

A Mormon believer's legacy to
Mesoamerican archaeology p. 264

An algorithm to
help refugees p. 325

Synthesizing genes
in oil drops p. 343

Science

\$15
19 JANUARY 2018
sciencemag.org

AAAS

NANOSCALE ROBO-ARM

Electric fields drive precise
rotary motion pp. 279 & 296



CONTENTS

19 JANUARY 2018 • VOLUME 359 • ISSUE 6373



264

The unexpected legacy
of a religious quest

NEWS

IN BRIEF

256 News at a glance

IN DEPTH

258 NEWBORN EXOPLANET EYED FOR MOONS AND RINGS

Dedicated microsatellite joins telescopes
watching for rare transit of β Pictoris b
By D. Clery

259 'LIQUID BIOPSY' FOR CANCER PROMISES EARLY DETECTION

Combining DNA and protein markers
brings researchers closer to a universal
cancer screening test *By J. Kaiser*

► REPORT BY J. D. COHEN ET AL.

10.1126/science.aar3247

260 TAMED IMMUNE REACTION AIDS PREGNANCY

Evolutionary studies show how dialing
back inflammation allows embryo
implantation *By E. Pennisi*

261 TENSIONS FLARE OVER ELECTRIC FISHING IN EUROPEAN WATERS

European Parliament calls for total
ban of a technique that saves fuel and
reduces damage to marine life
By E. Stokstad

262 ROCHESTER ROILED BY FALLOUT FROM SEXUAL HARASSMENT CASE

Report supports university's handling
of explosive charges involving linguist
T. Florian Jaeger, but president bows out
By M. Wadman

263 ARE ALGORITHMS GOOD JUDGES?

People are as good as machines in
predicting rearrest *By C. Maticic*

► SCIENCE ADVANCES RESEARCH ARTICLE BY

J. DRESSEL ET AL. 10.1126/sciadv.aao5580



257

New spider species



258

FEATURE

264 THE BELIEVER

How a Mormon lawyer transformed
Mesoamerican archaeology—and ended
up losing his faith *By L. Wade*

► VIDEO

INSIGHTS

POLICY FORUM

270 ASSESSING NATURE'S CONTRIBUTIONS TO PEOPLE

Recognizing culture, and diverse sources
of knowledge, can improve assessments
By S. Díaz et al.

PERSPECTIVES

273 THE ART OF MANUFACTURING MOLECULES

Additively manufactured monolithic
reactors allow on-demand synthesis of
drug molecules *By C. H. Hornung*

► REPORT P. 314

274 QUANTUM LIQUIDS GET THIN

A mix of two bosonic particles develops
attractive forces to create a quantum
liquid *By I. Ferrier-Barbut and T. Pfau*

► REPORT P. 301

276 A BACTERIAL COAT THAT IS NOT PURE COTTON

Biofilms formed by *E. coli* and
Salmonella contain a new form of
modified cellulose

By M. Y. Galperin and D. N. Shalaeva

► REPORT P. 334

277 TAKING DOWN DEFENSES TO IMPROVE VACCINES

A new approach to generating influenza
virus vaccines could improve responses

By J. R. Teijaro and D. R. Burton

► RESEARCH ARTICLE P. 290

279 REMOTE CONTROL OF NANOSCALE DEVICES

A DNA nanodevice can be manipulated
with an applied electric field

By B. Högberg

► RESEARCH ARTICLE P. 296

280 BEN BARRES (1954–2017)

A passionate neuroscientist and
advocate of equal opportunity in the
sciences *By M. Raff*

BOOKS ET AL.

281 QUARKS, CULTURE, COMBOGENESIS

A multidisciplinary tour of cosmic
history charts the “grand sequence”
of existence *By B. Wood*

282 ENRICO FERMI, FLAWS AND ALL

A revealing biography falls short when
it comes to the famous physicist's
problematic treatment of women
By M. Formato

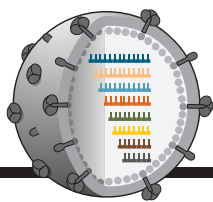
LETTERS

283 THE PITFALLS OF TAKING SCIENCE TO THE PUBLIC

By A. Antonelli and A. Perrigo

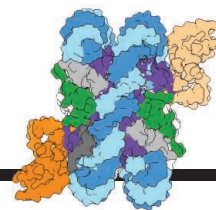
283 VACCINE MANDATES IN FRANCE WILL SAVE LIVES

By S. A. Plotkin et al.



277 & 290

Genomics for vaccine development



339

Assembling the kinetochore

284 LIFE IN SCIENCE: HAVE YOUR MOMOS AND EAT THEM, TOO

By L. E. Specker Sullivan

RESEARCH

IN BRIEF

285 From *Science* and other journals

REVIEW

288 BIOPHYSICS

Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer *E. Lerner* et al.

REVIEW SUMMARY; FOR FULL TEXT:

dx.doi.org/10.1126/science.aan1133

RESEARCH ARTICLES

289 MEMBRANE PROTEINS

Membrane protein insertion through a mitochondrial β -barrel gate *A. I. C. Höhr* et al.

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:

dx.doi.org/10.1126/science.aah6834

290 VACCINES

Genome-wide identification of interferon-sensitive mutations enables influenza vaccine design *Y. Du* et al.

► PERSPECTIVE P. 277

296 NANOROBOTICS

A self-assembled nanoscale robotic arm controlled by electric fields *E. Kopperger* et al.

► PERSPECTIVE P. 279

REPORTS

301 QUANTUM FLUIDS

Quantum liquid droplets in a mixture of Bose-Einstein condensates *C. R. Cabrera* et al.

► PERSPECTIVE P. 274

304 INDUCED SEISMICITY

Hydraulic fracturing volume is associated with induced earthquake productivity in the Duvernay play *R. Schultz* et al.

309 MAGNETIC MATERIALS

Chiomagnetic nanoparticles and gels *J. Yeom* et al.

314 CHEMICAL ENGINEERING

Digitization of multistep organic synthesis in reactionware for on-demand pharmaceuticals *P. J. Kitson* et al.

► PERSPECTIVE P. 273

320 BIOGEOGRAPHY

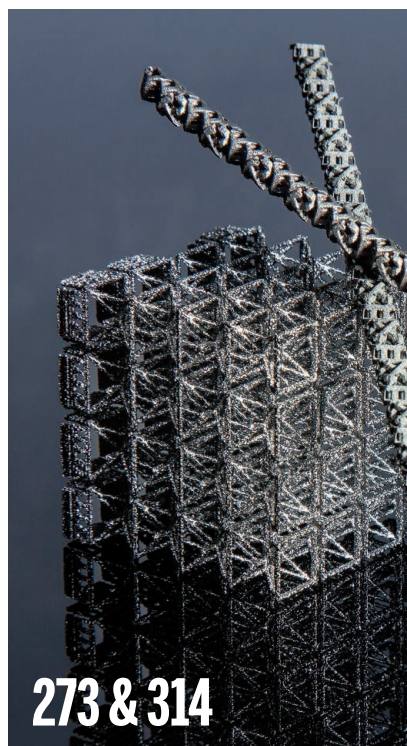
A global atlas of the dominant bacteria found in soil *M. Delgado-Baquerizo* et al.

325 SOCIAL SCIENCE

Improving refugee integration through data-driven algorithmic assignment *K. Bansak* et al.

329 MOLECULAR BIOLOGY

Dicer uses distinct modules for recognizing dsRNA termini *N. K. Sinha* et al.



273 & 314

334 CHEMICAL BIOLOGY

Phosphoethanolamine cellulose: A naturally produced chemically modified cellulose *W. Thongsomboon* et al.

► PERSPECTIVE P. 276

339 STRUCTURAL BIOLOGY

Structural mechanisms of centromeric nucleosome recognition by the kinetochore protein CENP-N *S. Chittori* et al.

343 SYNTHETIC BIOLOGY

Multiplexed gene synthesis in emulsions for exploring protein functional landscapes *C. Plesa* et al.

DEPARTMENTS

255 EDITORIAL

The 1918 flu, 100 years later

By Jessica A. Belser and Terrence M. Tumpey

362 WORKING LIFE

From parade ground to PI

By John Tregoning

ON THE COVER



Artistic rendering of a long chain of self-assembled DNA platforms bearing multiple rotor arms. The motion of the arms between docking stations (yellow) defined on the plates

can be precisely directed by externally applied electric fields. Such systems could form the basis for computer-controlled nanorobotic assembly lines in which the arms cooperate in the synthesis of nanoscale structures. See pages 279 and 296. *Illustration: C. Bickel/Science; Data: E. Kopperger and F. C. Simmel/Technical University of Munich*

Science Staff	254
New Products	348
AAAS Meeting Program	349
Science Careers	359

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2018 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): \$165 (\$74 allocated to subscription). Domestic institutional subscription (51 issues): \$1808; Foreign postage extra: Mexico, Caribbean (surface mail) \$55; other countries (air assist delivery): \$89 First class, airmail, student, and emeritus rates upon request. Canadian rates with GST available upon request. GST #R125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.** Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090-6178. **Single-copy sales:** \$15 each plus shipping and handling; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and others who use Copyright Clearance Center (CCC) Pay-Per-Use services provided that \$35.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for Science is 0036-8075. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.

Editor-in-Chief Jeremy Berg

Executive Editor Monica M. Bradford **News Editor** Tim Appenzeller

Deputy Editors Lisa D. Chong, Andrew M. Sugden(UK), Valda J. Vinson, Jake S. Yeston

Research and Insights

DEPUTY EDITOR, EMERITUS Barbara R. Jasny **SR. EDITORS** Gemma Alderton(UK), Caroline Ash(UK), Gilbert J. Chin, Julia Fahrenkamp-Uppenbrink(UK), Pamela J. Hines, Stella M. Hurlley(UK), Paula A. Kiberstis, Marc S. Lavine(Canada), Steve Mao, Ian S. Osborne(UK), Beverly A. Purnell, L. Bryan Ray, H. Jesse Smith, Jelena Stajic, Peter Stern(UK), Phillip D. Szuromi, Sacha Vignieri, Brad Wible, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Brent Grocholski, Priscilla N. Kelly, Seth Thomas Scanlon(UK), Keith T. Smith(UK) **ASSOCIATE BOOK REVIEW EDITOR** Valerie B. Thompson **LETTERS EDITOR** Jennifer Sills **LEAD CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmeck **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Chris Filiatreau, Cynthia Howe, Catherine Wolner **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Scott Miller, Jerry Richardson, Anita Wynn **PUBLICATIONS ASSISTANTS** Ope Martins, Nida Masiulis, Dona Mathieu, Hilary Stewart(UK), Alana Warnke, Alice Whaley(UK), Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ADMINISTRATIVE SUPPORT** Janet Clements(UK), Lizzanne Newton(UK)

News

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** Richard Stone **DEPUTY NEWS EDITORS** Elizabeth Culotta, Martin Enserink(Europe), David Grimm, Eric Hand, David Malakoff, Leslie Roberts **SR. CORRESPONDENTS** Daniel Clery(UK), Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brinard, Catherine Matakic **NEWS WRITERS** Adrian Co, Jon Cohen, Jennifer Couzin-Frankel, Jocelyn Kaiser, Kelly Servick, Robert F. Service, Erik Stokstad(Cambridge, UK), Paul Voosen, Meredith Wadman **INTERNS** Roni Dengler, Katie Langin, Matt Warren **CONTRIBUTING CORRESPONDENTS** John Bohannon, Warren Cornwall, Ann Gibbons, Mara Hvistendahl, Sam Kean, Eli Kintisch, Kai Kupferschmidt(Berlin), Andrew Lawler, Mitch Leslie, Eliot Marshall, Virginia Morell, Dennis Normile(Shanghai), Tania Rabesandratana(London), Emily Underwood, Gretchen Vogel(Berlin), Lizzie Wade(Mexico City) **CAREERS** Donisha Adams, Rachel Bernstein(Editor), Maggie Kuo **COPY EDITORS** Dorie Chevlen, Julia Cole (Senior Copy Editor), Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

Executive Publisher Rush D. Holt

Publisher Bill Moran **Chief Digital Media Officer** Josh Freeman

DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT Sarah Whalen **DIRECTOR, PRODUCT AND CUSTOM PUBLISHING** Will Schweitzer **MANAGER, PRODUCT DEVELOPMENT** Hannah Heckner **BUSINESS SYSTEMS AND FINANCIAL ANALYSIS** Director Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYST** Eric Knott **SENIOR SYSTEMS ANALYST** Nicole Mehmedovich **SENIOR BUSINESS ANALYST** Cory Lipman **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **BUSINESS ANALYSTS** Meron Kebede, Sandy Kim, Jourdan Stewart **FINANCIAL ANALYST** Julian Iriarte **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **SALES COORDINATOR** Shirley Young **DIRECTOR, COPYRIGHT, LICENSING, SPECIAL PROJECTS** Emilie David **DIGITAL PRODUCT ASSOCIATE** Michael Hardesty **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **RIGHTS, CONTRACTS, AND LICENSING ASSOCIATE** Lili Catlett **RIGHTS & PERMISSIONS ASSISTANT** Alexander Lee

MARKETING MANAGER, PUBLISHING Shawana Arnold **MARKETING ASSOCIATE** Steven Goodman **CREATIVE DIRECTOR** Scott Rodgers **SENIOR ART ASSOCIATES** Paula Fry **ART ASSOCIATE** Kim Huynh

INTERIM DIRECTOR, INSTITUTIONAL LICENSING Iquo Edim **ASSOCIATE DIRECTOR, RESEARCH & DEVELOPMENT** Elisabeth Leonard **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGERS** Marco Castellani, Chris Murawski **SENIOR OPERATIONS ANALYST** Lana Guz **MANAGER, AGENT RELATIONS & CUSTOMER SUCCESS** Judy Lillibridge

WEB TECHNOLOGIES PORTFOLIO MANAGER Trista Smith **TECHNICAL MANAGER** Chris Coleman **PROJECT MANAGER** Nick Fletcher **DEVELOPERS** Elissa Heller, Ryan Jensen, Brandon Morrison

DIGITAL MEDIA DIRECTOR OF ANALYTICS Enrique Gonzales **DIGITAL REPORTING ANALYST** Eric Hossinger **SR. MULTIMEDIA PRODUCER** Sarah Crespi **MANAGING DIGITAL PRODUCER** Kara Estelle-Powers **PRODUCER** Liana Birke **VIDEO PRODUCERS** Chris Burns, Nguyễn Khôi Nguyễn **DIGITAL SOCIAL MEDIA PRODUCER** Brice Russ

DIGITAL/PRINT STRATEGY MANAGER Jason Hillman **QUALITY TECHNICAL MANAGER** Marcus Spiegler **PROJECT ACCOUNT MANAGER** Tara Kelly **DIGITAL PRODUCTION MANAGER** Lisa Stanford **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT SPECIALISTS** Steve Forrester, Antoinette Hodal, Lori Murphy, Anthony Rosen **CONTENT SPECIALISTS** Jacob Hedrick, Kimberley Oster

DESIGN DIRECTOR Beth Rakouskas **DESIGN MANAGING EDITOR** Marcy Atarod **SENIOR DESIGNER** Chrystal Smith **DESIGNER** Christina Aycocock **GRAPHICS MANAGING EDITOR** Alberto Cuadra **SENIOR SCIENTIFIC ILLUSTRATORS** Valerie Altounian, Chris Bickel, Katharine Sutfill **SCIENTIFIC ILLUSTRATOR** Alice Kitterman **INTERACTIVE GRAPHICS EDITOR** Jia You **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **PHOTO EDITOR** Emily Petersen **IMAGE RIGHTS AND FINANCIAL MANAGER** Jessica Adams **INTERN** Emily Miah

SENIOR EDITOR, CUSTOM PUBLISHING Sean Sanders: 202-326-6430 **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst: 202-326-6463 **ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT** Justin Sawyers: 202-326-7061 science_advertising@aaas.org **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **SR. PRODUCTION SPECIALIST/GRAPHIC DESIGNER** Amy Hardcastle **SR. TRAFFIC ASSOCIATE** Christine Hall **DIRECTOR OF BUSINESS DEVELOPMENT AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu: +86-131 6136 3212, xchu@aaas.org **COLLABORATION/CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu + 81532-81-5142 asandhu@aaas.org **EAST COAST/E. CANADA** Laurie Faraday: 508-747-9395, FA617-507-8189 **WEST COAST/W. CANADA** Lynne Stickrod: 415-931-9782, FA415-520-6940 **MIDWEST** Jeffrey Dembski: 847-498-4520 x3005, Steven Loerch: 847-498-4520 x3006 **UK EUROPE/ASIA** Roger Goncalves: TEL/FAX +41 43 243 1358 **JAPAN** Kaoru Sasaki (Tokyo): +81 (3) 6459 4174 ksasaki@aaas.org

GLOBAL SALES DIRECTOR ADVERTISING AND CUSTOM PUBLISHING Tracy Holmes: +44 (0) 1223 326525 **CLASSIFIED** advertise@sciencecareers.org **SALES MANAGER, US, CANADA AND LATIN AMERICA** SCIENCE CAREERS Claudia Paulsen-Young: 202-326-6577 **EUROPE/ROW SALES** Sarah Lelarge **SALES ADMIN ASSISTANT** Kelly Grace +44 (0)1223 326528 **JAPAN** Miyuki Tani(Osaka): +81 (6) 6202 6272 mtani@aaas.org **CHINA/TAIWAN** Xiaoying Chu: +86-131 6136 3212, xchu@aaas.org **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING ASSOCIATE** Aimee Aponte

AAAS BOARD OF DIRECTORS, CHAIR Barbara A. Schaal **PRESIDENT** Susan Hockfield **PRESIDENT-ELECT** Margaret A. Hamburg **CHIEF EXECUTIVE OFFICER** Rush D. Holt **BOARD** Cynthia M. Beall, May R. Berenbaum, Carlos J. Bustamante, Kaye Husbands Fealing, Stephen P.A. Fodor, S. James Gates, Jr., Michael S. Gazzaniga, Laura H. Greene, Mercedes Pascual

SUBSCRIPTION SERVICES For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417. FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

INSTITUTIONAL SITE LICENSES 202-326-6730 **REPRINTS:** Author Inquiries 800-635-7181 **COMMERCIAL INQUIRIES** 803-359-4578 **PERMISSIONS** 202-326-6765, permissions@aaas.org **AAAS Member Central Support** 866-434-2227 www.aaas.org/membercentral

Science serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority or conflicting points of view. Accordingly, all articles published in Science—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

INFORMATION FOR AUTHORS See www.sciencemag.org/authors/science-information-authors

BOARD OF REVIEWING EDITORS (Statistics board members indicated with \$)

