box RNA helicases Ded1 and eIF4A. *Genome Res.* **25**, 1196–1205
- Lai, M. C., Chang, W. C., Shieh, S. Y., and Tarn, W. Y. (2010) DDX3 regulates cell growth through translational control of cyclin E1. *Mol. Cell Biol.* **30**, 5444–5453
- Lee, C. S., Dias, A. P., Jedrychowski, M., Patel, A. H., Hsu, J. L., and Reed, R. (2008) Human DDX3 functions in translation and interacts with the translation initiation factor eIF3. *Nucleic Acids Res.* **36**, 4708–4718
- de la Cruz, J., Iost, I., Kressler, D., and Linder, P. (1997) The p20 and Ded1 proteins have antagonistic roles in eIF4E-dependent translation in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5201–5206
- Hilliker, A., Gao, Z., Jankowsky, E., and Parker, R. (2011) The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex. *Mol. Cell* **43**, 962–972
- Shih, J. W., Wang, W. T., Tsai, T. Y., Kuo, C. Y., Li, H. K., and Wu Lee, Y. H. (2012) Critical roles of RNA helicase DDX3 and its interactions with eIF4E/PABP1 in stress granule assembly and stress response. *Biochem. J.* **441**, 119–129

12. Lai, M. C., Sun, H. S., Wang, S. W., and Tarn, W. Y. (2015) DDX3 functions in antiviral innate immunity through translational control of PACT. *FEBS J.* 10.1111/febs.13553
13. Soto-Rifo, R., Rubilar, P. S., Limousin, T., de Breyne, S., Décimo, D., and Ohlmann, T. (2012) DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. *EMBO J.* **31**, 3745–3756
14. Landau, D. A., Tausch, E., Taylor-Weiner, A. N., Stewart, C., Reiter, J. G., Bahlo, J., Kluth, S., Bozic, I., Lawrence, M., Böttcher, S., Carter, S. L., Cibulskis, K., Mertens, D., Sougnez, C. L., Rosenberg, M., Hess, J. M., Edlmann, J., Kless, S., Kneba, M., Ritgen, M., Fink, A., Fischer, K., Gabriel, S., Lander, E. S., Nowak, M. A., Döhner, H., Hallek, M., Neuberger, D., Getz, G., Stilgenbauer, S., and Wu, C. J. (2015) Mutations driving CLL and their evolution in progression and relapse. *Nature* **526**, 525–530
15. Ojha, J., Secreto, C. R., Rabe, K. G., Van Dyke, D. L., Kortum, K. M., Slager, S. L., Shanafelt, T. D., Fonseca, R., Kay, N. E., and Braggio, E. (2015) Identification of recurrent truncated DDX3X mutations in chronic lymphocytic leukaemia. *Br. J. Haematol.* **169**, 445–448
16. Wang, L., Lawrence, M. S., Wan, Y., Stojanov, P., Sougnez, C., Stevenson, K., Werner, L., Sivachenko, A., DeLuca, D. S., Zhang, L., Zhang, W., Vartanov, A. R., Fernandes, S. M., Goldstein, N. R., Folco, E. G., Cibulskis, K., Tesar, B., Sievers, Q. L., Shefler, E., Gabriel, S., Hacohen, N., Reed, R., Meyerson, M., Golub, T. R., Lander, E. S., Neuberger, D., Brown, J. R., Getz, G., and Wu, C. J. (2011) *SF3B1* and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365**, 2497–2506
17. Jiang, L., Gu, Z. H., Yan, Z. X., Zhao, X., Xie, Y. Y., Zhang, Z. G., Pan, C. M., Hu, Y., Cai, C. P., Dong, Y., Huang, J. Y., Wang, L., Shen, Y., Meng, G., Zhou, J. F., Hu, J. D., Wang, J. F., Liu, Y. H., Yang, L. H., Zhang, F., Wang, J. M., Wang, Z., Peng, Z. G., Chen, F. Y., Sun, Z. M., Ding, H., Shi, J. M., Hou, J., Yan, J. S., Shi, J. Y., Xu, L., Li, Y., Lu, J., Zheng, Z., Xue, W., Zhao, W. L., Chen, Z., and Chen, S. J. (2015) Exome sequencing identifies somatic mutations of *DDX3X* in natural killer/T-cell lymphoma. *Nat. Genet.* **47**, 1061–1066
18. Seiwert, T. Y., Zuo, Z., Keck, M. K., Khattri, A., Peadarall, C. S., Stricker, T., Brown, C., Pugh, T. J., Stojanov, P., Cho, J., Lawrence, M. S., Getz, G., Brägelmann, J., DeBoer, R., Weichselbaum, R. R., Langerman, A., Portugal, L., Blair, E., Stenson, K., Lingen, M. W., Cohen, E. E., Vokes, E. E., White, K. P., and Hammerman, P. S. (2015) Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. *Clin. Cancer Res.* **21**, 632–641
19. Stransky, N., Egloff, A. M., Tward, A. D., Kostic, A. D., Cibulskis, K., Sivachenko, A., Kryukov, G. V., Lawrence, M. S., Sougnez, C., McKenna, A., Shefler, E., Ramos, A. H., Stojanov, P., Carter, S. L., Voet, D., Cortés, M. L., Auclair, D., Berger, M. F., Saksena, G., Guiducci, C., Onofrio, R. C., Parkin, M., Romkes, M., Weissfeld, J. L., Seethala, R. R., Wang, L., Rangel-Escareño, C., Fernandez-Lopez, J. C., Hidalgo-Miranda, A., Melendez-Zajgla, J., Winckler, W., Ardlie, K., Gabriel, S. B., Meyerson, M., Lander, E. S., Getz, G., Golub, T. R., Garraway, L. A., and Grandis, J. R. (2011) The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157–1160
20. Bol, G. M., Vesuna, F., Xie, M., Zeng, J., Aziz, K., Gandhi, N., Levine, A., Irving, A., Korz, D., Tantravedi, S., Heerma van Voss, M. R., Gabrielson, K., Bordt, E. A., Polster, B. M., Cope, L., van der Groep, P., Kondaskar, A., Rudek, M. A., Hosmane, R. S., van der Wall, E., van Diest, P. J., Tran, P. T., and Raman, V. (2015) Targeting DDX3 with a small molecule inhibitor for lung cancer therapy. *EMBO Mol. Med.* **7**, 648–669
21. Jones, D. T., Jäger, N., Kool, M., Zichner, T., Hutter, B., Sultan, M., Cho, Y. J., Pugh, T. J., Hovestadt, V., Stütz, A. M., Rausch, T., Warnatz, H. J., Ryzhova, M., Bender, S., Sturm, D., Pleier, S., Cin, H., Pfaff, E., Sieber, L., Wittmann, A., Remke, M., Witt, H., Hutter, S., Tzaridis, T., Weischenfeldt, J., Raeder, B., Avci, M., Amstislavskiy, V., Zapatka, M., Weber, U. D., Wang, Q., Lasitschka, B., Bartholomae, C. C., Schmidt, M., von Kalle, C., Ast, V., Lawrenz, C., Eils, J., Kabbe, R., Benes, V., van Sluis, P., Koster, J., Volckmann, R., Shih, D., Betts, M. J., Russell, R. B., Coco, S., Tonini, G. P., Schüller, U., Hans, V., Graf, N., Kim, Y. J., Monoranu, C., Roggendorf, W., Unterberg, A., Herold-Mende, C., Milde, T., Kulozik, A. E., von Deimling, A., Witt, O., Maass, E., Rössler, J., Ebinger, M., Schuhmann, M. U., Frühwald, M. C., Hasselblatt, M., Jabado, N., Rutkowski, S., von Bueren, A. O., Williamson, D., Clifford, S. C., McCabe, M. G., Collins, V. P., Wolf, S., Wiemann, S., Lehrach, H., Brors, B., Scheurlen, W., Felsberg, J., Reifemberger, G., Northcott, P. A., Taylor, M. D., Meyerson, M., Pomeroy, S. L., Yaspo, M. L., Korbel, J. O., Korshunov, A., Eils, R., Pfister, S. M., and Lichter, P. (2012) Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**, 100–105
22. Kool, M., Jones, D. T., Jäger, N., Northcott, P. A., Pugh, T. J., Hovestadt, V., Piro, R. M., Esparza, L. A., Markant, S. L., Remke, M., Milde, T., Bourdeaut, F., Ryzhova, M., Sturm, D., Pfaff, E., Stark, S., Hutter, S., Seker-Cin, H., Johann, P., Bender, S., Schmidt, C., Rausch, T., Shih, D., Reimand, J., Sieber, L., Wittmann, A., Linke, L., Witt, H., Weber, U. D., Zapatka, M., König, R., Beroukchim, R., Bergthold, G., van Sluis, P., Volckmann, R., Koster, J., Versteeg, R., Schmidt, S., Wolf, S., Lawrenz, C., Bartholomae, C. C., von Kalle, C., Unterberg, A., Herold-Mende, C., Hofer, S., Kulozik, A. E., von Deimling, A., Scheurlen, W., Felsberg, J., Reifemberger, G., Hasselblatt, M., Crawford, J. R., Grant, G. A., Jabado, N., Perry, A., Coudrey, C., Croul, S., Zadeh, G., Korbel, J. O., Doz, F., Delattre, O., Bader, G. D., McCabe, M. G., Collins, V. P., Kieran, M. W., Cho, Y. J., Pomeroy, S. L., Witt, O., Brors, B., Taylor, M. D., Schüller, U., Korshunov, A., Eils, R., Wechsler-Reya, R. J., Lichter, P., Pfister, S. M., and ICGC PedBrain Tumor Project (2014) Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothed inhibition. *Cancer Cell* **25**, 393–405
23. Pugh, T. J., Weeraratne, S. D., Archer, T. C., Pomeranz Krummel, D. A., Auclair, D., Bochicchio, J., Carneiro, M. O., Carter, S. L., Cibulskis, K., Erlich, R. L., Greulich, H., Lawrence, M. S., Lennon, N. J., McKenna, A., Meldrim, J., Ramos, A. H., Ross, M. G., Russ, C., Shefler, E., Sivachenko, A., Sogoloff, B., Stojanov, P., Tamayo, P., Mesirov, J. P., Amani, V., Teider, N., Sengupta, S., Francois, J. P., Northcott, P. A., Taylor, M. D., Yu, F., Crabtree, G. R., Kautzman, A. G., Gabriel, S. B., Getz, G., Jäger, N., Jones, D. T., Lichter, P., Pfister, S. M., Roberts, T. M., Meyerson, M., Pomeroy, S. L., and Cho, Y. J. (2012) Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* **488**, 106–110
24. Robinson, G., Parker, M., Kranenburg, T. A., Lu, C., Chen, X., Ding, L., Phoenix, T. N., Hedlund, E., Wei, L., Zhu, X., Chalhoub, N., Baker, S. J., Huether, R., Krawicki, R., Curley, N., Thiruvakatam, R., Wang, J., Wu, G., Rusch, M., Hong, X., Becksfort, J., Gupta, P., Ma, J., Easton, J., Vadodaria, B., Onar-Thomas, A., Lin, T., Li, S., Pounds, S., Paugh, S., Zhao, D., Kawachi, D., Roussel, M. F., Finkelstein, D., Ellison, D. W., Lau, C. C., Bouffet, E., Hassall, T., Gururangan, S., Cohn, R., Fulton, R. S., Fulton, L. L., Dooling, D. J., Ochoa, K., Gajjar, A., Mardis, E. R., Wilson, R. K., Downing, J. R., Zhang, J., and Gilbertson, R. J. (2012) Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488**, 43–48
25. Epling, L. B., Grace, C. R., Lowe, B. R., Partridge, J. F., and Enemark, E. J. (2015) Cancer-associated mutants of RNA helicase DDX3X are defective in RNA-stimulated ATP hydrolysis. *J. Mol. Biol.* **427**, 1779–1796
26. Calo, E., Flynn, R. A., Martin, L., Spitale, R. C., Chang, H. Y., and Wysocka, J. (2015) RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature* **518**, 249–253
27. Ohnishi, S., Pääkkönen, K., Koshiba, S., Tochio, N., Sato, M., Kobayashi, N., Harada, T., Watanabe, S., Muto, Y., Güntert, P., Tanaka, A., Kigawa, T., and Yokoyama, S. (2009) Solution structure of the GUCT domain from human RNA helicase II/Guβ reveals the RRM fold, but implausible RNA interactions. *Proteins* **74**, 133–144
28. Schütz, P., Karlberg, T., van den Berg, S., Collins, R., Lehtiö, L., Högbom, M., Holmberg-Schiavone, L., Tempel, W., Park, H. W., Hammarström, M., Moche, M., Thorsell, A. G., and Schüller, H. (2010) Comparative structural analysis of human DEAD-box RNA helicases. *PLoS ONE* **5**, e12791
29. Banroques, J., Cordin, O., Doère, M., Linder, P., and Tanner, N. K. (2011) Analyses of the functional regions of DEAD-box RNA “helicases” with deletion and chimera constructs tested *in vivo* and *in vitro*. *J. Mol. Biol.* **413**, 451–472
30. Tarn, W. Y., and Chang, T. H. (2009) The current understanding of Ded1p/DDX3 homologs from yeast to human. *RNA Biol.* **6**, 17–20
31. Yedavalli, V. S., Neuveut, C., Chi, Y. H., Kleiman, L., and Jeang, K. T. (2004) Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. *Cell* **119**, 381–392
32. Putnam, A. A., Gao, Z., Liu, F., Jia, H., Yang, Q., and Jankowsky, E. (2015) Division of labor in an oligomer of the DEAD-Box RNA helicase Ded1p.