Adriano Aguzzi, *U. Hospital Zürich*
Takuzo Aida, *U. of Tokyo*
Leslie Aiello, *Wenner-Gren Foundation*
Judith Allen, *U. of Manchester*
Sebastian Amigorena, *Institut Curie*
Meinrat O. Andrae, *Max Planck Inst. Mainz*
Paola Ariotti, *Harvard U.*
Johan Auwerx, *EPFL*
David Awschalom, *U. of Chicago*
Clare Baker, *U. of Cambridge*
Nenad Ban, *ETH Zürich*
Franz Bauer, *Pontificia Universidad Católica de Chile*
Ray H. Baughman, *U. of Texas at Dallas*
Carlo Beenakker, *Leiden U.*
Kamran Behnia, *ESPCI*
Yasmine Belkaid, *NIAD, NIH*
Philip Benfey, *Duke U.*
May Berenbaum, *U. of Illinois at Urbana-Champaign*
Gabriele Bergers, *VIB*
Bradley Bernstein, *Massachusetts General Hospital*
Peer Bork, *EMBL*
Chris Bowler, *Ecole Normale Supérieure*
Ian Boyd, *U. of St. Andrews*
Emily Brodsky, *U. of California, Santa Cruz*
Ron Brookmeyer, *U. of California, Los Angeles (\$)*
Christian Büchel, *UKE Hamburg*
Dennis Burton, *The Scripps Res. Inst.*
Carter Tribley Butts, *U. of California, Irvine*
Gyorgy Buzsaki, *New York U. School of Medicine*
Blanche Capel, *Duke U.*
Mats Carlsson, *U. of Oslo*
Ib Chorkendorff, *Denmark TU*
James J. Collins, *MIT*
Robert Cook-Deegan, *Duke U.*
Lisa Coussens, *Oregon Health & Science U.*
Alan Cowman, *Walter & Eliza Hall Inst.*
Roberta Croce, *VU Amsterdam*
Janet Currie, *Princeton U.*
Jeff L. Dangl, *U. of North Carolina*
Tom Daniel, *U. of Washington*
Chiara Daraio, *Caltech*
Nicolas Dauphas, *U. of Chicago*
Frans de Waal, *Emory U.*
Stanislas Dehaene, *Collège de France*
Robert Desimone, *MIT*
Claude Desplan, *New York U.*
Sandra Díaz, *Universidad Nacional de Córdoba*
Dennis Discher, *U. of Pennsylvania*
Gerald W. Dorn II, *Washington U. in St. Louis*
Jennifer A. Doudna, *U. of California, Berkeley*
Bruce Dunn, *U. of California, Los Angeles*
William Dunphy, *Caltech*
Christopher Dye, *WHO*
Todd Ehlers, *U. of Tübingen*
Jennifer Elisseeff, *Johns Hopkins U.*
Tim Elston, *U. of North Carolina at Chapel Hill*
Barry Everitt, *U. of Cambridge*
Vanessa Ezenwa, *U. of Georgia*
Ernst Fehr, *U. of Zürich*
Anne C. Ferguson-Smith, *U. of Cambridge*
Michael Feuer, *The George Washington U.*
Toren Finkel, *NHLBI, NIH*
Kate Fitzgerald, *U. of Massachusetts*
Peter Fratzl, *Max Planck Inst. Potsdam*
Elaine Fuchs, *Rockefeller U.*
Eileen Furlong, *EMBL*
Jay Gallagher, *U. of Wisconsin*
Daniel Geschwind, *U. of California, Los Angeles*
Karl-Heinz Glassmeier, *TU Braunschweig*
Ramon Gonzalez, *Rice U.*
Elizabeth Grove, *U. of Chicago*
Nicolas Gruber, *ETH Zürich*
Kip Guy, *U. of Kentucky College of Pharmacy*
Taekjip Ha, *Johns Hopkins U.*
Christian Haass, *Ludwig Maximilians U.*
Sharon Hammes-Schiffer, *U. of Illinois at Urbana-Champaign*
Wolf-Dietrich Hardt, *ETH Zürich*
Michael Hasselmo, *Boston U.*
Martin Heimann, *Max Planck Inst. Jena*
Ykä Helariutta, *U. of Cambridge*
Janet G. Hering, *Swiss Fed. Inst. of Aquatic Science & Technology*
Kai-Uwe Hinrichs, *U. of Bremen*
David Hodell, *U. of Cambridge*
Lora Hooper, *UT Southwestern Medical Ctr. at Dallas*
Raymond Huey, *U. of Washington*
Fred Hughson, *Princeton U.*
Randall Hulet, *Rice U.*
Auke Ijspeert, *EPFL*
Stephen Jackson, *USGS Southwest Climate Science Ctr.*
Seema Jayachandran, *Northwestern U.*
Kai Johnson, *EPFL*
Peter Jonas, *Inst. of Science & Technology Austria*
Matt Kaeblerlein, *U. of Washington*
William Kaelin Jr., *Dana-Farber Cancer Inst.*
Daniel Kammen, *U. of California, Berkeley*
Abby Kavner, *U. of California, Los Angeles*
Hitoshi Kawakatsu, *U. of Tokyo*
Masashi Kawasaki, *U. of Tokyo*
V. Narry Kim, *Seoul Nat. U.*
Robert Kingston, *Harvard Medical School*

Etienne Kochlin, *Ecole Normale Supérieure*
Alexander Koldokin, *Johns Hopkins U.*
Thomas Langer, *U. of Cologne*
Mitchell A. Lazar, *U. of Pennsylvania*
David Lazer, *Harvard U.*
Thomas Lecuit, *IBDM*
Virginia Lee, *U. of Pennsylvania*
Stanley Lemon, *U. of North Carolina at Chapel Hill*
Ottoline Leyser, *U. of Cambridge*
Wendell Lim, *U. of California, San Francisco*
Marcia C. Linn, *U. of California, Berkeley*
Jianguo Liu, *Michigan State U.*
Luis Liz-Marzán, *CIC bioGUNE*
Jonathan Losos, *Harvard U.*
Ke Lu, *Chinese Acad. of Sciences*
Christian Lüscher, *U. of Geneva*
Laura Machesky, *Cancer Research UK Beatson Inst.*
Anne Magurran, *U. of St. Andrews*
Oscar Marín, *King's College London*
Charles Marshall, *U. of California, Berkeley*
Christopher Marx, *U. of Idaho*
C. Robertson McClung, *Dartmouth College*
Rodrigo Medellín, *U. of Mexico*
Graham Medley, *London School of Hygiene & Tropical Med.*
Jane Memmott, *U. of Bristol*
Tom Misteli, *NCI, NIH*
Yasushi Miyashita, *U. of Tokyo*
Mary Ann Moran, *U. of Georgia*
Richard Morris, *U. of Edinburgh*
Alison Motsinger-Reif, *NC State U. (\$)*
Daniel Neumark, *U. of California, Berkeley*
Kitty Nijmeijer, *TU Eindhoven*
Helga Nowotny, *Austrian Council*
Rachel O'Reilly, *U. of Warwick*
Joe Orenstein, *U. of California, Berkeley & Lawrence Berkeley Nat. Lab.*
Pilar Ossorio, *U. of Wisconsin*
Harry Orr, *U. of Minnesota*
Andrew Oswald, *U. of Warwick*
Isabella Pagano, *Istituto Nazionale di Astrofisica*
Margaret Palmer, *U. of Maryland*
Steve Palumbi, *Stanford U.*
Jane Parker, *Max Planck Inst. Cologne*
Giovanni Parmigiani, *Dana-Farber Cancer Inst. (\$)*
John H. J. Petrini, *Memorial Sloan Kettering Cancer Center*
Samuel Pfaff, *Stalk Inst. for Biological Studies*
Kathrin Plath, *U. of California, Los Angeles*
Martin Polman, *Ulm U.*
Albert Poljan, *FOM Institute for AMOLF*
Elvira Poloczanska, *Alfred-Wegener-Inst.*
Philippe Poulin, *CNRS*
Jonathan Pritchard, *Stanford U.*
David Randall, *Colorado State U.*
Sarah Reisman, *Caltech*
Félix A. Rey, *Institut Pasteur*
Trevor Robbins, *U. of Cambridge*
Amy Rosenzweig, *Northwestern U.*
Mike Ryan, *U. of Texas at Austin*
Mitinori Saitou, *Kyoto U.*
Shimon Sakaguchi, *Osaka U.*
Miquel Salmeron, *Lawrence Berkeley National Lab*
Jürgen Samdühler, *Medical U. of Vienna*
Alexander Schier, *Harvard U.*
Wolfram Schlenker, *Columbia U.*
Susanna Scott, *U. of California, Santa Barbara*
Vladimir Shalaev, *Purdue U.*
Beth Shapiro, *U. of California, Santa Cruz*
Jay Shendure, *U. of Washington*
Robert Siliciano, *Johns Hopkins School of Medicine*
Uri Simonsohn, *U. of Pennsylvania*
Alison Smith, *John Innes Centre*
Richard Smith, *U. of North Carolina at Chapel Hill (\$)*
Mark Smyth, *QIMR Berghofer*
Pam Solts, *U. of Florida*
John Speakman, *U. of Aberdeen*
Allan C. Spradling, *Carnegie Institution for Science*
Eric Steig, *U. of Washington*
Paula Stephan, *Georgia State U. & Nat. Bureau of Economic Res.*
V. S. Subrahmanian, *U. of Maryland*
Ira Tabas, *Columbia U.*
Sarah Teichmann, *U. of Cambridge*
Shubha Tole, *Tata Inst. of Fundamental Research*
Wim van der Putten, *Netherlands Inst. of Ecology*
Bert Vogelstein, *Johns Hopkins U.*
David Wallach, *Weizmann Inst. of Science*
Jane-Ling Wang, *U. of California, Davis (\$)*
David Waxman, *Fudan U.*
Jonathan Weissman, *U. of California, San Francisco*
Chris Wickle, *U. of Missouri (\$)*
Ian A. Wilson, *The Scripps Res. Inst. (\$)*
Timothy D. Wilson, *U. of Virginia*
Yu Xie, *Princeton U.*
Jan Zaanen, *Leiden U.*
Kenneth Zaret, *U. of Pennsylvania School of Medicine*
Jonathan Zehr, *U. of California, Santa Cruz*
Len Zon, *Boston Children's Hospital*
Maria Zuber, *MIT*

The 1918 flu, 100 years later

Combating a disease of unknown cause is a daunting task. One hundred years ago, a pandemic of poorly understood etiology and transmissibility spread worldwide, causing an estimated 50 million deaths. Initially attributed to *Haemophilus influenzae*, it was not until the 1930s that an H1 subtype was identified as the causative strain. Subsequent influenza pandemics in 1957, 1968, and 2009 did not approach levels of morbidity and mortality comparable to those of the 1918 “Spanish flu,” leaving unanswered for almost a century questions regarding the extraordinary virulence and transmissibility of this unique strain. Technological advances made reconstruction of the 1918 virus possible; now, continued research, vaccine development, and preparedness are essential to ensure that such a devastating public health event is not repeated.

Over the past 20 years, studies of individual genes and the fully reconstructed live 1918 virus have identified numerous features that likely contributed to its robustness and rapid global spread. Importantly, this research has often been conducted in tandem with viral isolates from recent human and zoonotic sources, enabling insights from the 1918 virus to inform evaluations of current pandemic risk. As we now know, wild birds are the natural reservoir for influenza A viruses. With extensive antigenic and genetic diversity inherent among influenza virus surface proteins, a strain to which humans are immunologically naïve could jump the species barrier at any time. A(H5N1) viruses and, more recently, A(H7N9) viruses, are two such examples. However, swine are also recognized as a “mixing vessel” for influenza viruses, and over the past two decades, there has been an increase in human cases following exposure to infected pigs. There is clearly, and alarmingly, a vast diversity of zoonotic sources of influenza A viruses that could acquire a transmissible phenotype in humans and cause a pandemic.

What is our readiness today? Many international health agencies and research laboratories collaborate to

track influenza virus evolution, evaluate antigenic drift among circulating and vaccine strains, and sequence viral genes to advance surveillance and preparedness. The production of improved vaccines and diagnostic tools, and better access to therapeutic agents represent resources that were not available a century ago. But influenza viruses are moving targets, and a pandemic virus could nevertheless emerge with as little warning in 2018 as in 1918. As evidenced by this current flu season, influenza viruses can rapidly acquire mutations that evade our most recent vaccine formulations. A universal, broadly protective influenza vaccine for seasonal epidemics—a goal of intense research efforts—would improve our preparedness for subsequent pandemics.

How, then, can we best study emerging pandemic threats? Looking to the past, elucidating the role of specific molecular determinants that confer virulence and transmissibility of prior pandemic viruses is one approach. But we must also look to the future. Advances in next-generation sequencing are improving our understanding of virus diversity. Investments in global partnerships and laboratory capacity worldwide are strengthening surveillance networks and diagnostic capabilities, and are also facilitating the identification of new viruses in humans and animals. The recent lifting of the U.S. moratorium on gain-of-function research on potential pandemic viruses further illustrates the contribution of unconventional, but responsible, research strategies to readiness.

Philosopher George Santayana pointed out, “Those who cannot remember the past are condemned to repeat it.” We are no doubt more prepared in 2018 for an infectious disease threat than in 1918. But it is critical to remember that preparation only stems from a global commitment to share data about viral isolates, support innovative research, and dedicate resources to assess the pandemic risk of new and emerging influenza viruses from zoonotic reservoirs.

—Jessica A. Belser and Terrence M. Tumpey



Jessica A. Belser is a microbiologist in the Influenza Division, National Center for Immunization and Respiratory Diseases, U.S. Centers for Disease Control and Prevention (CDC), Atlanta, GA 30333, USA. jbelser@cdc.gov



Terrence M. Tumpey is chief of the Immunology and Pathogenesis Branch in the Influenza Division of the CDC, Atlanta, GA 30333, USA. ttumpey@cdc.gov



“How, then, can we best study emerging pandemic threats?”

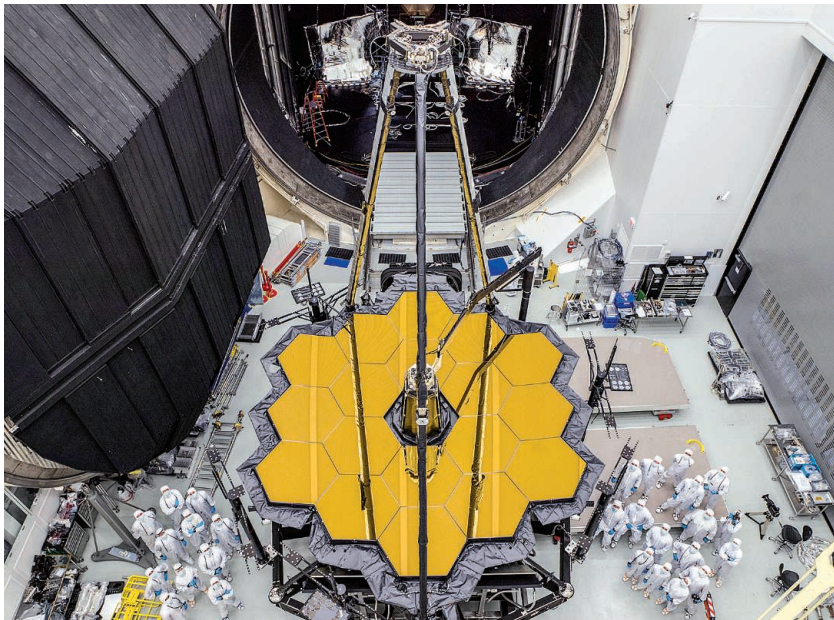
“The practice of plunging live lobsters into boiling water, which is common in restaurants, is no longer permitted.”

A Swiss government order to take effect 1 March, reported by *The Guardian*, revived the debate about whether lobsters feel pain.

IN BRIEF

Edited by Jeffrey Brainard

Space telescope's optics pass test



The Johnson Space Center's Chamber A, originally used in Apollo missions, tested the James Webb Space Telescope in space conditions.

NASA announced last week that the James Webb Space Telescope (JWST), its next big observatory, due for launch a year from now, successfully completed the final test of its optics and instruments. NASA workers inserted the 6.5-meter primary mirror of gold-coated beryllium, along with its secondary mirror and instruments, into a giant test chamber at the Johnson Space Center in Houston, Texas, last year. Over 100 days, engineers slowly cooled the hardware to -235°C in a vacuum to mimic conditions in space. They also shone simulated starlight through the mirror system into the detectors. The JWST team performed the test because it decided early on it didn't want to “make the Hubble mistake,” says project scientist John Mather, referring to the grinding error that initially marred the optics of the Hubble Space Telescope after NASA launched it in 1990. “We've confirmed the primary mirror is a great mirror,” he says. Meanwhile, at prime contractor Northrop Grumman in Redondo Beach, California, engineers tested the JWST's tennis court-size, multilayered sunshield for the first time in October 2017.

AROUND THE WORLD

‘Eugenics conference’ probed

LONDON | University College London (UCL) has opened an inquiry into a low-profile conference on the genetics of intelligence that included discussions of eugenics and was attended by some people who hold controversial views on race and intelligence. The London Conference on Intelligence has been held three times since 2014, according to a student newspaper that last week reported details of the small, invitation-only event. UCL says it did not endorse the meeting and administrators were not notified about the speakers and content, a requirement before conference rooms can be booked. James Thompson, an honorary lecturer in psychology at UCL, who hosted the conference, will not be allowed to organize conferences “of this nature” while the investigation is underway, and the university will examine how he received his lectureship. Thompson could not be reached for comment.

German R&D budget may rise

BERLIN | More than 3 months after voters went to the polls in Germany, it's still unclear whether Chancellor Angela Merkel's Christian Democrat party will form a government with the Social Democrats. But if the two end up in a coalition, scientists are likely to benefit. In a marathon negotiation session last week, the two political parties drew up a 28-page blueprint for a formal coalition. It includes a pledge to raise spending on R&D from 3% to 3.5% of gross domestic product (GDP) until 2025. That would put Germany's spending on par with Japan's. Only Israel and South Korea spend a higher percentage of their GDP on R&D. The parties also tackled the controversial herbicide glyphosate, agreeing to reduce its use in Germany.

Deadly virus in Florida monkeys

SILVER SPRINGS, FLORIDA | Scientists warn that bites or scratches from rhesus macaques roaming Silver Springs State Park in Florida could expose people to a deadly form of herpes. According to a study published online last week in *Emerging*

Infectious Diseases, many of the monkeys carry the herpes B virus, which can cause encephalitis and has killed almost half of all humans who have contracted it. So far, transmission of herpes B from macaques to people has been documented only in laboratories. Researchers from the University of Florida in Gainesville and the University of Washington in Seattle found herpes B in about one in four blood samples taken from the macaques, and, worryingly, as many as 30% of monkeys in one family group also had the virus in their saliva. A dozen macaques were introduced into the park in the 1930s to encourage tourism, and today that population is about 175. Previous efforts to control the monkey population have ended after public opposition.

EPA says less about risk reviews

WASHINGTON, D.C. | The U.S. Environmental Protection Agency (EPA) has curtailed the amount of information it releases on preliminary assessments of potentially hazardous new chemicals or new uses of existing chemicals, drawing complaints from environmental and government transparency advocates. Until recently, EPA routinely released notices on whether a preliminary review had concluded that the agency should scrutinize a new chemical or use for potential risks to human health or the environment. But on 5 January, the

agency revised the website that presented that information. It now indicates only that a “focus meeting occurred” regarding a new chemical or use, without reporting the outcome, E&E News reports. “Gosh, that’s helpful,” Richard Denison, senior scientist at the Environmental Defense Fund in Washington, D.C., sniped in a blog post. “This change dramatically limits the agency’s accountability to the public, not to mention transparency.”

New highway slices the Amazon

MANAUS, BRAZIL | The pending completion of a new highway through the Brazilian Amazon is alarming tropical ecologists, who fear it will open the way to increased habitat destruction as well as illegal hunting and development. The 870-kilometer road, known as BR319, connects Manaus, Brazil, in central Amazonia to Porto Velho, Brazil, in southern Amazonia. The project has been on the books for almost a decade, but in recent months work has accelerated such that now only the central section remains to be paved. If that occurs, BR319 and a second highway “will slice the Amazon in half along a north-south axis ... like a flayed fish,” says ecologist William Laurance, director of the Centre for Tropical Environmental and Sustainability Science at James Cook University in Cairns, Australia. Some natural reserves have been created along the highway, but “it’s now or

never in terms of increasing the measures to limit the highway’s impacts, or even halting the project,” says Laurance, who believes the project “is one of the biggest environmental issues for the planet in 2018.”