- Mol. Cell* **59**, 541–552
33. Högbom, M., Collins, R., van den Berg, S., Jenvert, R. M., Karlberg, T., Kotenyova, T., Flores, A., Karlsson Hedestam, G. B., and Schiavone, L. H. (2007) Crystal structure of conserved domains 1 and 2 of the human DEAD-box helicase DDX3X in complex with the mononucleotide AMP. *J. Mol. Biol.* **372**, 150–159
 34. Fairman, M. E., Maroney, P. A., Wang, W., Bowers, H. A., Gollnick, P., Nilsen, T. W., and Jankowsky, E. (2004) Protein displacement by DExH/D “RNA helicases” without duplex unwinding. *Science* **304**, 730–734
 35. Kabsch, W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
 36. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
 37. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
 38. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
 39. Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science* **336**, 1030–1033
 40. DeLano, W. L. (2010) *The PyMOL Molecular Graphics System*, version 1.3r1, Schrödinger, LLC, New York
 41. Jankowsky, E., and Putnam, A. (2010) Duplex unwinding with DEAD-box proteins. *Methods Mol. Biol.* **587**, 245–264
 42. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797
 43. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2: a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191
 44. Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780
 45. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854
 46. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera: a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612
 47. Fiser, A., Do, R. K., and Sali, A. (2000) Modeling of loops in protein structures. *Protein Sci.* **9**, 1753–1773
 48. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926
 49. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28
 50. Buchan, D. W., Minnici, F., Nugent, T. C., Bryson, K., and Jones, D. T. (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **41**, W349–W357
 51. Senissar, M., Le Saux, A., Belgareh-Touzé, N., Adam, C., Banroques, J., and Tanner, N. K. (2014) The DEAD-box helicase Ded1 from yeast is an mRNP cap-associated protein that shuttles between the cytoplasm and nucleus. *Nucleic Acids Res.* **42**, 10005–10022
 52. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., and Yokoyama, S. (2006) Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. *Cell* **125**, 287–300
 53. Jacewicz, A., Schwer, B., Smith, P., and Shuman, S. (2014) Crystal structure, mutational analysis and RNA-dependent ATPase activity of the yeast DEAD-box pre-mRNA splicing factor Prp28. *Nucleic Acids Res.* **42**, 12885–12898
 54. Zhang, Z. M., Yang, F., Zhang, J., Tang, Q., Li, J., Gu, J., Zhou, J., and Xu, Y. Z. (2013) Crystal structure of Prp5p reveals interdomain interactions that impact spliceosome assembly. *Cell Rep* **5**, 1269–1278
 55. Buchwald, G., Schüssler, S., Basquin, C., Le Hir, H., and Conti, E. (2013) Crystal structure of the human eIF4AIII-CWC22 complex shows how a DEAD-box protein is inhibited by a MIF4G domain. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E4611–E4618
 56. Kim, K. M., Cho, H., Choi, K., Kim, J., Kim, B. W., Ko, Y. G., Jang, S. K., and Kim, Y. K. (2009) A new MIF4G domain-containing protein, CTIF, directs nuclear cap-binding protein CBP80/20-dependent translation. *Genes Dev.* **23**, 2033–2045
 57. Montpetit, B., Thomsen, N. D., Helmke, K. J., Seeliger, M. A., Berger, J. M., and Weis, K. (2011) A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. *Nature* **472**, 238–242
 58. Craig, A. W., Haghghat, A., Yu, A. T., and Sonenberg, N. (1998) Interaction of polyadenylate-binding protein with the eIF4G homologue PAIP enhances translation. *Nature* **392**, 520–523
 59. Marcotrigiano, J., Lomakin, I. B., Sonenberg, N., Pestova, T. V., Hellen, C. U., and Burley, S. K. (2001) A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. *Mol. Cell* **7**, 193–203
 60. Collins, R., Karlberg, T., Lehtiö, L., Schütz, P., van den Berg, S., Dahlgren, L. G., Hammarström, M., Weigelt, J., and Schüler, H. (2009) The DEXD/H-box RNA helicase DDX19 is regulated by an α -helical switch. *J. Biol. Chem.* **284**, 10296–10300
 61. Mallam, A. L., Del Campo, M., Gilman, B., Sidote, D. J., and Lambowitz, A. M. (2012) Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature* **490**, 121–125
 62. Del Campo, M., and Lambowitz, A. M. (2009) Structure of the Yeast DEAD box protein Mss116p reveals two wedges that crimp RNA. *Mol. Cell* **35**, 598–609
 63. Schmitz, R., Young, R. M., Ceribelli, M., Jhavar, S., Xiao, W., Zhang, M., Wright, G., Shaffer, A. L., Hodson, D. J., Buras, E., Liu, X., Powell, J., Yang, Y., Xu, W., Zhao, H., Kohlhammer, H., Rosenwald, A., Kluin, P., Müller-Hermelink, H. K., Ott, G., Gascoyne, R. D., Connors, J. M., Rimsza, L. M., Campo, E., Jaffe, E. S., Delabie, J., Smeland, E. B., Olgwang, M. D., Reynolds, S. J., Fisher, R. I., Braziel, R. M., Tubbs, R. R., Cook, J. R., Weisenburger, D. D., Chan, W. C., Pittaluga, S., Wilson, W., Waldmann, T. A., Rowe, M., Mbulaiteye, S. M., Rickinson, A. B., and Staudt, L. M. (2012) Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature* **490**, 116–120

Different Epidermal Growth Factor Receptor (EGFR) Agonists Produce Unique Signatures for the Recruitment of Downstream Signaling Proteins^{*[S]♦}

Received for publication, December 11, 2015, and in revised form, January 19, 2016 Published, JBC Papers in Press, January 19, 2016, DOI 10.1074/jbc.M115.710087

Tom Ronan^{†1}, Jennifer L. Macdonald-Obermann[§], Lorel Huelsmann^{§2}, Nicholas J. Bessman^{†3}, Kristen M. Naegle^{†4}, and Linda J. Pike^{§5}

From the [†]Department of Biomedical Engineering and Center for Biological Systems Engineering and [§]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110 and the [†]Department of Biochemistry and Biophysics, Graduate Group in Biochemistry and Molecular Biophysics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104-6059

The EGF receptor can bind seven different agonist ligands. Although each agonist appears to stimulate the same suite of downstream signaling proteins, different agonists are capable of inducing distinct responses in the same cell. To determine the basis for these differences, we used luciferase fragment complementation imaging to monitor the recruitment of Cbl, CrkL, Gab1, Grb2, PI3K, p52 Shc, p66 Shc, and Shp2 to the EGF receptor when stimulated by the seven EGF receptor ligands. Recruitment of all eight proteins was rapid, dose-dependent, and inhibited by erlotinib and lapatinib, although to differing extents. Comparison of the time course of recruitment of the eight proteins in response to a fixed concentration of each growth factor revealed differences among the growth factors that could contribute to their differing biological effects. Principal component analysis of the resulting data set confirmed that the recruitment of these proteins differed between agonists and also between different doses of the same agonist. Ensemble clustering of the overall response to the different growth factors suggests that these EGF receptor ligands fall into two major groups as follows: (i) EGF, amphiregulin, and EPR; and (ii) betacellulin, TGF α , and epigen. Heparin-binding EGF is distantly related to both clusters. Our data identify differences in network utilization by different EGF receptor agonists and highlight the need to characterize network interactions under conditions other than high dose EGF.

The EGF receptor is an intrinsic membrane protein composed of an extracellular ligand-binding domain connected to an intracellular tyrosine kinase domain by a single transmembrane α -helix. In the absence of ligand, the EGF receptor is thought to exist as a monomer, although inactive “pre-dimers” are known to form (1–5). Upon binding an agonist ligand, the EGF receptor dimerizes leading to the activation of its tyrosine kinase and the phosphorylation of tyrosine residues in the C-terminal tail of the receptor (6–8).

The phosphorylated tyrosines on the EGF receptor serve as binding sites for a large number of signaling proteins that contain SH2 and/or phosphotyrosine-binding domains (9, 10). Some of these proteins, such as Cbl, possess an enzymatic activity (11). Others, such as Grb2 or Shc, serve as adapter proteins that bring other proteins into the EGF receptor-containing complex. For example, Grb2 recruits the scaffolding protein, Gab1, to the EGF receptor (12). Phosphorylation of Gab1 by the EGF receptor allows Gab1 to recruit additional proteins, such as Shp2 or PI3K-R1, to the signaling complex (13–16). The recruitment of these signaling proteins to the receptor ultimately triggers the activation of a variety of downstream signaling pathways, thereby mediating the intracellular effects of growth factor binding.

The EGF receptor binds seven different agonist ligands, including some of high affinity (EGF, TGF α , BTC,⁶ and HB-EGF) and some of low affinity (AREG, EPG, and EPR) (17). It has been reported that different EGF receptor ligands induce different responses when binding to the same cell line (18–21). Given that these agonists bind to the same receptor and stimulate similar downstream signaling molecules, it is difficult to explain how these divergent responses are achieved.

We have previously used a luciferase fragment complementation system to assess the ability of EGF to induce dimerization of the EGF receptor (22–24). In this study, we use our luciferase fragment complementation assay to visualize the recruitment of a variety of signaling proteins to the EGF receptor. The fine temporal resolution and quantitative nature of the split luciferase complementation system allowed us to continuously monitor the association of Cbl, CrkL, Gab1, Grb2, PI3K-R1, p52 Shc,

* This work was supported in part by National Institutes of Health Grant RO1GM099695 (to L. J. P.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

♦ This article was selected as a Paper of the Week.

[S] This article contains supplemental Tables S1 and S2 and Figs. S1–S7.

¹ Supported in part by the Center for Biological Systems Engineering, Washington University, St. Louis, MO.

² Present address: School of Medicine, University of Texas Southwestern, Dallas, TX 75390.

³ Supported by National Institutes of Health Grant RO1-CA079992 (to Mark Lemmon). Present address: Joan and Sanford I. Weill Depts. of Medicine and Microbiology and Immunology, Weill Cornell Medical College, Cornell University, New York, NY 10065.

⁴ To whom correspondence for computational analysis should be addressed. E-mail: knaegle@wustl.edu.

⁵ To whom correspondence should be addressed. E-mail: pike@biochem.wustl.edu.

⁶ The abbreviations used are: BTC, betacellulin; AREG, amphiregulin; EPG, epigen; EPR, epiregulin; HB-EGF, heparin-binding EGF; EGFR, EGF receptor; PCA, principal component analysis; PC, principal component.

p66 Shc, and Shp2 with the EGF receptor in response to increasing concentrations of all seven different EGF receptor ligands. Principal component analysis was applied to this large dataset to determine how the response to these growth factors differed. The data demonstrate that each growth factor produces a unique signature for the recruitment of signaling proteins, and this signature differs at different doses of the same growth factor. This suggests that each growth factor utilizes the signaling network differently, preferentially promoting flux through some pathways over others, which could readily lead to a different net biological outcome.

Experimental Procedures

Materials—EGF was purchased from Biomedical Technologies. TGF α and amphiregulin were from Leinco. Betacellulin was from ProSpec. Heparin-binding EGF was from Sigma. Epi-gen and epiregulin were synthesized and purified in the laboratory of Dr. Mark Lemmon (University of Pennsylvania). Fetal-Plex was from Gemini Bioproducts. The anti-EGF receptor antibody was from Cell Signaling. The PY20 anti-phosphotyrosine antibody was from BD Transduction Laboratories.

DNA Constructs—Full-length cDNA constructs for the signaling proteins were obtained from Addgene (CrkL PI3K-R1 and Shp2), Source Bioscience (Gab1), Thermo Fisher (p52 Shc, p66 Shc, and Grb2), or Sino Biologicals (c-Cbl). The stop codon in each was removed, and an in-frame BsiWI site was inserted through site-directed mutagenesis. The cDNAs were cut with BsiWI and fused to the C-terminal fragment of luciferase (CLuc). The construct was moved into the pcDNA3.1 Zeo expression vector where expression of the fusion protein is driven off the constitutive CMV promoter.

Cell Lines—CHO cells stably expressing the tetracycline-inducible EGF receptor C-terminally fused to the N-terminal fragment of firefly luciferase (EGFR-NLuc) (24) were used as the starting parental line. These cells were transfected with the pcDNA3.1 Zeo plasmids encoding the CLuc fusion of each of the eight signaling proteins. Eight (double) stable cell lines were selected by growth in 5 mg/ml Zeocin. Quantitation of EGF receptor expression in each line by ¹²⁵I-EGF saturation binding indicated that the number of cell surface EGF receptors expressed in each line is within $\pm 20\%$ of the average level of receptor expression (data not shown). Cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% FetalPlex, 100 μ g/ml G418, 100 μ g/ml hygromycin, and 100 μ g/ml Zeocin and maintained in an incubator at 37 °C in 5% CO₂.

Luciferase Assays—Double stable CHO cells were plated into 96-well black-walled dishes 2 days prior to use in medium containing 1.5 μ g/ml doxycycline to induce expression of the EGFR-NLuc fusion protein. For assay, cells were transferred into Dulbecco's phosphate-buffered saline supplemented with 5 mg/ml BSA and 20 mM MOPS, pH 7.2. Cells were incubated with 0.9 mg/ml D-luciferin for 30 min at 37 °C prior to the addition of growth factor and the start of imaging. Cell radiance (photons/s/cm²/steradian) was measured every 30 s for 25 min using a cooled charge-coupled device camera in the IVIS50 or IVIS Lumina imaging system. Assays were performed in hexuplicate. The lines through the data were drawn using Equation

1, which represents the sum of a logistics association equation and an exponential dissociation equation.

$$Y = (Y_0 / (1 + h \cdot \exp(-k_1 \cdot t)) + (\text{plateau} - \text{bottom}) \cdot \exp(-k_2 \cdot t) + \text{bottom}) \quad (\text{Eq. 1})$$

where Y = photons/s at time (t). k_1 represents the association rate constant, and k_2 is the dissociation rate constant. This curve drawing was not part of the principal component analysis and was used only for visual presentation of the dose-response curves.

Western Blotting—CHO cells expressing the wild type EGF receptor were treated without or with 5 μ M erlotinib or 10 μ M lapatinib for 1 h and then stimulated with the indicated concentrations of EGF for 5 min. Lysates were prepared, and Western blotting with anti-EGF receptor and anti-phosphotyrosine antibodies were performed as described previously (23).

PCA and Enrichment Analysis—Computational analysis was performed using the Python programming language. PCA utilized the scikit-learn package (25). PCA was performed on a 280 \times 44 matrix, with 280 unique combinations of protein, growth factor, and dose, each with 44 time points, normalized to the maximal response elicited for that agonist/protein pair. For PCA, a subset of five (out of seven) doses was chosen for each growth factor to bracket the EC₅₀ value for the recruited signaling proteins as follows: for BTC and EGF, the doses ranged from 0.03 to 3 nM; for TGF, the doses ranged from 0.1 to 10 nM; for HB-EGF, the doses ranged from 0.3 to 30 nM; and for AREG, EPG, and EPR, the doses ranged from 3 to 300 nM. References to "low" doses of growth factor (as used in Figs. 4 and 7) represent the second dose in the five-dose series, and references to "high" doses (as used for Fig. 7) represent the fourth dose in the five-dose series. Distances between protein pairs were calculated using Euclidean distance between the five-dimensional vector across doses in PC space. Top- and bottom-quartile enrichment was calculated using the hypergeometric test and Bonferroni-corrected.

For clustering of the growth factors based on protein recruitment across all doses, pairwise protein distances for each ligand were converted to a one-dimensional vector. The vectors for each ligand were then clustered using hierarchical clustering. An ensemble of 35 clustering results was generated by varying linkages (single, complete, average, and weighted) and distance metrics (Euclidean, Pearson correlation, city block, cosine, Bray-Curtis, Canberra, Chebyshev, and square Euclidean). The Euclidean metric was also used with median, centroid, and Ward linkage. The results for each ligand were assembled into a matrix and hierarchically clustered using single linkage and Euclidean distance. p66 Shc was not included in this analysis so as not to over-weight the results toward the contribution of Shc isoforms.

Results

Generation and Characterization of Stable Cell Lines—The split luciferase complementation assay utilizes an N-terminal (NLuc) and C-terminal (CLuc) fragment of firefly luciferase (26). Individually, the fragments have no enzymatic activity. However, when they are brought into proximity, they comple-

Signaling Protein Recruitment by EGFR Agonists

ment each other forming a catalytically active luciferase that produces light upon oxidation of luciferin. For our luciferase complementation assays, each of eight signaling proteins (Cbl, CrkL, Gab1, Grb2, PI3K-R1, p52 Shc, p66 Shc, and Shp2) was C-terminally fused to the CLuc fragment via a 16-amino acid flexible linker. The cDNA for the fusion protein was then transfected into a CHO cell line that stably expressed the EGF receptor C-terminally fused to the NLuc fragment (EGFR-NLuc) off a tetracycline-inducible promoter. Double stable cell lines were selected for use in these experiments. For assay, the CHO cells were cultured for 24 h in 1.5 $\mu\text{g/ml}$ doxycycline to induce expression of the EGFR-NLuc fusion protein. The signaling proteins were constitutively expressed from a CMV promoter.

Luciferase Complementation between the EGF Receptor and Signaling Proteins—All eight signaling proteins yielded an EGF-stimulated increase in luciferase activity when co-expressed in cells with EGFR-NLuc (Fig. 1). EGF-stimulated complementation between the EGF receptor and these signaling proteins was seen as early as 30 s after the addition of EGF. At low concentrations of EGF, essentially all of the pairings exhibited a rapid rise in luciferase activity, which peaked by \sim 5–8 min. For some pairings, such as the EGF receptor and PI3K-R1 (Fig. 1E), this level of complementation was maintained over the entire time course at all doses. In other pairings, such as Cbl (Fig. 1A) or CrkL (Fig. 1B), complementation plateaued at low concentrations of growth factor but declined after an early peak at high concentrations of EGF. Still other proteins demonstrated a bimodal response across doses. For example, for Grb2 (Fig. 1D) and Shp2 (Fig. 1H), the maximum complementation occurred at a relatively low dose of EGF, with higher doses of growth factor resulting in lower peak responses and a marked decrease at longer times.

In the EGF receptor/Shp2 pairing (Fig. 1H), the luciferase activity observed at the highest concentrations of EGF actually fell below the basal level after about 15 min. These data imply that Shp2 associates with the EGF receptor under non-stimulated conditions. This association is apparently disrupted upon stimulation with high doses of growth factor.

If these signaling proteins are being recruited to the EGF receptor via phosphotyrosine-dependent interactions, then the associations visualized through luciferase complementation should be sensitive to inhibition of the EGF receptor tyrosine kinase. As shown in Fig. 2, A–H, treatment of cells with 5 μM erlotinib (*green lines*) effectively inhibited EGF-stimulated complementation between the EGF receptor and each of these eight signaling proteins. Inhibition was essentially complete for all pairings with the exception of p52 Shc and p66 Shc, for which the inhibition was \sim 70%. The complementation between the EGF receptor and Shp2 actually showed an EGF-stimulated decline in luciferase activity, again consistent with there being a basal level of association between the two proteins, which is disrupted after ligand binding.

Despite the fact that lapatinib appeared to inhibit EGF receptor autophosphorylation to the same extent as erlotinib (Fig. 2I), pretreatment of the cells with 10 μM lapatinib (*red lines*) was far less effective than pretreatment with erlotinib at blocking the association of these signaling proteins with the EGF receptor. Although lapatinib was able to completely block com-

plementation between the EGF receptor and Cbl, CrkL, and Shp2, the other five signaling proteins all showed at least 20% residual EGF-stimulated luciferase activity in the presence of lapatinib. The association of Gab1 was particularly insensitive to lapatinib treatment. Thus, there is a significant difference between erlotinib and lapatinib in terms of their efficacy for inhibiting EGF-stimulated signaling complex assembly.

Recruitment Stimulated by Other EGF Receptor Ligands—The EGF receptor is activated by a family of homologous growth factors, including EGF, TGF α , BTC, HB-EGF, AREG, EPG, and EPR (17). To quantify the similarities and differences in the response of cells to stimulation by each of these ligands, the luciferase complementation assay was used to monitor the recruitment of the eight different signaling proteins to the EGF receptor in response to each of these agonist ligands.