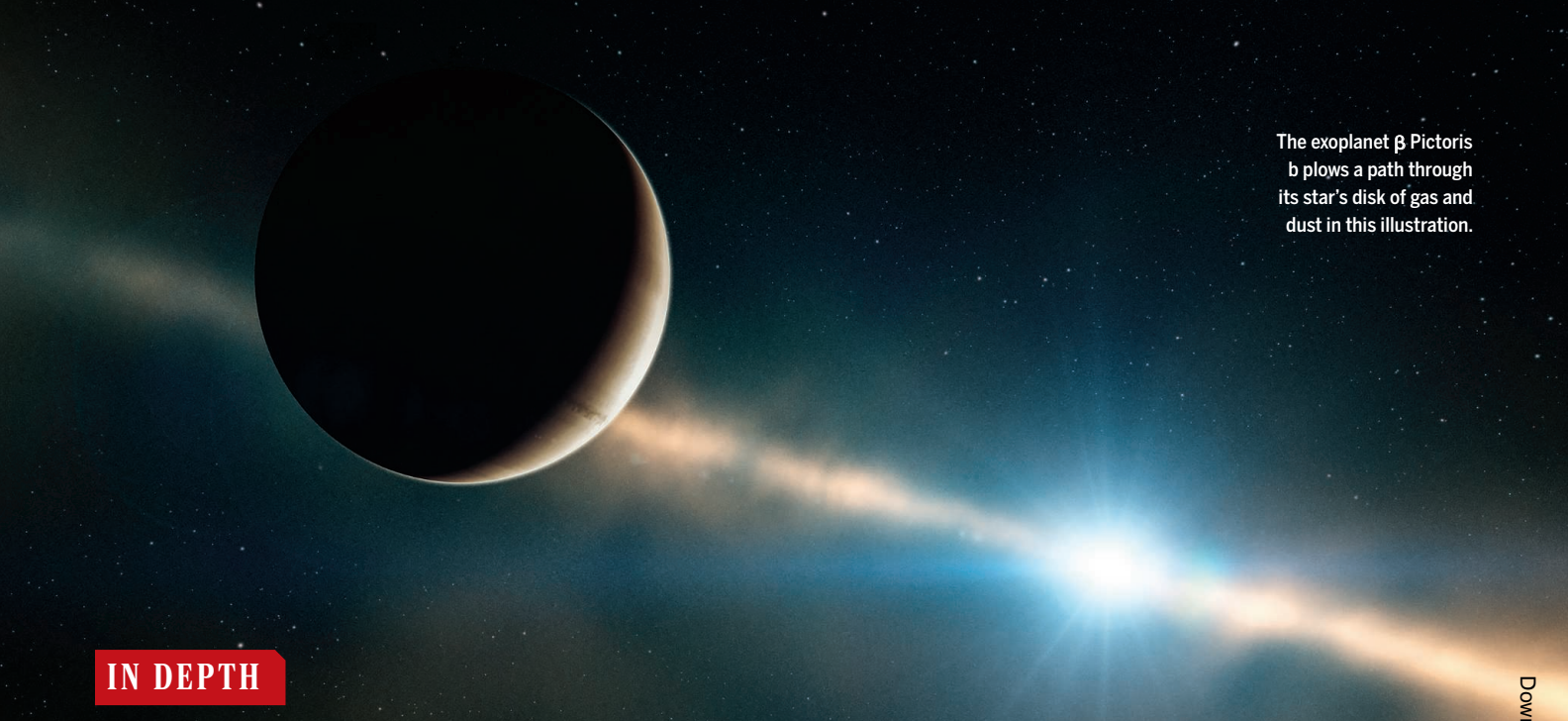
Shared credit, but men first

STORRS, CONNECTICUT | Papers with multiple authors more often list men as the first author, even when the male and female authors are noted as contributing equally, a study has found. Such imbalances may reflect conscious or unconscious biases, Nichole Broderick of the University of Connecticut in Storrs and Arturo Casadevall of Johns Hopkins University in Baltimore, Maryland, suggest in a preprint posted 31 December 2017 on the bioRxiv server. The disparity emerged when they analyzed about 3000 such articles on biomedicine published from 1995 to 2017 in journals including *Science*, *Nature*, and *PLOS Biology*. Since 2007, the disparity has not been statistically significant, indicating that unequal treatment may have lessened in recent years. Because first-author credits can influence a scientist’s funding and career advancement, journals should ask authors to explain the order of listing, Broderick and Casadevall propose. Following their own advice, they explain in their paper that their author order is alphabetical and by increasing seniority.



New spider species proliferate

A group of tiny arachnids called pelican spiders, named for their uncanny resemblance to the birds, uses beaklike mouthparts to spear other arachnids. Now, scientists have discovered 18 new species of these spiders living in Madagascar, they report in the 11 January issue of *ZooKeys*, doubling the number known to occur on the African island nation. The family of rice-size animals was first discovered in 1854 in a 50-million-year-old slab of amber. Scientists from the Smithsonian National Museum of Natural History in Washington, D.C., and the Natural History Museum of Denmark in Copenhagen identified the new species by looking at hundreds of pelican spiders under a microscope. Though all had “beaks” (left), some sported longer mouthparts and “necks,” and others had more spines—a telltale sign that they were members of different species. Pelican spiders use their elongated mouthparts to snatch small spiders and hold them at a distance while they inject them with venom. It’s a nifty strategy, but Madagascar’s pelican spiders face an uncertain future as the island’s rainforests where they live are disappearing. The birdlike arachnids also prowl the forests of South Africa and Australia.



The exoplanet β Pictoris b plows a path through its star's disk of gas and dust in this illustration.

IN DEPTH

ASTRONOMY

Newborn exoplanet eyed for moons and rings

Dedicated microsatellite joins telescopes watching for rare transit of β Pictoris b

By Daniel Clery

Astronomers are staring at a nearby star in hopes of seeing a giant baby of a planet pass across its face, perhaps accompanied by dust clouds, rings, or newborn moons. Last week, the newest and tiniest telescope joined the vigil, when the French-built PicSat rode into orbit on an Indian rocket. It will be able to continuously monitor the star, β Pictoris, until chances of seeing the once-in-20-year transit event diminish in a few months' time. "We can't miss this. We would be kicking ourselves," says astronomer Matthew Kenworthy of Leiden University in the Netherlands.

Astronomers have seen thousands of exoplanets transit, or cross the face of their stars, eclipsing a fraction of their light. But β Pictoris, a bright star just 63 light-years away, is special. It is a natural laboratory for how solar systems form because it is only 24 million years old—the "equivalent of a baby of a few weeks," says Sylvestre Lacour of the Paris Observatory.

In 1984, astronomers observed a disk of gas and dust around it, the first protoplanetary disk to be seen (*Science*, 21 December 1984, p. 1421). The disk, viewed nearly edge on, was warped and had gaps, a sign of planets in the making. But it wasn't until 2009 that researchers spied the faint glow of a hot, young

giant planet, 10 times the mass of Jupiter, in a roughly 20-year orbit. Now dubbed β Pictoris b, it is one of only a handful of exoplanets to be imaged directly.

The discovery could explain why, in 1981, β Pictoris's light dimmed erratically by up to 6% over 2 weeks, then brightened again. Another transit may have passed unnoticed 2 decades later, and the newfound planet appeared to be heading for yet another one in 2017 or 2018. Recent calculations suggest that the transit will be a near miss. But the planet's large "Hill sphere"—a zone of gravitational influence that may contain planetary rings, clouds of material, or newly formed moons—may yet reveal itself in dips in the light of β Pictoris.

To catch those dips, astronomers needed to monitor the star 24 hours a day over most of a year—too big a commitment for most observatories. So Lacour and his colleagues decided to build a small one of their own. In 3 years, with €1.5 million from the European Research Council, they built PicSat, a 5-centimeter space telescope in a satellite only slightly larger than a toaster. "It was risky and not everyone believed in it," Lacour says.

Kenworthy, along with Eric Mamajek of NASA's Jet Propulsion Laboratory in Pasadena, California, decided to observe from the ground. They built two washing machine-size robotic observatories, dubbed bRing, sited in Australia and South Africa.

A few existing telescopes have also joined the hunt: the Bright Target Explorer Constellation, five microsatellites designed to study luminous stars; and a 40-centimeter telescope that's part of the Antarctic Search for Transiting Exoplanets, which can watch β Pictoris continuously during the darkness of the southern winter. "We had to ensure we didn't drop the ball and had at least one scope on the star during the transit," Kenworthy says. The researchers also lined up agreements with larger telescopes to swing into action if they did see something.

PicSat, scheduled for launch in September 2017 but delayed by a launcher failure, finally reached orbit on 12 January. The Paris team is now checking its health. Sadly, it is joining the party toward its end; the transit of the Hill sphere is expected to end in February. "Maybe we won't see anything. We knew from the start it was risky," Lacour says.

Kenworthy remains optimistic. "We're not discounting anything," he says. Even a null result will imply that just 24 million years after its birth, the baby planet has already cleared out its Hill sphere. And once the predicted transit of β Pictoris b is over, the astronomers will keep watching the star and its planet nursery, hoping to see something else, like the fleeting transit of a smaller baby planet. ■

DIAGNOSTICS

'Liquid biopsy' for cancer promises early detection

Combining DNA and protein markers brings researchers closer to a universal cancer screening test

By Jocelyn Kaiser

A team of researchers has taken a major step toward one of the hottest goals in cancer research: a blood test that can detect tumors early. Their new test, which examines cancer-related DNA and proteins in the blood, yielded a positive result about 70% of the time across eight common cancer types in more than 1000 patients whose tumors had not yet spread—among the best performances yet for a universal cancer blood test. It also narrowed down the form of cancer, which previously published pancreatic blood tests have not.

The work, reported online today in *Science*, could one day lead to a tool for routinely screening people and catching tumors before they cause symptoms, when chances are best for a cure. Other groups, among them startups with more than \$1 billion in funding, are already pursuing that prospect. The new result could put the team, led by Nickolas Papadopoulos, Bert Vogelstein, and others at Johns Hopkins University in Baltimore, Maryland, among the front-runners.

"The clever part is to couple DNA with proteins," says cancer researcher Alberto Bardelli of the University of Turin in Italy, who was not involved in the work. The researchers have already begun a large study to see whether the test can pick up tumors in seemingly cancer-free women.

Genetic mutations drive the growth of cancer cells, and dying cells shed some of this mutated DNA into the blood. The Johns Hopkins group and others have shown that so-called liquid biopsies of blood-borne tumor DNA can reveal, for example, whether a patient's cancer should respond to a specific drug. But detecting the scant DNA released by early stage tumors is still challenging. Companies such as the \$1 billion Grail, launched in 2016 by sequencing giant Illumina, are using a big data approach, sequencing hundreds of genes in thousands of cancer patients' blood in search of a definitive set of DNA markers.

The Johns Hopkins researchers and collaborators found that gains in the detection rate tailed off when they added more genes to their test. They decided to sequence parts of just 16 genes often mutated in different types of cancer. They then added eight known protein biomarkers characteristic of specific kinds of cancer. This bumped up sensitivity and allowed the team to home in on the tissue type of the tumor.

In blood samples from 1005 patients with eight types of tumors that had evidently not yet metastasized, the test detected between 33% and 98% of cases, depending on the tumor type (see graph, below). The sensitivity was 69% or higher for ovarian, liver, stomach, pancreatic, and esophageal cancers—all types that are difficult to detect early.

The test rarely found cancer that wasn't there. Only seven of 812, or less than 1%, of healthy controls tested positive. And the test, called CancerSEEK, narrowed the origin of the cancer to two possible sites in about 80% of patients. The team, which is applying for patents on CancerSEEK, estimates the cost at less than \$500 per sample. "That's a very attractive number," says molecular pathologist Anirban Maitra of the MD Anderson Cancer Center in Houston, Texas, because it is in the range of other cancer screening tests such as colonoscopy.

Maitra and others point to caveats, however. One is that the cancer-related proteins used by the test reflect tissue damage and can also appear in people with inflammatory diseases such as arthritis. That means the 1% false positive rate will likely be higher in less healthy populations, notes proteomics researcher Lance Liotta of George Mason University in Manassas, Virginia. What's more, the 1005 patients already had cancer symptoms; CancerSEEK probably won't work as well in asymptomatic patients whose smaller tumors may shed less DNA. In fact, the test picked up only 43% of very early, stage 1 cancers. "We're still not there yet," Bardelli says.

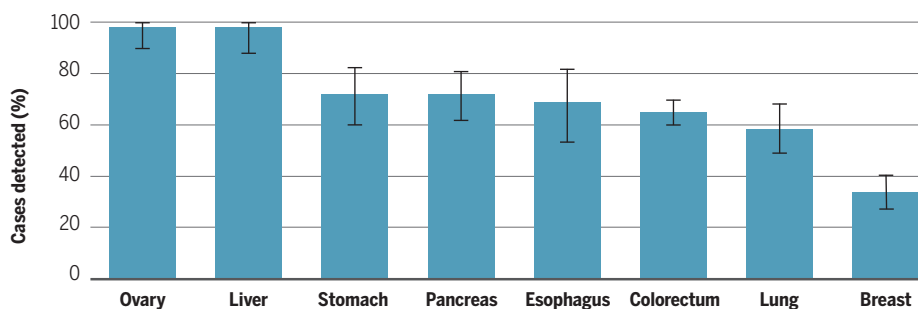
The Johns Hopkins team thinks CancerSEEK is ready for testing as a screening tool. "A test does not have to be perfect to be useful," Papadopoulos says. In collaboration with Johns Hopkins, the Geisinger Health System in Pennsylvania has already begun to use CancerSEEK on blood samples from female volunteers between ages 65 and 75 who have never had cancer. The planned \$50 million, 5-year study of up to 50,000 women is being funded by a private philanthropic group, The Marcus Foundation.

For those who test positive twice, the next step will be imaging to find the tumor. But that will bring up questions raised by other screening tests. Will the test pick up small tumors that would never grow large enough to cause problems yet will be treated anyway, at unnecessary cost, risk, and anxiety to the patient? Papadopoulos thinks the problem is manageable because an expert team will assess each case. "The issue is not overdiagnosis, but overtreatment," he says.

Still, others working on liquid biopsies say that it will take time to figure out whether widespread screening of healthy people with a universal blood test can reduce cancer deaths without doing harm. "If people expect to suddenly catch all cancers, they'll be disappointed," says cancer researcher Nitzan Rosenfeld of the University of Cambridge in the United Kingdom. "This is exciting progress," he says. "But evaluating it in the real world will be a long process." ■

A screening scorecard

A new cancer blood test worked better for some types than others, and caught only 43% of stage 1 cancers. (Error bars represent 95% confidence intervals.)



EVOLUTION

Tamed immune reaction aids pregnancy

Evolutionary studies show how dialing back inflammation allows embryo implantation

By **Elizabeth Pennisi**,
in San Francisco, California

The riskiest moment in any human pregnancy is arguably when the fertilized egg attaches to the womb wall and tries to establish a lifeline between embryo and mother. About half of in vitro pregnancies fail during this implantation stage, and many natural pregnancies end then as well. Now, researchers comparing pregnancy in opossums and several other mammals have shown how precise control of an immune process, inflammation, is critical to success or failure.

In work reported here this month at the annual meeting of the Society for Integrative

rejection of the embryo until it's ready to be born, Arun Chavan, a Yale graduate student in Wagner's lab, told the meeting.

Beyond solving a key mystery about pregnancy, the work could also point to treatments for infertility and miscarriage, says Tom Stewart, an evolutionary developmental biologist at the University of Chicago in Illinois. "The more we understand about pregnancy in other species, the more likely it is that we can treat medical issues that arise during human pregnancy."

Researchers have always puzzled over why the mother allows an embryo, which is basically a parasite, to settle in and grow. Yet implantation "was a critical first step in evolving pregnancy as humans experience it," says Julia Bowsher, an integrative biologist at North Dakota State University in Fargo.

This seeming paradox is even more perplexing because although a mother's inflammatory reaction to this "parasite" is the biggest threat to pregnancy, it also seems necessary for the pregnancy to be successful, Wagner, Chavan, and Yale postdoc Oliver Griffith pointed out last year. A woman's chance of implantation actually increases if her uterus has suffered mild trauma, for example, from a uterine biopsy as part of an in vitro fertilization (IVF) procedure. Studies have shown that the IVF embryo is more likely to settle in, particularly at the biopsy site. Furthermore, an immune "rejection" response helps create the contractions necessary for a baby's birth. Yet in between implantation and birth, the immune system is

held in check, allowing the fetus to thrive.

To understand the evolutionary basis for this interlude, Griffith recently led a study of gene activity in a marsupial, the gray short-tailed opossum (*Monodelphis domestica*). Marsupials have very short pregnancies. Early opossum embryos develop for about 12 days, enclosed as shelled eggs in the womb. They then shed their shells and try to attach to the uterine wall, activating placenta-promoting genes. But after about 2 days, the mother's immune system "re-

jects" the embryos, causing the birth of a litter that is still very immature compared with newborn placental mammals.

Griffith sampled opossum gene activity before pregnancy, during the egg-shell stage, and after implantation. The analysis revealed the array of immune system signaling molecules and steroid hormones taking part in the immune attack on the embryo. The gene activity also pointed to a role for immune cells such as neutrophils, which launch a full-fledged inflammatory reaction that includes molecules that stimulate contractions of the uterus. The timing and makeup of this response largely mirror what is seen in implantation in placental mammals, indicating that the process evolved in the common ancestor of placental and marsupial mammals, Griffith and colleagues reported in the 26 July 2017 issue of the *Proceedings of the National Academy of Sciences*.

But later in evolution, placental mammals apparently dialed back that inflammation to allow extended gestation. To find out how, Chavan compared implantation in the opossum with that in a range of placental mammals: rabbits, armadillos, and hyraxes, a 3-kilogram rodentlike mammal that's closely related to elephants. Based on studies of gene activity and immune cells, he found that these mammals have "domesticated" implantation's inflammatory response. At the implantation site, blood vessels proliferate in the uterine wall—the same hallmark of inflammation seen in the opossum—but the signaling molecule IL-17, which recruits neutrophils, is missing, Chavan reported at the meeting.

Specialized cells called decidual cells seem to be responsible, he found. These cells form in the uterine lining early in pregnancy and, in many placental mammals, disappear right after implantation. Chavan wondered whether these cells might have evolved to switch the inflammatory response into low gear. Supporting that notion, he found in tissue studies that secretions of those cells could keep immune cells from making IL-17.

"If that switch doesn't happen, there are miscarriages," says Gil Mor, a reproductive immunologist at Yale who was not involved with the work. "Understanding the evolution of decidual cells will be extremely helpful to those of us studying the nitty-gritty" of pregnancy. ■



An unrestrained inflammatory response triggers the birth of opossum young early in their development.

and Comparative Biology, a Yale University team led by evolutionary developmental biologist Günter Wagner concluded that human and other so-called placental mammals have tweaked an ancient inflammatory process to enable embryos to implant and persist in the womb. Placental mammals—named for the mass of tissue in the uterus that serves as the interface between mother and fetus—have specialized uterine cells that suppress the release of a key immune-stimulating molecule. This suppression may help delay the

Tensions flare over electric fishing in European waters

European Parliament calls for total ban of a technique that saves fuel and reduces damage to marine life

By Erik Stokstad

Bottom trawling is one of the most destructive types of fishing, decried for churning up massive swaths of sea bed and leaving dead sea urchins, mollusks, and other creatures in its wake. In the North Sea, Dutch fishing vessels are substituting a subtler technique for this brute-force method: using short bursts of electricity to get flatfish out of the sediment and into nets. But they are stirring up just as much controversy.

Dutch fishing companies say pulse trawling is less damaging to marine ecosystems and saves energy. But fishing groups in other EU countries are increasingly angry about competition from the Dutch pulse trawlers. And a coalition of environmental organizations worries about harm to non-target marine life. Other nongovernmental organizations, including Greenpeace Netherlands, say pulse trawling has promise and that ending it now would penalize the fishing industry for innovating.

"A train of emotion is now going full speed," says Marloes Kraan, an anthropologist at Wageningen Marine Research in IJmuiden, the Netherlands. And it appears likely to accelerate: On 16 January, the European Parliament voted to ban the technique as the first step in negotiations with the European Commission and member states over a large package of fisheries reforms. A ban, or even a major reduction in pulse trawling, would be a huge blow to the Dutch fishing industry.

Most bottom trawlers drag a net, held open by a wide metal beam, across the bottom to catch shrimp or fish. Trawlers targeting flatfish, such as sole or plaice, also use dangling iron chains to scare them out of the sediment. The beam and chains disturb or kill many bottom-dwelling organisms, the nets catch unwanted species, and all the tugging requires a lot of diesel.

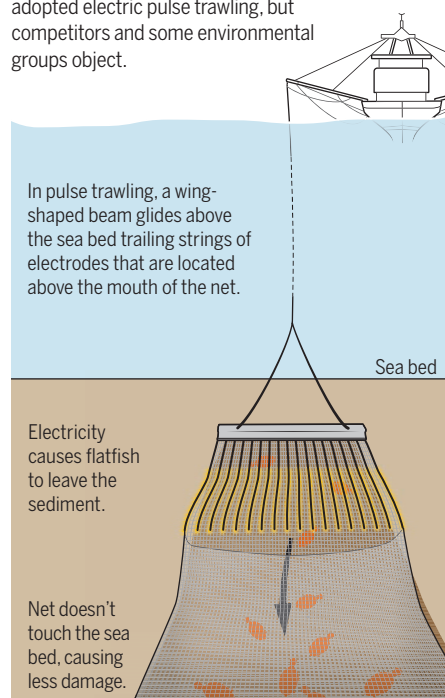
Pulse trawlers, by contrast, barely touch the bottom because they use bursts of low-voltage electricity to catch flatfish, particularly Dover sole (see graphic, right). After the current briefly cramps their muscles, they try to flee, and many end up in the net.

Because sole are more susceptible to electricity than other species, pulse trawling reduces bycatch. And the gear is lighter and can be towed slower, so the boats burn half as much fuel. "We catch with a lesser environmental impact and greater economic returns," says Pim Visser of VisNed, a trawling trade group in Urk, the Netherlands. He credits the gear with saving many fishing companies from bankruptcy.

Encouraged by initial studies, the Dutch government in 2006 successfully lobbied the European Commission to allow 5% of each country's fleet to use pulse trawling, exempting them from the European Union's 1988 general ban on electrical fishing. By 2009, Dutch companies had embraced the opportunity. As demand grew, they received additional licenses for reducing bycatch or for research, with the condition that they provide detailed data on their catches. Now, 75 vessels—about 28% of Dutch

A charged approach

Many Dutch fishing vessels have adopted electric pulse trawling, but competitors and some environmental groups object.



trawlers—use pulse gear. Fishing companies outside the Netherlands fish for sole, too, but don't specialize in it; as a result, few have invested in the expensive technology.

BLOOM Association, an environmental group in Paris, argues that the research and bycatch licenses are illegal and a guise for commercial fishing, and that pulse trawling puts small-scale fishing at an even bigger disadvantage than conventional trawling does. BLOOM advocates catching flatfish with gillnets, stationary curtains of netting that have a much lower bycatch rate than either kind of trawling and do less damage to the sea floor. "There shouldn't be any use of electric current," says BLOOM Director Claire Nouvian. "We've got enough evidence to know this is nonsense."

Scientists have so far found little evidence that the electrical currents cause serious harm. Last year, a working group with the International Council for the Exploration of the Sea (ICES) highlighted harm to large cod and whiting as the only known irreversible effect. Although not many cod are accidentally caught by pulse trawlers, about 10% of them suffer vertebral fractures and hemorrhages when their muscles over-contraction from the shocks. Initial laboratory research on other organisms has not shown lasting, serious effects, but the ICES group says questions remain, for instance about the effects on sharks and rays.

Nevertheless, "We know enough to continue with pulse trawling in the present context," says Adriaan Rijnsdorp, a fisheries biologist at Wageningen Marine Research and a co-chair of the ICES working group. But he says a decision on the future of pulse trawling should wait until 2019, when a 4-year, EU-funded research program on ecological impacts, which he coordinates, is due to wrap up.

Any decision will have to be agreed on by the European Parliament, the commission, and member states, in this case represented by their fisheries ministers. The commission has proposed removing the cap on licenses in the southern North Sea, where pulse trawling now occurs; other areas could follow after further studies. The ministers, by contrast, would de facto remove licenses beyond the 5% limit of a country's fleet, which would force most Dutch vessels to give up pulse trawling. A compromise in which the technique is greatly curtailed is the most likely outcome, says Irene Kingma, director of the Dutch Elasmobranch Society in Amsterdam, which promotes the study and conservation of sharks and rays. "There might be carnage within the Dutch fishing sector," Kingma says. "And if they change back to beam trawling, we have all the environmental problems from that." ■



SCIENTIFIC COMMUNITY

Rochester roiled by fallout from sexual harassment case

Report supports university's handling of explosive charges involving linguist T. Florian Jaeger, but president bows out

By **Meredith Wadman**

After months of turmoil that has riven the University of Rochester in New York and its highly esteemed cognitive science department, the controversy took a pivotal turn last week. An outside attorney hired by the university released a 207-page investigation that largely approved of the university's handling of explosive sexual harassment claims, which have sparked formal complaints, campus protests, and a boycott effort. The report also concluded that no individual woman had experiences with professor T. Florian Jaeger that met what it described as the "demanding" legal standard for sexual harassment, which is that it is severe or pervasive enough to create a hostile environment.

The turmoil is far from over. The university still faces a federal lawsuit, filed last month by seven current or former faculty, a former postdoc, and a former graduate student. They allege that the university allowed Jaeger to create a hostile environment and retaliated against them when they complained. "This is what it looks like when an institution protects a sexual harasser," says plaintiff Steven Piantadosi, an assistant professor in the De-

partment of Brain and Cognitive Sciences (BCS). Adding to the uncertainty, just as the report was released on 11 January, university head Joel Seligman, announced he was leaving, effective 28 February. In a statement, Seligman, who is a defendant in the lawsuit, said, "It is clear to me that the best interests of the University are best served with new leadership, and a fresh perspective to focus on healing our campus." As for Jaeger, he declined through his lawyer to say whether he has discussed his employment status with the university.

The report was written by Mary Jo White, a partner at the law firm Debevoise & Plimpton in New York City who is a former U.S. attorney and former chair of the Securities and Exchange Commission. The university's Board of Trustees hired her in September 2017 to investigate complaints about Jaeger, as well as the university's previous probes of the case and the claims of retaliation. (The university paid her firm \$4.5 million for its work.) She concluded that although the behavior of Jaeger, a professor in the BCS department, was "offensive," "inappropriate," and "disturbing," he violated neither university policies then in place nor federal or state sexual harassment laws.

Last fall, campus protesters targeted University of Rochester President Joel Seligman.

The report concedes that Jaeger sent an unwanted photo of his penis to a student; made inappropriate sexual comments in social and academic settings, including commenting on a student's attractiveness in front of peers and faculty; flirted with students; served on a thesis committee for a student he had had sex with; submitted a letter of recommendation for an undergraduate while in a sexual relationship with her; and engaged in multiple consensual relationships with current, former, or prospective students. "A number of female graduate students from that time period told us that, as a result of Jaeger's reputation or behavior, they made a conscious decision to avoid him and the educational opportunities he offered, which we found to be very troubling," the report states.

But the report finds that Jaeger did not sexually harass any individual woman or breach university policies in place at the time. Jaeger arrived at the university in 2007, and the complaints center on his behavior from that time until 2013. The report found that he engaged in no sexual relationships with current or former students after 2011. The university did not bar intimate relationships between faculty and undergraduates or between faculty and graduate students over whom they have authority until 2014.

White's team at Debevoise & Plimpton, where her group is deployed to help institutions in crisis, also found that the school's investigations of Jaeger's behavior were impartial and that no retaliation occurred. It called the accounts by Jaeger's accusers "exaggerated and misleading in many respects." It adds: "We think that the University acted in good faith and appropriately under its then-current policies and that the steps it took in an effort to navigate an unusually difficult situation were reasonable."

However, the report does conclude that the school handled some matters poorly, such as sharing complaining faculty members' private emails with the BCS chair and promoting Jaeger to full professor in 2016, while he was under investigation. It's policies and procedures for addressing sexual harassment complaints "can and should be enhanced," the report states.

Jaeger responded in a statement: "This report does not exonerate me, but neither does it give merit to many of the worst accusations made against me. ... I appreciate their commitment to seeking out the truth."

Some of his colleagues agree. "The White report largely hits the mark and affirms what the majority of my faculty colleagues have been thinking and experienced," Ralf Haefner, an assistant professor in the BCS

department since 2014, wrote in an email to *Science*. “While some of Jaeger’s behavior as a junior faculty member was inappropriate, he clearly wasn’t a sexual predator.”

The complainants, who refused to be interviewed by White’s firm because of the ongoing litigation, strongly disagreed. “The thrust of their report is to admit that many bad things happened at UR—but miraculously ... no legal liability attaches to the university,” the plaintiffs’ lawyer, Ann Olivarius of McAllister Olivarius in Maidenhead, U.K., said at an 11 January news conference. “In fact, there is substantial case law the report ignores that strongly supports the idea that the university is absolutely liable for the hostile environment created by Jaeger’s actions.”

The report does not evaluate the effect of Jaeger’s behavior in the aggregate, which the court might do, notes Alexandra Tracy-Ramirez, a lawyer at HopkinsWay in Phoenix. The plaintiffs also dispute details of White’s account. On 13 January, they wrote to the faculty Senate Executive Committee, saying the report suppresses and misrepresents evidence. “In court, our audio recordings ... will establish blatant factual inaccuracies in Ms. White’s narrative,” they add. Their suit names the university, Seligman, and Provost Robert Clark. In addition to retaliation and defamation, it charges that the school allowed a hostile environment for three plaintiffs. University attorneys are expected to file their first response in early February.

The BCS department now faces the task of healing and rebuilding. Of the nine plaintiffs, seven have left or will soon leave the university. Two, Piantadosi and Celeste Kidd, assistant professors who are married to each other, are searching for positions. The BCS website lists 40 remaining faculty.

Recruiting students may be difficult, given an open letter signed late last year by 454 professors in cognitive neuroscience and other disciplines. The signatories vowed not to encourage students to attend the university because it had “abrogated” its duty to students “by supporting the predator and intimidating the victims.”

Few dispute that the department has been damaged by the case. “It was known as one of the top cognitive neuroscience departments in the country,” says cognitive neuroscientist Timothy Verstynen of Carnegie Mellon University in Pittsburgh, Pennsylvania, who signed the letter. “They really did take a hit by losing some strong key faculty.” The Board of Trustees said last week it will need time to digest the lengthy report before taking action. It also faces the task of replacing Seligman, whom it called a “brilliant, transformative leader.” ■

CRIMINAL JUSTICE

Are algorithms good judges?

People are as good as machines in predicting rearrest

By Catherine Matacic

Every day, judges across the United States face an important decision: Should they jail a defendant whose innocence or guilt has not yet been determined, or should they release that person back into the community, where he or she might commit a crime?

Increasingly, courts are turning to computer-based tools to help make those decisions, lured by the promise of complex algorithms that use an array of factors to spit out risk scores. But a new study suggests that at least one widely used algorithm produces risk assessments that are no better than those reached by people given just a few key pieces of information about a defendant.

The result, published this week in *Science Advances*, challenges widespread assumptions that algorithms are better at calculating risk than humans, researchers say. “A fancy model isn’t necessarily a better model,” says David Robinson, a legal scholar at Georgetown University in Washington, D.C.

Julia Dressel, a computer science major at Dartmouth College, got interested in risk scoring algorithms after reading a 2016 series by investigative reporters at ProPublica about one popular system, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), which is used in at least five states and can cost up to \$22,000 a year. After examining COMPAS scores for 10,000 defendants awaiting trial in Broward County in Florida, as well as their arrest records over the next 2 years, ProPublica concluded that the tool disproportionately classified black offenders as being at high risk of rearrest.

That finding concerned civil rights advocates, but Dressel went on to ask another question: Are humans or machines better at assessing the risk of rearrest? To find out, she randomly selected 1000 of the defendants and recorded seven pieces of information about each, including their age, sex, and number of previous arrests. She then recruited 400 people using the online crowdsourcing service Amazon Mechanical Turk. Each received profiles of 50 defendants and was asked to predict whether

they would be rearrested within 2 years, the same standard COMPAS uses. The human judges were right about 63% to 67% of the time, compared with about 65% of the time for COMPAS.

Dressel was surprised. So was Megan Stevenson, an economist and legal scholar at George Mason University in Arlington, Virginia, who calls the study the first to run a “horse race” between human and algorithm. She always assumed that algorithms were somewhat better, so the results left her “quite shocked.”

In a second experiment, Dressel and her adviser, Dartmouth computer scientist Hany Farid, explored whether a simpler algorithm could beat COMPAS’s, which is proprietary but described in technical literature. They created their own, ultimately settling on just two factors: age and number of prior convictions. Plugging that information into a simple formula yielded predictions that were about 67% accurate, roughly matching COMPAS.

Robinson says that result reflects something that has long been known in criminology: If you’re young, you’re risky.

Mathematician Tim Brennan, who created COMPAS in 1998 when working at Northpointe (now Equivant) in Canton, Ohio, says that far from undercutting COMPAS, the new study validates his approach. Seventy percent accuracy, he says, has long been considered the “speed limit” of such prediction systems, and the fact that humans did no better is encouraging.

But humans are no better than machines at eliminating bias, notes mathematician Cathy O’Neil, founder of the risk consulting and auditing firm O’Neil Risk Consulting & Algorithmic Auditing in New York City. Dressel’s study, for example, found that people were just as likely as COMPAS to overstate rearrest risks for black defendants and understate risks for white defendants. That’s troubling, given that similar algorithms are increasingly influencing not just court decisions, but also loan approvals, teacher evaluations, and even whether child abuse charges are investigated by the state. “People get awed by mathematical sophistication,” O’Neil says, “but it’s mostly a distraction.” ■

“A fancy model isn’t necessarily a better model.”

David Robinson,
Georgetown University

FEATURES



THE BELIEVER

How a Mormon lawyer transformed Mesoamerican archaeology—and ended up losing his faith

By **Lizzie Wade**, in San Cristóbal de las Casas, Mexico

Thomas Stuart Ferguson lay in his hammock, certain that he had found the promised land. It had been raining for 5 hours in his camp in tropical Mexico on this late January evening in 1948, and his three campmates had long since drifted off to sleep. But Ferguson was vibrating with excitement. Eager to tell someone what he had seen, he dashed through the downpour to retrieve paper from his supply bag. Ensnared in his hammock's cocoon of mosquito netting, he clicked on his flashlight and began to write a letter home.

"We have discovered a very great city here in the heart of 'Bountiful' land," Ferguson wrote. According to the Book of Mormon, Bountiful was one of the first areas settled by the Nephites, ancient people who supposedly sailed from Israel to the Americas around 600 B.C.E. Centuries later, according to the

scripture, Jesus appeared to the Nephites in the same region after his resurrection. Mormons like Ferguson were certain that these events had happened in the ancient Americas, but debates raged over exactly how their sacred lands mapped onto real-world geography. The Book of Mormon gave only scattered clues, speaking of a narrow isthmus, a river called Sidon, and lands to the north and south occupied by the Nephites and their enemies, the Lamanites.

After years of studying maps, Mormon scripture, and Spanish chronicles, Ferguson had concluded that the Book of Mormon took place around the Isthmus of Tehuantepec, the narrowest part of Mexico (see map, p. 266). He had come to the jungles of Campeche, northeast of the isthmus, to find proof.

As the group's local guide hacked a path through the undergrowth with his machete, that proof seemed to materialize before Ferguson's eyes. "We have explored four days

and have found eight pyramids and many lesser structures and there are more at every turn," he wrote of the ruins he and his companions found on the western shore of Laguna de Términos. "Hundreds and possibly several thousand people must have lived here anciently. This site has never been explored before."

Ferguson, a lawyer by training, did go on to open an important new window on Mesoamerica's past. His quest eventually spurred expeditions that transformed Mesoamerican archaeology by unearthing traces of the region's earliest complex societies and exploring an unstudied area that turned out to be a crucial cultural crossroads. Even today, the institute he founded hums with research. But proof of Mormon beliefs eluded him. His mission led him further and further from his faith, eventually sapping him of religious conviction entirely. Ferguson placed his faith in the hands of science, not



Stela 5 from Izapa in Mexico—an early site first extensively excavated by New World Archaeological Foundation archaeologists—shows a mythical tree; some Mormons believe it reflects a prophetic dream from the Book of Mormon.

realizing they were the lion's jaws.

But that night, lying in his hammock listening to the rain and the occasional roar of a jaguar in the distance, Ferguson felt surer than ever that Mesoamerican civilizations had been founded by migrants from the Near East, just as his religion had taught him. Now, he thought, how would he convince the rest of the world?

THE CHURCH OF JESUS CHRIST of Latter-day Saints (LDS) doesn't take an official position on where the events in the Book of Mormon occurred. But the faithful have been trying to figure it out practically since 1830, when church founder Joseph Smith published what he said was a divinely inspired account of the ancient Americas. Smith said an angel had led him to buried ancient golden plates, which he dug up and translated into the Book of Mormon. Smith's account of buried wonders was one of many in the United

States at the time. As white settlers moved west, they encountered mounds filled with skeletons and artifacts, including beautiful pottery and ornaments. Newspapers, including those in Smith's hometown of Palmyra, New York, buzzed with speculation about who the "mound builders" were and how they came by their refined culture. Many settlers, blinded by racism, concluded that the mound builders—now known to be indigenous farming societies—were a lost people who had been exterminated by the violent ancestors of Native Americans. The Book of Mormon, with its saga of righteous, white Nephites and wicked, dark-skinned Lamanites, echoed these ideas.

The Book of Mormon also spoke of sprawling ancient cities, none of which had been identified in the United States. So in the 1840s, Mormons, including Smith himself, took notice of a U.S. explorer's best-selling accounts of visits to the ruins of

Mayan cities in Mexico and Guatemala. In 1842, as editor of a Mormon newspaper, Smith published excerpts from a book about the ruins of the Mayan city of Palenque in Mexico, with the commentary: "Even the most credulous cannot doubt ... these wonderful ruins of Palenque are among the mighty works of the Nephites—and the mystery is solved."

But non-Mormons continued to doubt, and church authorities gradually retreated from explicit statements about Book of Mormon locations. By the 1930s, when Ferguson learned about Mesoamerican civilizations as an undergraduate at the University of California (UC), Berkeley, the matter had been largely ceded to amateurs who pored over maps and the Book of Mormon looking for correspondences.

Ferguson wasn't impressed by their efforts. "The interested and inquiring mind of the modern investigator is not satisfied with explanations which are vague, unsound, and illogical," he wrote in an article in a church magazine in 1941. By then he was a law student at UC Berkeley and intrigued by the idea of scientifically testing Smith's revelation. In a later letter, he wrote, "It is the only Church on the face of the earth which can be subjected to this kind of investigation and checking." And in another, to the LDS leadership, he declared, "The Book of Mormon is either fake or fact. If fake, the [ancient] cities

described in it are non-existent. If fact—as we know it to be—the cities will be there."

TALL AND HANDSOME, with a lawyer's practiced authority, Ferguson trusted that the tools of science could persuade the world of the truth of the Book of Mormon. Soon after he finished college, he began searching for clues in colonial documents that recorded some of Latin America's indigenous traditions. One, written around 1554 by a group of K'iche' Mayan villagers in the Guatemala highlands, stated that their ancestors—"sons of Abraham and Jacob"—had sailed across a sea to reach their homeland. The K'iche' were defeated by Spanish conquistadors in 1524, and the biblical references were likely the product of contact with Catholic priests, who enthusiastically converted allies and former foes alike.

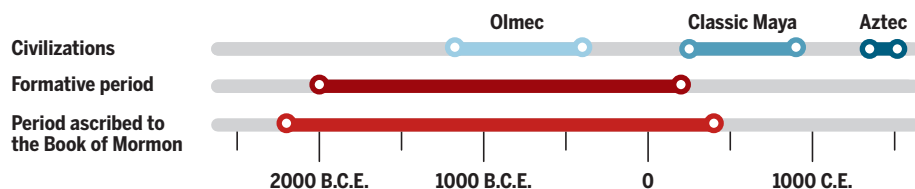
But Ferguson, who had grown up in a Mormon family in Idaho, eagerly took such

Ferguson's holy land

His quest spurred digs in central and coastal Chiapas in Mexico, previously overlooked in favor of Olmec and Mayan lands.