Supplemental Figs. 1–6 show the time courses of the recruitment of these eight signaling proteins to the EGF receptor in response to increasing doses of each of these additional six growth factors. Like EGF, all of these growth factors stimulated the recruitment of all eight signaling proteins in a dose-dependent manner. However, the patterns of the dose response curves for all seven growth factors for each individual pairing were similar. For example, for all growth factors, PI3K-R1 recruitment plateaued early, and the level of signal was maintained over the entire time course. Similarly, the bimodal response for the recruitment of Grb2 and Shp2 was observed for all growth factors.

Table 1 reports the estimated EC₅₀ values for each ligand stimulating the recruitment of each protein. As expected from their low binding affinities, AREG, EPG, and EPR required \sim 30–100-fold greater concentrations of ligand to stimulate a maximal response than did EGF, TGF α , BTC, or HB-EGF. Surprisingly, the EC₅₀ values for a given growth factor for stimulating the recruitment of the different signaling proteins differed up to 18-fold.

Fig. 3 compares the extent of recruitment of the eight signaling proteins in response to an optimal concentration of each of the seven growth factors. The concentrations compared were those that gave the maximal peak response for that particular pairing (Fig. 1 and supplemental Figs. 1–6). For most of the pairs, all seven ligands stimulated a similar maximal response. However, HB-EGF routinely elicited a slightly lower response than the other growth factors. The greatest difference in response was observed for the recruitment of Grb2 for which the response to EPG and EPR was \sim 30% higher than that to EGF, while the response to HB-EGF was \sim 30% lower than that to EGF. Consequently, there was nearly a 2-fold difference in the relative extent of Grb2 recruitment between the high of EPG/EPR and the low of HB-EGF.

Fig. 4 compares the ability of a fixed (comparable) low dose of each growth factor to stimulate the recruitment of all eight signaling proteins. The responses have been normalized to the maximal response observed for that EGFR/protein pair at the optimal dose of that growth factor. For all growth factors, Grb2 appears to have the fastest relative response time. Cbl and CrkL most frequently have the slowest relative response time. The recruitment of PI3K-R1 shows the most variability being similar to Cbl and CrkL for the low affinity ligands but closer to

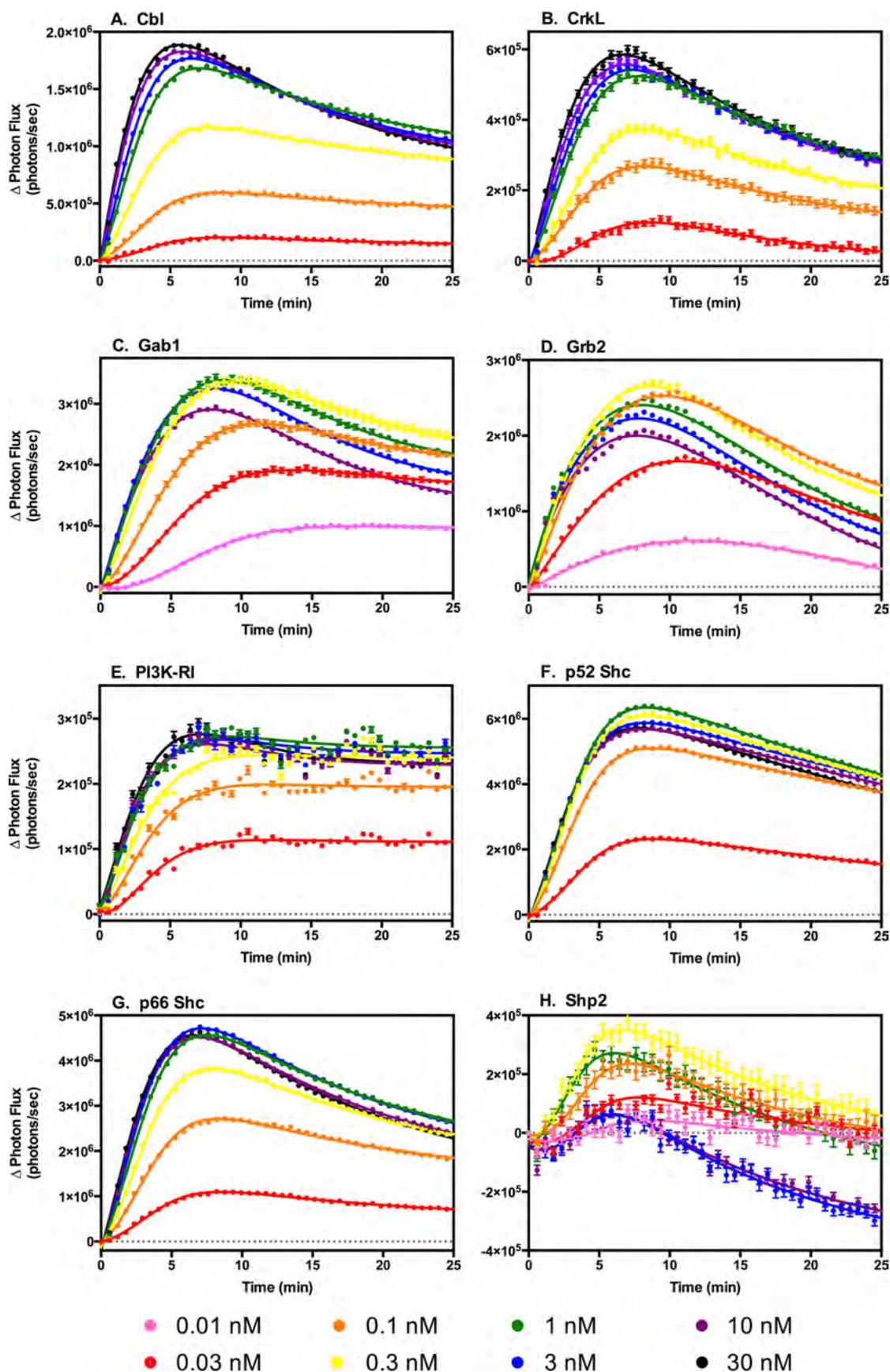


FIGURE 1. EGF-stimulated association of eight signaling proteins with the EGF receptor measured using luciferase fragment complementation imaging. CHO cells stably co-expressing EGFR-NLuc and the CLuc-fused version of one of eight signaling proteins were assayed for EGF-stimulated light production in the presence of luciferin. Cells were stimulated with the indicated concentration of EGF at time $t = 0$, and light production was monitored for 25 min.

Signaling Protein Recruitment by EGFR Agonists

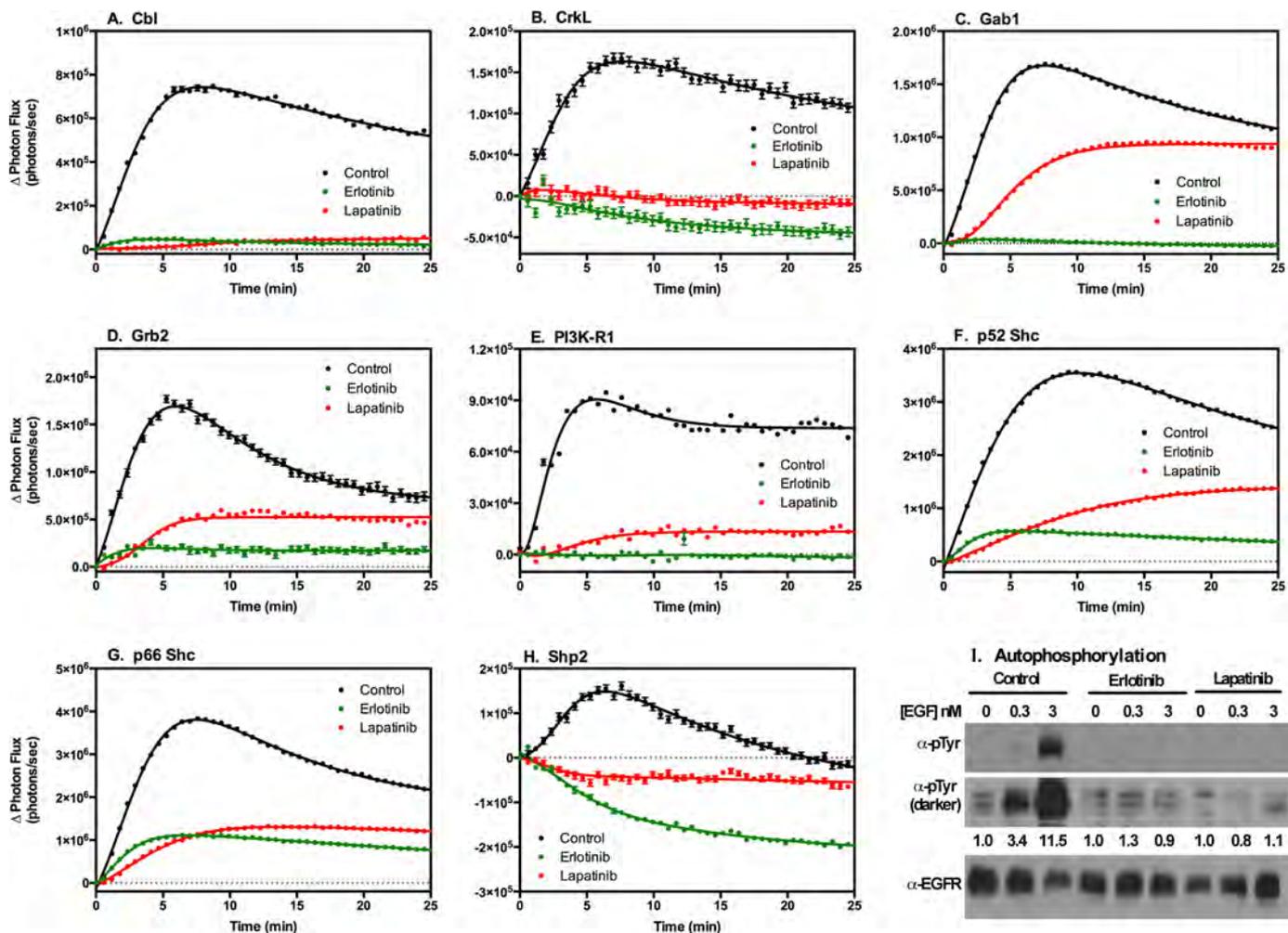


FIGURE 2. Effect of erlotinib and lapatinib on EGF-stimulated association of eight signaling proteins with the EGF receptor. A–H, CHO cells stably co-expressing EGFR-NLuc and the CLuc-fused version of one of eight signaling proteins were treated with 5 μM erlotinib (green lines) or 10 μM lapatinib (red lines) for 60 min prior to stimulation without or with 0.3 nM EGF. EGF-stimulated light production was monitored for 25 min after addition of EGF. I, CHO cells expressing wild type EGF receptor were treated with 5 μM erlotinib or 10 μM lapatinib for 60 min prior to stimulation with 0.3 or 3.0 nM EGF. Lysates were prepared and equal amounts of protein analyzed by SDS-gel electrophoresis and Western blotting with an anti-phosphotyrosine antibody or an anti-EGF receptor antibody. Quantitation of anti-phosphotyrosine blot is shown.