Mesoamerican timeline



syncretism as proof that Israelites had once settled in the Americas. He was also taken by the myth of Quetzalcóatl, the feathered serpent deity that some colonial priests described as a bearded white man. Ferguson concluded that he was Jesus, appearing in Bountiful after his resurrection just as the Book of Mormon recorded. His library research spurred his first hunt for archaeological evidence, in Campeche in 1948.

Ferguson realized, however, that colonial sources represented circumstantial evidence at best. Nor was it enough to find ruins of past civilizations in more or less the right location, as he had done in Campeche. To persuade and convert outsiders—a priority for Mormons—he sought objects mentioned in the Book of Mormon that archaeologists hadn't found in Mesoamerica: horses, wheeled chariots, steel swords, and, most important, Hebrew or Egyptian script. "The final test of our views of Book of Mormon geography will be archaeological work in the ground itself," Ferguson wrote in 1951 to his friend J. Willard Marriott, the wealthy founder of the Marriott hospitality chain and a powerful figure in the church.

Ferguson's idea that Mesoamerican societies were seeded by Western ones is widely recognized as racist today. But it fit right into the archaeological thinking of the time, when Mesoamerican archaeologists were consumed by the question of whether civilizations had evolved independently in the Americas or had roots elsewhere. "In the 1940s and 1950s, these were the questions everyone was investigating," says Robert Rosenswig, an archaeologist at the State University of New York (SUNY) in Albany.

Ferguson never received a formal education in archaeology. He practiced law to support his growing family—he eventually had five children—as well as his research. But in 1951, he recruited leading archaeologists to explore the origin of Mesoamerican civilization as part of a new institution, the New World Archaeological Foundation (NWF). First on board was renowned researcher Alfred Kidder of Harvard University and the Carnegie Institution for Science in Washington, D.C. Kidder thought Mesoamerican civilizations had developed independently, but he and Ferguson had met at a museum in Guatemala City in 1946 and struck up a correspondence.

Kidder "is recognized as the best [Mesoamerican] archaeologist of the 20th century," says archaeologist John Clark of Brigham Young University (BYU) in Provo, Utah, who directed NWF from 1987 to 2009. To get Kidder on the project, Clark says, "There's no question that Ferguson had to be some charismatic guy." Also recruited was Gordon Ekholm, an anthropologist at the American Museum of Natural History in New York City, who thought that Mesoamerican civilizations had their roots in advanced Asian cultures.

Their timing was good. Radiocarbon dating had just been invented, and Ferguson immediately recognized its potential for tracing the origins of Mesoamerican cultures. "This is the greatest development since the beginning of archaeology," he wrote to LDS leadership. "I am of the personal opinion that the Lord inspired [radiocarbon dating] that it might be used effectively in connection with the Book of Mormon."

Yet the first years of NWF were a desperate scramble for money. Ferguson contributed thousands himself and raised funds from wealthy Mormons and the audiences of his lectures about Book of Mormon geography. In 1952, NWF managed to send a handful of U.S. and Mexican archaeologists to survey the drainage basin of the Grijalva River in Tabasco and Chiapas, which Ferguson believed to be the Book of Mormon's River Sidon.

By this point, Ferguson had become more

"The Book of Mormon is either fake or fact. If fake, the [ancient] cities described in it are non-existent. If fact—as we know it to be—the cities will be there."

Thomas Stuart Ferguson,
in a 1958 letter

discerning about time periods than he had been in the jungles of Campeche. The ruins he found there were likely Classic or post-Classic Mayan, from between 250 C.E. and the Spanish conquest—much too late to be Mesoamerica's earliest civilization or the period mentioned in the Book of Mormon, believed to be about 2200 B.C.E. to 400 C.E. "We'll never solve pre-Maya origins by digging up more Mayas," Ferguson wrote to Kidder in April 1953. They needed Formative period sites, dating from about 2000 B.C.E. to 200 C.E., roughly matching the dates associated with the Book of Mormon.

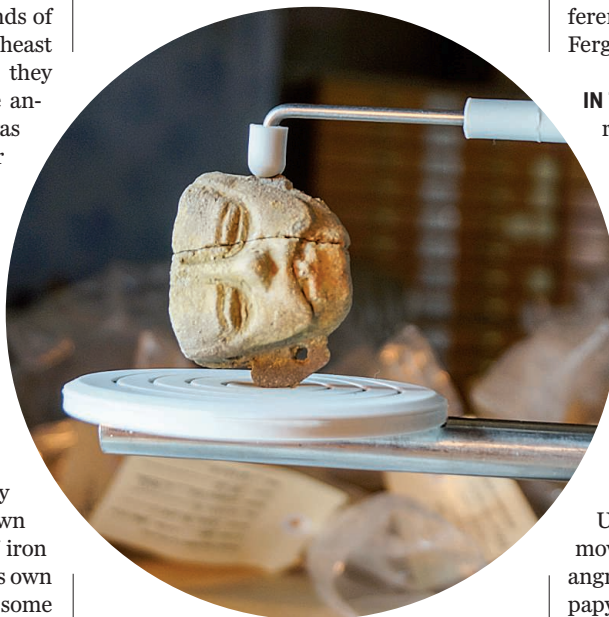
In May 1953, Ferguson arrived in Chiapas to lend a hand. “He was rather alarmed that we hadn’t found anything notable, because he felt he had to have something pretty spectacular to go and get more money for another year,” recalls John Sorenson, then a master’s student in archaeology at BYU (and a Mormon). To jump-start the search, Ferguson chartered a small plane, and he and Sorenson flew over the lush lowlands of central Chiapas. Fifteen kilometers southeast of the state capital, Tuxtla Gutiérrez, they spotted the mounds and plazas of the ancient site of Chiapa de Corzo—which was then unknown to archaeologists. Later NWAf excavations dated the city to the Formative period.

Back on the ground, Ferguson and Sorenson set out by jeep for a 10-day survey to see what else they could find. “We’d go from site to site, town to town, asking ‘Are there any ruins around here?’” says Sorenson, who went on to receive a Ph.D. in anthropology from UC Los Angeles (UCLA) and is now a professor emeritus at BYU. Ferguson also asked locals whether they had found figurines of horses—unknown in ancient Mesoamerica—or sources of iron ore, which Sorenson found naïve. But his own archaeological training paid off, and at some sites he was able to identify the polished, monochrome pottery and hand-sculpted, irregular human figurines of the Formative period, so different from the intricate but standardized figurines the Classic Maya had made from molds. In all, Sorenson and Ferguson surveyed 22 sites on that journey and collected an astounding number of Formative artifacts. “In my humble opinion there is little or no question about it—they are Nephite making,” Ferguson wrote to his church funders.

In 1954, LDS authorities granted NWAf \$250,000 for 5 years of work. Intensive excavations at Chiapa de Corzo uncovered stone pyramids and tombs, and a wealth of pottery that impressed University of Pennsylvania anthropologist John Alden Mason, then working with NWAf. “Since pre-Classic pottery is not very common anywhere, and that of this region is entirely new, it is of course a very great scientific contribution,” Mason wrote to Ferguson. Eventually, archaeologists reported that the site was settled around 1200 B.C.E., likely by people connected to the Olmec, an early civilization that dominated the gulf coast of Mexico from 1200 B.C.E. to 400 B.C.E., centuries before the Classic Maya arose.

Then, in the early 1960s, NWAf archaeologists became the first to extensively excavate at Izapa, near the Chiapas coast and the Guatemalan border. They were drawn to

the site in part because of a monument that apparently depicts a myth involving a tree; Ferguson’s friend and founder of BYU’s archaeology department, M. Wells Jakeman, argued that the carving shows visions received in a dream by the Mormon prophet Lehi. NWAf archaeologists, some of whom were Mormon, later soundly rebuffed that



A ritual figurine from the site of Los Horcones is scanned at New World Archaeological Foundation headquarters.

interpretation. But Izapa turned out to be a key site in the Soconusco, the Pacific coast region from which every Mesoamerican political power, from the Olmec in 1200 B.C.E. to the Aztec empire in the early 1500s C.E., sourced key luxury goods such as cacao and quetzal feathers. NWAf spearheaded excavations throughout this region. Pottery finds and dates from Izapa and elsewhere formed the basis of the ceramic chronologies for the Formative period that are still used by every archaeologist working in central and coastal Chiapas today.

“They were working in a part of Mesoamerica that was really unknown,” says Michael Coe, an influential Mesoamerican archaeologist and professor emeritus at Yale University who, at the time, was surveying Formative sites just over the border in Guatemala. “NWAf put it on the map.”

But even as NWAf grew in scientific stature, and was finally assured continued existence when BYU took it over in 1961, Ferguson was quietly becoming frustrated. The smoking gun he had been certain he would find—Egyptian or Hebrew script—proved elusive. He once had promised that archaeological evidence for the Book of Mormon would be

found within 10 years of NWAf starting excavations. But in 1966 he wrote, “My number one goal of establishing that Christ appeared in Mexico following the crucifixion will never be achieved until significant ancient manuscript discoveries are made. I hope it happens during our lifetimes.”

When an ancient manuscript discovery did come, however, it was from a different quarter of the world—and it shook Ferguson’s faith to its core.

IN THE SUMMER OF 1835, Joseph Smith had received a curious visitor in Kirtland, Ohio, then the headquarters of his burgeoning LDS church: a traveling showman, with four Egyptian mummies and some hieroglyphic texts in tow. The church bought the mummies and texts, and Smith said he translated the hieroglyphics, resulting in the Book of Abraham, which lays out Smith’s cosmic vision of the afterlife. (Although Egyptian hieroglyphics had been deciphered in France in 1822 with the help of the Rosetta Stone, the news had barely reached U.S. shores.) As Smith and his followers moved around the Midwest, often fleeing angry mobs, they carried the mummies and papyri with them. After Smith’s death at the hands of one of those mobs in Nauvoo, Illinois, they were sold by his family.

The fate of the mummies remains a mystery. But in 1966, a University of Utah professor examining artifacts at the Metropolitan Museum of Art in New York City came across 11 Egyptian papyri with an 1856 certificate of sale signed by Smith’s widow, Emma. The professor realized he was looking at the Book of Abraham papyri, and the documents were returned to the Mormon church.

Ferguson learned the news from a front-page article in the newspaper *Deseret News* on 27 November 1967. Within days, he wrote to a friend in the church leadership, begging to know whether the papyri would be studied. Hearing that no studies were planned, Ferguson, as ever, took matters into his own hands. He received photos of the documents from the church and hired Egyptologists at UC Berkeley to translate them. He told the scholars nothing about the religious significance of the papyri. “He was conducting a clearly blind test,” Clark says.

The results started coming in 6 weeks later. “I believe that all of these are spells from the Egyptian Book of the Dead,” UC Berkeley Egyptologist Leonard Lesko wrote to Ferguson. Three other scholars independently gave Ferguson the same result: The texts were authentic ancient Egyptian, but represented one of the most common documents in that culture.

After decades of stressing the importance of the scientific method and using it to shore up his own faith, Ferguson now found himself at its mercy. “I must conclude that Joseph Smith had not the remotest skill in things Egyptian-hieroglyphics,” he wrote to a fellow doubting Mormon in 1971. What’s more, he wrote to another, “Right now I am inclined to think that all of those who claim to be ‘prophets,’ including Moses, were without a means of communication with deity.”

This doubt ultimately spread to Ferguson’s archaeological quest. In 1975, he submitted a paper to a symposium about Book of Mormon geography outlining the failure of archaeologists to find Old World plants, animals, metals, and scripts in Mesoamerica. “The real implication of the paper,” he wrote in a letter the following year, “is that you can’t set Book of Mormon geography down anywhere—because it is fictional.”

Although open about his doubts in his private letters, Ferguson didn’t discuss his loss of faith with his family. He continued attending church, singing in the choir, and even giving blessings. “[Mormons] are so immersed in that culture ... [that] to lose your faith, it’s like you’re being expelled from Eden,” Coe says. “I felt sorry for him.”

Ferguson continued to visit Mexico and from time to time stopped by NWAf head-

de la Amada on the Chiapas coast, home to Mesoamerica’s first known ball court and elite residences. With NWAf support, Lesure has spent nearly 3 decades studying why mobile, egalitarian hunter-gatherers settled down here and created the oldest complex society in Mesoamerica around 1900 B.C.E., before even the Olmec rose to power.



At the New World Archaeological Foundation, Richard Lesure studies artifacts from Mesoamerica’s earliest complex society.

“I am of the personal opinion that the Lord inspired [radiocarbon dating] that it might be used effectively in connection with the Book of Mormon.”

Thomas Stuart Ferguson,
in a 1956 letter

quarters in Chiapas, where he spoke frankly with Clark in 1983. “He resented that he spent so much time trying to prove the Book of Mormon. He said it was a fraud,” remembers Clark, who is Mormon. The next month, Ferguson died of a heart attack while playing tennis. He was 67.

ON A RECENT AFTERNOON AT NWAf headquarters here, scholars wander among buildings, sheltered patios, and a courtyard brimming with flowers and citrus trees. UCLA archaeologist Richard Lesure sorts through ceramics he excavated 27 years ago at Paso

Upstairs, Claudia García-Des Lauriers, an archaeologist at California State Polytechnic University in Pomona, watches as an undergraduate student carefully positions an opossum-shaped ceramic whistle in the thin red laser beams of a 3D scanner. The researchers are creating a digital version of the ritual object, which García-Des Lauriers discovered at the Classic period site of Los Horcones on the Chiapas coast. Meanwhile, in the backyard, Clark leads an impromptu flint knapping lesson, using obsidian nodules strewn about the lawn.

“It’s such a stimulating place to work,” says Janine Gasco, an archaeologist at California State University in Dominguez Hills, who began working with NWAf in 1978. “It’s been a force in my life.”

In the years after Ferguson drifted away from the church and the foundation, NWAf continued to lead excavations, fund graduate students, publish an impressive amount of raw data, and store archaeological collections. Thanks to its work, a region that once seemed an archaeological backwater compared with the nearby Classic Mayan heartland in the Yucatán, Guatemala, and Belize has been revealed as the birthplace of

Mesoamerican civilization and an economic and cultural hot spot, where people from all over the region crossed paths. “We wouldn’t know anything about [central and coastal] Chiapas if it wasn’t for [NWAf],” García-Des Lauriers says.

“Their work set the stage for everything I’ve done,” says SUNY Albany’s Rosenswig, who led recent excavations at Izapa to study the origins of urban life in Mesoamerica. When his graduate student Rebecca Mendelsohn, now a postdoc at the Smithsonian Tropical Research Institute in Panama City, excavated in Izapa in 2014, NWAf’s original map of its mounds and monuments served as a vital field reference (*Science*, 16 May 2014, p. 684). “I’ve been surprised at how sound the work from the 1960s still is,” she says.

NWAf is still run by BYU, which means its funding comes from the Mormon church and all its directors have been Mormons. But aside from a ban on coffee at headquarters, the archaeologists who work here barely notice its religious roots. “There aren’t conversations about religion,” Gasco says. “The archaeological community has a lot of respect for the work done here.”

Ferguson had hoped the Chiapas coast would turn out to be a crossroads not just for Mesoamerica, but the world. But the more NWAf and its collaborators excavated and analyzed sites in the region, the more they confirmed that Mesoamerican civilization sprang up from entirely New World origins. For archaeologists today, that makes the field all the more exciting. “That’s one of the most amazing things about studying Mesoamerican archaeology—it’s one of a half-dozen or so cases of independent development of agriculture, development of complexity, development of cities,” Rosenswig says.

It is hard to know whether Ferguson would have shared that excitement. For all his trust in science, his goal was to serve his faith. Some believing Mormons still read his books and trust his early, enthusiastic ideas about Mesoamerica. Others who came to doubt their religion also found hope in his story. His loss of faith gave them conviction and strength as they began their own journey down a difficult road, as shown by many who wrote him anguished letters in his later years.

But it is his scientific legacy, long unrecognized, that is perhaps most significant. “Facts are facts and truth is truth,” Ferguson once wrote about the archaeological evidence for the Book of Mormon that he was sure was about to be discovered in southern Mexico. His belief in that principle never wavered. ■

BIODIVERSITY AND ECOSYSTEMS

Assessing nature's contributions to people

Recognizing culture, and diverse sources of knowledge, can improve assessments

By Sandra Díaz, Unai Pascual, Marie Stenseke, Berta Martín-López, Robert T. Watson, Zsolt Molnár, Rosemary Hill, Kai M. A. Chan, Ivar A. Baste, Kate A. Brauman, Stephen Polasky, Andrew Church, Mark Lonsdale, Anne Larigauderie, Paul W. Leadley, Alexander P. E. van Oudenhoven, Felice van der Plaats, Matthias Schröter, Sandra Lavorel, Yildiz Aumeeruddy-Thomas, Elena Bukvareva, Kirsten Davies, Sebsebe Demissew, Gunay Erpul, Pierre Failler, Carlos A. Guerra, Chad L. Hewitt, Hans Keune, Sarah Lindley, Yoshihisa Shirayama

A major challenge today and into the future is to maintain or enhance beneficial contributions of nature to a good quality of life for all people. This is among the key motivations of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), a joint global effort by governments, academia, and civil society to assess and promote knowledge of Earth's biodiversity and ecosystems and their contribution to human societies in order to inform policy formulation. One of the more recent key elements of the IPBES conceptual framework (1) is the notion of nature's contributions to people (NCP), which builds on the ecosystem service concept popularized by the Millennium

Ecosystem Assessment (MA) (2). But as we detail below, NCP as defined and put into practice in IPBES differs from earlier work in several important ways. First, the NCP approach recognizes the central and pervasive role that culture plays in defining all links between people and nature. Second, use of NCP elevates, emphasizes, and operationalizes the role of indigenous and local knowledge in understanding nature's contribution to people.

The broad remit of IPBES requires it to engage a wide range of stakeholders, spanning from natural, social, humanistic, and engineering sciences to indigenous peoples and local communities in whose territories lie much of the world's biodiversity. Being an intergovernmental body, such inclusiveness

is essential not only for advancing knowledge but also for the political legitimacy of assessment findings (3).

FROM SERVICES TO CONTRIBUTIONS

NCP are all the contributions, both positive and negative, of living nature (diversity of organisms, ecosystems, and their associated ecological and evolutionary processes) to people's quality of life (4). Beneficial contributions include, for example, food provision, water purification, and artistic inspiration, whereas detrimental contributions include disease transmission and predation that damage people or their assets. Many NCP may be perceived as benefits or detriments depending on the cultural, socioeconomic,



Nature in the form of a living root bridge in Meghalaya, India, contributes to people by connecting both sides of the river.

temporal, or spatial context. For example, some carnivores are recognized—even by the same people—as beneficial for control of wild ungulates but as harmful because they may attack livestock.

At first inspection, the notion of NCP does not appear to differ much from the original MA definition of ecosystem services (2), which was broad and contemplated links to many facets of well-being. However, the detailed conceptualization and the practical work on ecosystem services following on the MA were dominated by knowledge from the natural sciences and economics. The natural sciences, and ecology in particular, were used to define “ecological production functions” to determine the supply of services, conceptualized as flows stemming from ecosystems (stocks of natural capital) (5). Economics was used to estimate the monetary value of those ecosystem services flows so as to identify trade-offs among them and their impacts on well-being. Aided by ecology and economics having readily available tools, the ecosystem services approach developed into a vibrant research field, influ-

enced policy discourse, and advanced the sustainability agenda.