TABLE 1

EC₅₀ values for agonist-stimulated EGF receptor/protein association

Table compares the EC₅₀ values for each growth factor for stimulating the recruitment of the eight signaling proteins. These values were estimated based on the response to each dose of growth factor at *t* = 2.5 min. This largely eliminates the effects of the declines in signal at longer times and means that these values reflect mainly the initial association of the two proteins. The EC₅₀ values differed for the recruitment of different proteins by the same growth factor. So, for example, EGF exhibited an EC₅₀ of ~0.03 nM for recruiting Grb2 and PI3K-R1 but an EC₅₀ about 10-fold higher for recruiting Cbl. EPG exhibited the widest range of EC₅₀ values (~18-fold difference), whereas HB-EGF showed the smallest range of EC₅₀ values (~3-fold).

EC ₅₀	EGF	TGF	BTC	HB-EGF	AREG	EPG	EPR
				<i>nM</i>			
Cbl	0.31	0.73	0.40	0.85	36.0	33.0	21.0
CrkL	0.14	0.40	0.25	1.5	26.0	19.0	21.0
Gab1	0.08	0.11	0.06	0.47	5.9	4.8	2.7
Grb2	0.03	0.06	0.04	0.64	4.0	1.8	2.7
PI3K-R1	0.03	0.24	0.11	1.2	21.0	22.0	18.0
p52 Shc	0.06	0.12	0.15	0.85	3.7	3.7	2.6
p66 Shc	0.10	0.21	0.05	1.7	13.0	6.1	4.0
Shp2	0.09	0.11	0.09	0.50	13.0	3.3	4.0

Grb2 and Gab1 for the high affinity ligands. Interestingly, the recruitment of p52 Shc and p66 Shc differs noticeably from each other. In many cases, p52 Shc shows a shorter relative response time than p66 Shc, often significantly shorter, as for AREG and HB-EGF. However, this order is reversed for BTC where p66 Shc is recruited more rapidly than p52 Shc. Thus, at

the earliest times of signal transduction, differences in response to the different the growth factors can be identified and would contribute to a different biological outcome.

Global Behaviors Observed via Reduced Dimensionality—The foregoing data represent an extremely rich set of measurements of the recruitment of eight different signaling proteins by

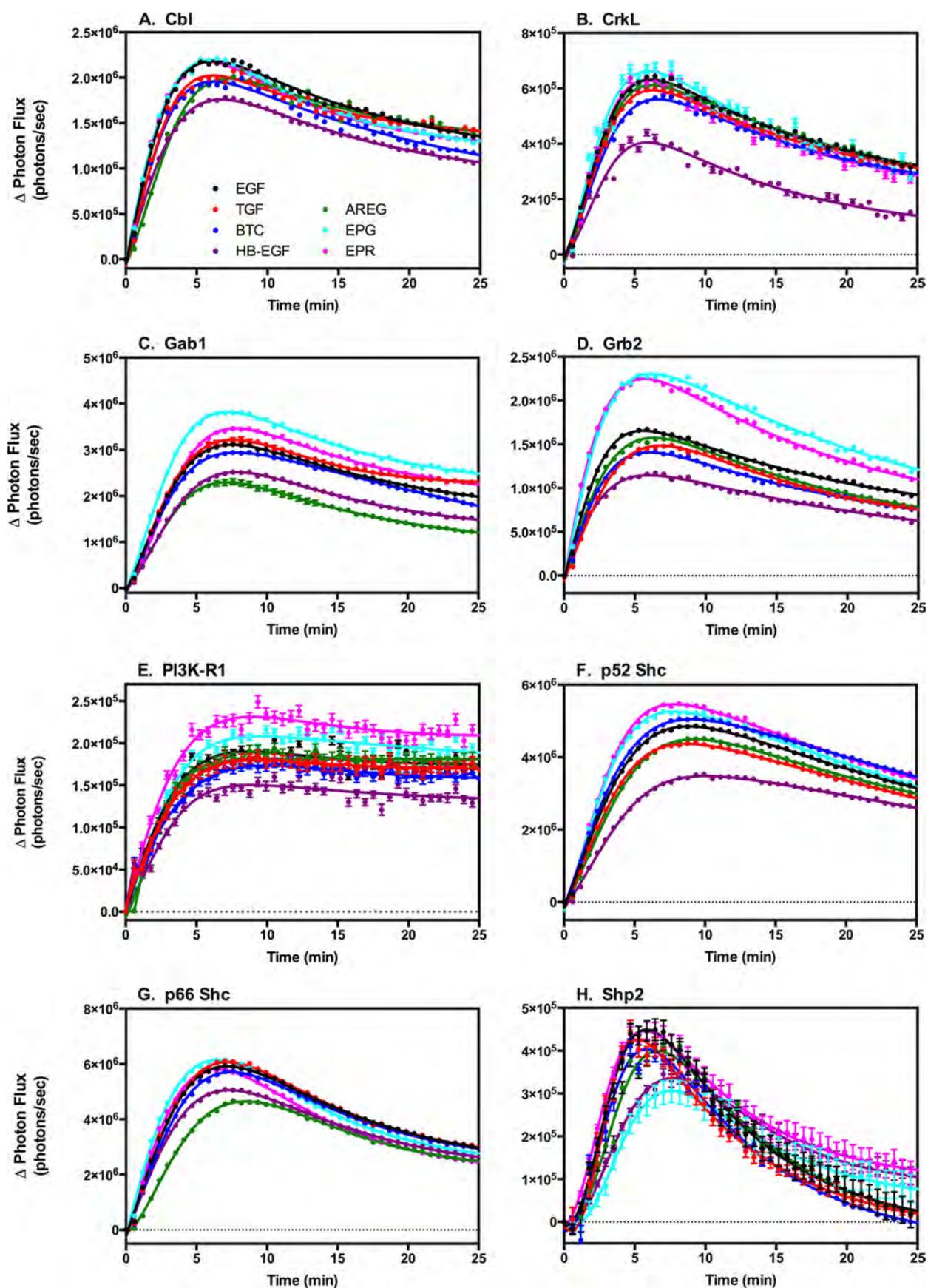


FIGURE 3. Comparison of the association of eight signaling proteins with the EGF receptor stimulated by optimal concentrations of the seven EGF receptor agonists. CHO cells stably co-expressing EGFR-NLuc and the CLuc-fused version of one of eight signaling proteins were stimulated at $t = 0$ with the concentration of each growth factor that yielded maximal peak complementation for a given receptor/protein pair and light production monitored for 25 min.

Signaling Protein Recruitment by EGFR Agonists

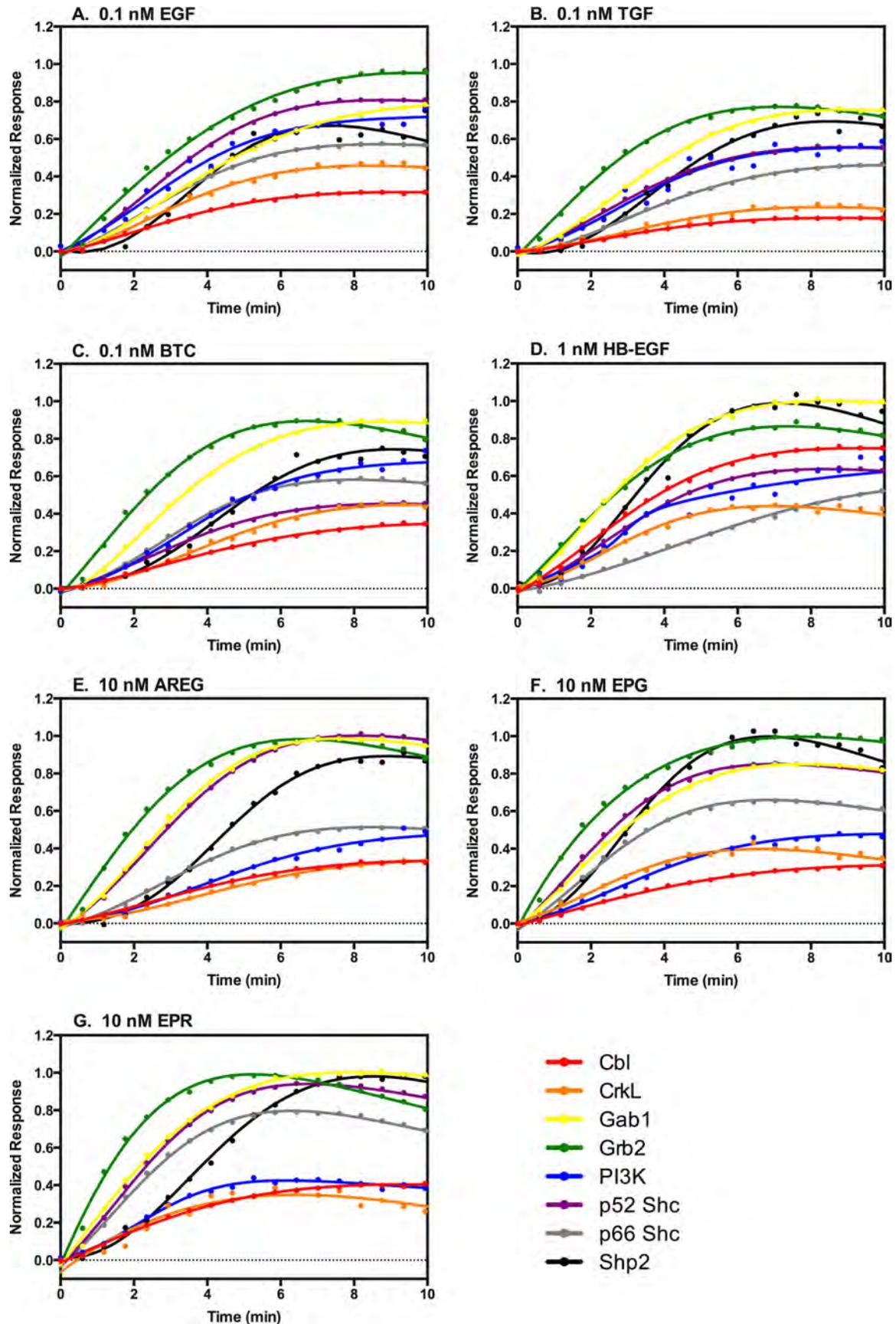


FIGURE 4. **Relative response times for the recruitment of the eight signaling proteins by comparable low doses of each of the EGFR ligands.** The response to the indicated low concentration of each of the seven agonists was normalized to the maximal response elicited by that agonist for that EGFR/protein pair. The normalized responses for all signaling proteins stimulated by a single agonist were then plotted on the same graph.

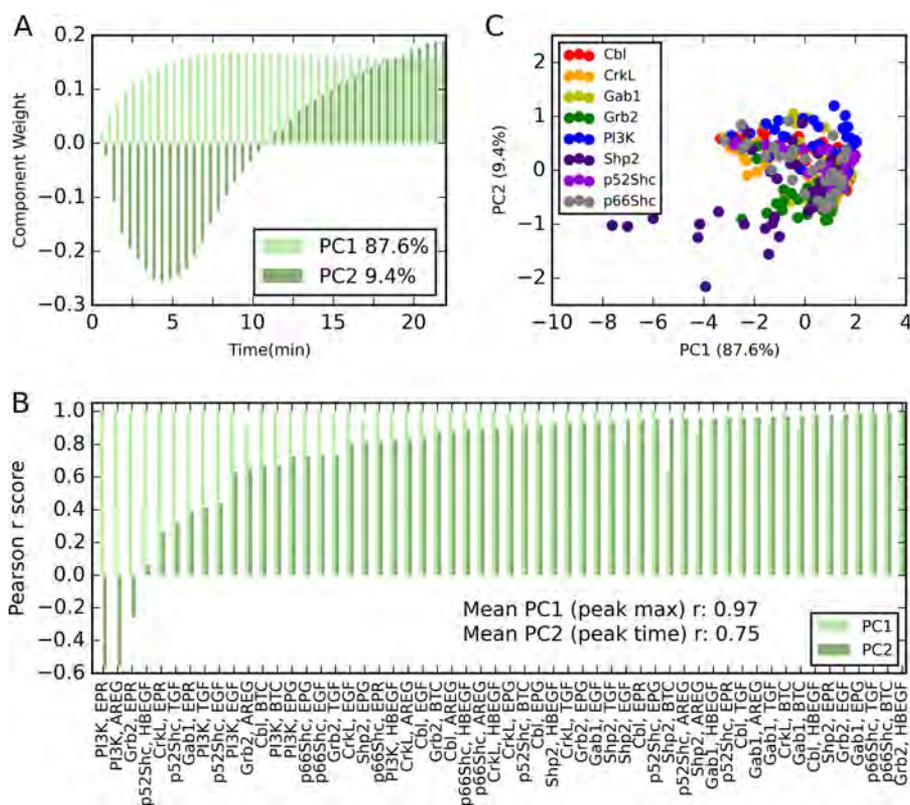


FIGURE 5. **Dimensionality can be reduced using principal component analysis.** *A*, plot of the loadings of the first two principal components, which accounts for 97.0% of the covariance in the dataset (PC1 accounts for 87.6%; PC2 accounts for 9.4%). *B*, correlation of PC1 with maximum peak magnitude and PC2 with peak time was calculated across five doses for each protein-growth factor pair using Pearson correlation. Mean correlation for PC1 with maximum peak magnitude was 0.97, whereas mean correlation for PC2 with peak time was 0.75. *C*, dataset is plotted based on projections onto the first two principal components, capturing 97.0% of the variance. Individual points are colored according to the signaling protein measured.

the EGF receptor. Within this dataset, discovering relationships among the proteins and growth factors is difficult due to the high dimensionality of the problem. To reduce the dimensionality of the dataset, while keeping the relationships within the data intact, we used PCA.

For this analysis, a fixed subset of five (out of the seven) doses of each growth factor was used (Supplemental Table 1). The subset of doses was chosen so that we captured a comparable range of response above and below the EC_{50} values for each of the different growth factors. As a result, the doses that were not included in the analysis were either the very lowest concentrations that elicited a weak or no response or the very highest concentrations that were super-saturating. This approach allows us to compare the behavior of the same dose of a single growth factor across all eight signaling proteins and to compare the behavior of a single signaling protein at comparable concentrations across the different growth factors.

We could account for 97.0% of the co-variation within the entire dataset by projecting the original data into the first two dimensions of the principal component space. Principal component 1 (PC1) captures 87.6% of the variance, whereas PC2 captures 9.4% of the variance. As a result, each time series for one growth factor dose and signaling protein response can be plotted as a single point in two-dimensional PC space while retaining almost all of the variation that exists in the original 44 dimensions (*i.e.* the 44 time points per curve). The loading plots (Fig. 5A) indicate that PC1 represents a positive integration of information across

most time points. By contrast, PC2 negatively weights the earliest time points while positively weighting the latter half of the time course.

Based on our observations of the data, we thought the latent dimensions of the principal component analysis might describe physical features of the data, specifically information about the relative maximum signal achieved and the rate at which this signal was achieved. To test this hypothesis, for each protein-ligand pair we determined the correlation between the dose-response vector in PC1 and the magnitude of the peak for each dose in the original normalized data. We also determined the correlation between the PC2 dose-response vector and the time of peak signal for each dose. As shown in Fig. 5B, there is an extremely high correlation between the PC1 value and the relative magnitude of the peak response (mean $r = 0.97$). Similarly, with the exception of a few outliers in PC2 space, there is a high correlation between the value in PC2 space and the time to peak response in the original data (mean $r = 0.76$). The high correlation indicates that we can ascribe physical meaning to our principal component axes. Specifically, higher values along PC1 indicate that the signal achieves a higher relative maximum value. Higher values in PC2 space indicate that the signal achieves its maximum value at a later time. Lower values in PC2 space indicate that the signal achieves its maximum value at an earlier time.

Fig. 5C shows the entire dataset reduced to the first two dimensions of PCA space. Each point represents a time course for a particular signaling protein at a single dose of a single

Signaling Protein Recruitment by EGFR Agonists

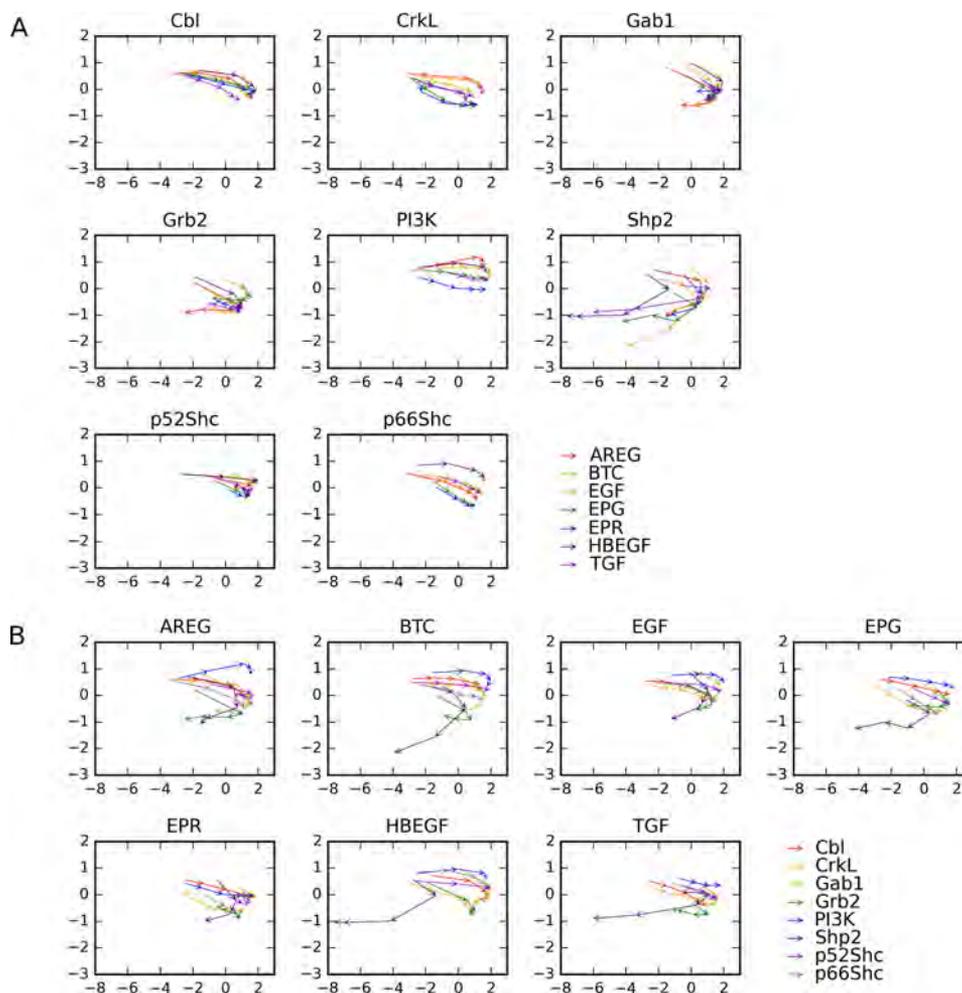


FIGURE 6. Global trends based on dose response. The individual points in two-dimensional PC space representing a protein-ligand pair at a particular dose were organized into a dose-response series for the five chosen doses, by connecting the response at lower doses to the next higher dose using a *directed arrow*. *A*, resulting vectors are grouped by signaling protein and colored according to the growth factor. *B*, resulting vectors are grouped by growth factor and colored according to signaling protein.

growth factor. Points are colored to indicate the signaling protein being recruited to the EGF receptor. Points close together in PCA space represent responses that are similar to each other across the entire time course. The responses of Shp2 and PI3K-R1 are the most separated in both PCs indicating that they are the most different. The remaining points are densely packed in the intermediate region between the extremes of the PI3K-R1 and Shp2 signals.

To identify trends in the data, the measurements were organized into a dose series for each ligand/protein pair. This was visualized by connecting the PCA point representing the curve at the lowest dose of one growth factor to the point representing the curve at the next higher dose of that same growth factor with a directed arrow, continuing on for the five doses of each growth factor (see [supplemental Fig. 7](#) for an example). This approach reveals significant trends in the evolution of signals across the dose range, despite the density initially observed in PC space (Fig. 6). Overall, the major mode of behavior for a given signaling protein is dominated by the identity of the protein rather than the identity of the growth factor. Therefore, the curves describing the recruitment of the same signaling protein stimulated by any of the seven ligands (Fig. 6A) are more similar

to each other than they are to the curves that describe the recruitment of a different signaling protein stimulated by the same growth factor (Fig. 6B). As is apparent from Fig. 6B, there are differences in how the individual protein responses evolve based on the stimulatory ligand.

Aside from these general observations, each PC shows two contrasting trends in a subset of proteins. First, for Cbl, CrkL, PI3K-R1, p52 Shc, and p66 Shc, there is a monotonic increase in PC1 (relative maximal signal) as the dose of growth factor increases. This is what is expected in a traditional dose-response curve, *i.e.* the signal increases with increasing dose. By contrast, Gab1, Grb2, and Shp2 show a bimodal response in PC1 space, reflecting an initial increase in response followed by a marked decrease in peak signal at the highest doses of most of the growth factors.

A second trend is that for most of the signals there is a monotonic progression down the PC2 axis. This indicates that the peak response is achieved more rapidly at higher concentrations of growth factor. An exception to this rule is the recruitment of p52 Shc in response to EGF, BTC, and HB-EGF, where there is little change in the time to peak response over the entire dose range tested.

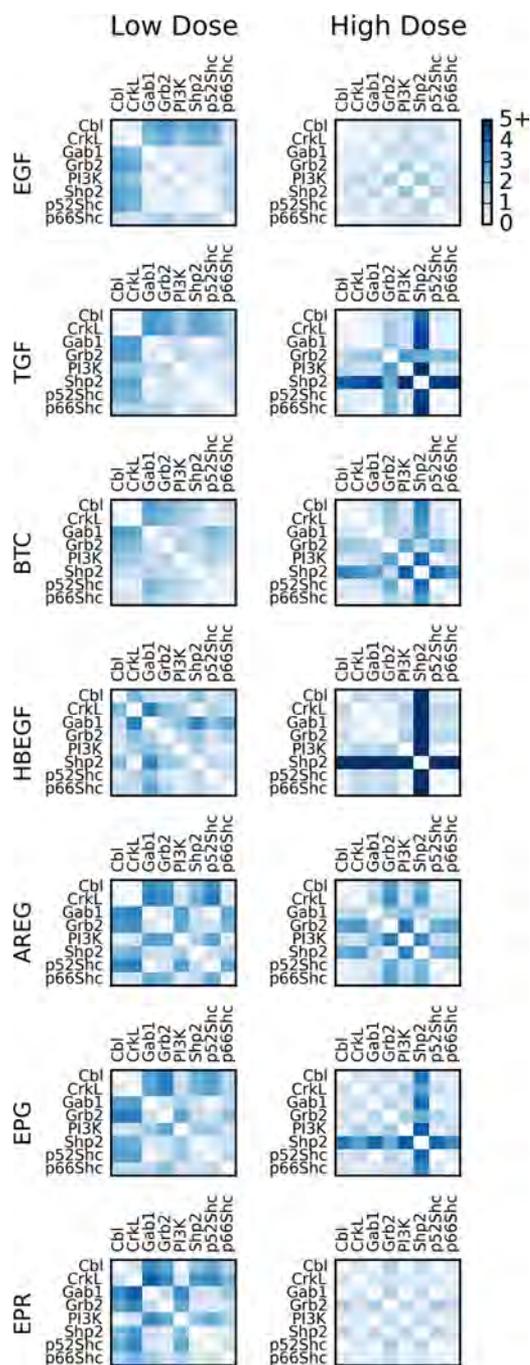


FIGURE 7. Signaling protein response varies by growth factor dose. To quantify differences in protein response, Euclidean distances were calculated between proteins for each growth factor at both a low and high dose and visualized as a heat map. The response pattern for each growth factor at low dose is in the left column, and the response pattern for a high dose is in the right column.