However, this predominantly stock-and-flow framing of people-nature relationships largely failed to engage a range of perspectives from the social sciences (6), or those of local practitioners, including indigenous peoples. This reinforced a mutual alienation process in which MA-inspired studies and policies became increasingly narrow, which in turn led to voluntary self-exclusion of disciplines, stakeholders, and worldviews. As a consequence, the ecosystem services research program proceeded largely without benefiting from insights and tools in social sciences and humanities. For example, the unpacking and valuation of some “cultural ecosystem services” not readily amenable to biophysical or monetary metrics have lagged behind (7), and so has their mainstreaming into policy. In addition, as diverse disciplines and stakeholders remained at the margins, the initial skepticism toward the ecosystem services framework turned into active opposition, often based on the perceived risks of commodification of nature (8) and associated social equity concerns (9).

The need to be inclusive, both in terms of the strands of knowledge incorporated and representation of worldviews, interests and values (10), required IPBES to move to using NCP. Although still rooted in the MA ecosystem services framework (fig. S1), this new approach has the potential to firmly embed and welcome a wider set of viewpoints and stakeholders. It should also be less likely to be subsumed within a narrow economic (such as market-based) approach as the mediating factor between people and nature.

AN INCLUSIVE SYSTEM

The NCP approach explicitly recognizes that a range of views exist. At one extreme, humans and nature are viewed as distinct (2); at the other, humans and nonhuman entities are interwoven in deep relationships of kinship and reciprocal obligations (11, 12). In addition, the way NCP are coproduced by nature and people is understood through different cultural lenses. For instance, coproduction of food in high-diversity agriculture can be framed as a process that combines a set of biological and technological inputs aimed at maximizing coexistence between useful plants and animals in order to achieve higher yields.

Alternatively, coproduction of food can be seen as a “practice of care” (12, 13) through social relationships and connection with spiritual entities. Therefore, we propose two lenses through which to view NCP: a generalizing perspective and a context-specific perspective. Although presented here as extremes, these two perspectives are often

blended and interwoven (14), enabling co-construction of knowledge among disciplines and knowledge systems (fig. S2).

Generalizing perspective

Typical of the natural sciences and economics, this perspective (represented in green at the bottom of fig. S2) is fundamentally analytical in purpose; it seeks a universally applicable set of categories of flows from nature to people. Distinction between them is often sharp, and agency is acknowledged only in the case of people. NCP categories can be seen at finer or coarser resolution but can still be organized into a single, self-consistent system.

We identify 18 such categories for reporting NCP within the generalizing perspective, organized in three partially overlapping groups: regulating, material, and nonmaterial NCP (fig. S3 and table S1), defined according to the type of contribution they make to people’s quality of life.

Material contributions are substances, objects, or other material elements from nature that directly sustain people’s physical existence and material assets. They are typically physically consumed in the process of being experienced—for example, when organisms are transformed into food, energy, or materials for ornamental purposes.

Nonmaterial contributions are nature’s effects on subjective or psychological aspects underpinning people’s quality of life, both individually and collectively. Examples include forests and coral reefs providing opportunities for recreation and inspiration, or particular animals and plants being the basis of spiritual or social-cohesion experiences.

Regulating contributions are functional and structural aspects of organisms and ecosystems that modify environmental conditions experienced by people and/or regulate the generation of material and nonmaterial contributions. Regulating contributions frequently affect quality of life in indirect ways. For example, people directly enjoy useful or beautiful plants but only indirectly benefit from the soil organisms that are essential for the supply of nutrients to such plants.

Culture permeates through and across all three broad NCP groups (fig. S1) rather than being confined to an isolated category (the “cultural ecosystem services” category in the MA framework). In addition, the three broad groups—rather than being independent compartments, as typically framed within the ecosystem services approach—explicitly overlap. We distinguish them for practical reporting reasons, acknowledging that many of the 18 NCP categories do not fit squarely into a single group (fig. S3). For example, food is primarily a material NCP because calories and nutrients are essential for physi-

A complete listing of affiliations is provided in the supplementary materials. Email: sandra.diaz@unc.edu.ar; unai.pascual@bc3research.org

cal sustenance. However, food is full of symbolic meaning well beyond physical survival. Indeed, nonmaterial and material contributions are often interlinked in most, if not all, cultural contexts (7).

Context-specific perspective

This is the perspective typical, but not exclusive, of local and indigenous knowledge systems (represented in blue at the top of fig. S2). In local and indigenous knowledge systems, the production of knowledge typically does not explicitly seek to extend or validate itself beyond specific geographical and cultural contexts (14). Indeed, the context-specific perspective on NCP often tends to resist the scientific goal of attaining a universally applicable schema.

Although subdivision into internally consistent systems of categories is common in many local knowledge systems, a universally applicable classification—such as the one proposed in the generalizing perspective on NCP (table S1)—is not currently available and may be inappropriate because of cultural incommensurability and resistance to universal perspectives on human-nature relations. The context-specific perspective may instead present NCP as bundles that follow from distinct lived experiences such as fishing, farming, or hunting or from places, organisms, or entities of key spiritual significance, such as sacred trees, animals, or landscapes (11, 13).

Providing space for context-specific perspectives recognizes that there are multiple ways of understanding and categorizing relationships between people and nature and avoids leaving these perspectives out of the picture or forcing them into the 18 generalizing NCP categories. The NCP approach thus facilitates respectful cooperation across knowledge systems in the co-construction of knowledge for sustainability.

NURTURING A PARADIGM SHIFT

The NCP concept extends beyond the highly influential yet often contested notion of ecosystem services, incorporating a number of interdisciplinary insights and tools. Most of them were called for during the past decade (9, 10, 12, 14) but only now are enshrined explicitly in an environmental assessment framework.

The implementation of the NCP approach and its reporting categories (tables S1 and S2) is still in its infancy and is expected to be fully fledged only in the IPBES Global Assessment, but the NCP approach is already changing assessment procedures and their outcomes. For example, the ongoing IPBES regional assessments include an unprecedented effort to tap indigenous and local knowledge, from the literature and also from dialogues with indigenous and local knowledge-holders, to which

they contributed information presented in their own narratives. In the Europe and Central Asia assessment, these narratives (15) revealed complex interactions between detrimental (predation on livestock) and beneficial NCP (carcass removal or protection by shepherd/guard dogs) that were not considered in previous national ecosystem assessments. This kind of evidence also enhanced the confidence about the status and trends of other NCP in cases in which the evidence based on published literature was scarce (such as for NCP “Supporting identities”). In this regional assessment, it was relatively easy to fit most narratives into the 18 categories of the generalizing perspective on NCP.

In assessing pollinators, pollination, and food production (16), the dialogue with local and indigenous knowledge-holders highlighted some NCP that were defined as practices of care gifted to people, such as fostering pollinator nesting resources in forests, totemic relationships requiring reciprocal obligations between people and

“The NCP approach aims at ... products that are ... more likely to be incorporated into policy and practice.”

pollinators, and traditional governance that depends on ongoing presence of bees and butterflies in the landscape (table S2) (13). These context-specific NCP do not fit easily in the 18 generalizing NCP categories. Nevertheless, these knowledge sources underpinned innovative strategic responses highlighted in the main messages to policy-makers that were agreed on among all the member countries of IPBES (16): to strengthen traditional governance and tenure systems that support pollinators, which are critical in many places where these systems are being eroded through rapid industrialization.

These examples illustrate how the interweaving of epistemologically diverse lines of evidence (14) about specific subjects can result in richer solutions for people and nature, even within the context of large-scale assessments. But regardless of the outcomes of the assessments, the consideration of different knowledge systems—and the fact that generalizing, context-specific, and mixed perspectives are considered as equally useful—matters in terms of making IPBES procedures and outcomes more equitable. This should help overcome existing power asymmetries between western science and indigenous and local knowledge, and among

different disciplines within western science, in the science-policy interface. The NCP approach aims at coming up with products that are better and also more legitimate and therefore more likely to be incorporated into policy and practice.

In addition to assessments, environmental governance and associated policies would likely increase their effectiveness and social legitimacy by drawing on the NCP approach. This is because it facilitates much more than previous framings the connection with rights-based approaches to conservation and sustainable use of nature and their implications for quality of life. The presence of multiple worldviews and diverse ways of expressing them in the wording of the Convention on Biological Diversity’s strategic plan for biodiversity and specific objectives, such as the Aichi Targets, further illustrates how important inclusive framings are to the broad political legitimacy of these international objectives and their implementation instruments. ■

REFERENCES AND NOTES

1. S. Díaz et al., *Curr. Op. Environ. Sustain.* **14**, 1 (2015).
2. *Millennium Ecosystem Assessment* Washington, DC (Island Press, 2005).
3. E. S. Brondizio, F.-M. L. Tourneau, *Science* **352**, 1272 (2016).
4. IPBES Plenary 5 Decision IPBES-5/1: Implementation of the First Work Programme of the Platform, page 23; www.ipbes.net/event/ipbes-5-plenary.
5. S. Polasky, K. Segerson, *Ann. Rev. Resour. Econ.* **1**, 409 (2009).
6. R. B. Norgaard, *Ecol. Econ.* **69**, 1219 (2010).
7. K. M. A. Chan et al., *Bioscience* **62**, 744 (2012).
8. S. Lele et al., *Conserv. Soc.* **11**, 343 (2013).
9. U. Pascual et al., *BioScience* **64**, 1027 (2014).
10. U. Pascual et al., *Curr. Op. Environ. Sustain.* **26**, 7 (2017).
11. F. Berkes, *Sacred Ecology* (Routledge, ed. 3, 2012).
12. C. Comberti et al., *Glob. Environ. Change* **34**, 247 (2015).
13. R. Hill et al., in *Pollinators, Pollination and Food Production: A Global Assessment*, S. G. Potts et al., Eds. (IPBES, 2016).
14. M. Tengö et al., *Curr. Op. Environ. Sustain.* **26–27**, 17 (2017).
15. M. Roué, Z. Molnár, Eds., *Knowing Our Lands and Resources: Indigenous and Local Knowledge of Biodiversity and Ecosystem Services in Europe and Central Asia*. *Knowledge of Nature 9* (UNESCO, 2017).
16. IPBES, *Summary for Policymakers of the Assessment Report of the IPBES on Pollinators, Pollination and Food Production*, S. G. Potts et al., Eds. (Secretariat of IPBES, 2016).

ACKNOWLEDGMENTS

We acknowledge the following experts participating in IPBES assessments: C. Anderson, P. Balvanera, B. Baptiste, N. Bennas, F. Berkes, M. Carneiro da Cunha, C. Chenu, M.-C. Cornier-Salem, B. Czúcz, P. Elias, B. Erasmus, S. Fennessy, J. Fisher, C. Fürst, S. Jacobs, O. Osano, D. Pacheco, M. Potts, S. Preston, A. Purvis, A. Rajwanshi, J. Rice, M. Rosales-Benites, C. S. Seixas, M. Solan, J. Tassin, W. Townsend, G. von Maltitz, T. Yahara, C.-Y. Yao, and Y.-C. Yoon. We thank C. Broshi, M. Colloff, H. T. Ngo, and D. Singer for useful input during the development of this work; V. Falczuk for help with the bibliography; and Y. Estrada for preparing the figures. S.D. was partially supported by the Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad Nacional de Córdoba, and Fondo para la Investigación Científica y Tecnológica. U.P. was supported by the Basque Foundation for Science, IKERBASQUE.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6373/270/suppl/DC1

10.1126/science.aap8826



The two 3D-printed cylindrical static mixers (diameter of 6 mm and length of 150 mm) were coated with metal catalysts for hydrogenation reactions (10).

PERSPECTIVES

CHEMICAL ENGINEERING

The art of manufacturing molecules

Additively manufactured monolithic reactors allow on-demand synthesis of drug molecules

By Christian H. Hornung

The way we manufacture many of the products used in everyday life, such as the ingredients in shampoo, the plastic components of smartphones, the vitamins and pharmaceuticals we take, and the packaging that all of them come in, has not changed in a significant way over the last hundred years. Arguably, these methods of manufacturing are even older and were already applied in the first large-scale chemical processes in the 19th century, in which new products such as vulcanized rubber, synthetic dyes, or industrial fertilizers were first produced on scales

unknown to society at the time. The development of these industrial processes was driven by the benefits of economy of scale, with the aim of centralizing, optimizing, maximizing, and integrating production. In recent years, efforts were made by a series of research groups to reverse this trend and decentralize, miniaturize, and even digitize chemical manufacturing. On page 314 of this issue, Kitson *et al.* (1) report the synthesis of active pharmaceutical ingredients (APIs) on demand in a three-dimensional (3D)-printed, miniaturized reactor cascade. A complete multistep synthesis of the muscle relaxant baclofen was developed and digitized for remote bench-scale manufacture.

Chemical production, whether for petrochemicals or pharmaceuticals, requires specialized equipment and facilities, and its

hazardous nature demands highly trained operators. Centralizing a series of chemical operations within one large dedicated plant is generally a much more sensible approach than dividing them up between small, distributed manufacturers or single-person craftsmanship. However, centralized production also has its downsides, such as the distance to the point of use and the associated transport and storage issues, which are a concern for high-value products with a limited shelf life (e.g., many pharmaceuticals). Recent technological advances are allowing us to challenge this old way of thinking, and to propose new answers to the question of how and where the molecules and materials we use day in, day out should be made. Miniaturization and additive manufacture are two key elements of these efforts to de-

CSIRO Manufacturing, Private Bag 10, Clayton South, Victoria 3169, Australia. Email: christian.hornung@csiro.au

centralize chemical production, and Kitson *et al.*'s innovative method of synthesizing APIs in a 3D-printed, miniaturized reactor incorporates both of these concepts.

The miniaturization of chemical synthesis and analysis, and of the apparatus and devices used, has changed laboratory methodologies during the past few decades. Mini- and microfabrication techniques introduced in the 1980s and 1990s have led to the development of new plate- and chip-type continuous-flow microreactors that can better control heat and mass transfer, and in turn improve conversion, efficiency, or safety of a chemical reaction by orders of magnitude compared with conventional reactor geometries (2–4). Microreactor (or flow-chemistry) technology makes ultrafast, highly exothermic reactions controllable and practicable on laboratory and industrial scales. Device miniaturization has also changed the way chemicals are analyzed. Lab-on-a-chip devices made from glass, plastic, or even paper can simplify, accelerate, and reduce the cost of clinical diagnostics and remote chemical analysis (5–7).

Additive manufacturing techniques, such as 3D printing of polymers, ceramics, or metals, overcome limitations of conventional subtractive manufacturing methods, resulting in nearly complete freedom of design. Recently, chemical engineers, materials scientists, and others have utilized this potential of 3D printing for building more efficient chemical reactors with geometries that are otherwise not accessible. For example, the use of 3D-printed metal structures as more efficient and tailor-made mixers, catalysts, or both in the synthesis of organic compounds is currently being investigated by several research groups (see the photo) (8–10). The ability to design the reactor geometry specifically for a given fluidic or chemical application, and the ability to rapidly prototype the device, can give 3D-printed “reactorware” a pivotal advantage over traditional methods.

Kitson *et al.* developed a blueprint that digitizes the bench-scale synthesis of APIs into a sequence of batch operations conducted in a monolithic polymer reactor. The originality of this approach is twofold. First, it combines several processing steps, including four reactions, two liquid-liquid extractions, and a set of evaporations and filtrations inside one tailor-made reactor device containing several interconnected modules. Second, said device can be printed on demand for use in a distributed setting. All that is needed is access to a 3D printer, a library of common chemical starting materials, and a set of instructions (i.e., design files and chemical synthesis protocols), and in principle, it would be possible to synthesize small amounts of APIs or other

complex, high-value compounds with a limited shelf life anywhere. Such an approach is strongly aligned with current efforts by other researchers working on distributed and remote manufacturing of chemicals and pharmaceuticals (11), which usually have been executed in larger and more intensified processing equipment. The concept of designing the operations such that they can be housed inside mobile shipping containers has been adopted by several industrial research and development groups in recent years, such as the integrated chemical plants developed by the F³ Factory consortium (12).

By demonstrating the multistep synthesis of baclofen in this integrated, benchtop device, the door has been opened to making complex molecules, such as APIs, on demand in nontraditional manufacturing environments such as hospitals or even doctors' offices, bringing manufacturing closer to the point of use. These manufacturing scenarios might also include remote settings, synthesis of personalized medicines, small-scale production of abandoned pharmaceuticals, or even space missions.

For this technology to be put into practice, a range of regulatory hurdles would have to be considered. How would quality control and chemical analysis be approached in remote settings? Furthermore, how do the costs of remote on-demand synthesis compare with shipping, storage, and inventory of traditionally manufactured drugs? Nonetheless, new technological solutions for the manufacture of chemicals and pharmaceuticals in a decentralized setting are needed, and these may have economic, environmental, and societal benefits, particularly for rural communities and industries (11). Distributed manufacturing is a promising approach for sustainable and socially responsible production of goods close to their point of use, and Kitson *et al.* are among the pioneers bringing us closer to making flexible and potentially movable, portable, or even printable minifactories a reality. ■

REFERENCES

1. P. J. Kitson *et al.*, *Science* **359**, 314 (2018).
2. V. Hessel, D. Kralisch, N. Kockmann, T. Noël, Q. Wang, *ChemSusChem* **6**, 746 (2013).
3. B. Gutmann, D. Cantillo, C. O. Kappe, *Angew. Chem. Int. Ed. Engl.* **54**, 6688 (2015).
4. D. E. Fitzpatrick, S. V. Ley, *Tetrahedron* **10.1016/j.tet.2017.08.050** (2017).
5. G. M. Whitesides, *Nature* **442**, 368 (2006).
6. A. K. Yetisen, M. S. Akram, C. R. Lowe, *Lab Chip* **13**, 2210 (2013).
7. M. I. Mohammed *et al.*, *Procedia Tech.* **20**, 54 (2015).
8. W. Peters *et al.*, *Energy Environ. Sci.* **8**, 641 (2015).
9. C. H. Hornung *et al.*, *Org. Process Res. Dev.* **21**, 1311 (2017).
10. C. Parra-Cabrera, C. Achille, S. Kuhn, R. Ameloot, *Chem. Soc. Rev.* **47**, 209 (2018).
11. E. Rauch, P. Dallasega, D. T. Matt, *J. Cleaner Prod.* **135**, 127 (2016).
12. www.f3factory.com/scripts/pages/en/home.php

10.1126/science.aar4543

QUANTUM FLUIDS

Quantum liquids get thin

A mix of two bosonic particles develops attractive forces to create a quantum liquid

By Igor Ferrier-Barbut and Tilman Pfau

A liquid exists when interactions that attract its constituent particles to each other are counterbalanced by a repulsion acting at higher densities. Other characteristics of liquids are short-range correlations and the existence of surface tension (1). Ultracold atom experiments provide a privileged platform with which to observe exotic states of matter, but the densities are far too low to obtain a conventional liquid because the atoms are too far apart to create repulsive forces arising from the Pauli exclusion principle of the atoms' internal electrons. The observation of quantum liquid droplets in an ultracold mixture of two

“...in this configuration, the droplet does not expand like a gas would do but stays self-bound and behaves like a liquid.”

quantum fluids is now reported on page 301 of this issue by Cabrera *et al.* (2) and a recent preprint by Semeghini *et al.* (3). Unlike conventional liquids, these liquids arise from a weak attraction and repulsive many-body correlations in the mixtures.

In ordinary liquids, the attraction between the constituents emerge from weak forces such as hydrogen bonds or van der Waals interactions, and the repulsion at higher density stems from the Pauli exclusion principle for electrons. The ultracold-atom samples that were studied were

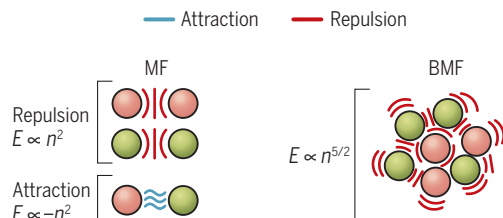
Physikalisches Institut und Center for Integrated Quantum Science and Technology, Universität Stuttgart, Germany.
Email: i.ferrier-barbut@physik.uni-stuttgart.de

Mixing up a quantum liquid

Cabrera *et al.* and Semeghini *et al.* studied mixtures of two species of trapped, ultracold bosonic potassium atoms and show that they formed a quantum liquid. Single species have only repulsive interactions and form quantum gases, but the interaction between species is attractive and allows a liquid to form.