Pairwise Interactions—The proteins chosen for this study were selected because they are involved in well documented interactions with the EGF receptor and with each other. Therefore, we would expect that the behaviors of some of these proteins should correlate in PC space. To quantify these relationships, we calculated the distances between interacting protein pairs in PC space for each dose of a single growth factor. Fig. 7 shows heat maps of protein-pair distances for a low dose and a high dose for each growth factor. Two important features are

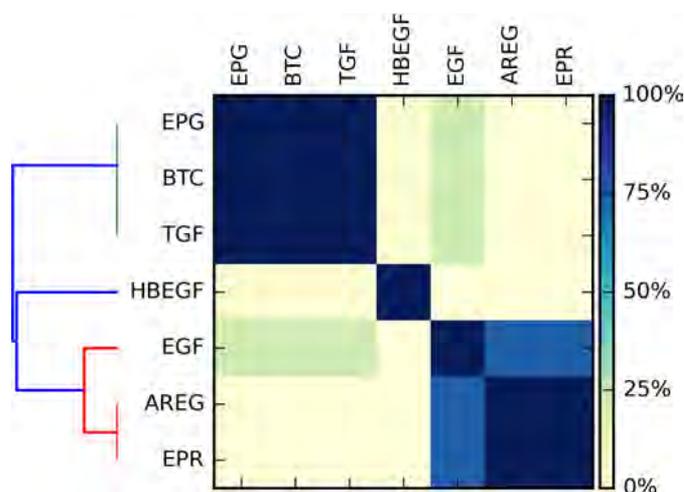


FIGURE 8. Heat map and dendrogram showing the results of clustering of the responses to the seven growth factors. The pairwise protein distances for each ligand were converted to a vector, and the vectors for each ligand were hierarchically clustered via the ensemble approach described under “Experimental Procedures.” The results are visualized via a heat map displaying the fraction of time each ligand pair clustered together across the ensemble.

immediately apparent from these heat maps. First, the patterns seen for the low and high doses of the same growth factor are distinctly different. This suggests that the same growth factor utilizes the network differently when applied at different concentrations. Second, the heat maps for each growth factor are very different, suggesting that the different growth factors activate the network in a manner that is specific to that growth factor.

To evaluate the similarity of protein recruitment dynamics across the different growth factors, distances were calculated between protein pairs in PC space across the five doses of each individual growth factor. The complete set of these cumulative distances was then rank-ordered, and both the top and bottom quartiles were probed for statistical enrichment for individual proteins or specific protein pairs (supplemental Table 2).

Several specific protein pairs were strongly represented in the top quartile. The pairwise distances of Cbl with CrkL and p52 Shc with p66 Shc were the most significantly enriched protein pairs in the top quartile ($p < 0.005$, Bonferroni-corrected), whereas the interaction of Gab1 with p52 Shc was also significantly enriched ($p < 0.05$, Bonferroni-corrected).

Enrichment in the bottom quartile was also calculated to identify proteins and protein pairs that rarely exhibited similar dynamics. Shp2-based interactions as a group were identified as being enriched in this quartile ($p < 0.0005$, Bonferroni-corrected).

To compare the global response of the network to each of the seven different growth factors, we performed ensemble clustering on the pairwise distances between proteins, as described under “Experimental Procedures.” The percentage of time each growth factor clustered with another growth factor in the ensemble of clustering solutions is visualized in Fig. 8 as a heat map. The growth factors were then hierarchically clustered. As can be seen from this figure, BTC, EPG, and TGF α form a strong cluster (the BTC cluster) because they cluster together in every clustering solution in the ensemble. AREG and EPR form

Signaling Protein Recruitment by EGFR Agonists

a second strong cluster (the AREG cluster). EGF clusters most frequently with the AREG cluster (77%) but shares some membership in the BTC cluster (23%). HB-EGF is rather unique and is far from both the BTC and AREG clusters.

Discussion

We report here on the use of luciferase fragment complementation to study the association of downstream signaling proteins with the EGF receptor. The advantages of this system include the ease of assay and the fact that it can be done in live cells with continuous monitoring. In addition, the signals generated from the eight signaling proteins examined here were robust, allowing detection of differences associated with changes in the concentration of growth factor. Finally, the approach is scalable and useful for screening applications.

Using this system, we found that all eight of the selected signaling proteins are rapidly recruited to an EGF receptor-containing complex, with association being apparent by 30 s. The peak extent of association occurred between 5 and 7 min, depending on the pairing. This is consistent with the time course of assembly of Shc-containing complexes after stimulation of cells with EGF, as documented through quantitative mass spectrometry (27).

In most of the pairings, the luciferase signal decreased slowly over time particularly at the higher doses of growth factor. As internalization of the EGF receptor begins almost immediately after addition of growth factor (28), it seems likely that at least part of the decrease in signal at longer times is due to internalization and degradation of the receptor and its associated signaling proteins. Nevertheless, at least some fraction of the agonist-induced increase in luciferase activity is maintained for as long as 25 min after the addition of EGF. These data imply that these signaling proteins remain associated with the EGF receptor even after it has been internalized. Thus, some aspects of signaling probably continue to occur well after the receptor has been removed from the cell surface.

Receptor internalization is unlikely to account for the decrease in peak signal observed for the recruitment of Gab1, Grb2, and Shp2 at high concentrations of all the growth factors. This decrease could reflect increased competition between signaling proteins for binding to sites on the EGF receptor when the signal is strong. It could also arise from depletion of a common pool of adapter or scaffold proteins when the stimulatory signal is maximal. Finally, it is possible that there is steric interference with luciferase complementation when the signal is strong, and Gab1, Grb2, and Shp2 bind to the EGF receptor in a multiprotein complex.

All EGF receptor/signaling protein pairs showed a dose dependence on the concentration of growth factor. However, the EC_{50} for any given growth factor varied as much as 18-fold for the recruitment of different proteins. Knudsen *et al.* (29) reported similar differences in the EC_{50} values of four EGF receptor ligands for inducing the phosphorylation of the EGF receptor and several signaling proteins. The molecular basis for this observation is not known, but it may reflect differences in the order or extent of phosphorylation of sites in response to these seven agonists (21, 30–32).

Surprisingly, there were significant differences in the ability of saturating concentrations of erlotinib and lapatinib to inhibit the recruitment of these signaling proteins. This is likely due to differences in residual phosphorylation of the EGF receptor. These findings clearly identify erlotinib as a more effective inhibitor of signaling in this system than lapatinib and suggest that these complementation assays may be useful for identifying residual signaling pathways that could be targeted for therapeutic benefit.

The overarching message from the principal component analysis is that there are significant differences in signaling protein recruitment depending on both ligand and dose. These variable responses likely reflect different signaling protein recruitment strategies employed by the individual ligands over their entire dose range. Although the observed differences are subtle at the level of individual proteins recruited, collectively they could readily give rise to a distinctly different biological outcome for each of the agonist ligands.

In vivo levels of EGF and other ErbB family growth factors vary widely from tissue to tissue, being low in plasma but 10–100-fold higher in secretions such as saliva and tears (33–35). Given the differences in network behavior identified here, our data imply that therapeutic agents that target one particular node in the signaling pathway could be efficacious in one tissue but not in another, simply because of differences in network utilization based on the identity of the stimulating growth factor and/or the dose involved. This underscores the need to understand the signaling network at all doses of growth factor, as different tissues will likely be responding to vastly different doses of EGF or other EGF receptor agonists. Because many of the experiments that have defined our understanding of this network have been carried out using high dose EGF (36–40), our current appreciation of the network may not reflect the actual flux through the pathways under all physiological conditions.

Ensemble clustering of the responses to the growth factors demonstrated that the seven different EGF receptor ligands basically cluster into three groups as follows: (i) BTC, TGF α , and EPG; and (ii) EGF, AREG, and EPR. HB-EGF is distantly related to both clusters. Thus, based on their ability to recruit these signaling proteins to the EGF receptor, these ligands do not fall neatly into groups defined by high *versus* low affinity nor do they fall into groups based on whether they bind only to the EGF receptor or to both the EGF receptor and ErbB4 (17). Whether there is some specific functional difference that distinguishes the two main groups of EGF receptor agonists, such as temporal or spatial differences in expression, remains to be determined.

With respect to similarities in the utilization of the network by the different growth factors, our analysis identified a strong correlation between the recruitment of Cbl and the recruitment of CrkL to the EGF receptor. As CrkL is known to bind directly to Cbl (41, 42), the detection of a correlation between Cbl and CrkL binding to the EGF receptor suggests that the primary mechanism through which CrkL associates with the EGF receptor may be through binding to tyrosine-phosphorylated Cbl. The fact that this relationship is clearly observed in our dataset suggests that this analysis is capable of identifying inter-

actions between proteins that associate within this signaling network. Viewed in this light, the significant correlation between p52 Shc and Gab1 suggests that this also represents a preferred interaction in this network. The direct binding of p52 Shc to Gab1 has been reported (43, 44).

The finding that other canonical network interactions, such as Grb2/Shc or Grb2-Gab1, were not detected in this analysis likely reflects the complex and dynamic behavior of the network. Grb2 is an adapter protein that recruits a number of proteins, including Cbl, Gab1, and Shp2, to the EGF receptor. It can bind directly to the EGF receptor or indirectly through Shc. As a result, the interaction of Grb2 with the EGF receptor represents the summation of a multiplicity of different binding events. Variation in the dynamics of the different binding events, such as Grb2-Cbl *versus* Grb2-Gab1, could easily obscure any correlations between the binding of the individual partners in the protein pairs. Thus, it will be necessary to assess these interactions more directly to determine whether their association is differentially affected by the seven EGF receptor agonists.

Ultimately, we would like to be able to determine which path through the network is used to recruit a particular protein to the EGF receptor signaling complex by a particular growth factor at a particular dose. Prediction on this level is likely to require careful modeling of network behavior. To this end, these data can be used, in conjunction with other information, to build mechanistic models of the network interactions to determine the dose-dependent network paths of a given signaling protein.

Author Contributions—L. J. P. conceived and designed the study, oversaw all the work, and wrote the manuscript. J. L. M. and L. H. performed the experiments. N. J. B. synthesized and characterized the epigen and epiregulin. T. R. and K. M. N. performed and interpreted the principal component analysis and ensemble clustering, and wrote the manuscript.

References

1. Clayton, A. H., Walker, F., Orchard, S. G., Henderson, C., Fuchs, D., Rothacker, J., Nice, E. C., and Burgess, A. W. (2005) Ligand-induced dimer-tetramer transition during the activation of cell surface epidermal growth factor receptor. A multidimensional microscopy analysis. *J. Biol. Chem.* **280**, 30392–30399
2. Hofman, E. G., Bader, A. N., Voortman, J., van den Heuvel, D. J., Sigismund, S., Verkleij, A. J., Gerritsen, H. C., and van Bergen en Henegouwen, P. M. (2010) Ligand-induced epidermal growth factor receptor (EGFR) oligomerization is kinase-dependent and enhances internalization. *J. Biol. Chem.* **285**, 39481–39489
3. Moriki, T., Maruyama, H., and Maruyama, I. N. (2001) Activation of preformed EGF receptor dimers by ligand-induced rotation of the transmembrane domain. *J. Mol. Biol.* **311**, 1011–1026
4. Saffarian, S., Li, Y., Elson, E. L., and Pike, L. J. (2007) Oligomerization of the EGF receptor investigated by live cell fluorescence intensity distribution analysis. *Biophys. J.* **93**, 1021–1031
5. Yu, X., Sharma, K. D., Takahashi, T., Iwamoto, R., and Mekada, E. (2002) Ligand-independent dimer formation of epidermal growth factor receptor (EGFR) is a step separable from ligand-induced EGFR signaling. *Mol. Biol. Cell* **13**, 2547–2557
6. Garrett, T. P., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Zhu, H.-J., Walker, F., Frenkel, M. J., Hoyne, P. A., Jorissen, R. N., Nice, E. C., Burgess, A. W., and Ward, C. W. (2002) Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α . *Cell* **110**, 763–773
7. Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.-H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., and Yokoyama, S. (2002) Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* **110**, 775–787
8. Zhang, X., Gureasko, J., Shen, K., Cole, P. A., and Kuriyan, J. (2006) An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* **125**, 1137–1149
9. Citri, A., and Yarden, Y. (2006) EGF-ErbB signalling: toward the systems level. *Nat. Rev. Mol. Cell Biol.* **7**, 505–516
10. Pawson, T. (2004) Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* **116**, 191–203
11. Levkowitz, G., Waterman, H., Ettenberg, S. A., Katz, M., Tsygankov, A. Y., Alroy, I., Lavi, S., Iwai, K., Reiss, Y., Ciechanover, A., Lipkowitz, S., and Yarden, Y. (1999) Ubiquitin ligase activity and tyrosine phosphorylation underlie suppression of growth factor signaling by c-Cbl/Sli-1. *Mol. Cell* **4**, 1029–1040
12. Holgado-Madruga, M., Emllet, D. R., Moscatello, D. K., Godwin, A. K., and Wong, A. J. (1996) A Grb2-associated docking protein in EGF- and insulin-receptor signalling. *Nature* **379**, 560–564
13. Gu, H., and Neel, B. G. (2003) The ‘Gab’ in signal transduction. *Trends Cell Biol.* **13**, 122–130
14. Rodrigues, G. A., Falasca, M., Zhang, Z., Ong, S. H., and Schlessinger, J. (2000) A novel positive feedback loop mediated by the docking protein Gab1 and phosphatidylinositol 3-kinase in epidermal growth factor receptor signaling. *Mol. Cell Biol.* **20**, 1448–1459
15. Simister, P. C., and Feller, S. M. (2012) Order and disorder in large multi-site docking proteins of the Gab family—implications for signalling complex formation and inhibitor design strategies. *Mol. Biosyst.* **9**, 33–46
16. Wang, W., Xu, S., Yin, M., and Jin, Z. G. (2015) Essential roles of Gab1 tyrosine phosphorylation in growth factor-mediated signaling and angiogenesis. *Int. J. Cardiol.* **181**, 180–184
17. Wilson, K. J., Gilmore, J. L., Foley, J., Lemmon, M. A., and Riese, D. J., 2nd (2009) Functional selectivity of EGF family peptide growth factors: implications for cancer. *Pharmacol. Ther.* **122**, 1–8
18. Saito, T., Okada, S., Ohshima, K., Yamada, E., Sato, M., Uehara, Y., Shimizu, H., Pessin, J. E., and Mori, M. (2004) Differential activation of epidermal growth factor (EGF) receptor downstream signaling pathways by betacellulin and EGF. *Endocrinology* **145**, 4232–4243
19. Shin, H. S., Lee, H. J., Nishida, M., Lee, M.-S., Tamura, R., Yamashita, S., Matsuzawa, Y., Lee, I.-K., and Koh, G. Y. (2003) Betacellulin and amphiregulin induce upregulation of cyclin D1 and DNA synthesis activity through differential signaling pathways in vascular smooth muscle cells. *Circ. Res.* **93**, 302–310
20. Streicher, K. L., Willmarth, N. E., Garcia, J., Boerner, J. L., Dewey, T. G., and Ethier, S. P. (2007) Activation of a nuclear factor- κ B/interleukin-1 positive feedback loop by amphiregulin in human breast cancer cells. *Mol. Cancer Res.* **5**, 847–861
21. Wilson, K. J., Mill, C., Lambert, S., Buchman, J., Wilson, T. R., Hernandez-Gordillo, V., Gallo, R. M., Ades, L. M., Settleman, J., and Riese, D. J., 2nd (2012) EGFR ligands exhibit functional differences in models of paracrine and autocrine signaling. *Growth Factors* **30**, 107–116
22. Macdonald-Obermann, J. L., Adak, S., Landgraf, R., Piwnicka-Worms, D., and Pike, L. J. (2013) Dynamic analysis of the epidermal growth factor (EGF) receptor-ErbB2-ErbB3 protein network by luciferase fragment complementation imaging. *J. Biol. Chem.* **288**, 30773–30784
23. Macdonald-Obermann, J. L., and Pike, L. J. (2014) Different epidermal growth factor (EGF) ligands show distinct kinetics and biased or partial agonism for homodimer and heterodimer formation. *J. Biol. Chem.* **289**, 26178–26188
24. Yang, K. S., Ilagan, M. X., Piwnicka-Worms, D., and Pike, L. J. (2009) Luciferase fragment complementation imaging of conformational changes in the EGF receptor. *J. Biol. Chem.* **284**, 7474–7482
25. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.

- (2011) Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830
26. Luker, K. E., Smith, M. C., Luker, G. D., Gammon, S. T., Piwnicka-Worms, H., and Piwnicka-Worms, D. (2004) Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12288–12293
 27. Zheng, Y., Zhang, C., Croucher, D. R., Soliman, M. A., St-Denis, N., Pasculescu, A., Taylor, L., Tate, S. A., Hardy, W. R., Colwill, K., Dai, A. Y., Bagshaw, R., Dennis, J. W., Gingras, A.-C., Daly, R. J., and Pawson, T. (2013) Temporal regulation of EGF signalling networks by the scaffold protein Shc1. *Nature* **499**, 166–173
 28. Beguinot, L., Lyall, R. M., Willingham, M. C., and Pastan, I. (1984) Down-regulation of the epidermal growth factor receptor in KB cells is due to receptor internalization and subsequent degradation in lysosomes. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2384–2388
 29. Knudsen, S. L., Mac, A. S., Henriksen, L., van Deurs, B., and Grøvdal, L. M. (2014) EGFR signaling patterns are regulated by its different ligands. *Growth Factors* **32**, 155–163
 30. Gilmore, J. L., Scott, J. A., Bouizar, Z., Robling, A., Pitfield, S. E., Riese, D. J., 2nd, and Foley, J. (2008) Amphiregulin-EGFR signaling regulates PTHrP gene expression in breast cancer cells. *Breast Cancer Res. Treat.* **110**, 493–505
 31. Curran, T. G., Zhang, Y., Ma, D. J., Sarkaria, J. N., and White, F. M. (2015) A multiplex method for absolute quantification of peptides and posttranslational modifications. *Nat. Commun.* **6**, 5924
 32. Guo, L., Kozlosky, C. J., Ericsson, L. H., Daniel, T. O., Cerretti, D. P., and Johnson, R. S. (2003) Studies of ligand-induced site-specific phosphorylation of epidermal growth factor receptor. *J. Am. Soc. Mass Spectrom.* **14**, 1022–1031
 33. Grayson, L. S., Hansbrough, J. F., Zapata-Sirvent, R. L., Dore, C. A., Morgan, J. L., and Nicolson, M. A. (1993) Quantitation of cytokine levels in skin graft donor site wound fluid. *Burns* **19**, 401–405
 34. Lemos-González, Y., Rodríguez-Berrocal, F. J., Cordero, O. J., Gómez, C., and Páez de la Cadena, M. (2007) Alteration of the serum levels of the epidermal growth factor receptor and its ligands in patients with non-small cell lung cancer and head and neck carcinoma. *Br. J. Cancer* **96**, 1569–1578
 35. Marti, U., Burwen, S. J., and Jones, A. L. (1989) Biological effects of epidermal growth factor, with emphasis on the gastrointestinal tract and liver: an update. *Hepatology* **9**, 126–138
 36. Blagoev, B., Ong, S.-E., Kratchmarova, I., and Mann, M. (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat. Biotechnol.* **22**, 1139–1145
 37. Dengjel, J., Akimov, V., Olsen, J. V., Bunkenborg, J., Mann, M., Blagoev, B., and Andersen, J. S. (2007) Quantitative proteomic assessment of very early cellular signaling events. *Nat. Biotechnol.* **25**, 566–568
 38. Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648
 39. Zhang, Y., Wolf-Yadlin, A., Ross, P. L., Pappin, D. J., Rush, J., Lauffenburger, D. A., and White, F. M. (2005) Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell. Proteomics* **4**, 1240–1250
 40. Steen, H., Kuster, B., Fernandez, M., Pandey, A., and Mann, M. (2002) Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway. *J. Biol. Chem.* **277**, 1031–1039
 41. Kyono, W. T., de Jong, R., Park, R. K., Liu, Y., Heisterkamp, N., Groffen, J., and Durden, D. (1998) Differential interaction of CrkL with Cbl or C3G, Hef-1 and γ subunit immunoreceptor tyrosine-based activation motif in signaling of myeloid high affinity Fc receptor for IgG (Fc γ RI). *J. Immunol.* **161**, 5555–5563
 42. Sattler, M., Salgia, R., Shrikhande, G., Verma, S., Uemura, N., Law, S. F., Golemis, E. A., and Griffin, J. D. (1997) Differential signaling after β 1 integrin ligation is mediated through binding of CrkL to p120Cbl and p110HEF1. *J. Biol. Chem.* **272**, 14320–14326
 43. Ingham, R. J., Holgado-Madruga, M., Siu, C., Wong, A. J., and Gold, M. R. (1998) The Gab1 protein is a docking site for multiple proteins involved in signaling by the B cell antigen receptor. *J. Biol. Chem.* **273**, 30630–30637
 44. Lecoq-Lafon, C., Verdier, F., Fichelson, S., Chrétien, S., Gisselbrecht, S., Lacombe, C., and Mayeux, P. (1999) Erythropoietin induces the tyrosine phosphorylation of Gab1 and its association with Shc, Shp2, SHIP and phosphatidylinositol 3-kinase. *Blood* **93**, 2578–2585