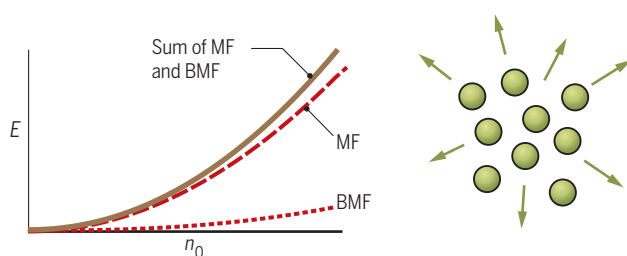
Repulsion versus attraction

The mean-field (MF) approximation predicts that energy (E) varies as n^2 , where n is density. Beyond the mean-field (BMF) corrections are always positive, with stronger $n^{5/2}$ dependence.



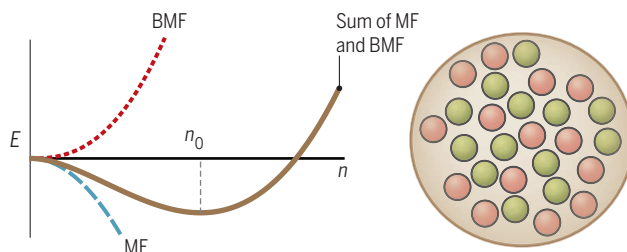
Gas

For a single atomic species, the ensemble MF energy is positive, and BMF corrections are weak. A gas forms that expands in free space to minimize its energy.



Quantum liquid

When two types of atoms are mixed, MF effects nearly cancel out, creating a weak attraction that is counterbalanced by BMF corrections. A liquid forms at a particular density n_0 that minimizes energy.



at about 100 nK, well below the freezing points of ordinary liquids, but had densities about eight orders of magnitude lower than that of water. In this regime, atoms behave as coherent matter waves, interfering with each other, and in the case of so-called bosonic atoms, they form a Bose-Einstein condensate (BEC).

Because BECs are very dilute, they exist mostly in the weakly interacting regime where the mean-field (MF) approximation is sufficient to calculate the energy of the condensate accurately. This approximation assumes that fluctuations are not relevant in the system. By properly taking into account many-body effects stemming from quantum fluctuations, corrections to the energy beyond the mean-field approximation (BMF) can be obtained. These corrections create a shift in energy that grows with interaction strength but is typically very small compared with the MF energy.

However, this BMF energy shift exhibits a stronger dependence with density than the MF energy. In a large and stable single-species BEC, the MF energy is already positive so that both effects are repulsive

(see the figure). To obtain a balance between attraction and repulsion, Petrov (4) suggested mixing two species of atoms together. Within each species, the atoms repel each other and form a BEC, but the interspecies interaction is attractive and slightly stronger. For these conditions, the MF energy is very small and negative and acts effectively as an attractive force, whereas the many-body BMF correction remains repulsive and sizable. The system then forms a self-bound liquid.

The experiments of Cabrera *et al.* and Semeghini *et al.* trap potassium atoms of the bosonic isotope ^{39}K . They prepared the atoms in two different states with the interatomic interactions tuned so that this liquid state would form. To keep the sample from forming a gas, the two teams used specially designed optical traps to hold the droplet against gravity while allowing it to expel atoms in the horizontal plane. They show that in this configuration, the droplet does not expand like a gas would do but stays self-bound and behaves like a liquid. One defining particularity of this quantum liquid is that below a certain atom number,

it evaporates into a gas because quantum kinetic energy overcomes interactions. Both teams explored the resulting liquid-gas phase diagram by varying atom number and interactions.

The original theoretical prediction of quantum droplets was formulated for BEC mixtures, so the observation by Cabrera *et al.* and Semeghini *et al.* marks an important benchmark. However, quantum droplets also occur in other Bose-condensed systems in which two types of interactions exist. In particular, they have been observed in BECs where magnetic interactions occur on top of the usual zero-range interactions, such as the ones taking place in potassium (5–7). These interactions are repulsive, whereas the dipolar interaction can be made attractive, thus fulfilling the conditions for the liquid state. These magnetic droplets were anisotropic and longer-lived. This diversity of self-bound liquids opens possibilities that so far were out of reach with BECs. As Cabrera *et al.* show, no theory yet describes the precise properties of quantum droplets. Thus, these studies will serve as benchmarks for quantum many-body theories.

A question to now address is whether these droplets undergo evaporation and cool to nearly 0 K, as suggested theoretically in (4). This very peculiar property comes from the fact that the self-bound liquid should be the only bound solution so that any higher-energy excitation should be unbound and evaporated. Furthermore, constraining the liquids in lower dimensions should lead to new physics. It is indeed a known result from quantum mechanics that in three dimensions, an attractive potential cannot always bind a particle, but in two and one dimensions, it always can. Thus, the gas-liquid phase diagram would be modified, also because fluctuations play a stronger role (8, 9). Future studies can also be inspired by the rich physics of helium nanodroplets (10). For example, quantum droplets could act as solvents to trap and study impurity atoms one by one. ■

REFERENCES

1. J. Barker, D. Henderson, *Rev. Mod. Phys.* **48**, 587 (1976).
2. C. R. Cabrera *et al.*, *Science* **359**, 301 (2018).
3. G. Semeghini *et al.*, arXiv:1710.10890 (2017).
4. D. Petrov, *Phys. Rev. Lett.* **115**, 155302 (2015).
5. I. Ferrier-Barbut, H. Kadau, M. Schmitt, M. Wenzel, T. Pfau, *Phys. Rev. Lett.* **116**, 215301 (2016).
6. L. Chomaz *et al.*, *Phys. Rev. Lett.* **X 6**, 041039 (2016).
7. M. Schmitt, M. Wenzel, F. Böttcher, I. Ferrier-Barbut, T. Pfau, *Nature* **539**, 259 (2016).
8. D. Petrov, G. Astrakharchik, *Phys. Rev. Lett.* **117**, 100401 (2016).
9. D. Edler *et al.*, *Phys. Rev. Lett.* **119**, 050403 (2017).
10. P. Toennies, A. Vilesov, B. Whaley, *Phys. Today* **54**, 31 (2001).

10.1126/science.aar3785

MICROBIOLOGY

A bacterial coat that is not pure cotton

Biofilms formed by *E. coli* and *Salmonella* contain a new form of modified cellulose

By Michael Y. Galperin¹ and
Daria N. Shalaeva^{2,3}

Cellulose, a linear polymer of glucose residues, is the main component of plant cell walls and the most abundant biomolecule on the planet. Cellulose fibers from wood, cotton, and linen are mostly used as such, but can also be chemically modified to make rayon, viscose, and other textiles. Many bacteria also synthesize cellulose. Cellulose fibers produced by the model organism *Komagataeibacter (Gluconacetobacter) xylinus* are very similar to those found in plants (1) and are increasingly used in biotechnology and nanotechnology (2, 3). *Escherichia coli* and many other bacteria produce cellulose as a key component of the extracellular matrix that coats the cells to form a biofilm, a complex multicellular community consisting of numerous bacteria, exopolysaccharides (like cellulose), protein fibers, and DNA (4–6). The cellulose in biofilms was assumed to be the same as that produced by *G. xylinus*, owing to the same pattern of staining with Congo red dye and the same cellulose synthase enzyme (4–6). However, on page 334 of this issue, Thongsomboon *et al.* (7) report that *E. coli* and *Salmonella enterica* serovar Typhimurium produce modified cellulose, in which every other glucosyl residue carries an additional phosphoethanolamine (pEtN) group. These findings have important implications for a wide variety of disciplines, from microbiology to materials science.

It appears that previous studies simply overlooked the presence of this cellulose modification. Bacterial cellulose is a stable polymer that is resistant to a variety of harsh treatments. When such treatments were used to purify cellulose fibers from microbial biofilms, the pEtN group was lost. Thongsomboon *et al.* used mass spectrometry and several versions of solid-state

nuclear magnetic resonance spectroscopy to detect and study the pEtN modification in vivo. They also showed that the pEtN group comes from the cell membrane lipid phosphatidylethanolamine and identified BcsG, a subunit of the cellulose synthase complex, as the pEtN transferase responsible for catalyzing the pEtN modification.

Microbiologists might be most interested in the wide phylogenetic distribution of the pEtN modification. In addition to *E. coli* and *Salmonella*, BcsG is encoded in cellulose synthase operons of some important pathogens, including certain species of *Klebsiella*, *Shigella*, *Enterobacter*, and *Burkholderia*, as well

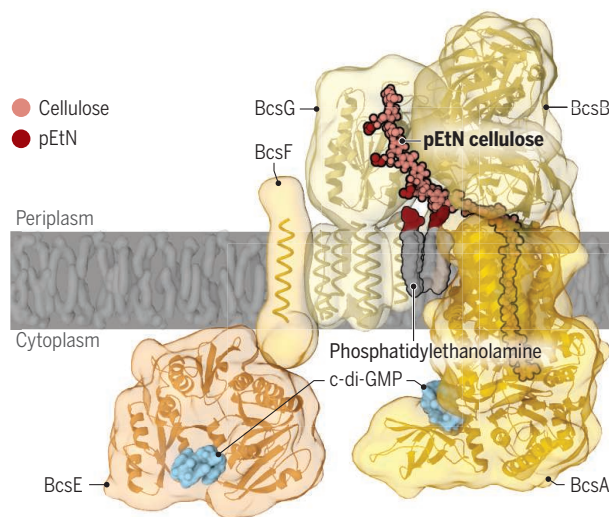
to detergents and enzymes that do not affect standard cellulose fibers but that might be active on pEtN-modified cellulose. Thus, the goal of preventing infection by disrupting such biofilms might be within closer reach.

The work by Thongsomboon *et al.* should also help with understanding the exact role of the cellulose-containing biofilm as a virulence factor. Curiously, some of the worst disease-causing strains of *E. coli* and *Shigella* do not produce biofilms in acute infections (4, 8). It appears that the ability to form a biofilm allows these pathogens to better adapt to the host environment and shifts the infection from an acute to chronic state. Thus, improved understanding of pEtN cellulose biosynthesis is an important step toward fighting bacterial diseases.

More generally, Thongsomboon *et al.* provide insight into the regulation of polymer export in bacteria. Bacterial cellulose biosynthesis depends on the second messenger cyclic diguanosine monophosphate (c-di-GMP) (4, 9–11). The structure of cellulose synthase (10) showed that c-di-GMP is required to allow the substrate access to the enzyme active site. The amount of cellulose produced by *E. coli* also depends on the expression of the *bcsEFG* operon and one of its products, BcsE, which binds c-di-GMP (12). However, the exact roles of these proteins have remained obscure. The finding that the catalytic domain of BcsG is located in the periplasm (between the inner and outer membranes) and modifies glucosyl residues of the nascent cellulose chain (7) allowed Thongsomboon *et al.* to predict the organization of the entire mem-

E. coli cellulose synthase complex

The findings of Thongsomboon *et al.* have helped to clarify the structure of the cellulose synthase enzyme, which consists of multiple subunits and is regulated by c-di-GMP. Cellulose can be modified with pEtN derived from the membrane lipid phosphatidylethanolamine, a reaction catalyzed by BcsG. This model is based on the data from (4, 7, 10–12, 15).



brane-bound cellulose synthase complex (see the figure). This not only resolves the role of the *bcsEFG* operon, but also highlights the complexity of c-di-GMP-mediated regulatory processes. At least in *E. coli* and *Salmonella*, c-di-GMP seems to regulate cellulose formation at several different levels. This multilevel regulation of important life-cycle decisions, such as biofilm formation, has been dubbed “sustained sensing” (13). It appears to be common in the microbial world but is difficult to disentangle without a detailed analysis of regulatory interactions.

¹National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA. ²School of Bioengineering and Bioinformatics, Moscow State University, Moscow 119992, Russia. ³School of Physics, University of Osnabrück, Osnabrück D-49069, Germany. Email: galperin@ncbi.nlm.nih.gov; dshalaeva@uos.de

Finally, the work by Thongsomboon *et al.* will benefit efforts to find new applications for bacterially synthesized cellulose and develop new cellulose-based compounds. *Gluconacetobacter*-produced cellulose microfibrils and crystals, commonly referred to as nanocellulose, have numerous applications (2). The apparent biocompatibility—lack of toxicity, immunogenicity, and proinflammatory response—of unmodified cellulose makes it an attractive choice for a variety of biomedical applications, such as drug delivery, wound dressing, replacement of blood vessels, and tissue engineering of bone and cartilage (2, 3). Thus, pEtN-modified cellulose would have to undergo rigorous biocompatibility testing. Nevertheless, the availability of genetic tools to manipulate *E. coli* opens numerous possibilities for using cellulose synthase genes for synthetic biology. A particularly interesting development could be the ability to produce entirely new kinds of cellulose films for applications ranging from optoelectronics to packaging.

In principle, one could imagine production of cellulose microfibrils with new modifications. The catalytic domain of BcsG is a metalloenzyme of the alkaline phosphatase and sulfatase superfamily (14), which is anchored in the membrane by five predicted transmembrane helices (see the figure). Replacing the catalytic domain of BcsG with a different enzyme, such as an acyltransferase or a glycosyltransferase, could allow biosynthesis of cellulose nanocrystals with new optical properties, increased conductivity, or the ability to bind metal ions. ■

REFERENCES AND NOTES

- P. Ross, R. Mayer, M. Benizman, *Microbiol. Rev.* **55**, 35 (1991).
- N. Lin, A. Dufresne, *Eur. Polymer J.* **59**, 302 (2014).
- T. Abitbol *et al.*, *Curr. Opin. Biotechnol.* **39**, 76 (2016).
- U. Römling, M. Y. Galperin, *Trends Microbiol.* **23**, 545 (2015).
- X. Zogaj, M. Nimtz, M. Rohde, W. Bokranz, U. Römling, *Mol. Microbiol.* **39**, 1452 (2001).
- D. O. Serra, A. M. Richter, R. Hengge, *J. Bacteriol.* **195**, 5540 (2013).
- W. Thongsomboon *et al.*, *Science* **359**, 334 (2018).
- A. M. Richter, T. L. Povolotsky, L. H. Wieler, R. Hengge, *EMBO Mol. Med.* **6**, 1622 (2014).
- U. Römling, M. Y. Galperin, M. Gomelsky, *Microbiol. Mol. Biol. Rev.* **77**, 1 (2013).
- J. L. Morgan *et al.*, *Nature* **531**, 329 (2016).
- P. V. Krasteva *et al.*, *Nat. Commun.* **8**, 2065 (2017).
- X. Fang *et al.*, *Mol. Microbiol.* **93**, 439 (2014).
- M. W. Orr, M. Y. Galperin, V. T. Lee, *Curr. Opin. Microbiol.* **34**, 119 (2016).
- M. Y. Galperin, A. Bairoch, E. V. Koonin, *Protein Sci.* **7**, 1829 (1998).
- M. A. Lomize, I. D. Pogozheva, J. Joo, H. I. Mosberg, A. L. Lomize, *Nucleic Acids Res.* **40**, D370 (2012).

ACKNOWLEDGMENTS

M.Y.G. is supported by the U.S. National Institutes of Health Intramural Research Program at the National Library of Medicine. D.N.S. is supported by the German Academic Exchange Service (DAAD).

MICROBIOLOGY

Taking down defenses to improve vaccines

A new approach to generating influenza virus vaccines could improve responses

By John R. Teijaro¹ and Dennis R. Burton^{1,2}

Vaccines have been spectacularly successful in durable protection against a range of pathogens. However, they have been less successful against pathogens that have evolved immune escape mechanisms (1). For example, the influenza virus surface glycoprotein hemagglutinin (HA), which is the main target (antigen) for protective antibodies, shows enormous sequence diversity between different strains, meaning that antibodies induced by immune responses to one strain of the virus tend to be either inefficient or ineffective against other strains. This observation is often associated with the need for a new influenza vaccine every year. However, the escape mechanisms of influenza virus extend beyond antigenic variation of surface proteins. For example, wild-type viruses typically encountered in natural infection can suppress the host type I interferon (IFN-I) response, which provides the first line of defense against viral infections and promotes stimulation of an optimal immune response (2). On page 290 of this issue, Du *et al.* (3) describe the generation of a variant influenza virus that, in contrast to the wild type, is hyper-interferon-sensitive (HIS) and therefore attenuated (reduced in virulence). Attenuated viruses typically have lower immune responses than their wild-type counterparts but, in this case, the level of attenuation still resulted in robust immune responses. The authors propose that the HIS approach could form the basis for a more effective influenza vaccine.

Following infection, viral nucleic acids are sensed by the host innate immune system through multiple pattern recognition receptors. Among the first antiviral proteins produced are IFN-I, a family of cytokines that signal to surrounding host cells and induce multiple interferon-stimulated genes (ISGs) that act to prevent viral amplification and dissemination. In addition

to inhibiting viral replication, IFN-I signaling also promotes the optimal induction of both the innate and adaptive arms of the immune response. To counteract a potent IFN-I response, many viruses encode proteins that inhibit IFN-I production and/or signaling, highlighting the evolutionary importance of host IFN-I signaling in controlling early virus infection (4). Influenza virus is no exception and encodes nonstructural protein 1 (NS1), which exerts

“The authors argue that the approach may be broadly applicable in creating efficacious live attenuated vaccines for a plethora of viral infections.”

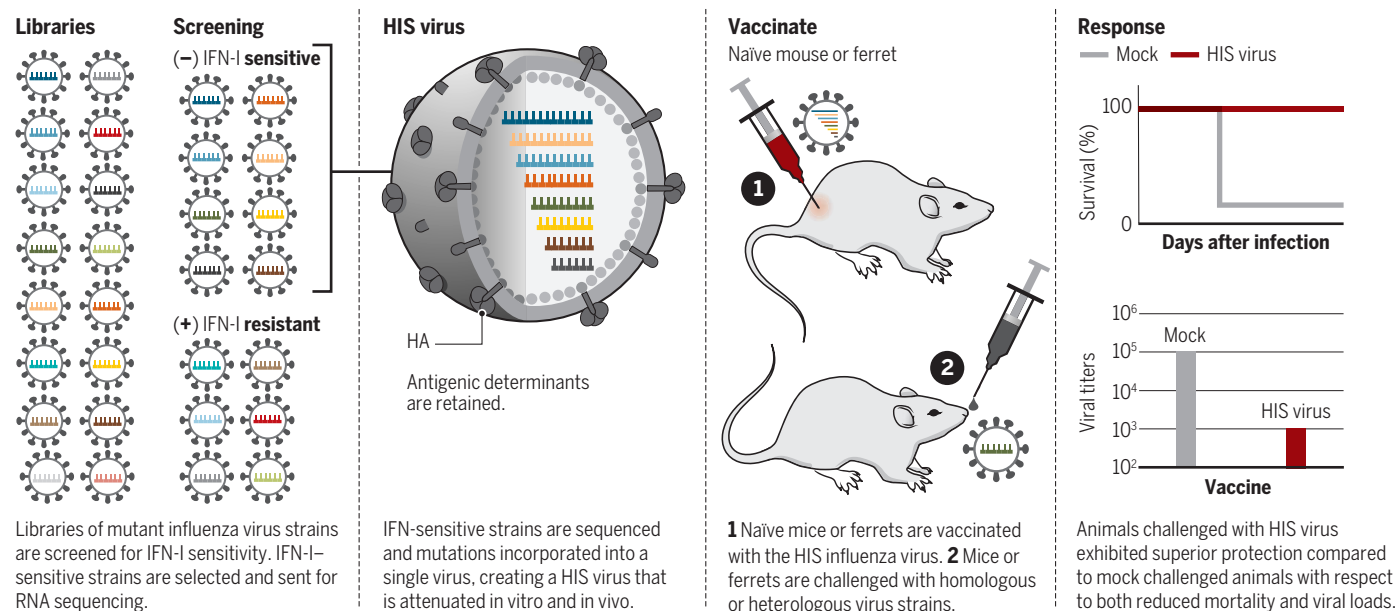
potent anti-IFN-I effects and is essential for viral fitness (5).

To investigate the possibility of harnessing IFN-I sensitivity for attenuated vaccine design, Du *et al.* employed a quantitative high-throughput genetic mutagenesis system coupled to next-generation sequencing (6, 7) to simultaneously measure loss-of-function mutations in influenza virus in the presence or absence of IFN-I signaling. Using this approach, they identified multiple IFN-I-sensitive mutations across the influenza virus genome, including mutations outside of the gene encoding NS1. The authors then combined an assortment of eight IFN-I-sensitive mutations in multiple viral genes to create a HIS virus. The HIS influenza virus increased both IFN-I sensitivity and production, attenuated viral fitness in vitro and in vivo, and did not produce disease pathology after high-dose administration to mice and ferrets (commonly used models of influenza infection). Importantly, despite multiple mutations in viral genes in the HIS virus, enough antigenic determinants were conserved to generate antiviral T and B cell-mediated immune responses (which involve antibodies) to promote protection after challenge with multi-

¹Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037, USA. ²Ragon Institute of MGH, Harvard and MIT, Cambridge, MA 02139, USA. Email: teijaro@scripps.edu; burton@scripps.edu

Generating live attenuated HIS virus vaccines

HIS viruses confer improved vaccination against influenza virus strains. This approach could be used to generate more effective live attenuated vaccines against other viruses.



ple strains of influenza virus when mice were vaccinated with HIS viruses derived from the H1N1 subtype and subsequently challenged with either H1N1 or H3N2 subtypes in mice and ferrets (see the figure). The protection generated by the HIS influenza virus vaccine is likely mediated through the generation of cross-reactive T cells, which can react with multiple viral strains. This is especially likely to be the case because mice vaccinated with the H1N1-HIS virus and subsequently challenged with heterologous H3N2 influenza viruses produce potent protective antibodies against the homologous H1N1 virus that will not have substantial cross-reactivity with H3N2 viruses.

The authors argue that the approach may be broadly applicable in creating efficacious live attenuated vaccines for a plethora of viral infections. Although a concern with live attenuated vaccines is often their potential to revert to virulence (8), attenuating the virus with multiple point mutations, as in Du *et al.*, is likely to avert this outcome. In addition to increasing safety, the use of mutations scattered throughout the viral genome should provide a barrier to the development of viral resistance. Moreover, the avoidance of mutations in HA should help preserve robust antibody responses.

The holy grail of the influenza virus vaccine field is a universal influenza vaccine (9). Such a vaccine would obviate the need for an annual vaccine. Furthermore, if providing durable protection, such a vaccine could be given earlier in life, when the induction of immune responses

is more optimal, rather than later in life, when susceptibility to serious infection is higher (10) and vaccine efficacy significantly declines (11). The approach of Du *et al.* may be a step toward a universal influenza vaccine in that it will be safe and retain sufficient antigenic determinants to promote immune protection to multiple viral strains. However, many challenges remain. Data on cross-protection are limited to exposure to a small set of strains from the H1N1 and H3N2 subtypes of influenza viruses. It would be valuable to test additional viruses, including highly virulent avian subtypes such as H5N1 and H7N9, during subsequent challenge studies.

Another limitation is that the approach relies heavily on T cell immunity for cross-strain protection. However, much evidence suggests that B cell immunity in the form of antibodies can be very important in protection against influenza virus and indeed antibodies are the oft-used correlate of likely vaccine efficacy (12). In terms of the HIS approach and antibody-mediated responses, the diversity of the HA molecule is a problem because the approach is largely expected to induce strain-specific antibodies and the HA molecule rapidly mutates under immune selection pressure. It might be beneficial to combine the HIS approach with one that targets the induction of broadly neutralizing antibodies (bnAbs), which can bind and neutralize diverse influenza virus strains. For example, one could use emerging constructs being developed that promote the generation of HA bnAbs that recognize the stem re-

gion of the HA molecule, which is relatively conserved for different viral strains (9). Using a combined approach might increase the titer of difficult-to-induce bnAbs. Moreover, the added T cell immunity could further improve overall efficacy in humans. Indeed, there is evidence that both T and B cell immunity can contribute to vaccine protection against viral infections (12).

It will also be important to determine whether the approach of Du *et al.* can be extended to the generation of more effective live attenuated vaccines against other viruses. As part of such studies, the genetic analysis could be applied to other critical innate immune signaling pathways such as mutating viral proteins that attenuate the viral RNA-sensing retinoic acid-inducible gene-1 (RIG-I) and melanoma differentiation associated protein-5 (MDA-5) (13) as well as viral proteins that antagonize IFN-I signaling. ■

REFERENCES

1. D. R. Burton, *Cold Spring Harb. Perspect. Biol.* **9**, a030262 (2017).
2. J. R. Teijaro, *Adv. Immunol.* **132**, 135 (2016).
3. Y. Du *et al.*, *Science* **359**, 290 (2018).
4. A. García-Sastre, C. A. Biron, *Science* **312**, 879 (2006).
5. B. G. Hale *et al.*, *J. Gen. Virol.* **89**, 2359 (2008).
6. N. C. Wu *et al.*, *Sci. Rep.* **4**, 4942 (2014).
7. N. C. Wu *et al.*, *J. Virol.* **88**, 10157 (2014).
8. K. A. Hanley, *Evolution (N. Y.)* **4**, 635 (2011).
9. F. Krammer, P. Palese, *Nat. Rev. Drug. Discov.* **14**, 167 (2015).
10. K. Dorshkind, E. Montecino-Rodriguez, R. A. J. Signer, *Nat. Rev. Immunol.* **9**, 57 (2009).
11. J. E. McElhaney, J. P. Dutz, *J. Infect. Dis.* **198**, 632 (2008).
12. I. J. Amanna, M. K. Slifka, *Virology* **411**, 206 (2011).
13. D. C. Beachboard, S. M. Horner, *Curr. Opin. Microbiol.* **32**, 113 (2016).

10.1126/science.aar5421

Remote control of nanoscale devices

A DNA nanodevice can be manipulated with an applied electric field

By Björn Högberg

Processes that occur at the nanometer scale have a tremendous impact on our daily lives. Sophisticated evolved nanomachines operate in each of our cells; we also, as a society, increasingly rely on synthetic nanodevices for communication and computation. Scientists are still only beginning to master this scale, but, recently, DNA nanotechnology (1)—in particular, DNA origami (2)—has emerged as a powerful tool to build structures precise enough to help us do so. On page 296 of this issue, Kopperger *et al.* (3) show that they are now also able to control the motion of a DNA origami device from the outside by applying electric fields.

Self-assembled in a test tube from hundreds of single strands of DNA, DNA origami devices can form almost any conceivable shape at the nanoscale. Shapes in and of themselves, however, do not make a machine. There is increasing interest in nanotechnological approaches that can perform tasks on the nanoscale, particularly within the human body itself. Kopperger *et al.* make an impressive stride in this direction by creating a dynamic DNA origami structure that they can directly control from the macroscale with easily tunable electric fields—similar to a remote-controlled robot.

Kopperger *et al.* immobilized a typical three-dimensional (3D) DNA origami (4) platform attached to a flexible rod on a glass slide. This rod essentially looks like a gear stick that protrudes out of a square platform locked in place on the glass (see the figure). Most of the structure is made out of relatively rigid DNA double helices that are bundled together, but the stick itself is attached to the platform by flexible single-stranded DNA pieces and can thus rotate with respect to the fixed platform.

Under normal conditions, when dispersed in common water-based buffers, DNA is a charged molecule. Therefore, applying an electric field to the buffer solution surrounding the platform will force the charged DNA stick to move in the direction of the field (see the figure). This simple idea is the basis for the work. Earlier work has

used similar concepts (5), but the flexible arm in conjunction with the fixed platform in the present work allows the authors to show precise positional control with respect to the platform. To achieve this, the authors use a system of latches. These are short, protruding single-stranded DNAs that momentarily grab the arm and lock it in place at predefined positions. With these in place, the arm can only be forced into a new position by using enough applied electric force.

Again, the analogy with a gear stick might be appropriate. Think of these latches as the predefined shift pattern on a manual-transmission lever: once in position 1, some force is required to shift the stick to position 2, and so on. This construction allowed the researchers to precisely move the lever from position to position, despite the relatively rough alignment of the external electric field. Kopperger *et al.* demonstrate these and other capabilities of their system using total internal reflection fluorescence (TIRF) microscopy, an imaging technique that allows researchers to precisely track molecular motions.

Although the experiment in itself is fascinating, a lot of the promise lies in what could potentially be achieved when this approach is combined with other recently pub-

lished techniques. Combining the latching system with precise lithographic patterning and optical readout (6) could possibly form the basis for a new type of digital memory. Remote-controlled picking up and placing of molecular components could potentially be accomplished if this concept was combined with cargo-transfer techniques (7, 8). If this became possible, one could imagine a future where molecules could be put together following orders from an outside operator. DNA origami is effectively already a 3D-printer system for the nanoscale (9, 10), but with some of these extensions, Kopperger *et al.*'s technique could enable true 3D printing of molecules. ■

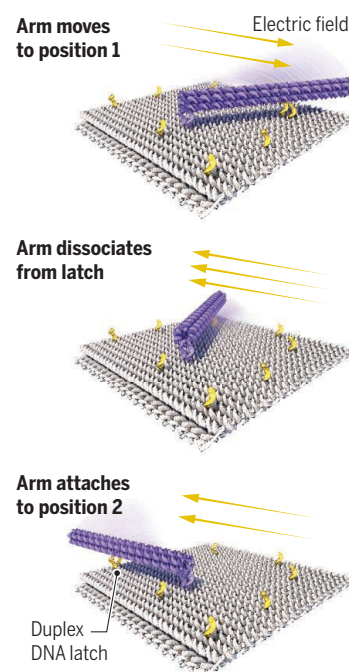
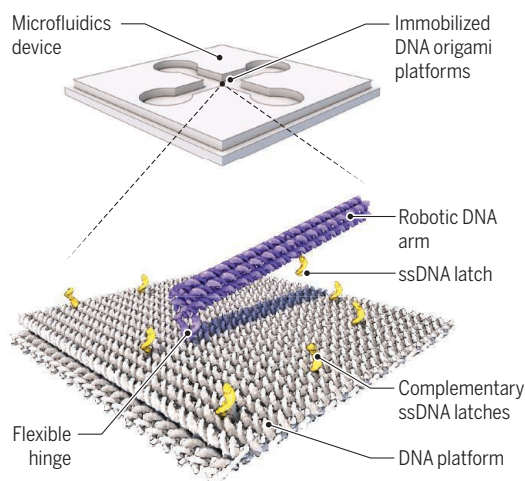
REFERENCES

1. N. C. Seeman, *J. Theor. Biol.* **99**, 237 (1982).
2. P. W. K. Rothemund, *Nature* **440**, 297 (2006).
3. E. Kopperger *et al.*, *Science* **359**, 296 (2018).
4. S. M. Douglas *et al.*, *Nature* **459**, 414 (2009).
5. Y. Klapper, N. Sinha, T. W. S. Ng, D. Lubrich, *Small* **6**, 44 (2010).
6. A. Gopinath, E. Miyazono, A. Faraon, P. W. K. Rothemund, *Nature* **535**, 401 (2016).
7. A. J. Thubagere *et al.*, *Science* **357**, eaan6558 (2017).
8. S. Kassem, A. T. L. Lee, D. A. Leigh, A. Markevicius, J. Solà, *Nat. Chem.* **8**, 138 (2016).
9. S. M. Douglas *et al.*, *Nucleic Acids Res.* **37**, 5001 (2009).
10. E. Benson *et al.*, *Nature* **523**, 441 (2015).

10.1126/science.aar6580

Controlled nanoscale motion

Kopperger *et al.* have created a DNA arm that can be rotated with an external electric field. Single-stranded DNAs (ssDNAs) that hybridize to each other act as latches that fix the arm in position until sufficient electric force is applied to rotate the arm.



Department of Medical Biochemistry and Biophysics,
Karolinska Institutet, S-171 77 Stockholm, Sweden.
Email: bjorn.hogberg@ki.se

RETROSPECTIVE

Ben Barres (1954–2017)

A passionate neuroscientist and advocate of equal opportunity in the sciences

By **Martin Raff**

Ben Barres transformed our understanding of brain glial cells. He lived an extraordinary life and died too young after a 2-year battle with pancreatic cancer. From early childhood, he suffered greatly from gender dysphoria, until transitioning from Barbara to Ben at age 43. This provided enormous relief, as well as rare insight into how differently females and males are treated, fueling a passionate intolerance to prejudice. He was the first openly transgender person elected to the U.S. National Academy of Sciences.

Barres was born and raised in West Orange, New Jersey, with the given name Barbara. Neither of his parents went to university, and they and their three other children showed no interest in science, whereas Barres wanted to be a scientist by age 5, excelled in science and math at school, and took extra classes in these subjects at nearby universities. As a scholarship undergraduate student at the Massachusetts Institute of Technology, he took a brain science course that sparked his interest in neuroscience. After receiving a medical degree at Dartmouth Medical School, he completed a neurology residency at the Cornell Cooperating Hospitals in New York.

Having decided to become a neuroscientist, he entered the neuroscience Ph.D. Program at Harvard Medical School. After several lab rotations, he chose to work with David Corey, an auditory hair-cell physiologist with expertise in the patch-clamp technique that Barres wanted to use to hunt for ion channels in glial cells in the mammalian central nervous system (CNS)—cells that had become his passion since his neurology residency. While the Corey lab was a bold choice, as it had no interest or experience in glial cells, Barres's earlier decision to devote his life to understanding these cells

was even bolder, as they were thought to be uninteresting and relatively unimportant. Barres, however, suspected that they were far more important than recognized and spent the next 34 years proving it.

Early in his graduate work, Barres realized the advantages of identifying and isolating the different types of CNS cells and using cell-type-specific antibodies to do so. This approach enabled him to use patch-clamp recording to discover an unexpected variety of ion channels in the plasma membrane of three types of glial cells—astrocytes, whose functions were mostly unknown; oligodendrocytes, which myelinate CNS axons; and oligodendrocyte precursor cells (OPCs). The findings were new and important, making him an internationally recognized player in the glial field.

Barres next spent 3 years as a postdoc in my lab at University College London, where he studied the early development of oligodendrocytes. Working largely independently, he devised a method to purify OPCs from developing rat optic nerve and to culture them at low density under conditions where they would survive, proliferate, and differentiate into oligodendrocytes. This allowed him to show how various known extracellular signaling molecules influence these three processes at the level of individual cells. Among his many discoveries at University College London was that newly formed oligodendrocytes in the developing rat optic nerve normally die in large numbers in a competition for limiting amounts of survival signals on axons—an elegant mechanism for matching the number of oligodendrocytes to the number and length of axons during CNS development.

The most important discoveries, however, came after starting his own lab at Stanford University at age 39, where Barres remained for the rest of his career, attracting many outstanding students and postdocs and transitioning to Ben after 4 years. The Barres lab made a seemingly endless series of astonishing discoveries about how

glial cells function in the normal and diseased rodent and human brain.

Most of the lab members studied glial-neuronal interactions, working out ways to purify different cell types and keep them alive in culture, which took years of pioneering work. Once they had demonstrated a glial cell influence on a neuron, it took even longer to identify the molecules mediating the interaction—usually signaling molecules secreted by the glial cell and the complementary receptors on the neuronal surface. They then used mice deficient in these molecules to verify their roles *in vivo*.

I can mention only a few of the many transforming discoveries from the Barres lab. They showed that astrocytes are required for many CNS neuronal synapses to form, mature, and function normally, and they identified the signal molecules and receptors mediating these interactions. The Barres lab also demonstrated that both astrocytes and microglia (the specialized macrophage-like glial cells in the CNS) have crucial and distinct roles in synapse elimination, which occurs during normal neuronal development and in many neurodegenerative disorders; moreover, they identified distinct phagocytic pathways involved in each case, including the surprising role of complement proteins in the microglial pathway. Recently, they discovered a type of reactive astrocyte that develops in some forms of brain injury, which secretes toxins that kill injured neurons and oligodendrocytes. This type of astrocyte is greatly increased in various neurodegenerative diseases, where it is thought to contribute to neuronal cell death, making it and its toxins potential targets for drug discovery.

Barres's influence extended well beyond his lab's discoveries. He was an esteemed teacher, mentor, and chair of Stanford's Neurobiology Department. He organized glial meetings and courses and wrote many perceptive reviews. He made his lab's invaluable experimental protocols, reagents, and cell-type-specific transcriptome databases freely available, which greatly accelerated progress in the field. He fought energetically and effectively for the many issues he was passionate about, including the plight of women, minorities, and transgender individuals in society and especially in science.

Ben's death leaves a gaping hole in the glial field he dominated for so long. He is irreplaceable: His intelligence, passion, generosity, and childlike enthusiasm and frankness will be deeply missed by the remarkably large number of us who cherished him and owe him so much. ■



MRC Laboratory for Molecular Cell Biology, University College London, London WC1E 6BT, UK. Email: m.raff@ucl.ac.uk

10.1126/science.aas9270



A complex, colonial organism forms when individual red-spotted siphonophores come together.

COMPLEXITY

Quarks, culture, combogenesis

A multidisciplinary tour of cosmic history charts the “grand sequence” of existence

By **Barry Wood**

The value of Tyler Volk’s *Quarks to Culture* is evident when the book is placed against popular histories of the universe, dozens of which have provided evidence for an immense cosmic past. But such histories are often anecdotal, like early British histories of the kings of England. Unlike these works, Volk artfully presents the case for structural continuity and systematic creativity across 13.8 billion years of cosmic history.

He begins with a simple observation: “As you go down into the body, you go back in time: from the body inward to cells, to molecules, and then to atoms. Passing from life to physics, each first type in this series of nested things came into existence earlier.” The primary achievement of his book is the clear articulation of the temporal sequence in which the smallest particles combine to form atoms, which form molecules, then cells, and eventually organisms. He further extends his analysis to dynamically related tiers that ascend to animal herds and hives, tribal associations, and eventu-

ally large societies. Volk goes beyond current concepts of emergence, complexity, self-organization, and autopoiesis with a sustained and impressive presentation of “combogenesis,” his own term for the innovative creativity of the physical, biological, and cultural realms. The fact that he accomplishes this grand sweep within just

250 pages makes the book a superb contribution deserving of wide readership.

Volk’s study traces the sequence of combogenesis in 12 chapters. A series of parallel diagrams at the head of each chapter clarifies the combinatory steps from each level to the next. Acknowledging familiar meta-patterns of layers, hierarchies, thresholds, and domains, he emphasizes the structural relation of these tiers as “nested.” The

term is most obviously applicable to quarks, nucleons, and atoms but takes on metaphoric richness in the higher biological and cultural realms. Principles that may seem specific to agrovillages or geopolitical states are nested within the governing limitations of earlier structures while manifesting the creative integration evident at higher levels.

A recent emphasis on “big history,” which ranges over similar territory, is recast as “grand sequence” in *Quarks to Culture*—a term that bypasses the march of events in



Quarks to Culture
How We Came to Be
Tyler Volk
Columbia University
Press, 2017. 280 pp.

order to highlight the mystery of temporal change as the fundamental reality. Volk skips the usual rehearsal of geological eras, land colonization, continental drift, and dinosaur extinction, focusing instead on the big picture. In doing so, he manages to frame a complex welter of multidisciplinary information as a “narrative of the universe” that is creative and compelling.

Despite the book’s emphasis on storytelling, Volk must occasionally resort to numbers to impress upon the reader the complexity of contemporary existence. A suite of fundamental particles is organized into 92 elements, he writes, which are combined in hundreds of molecules before the rise of life. Millions more come into being after the rise of life, and perhaps 100 billion can be found in what Volk calls the “protein sequence space.” His rigorous philosophical emphasis assumes that his readers are already aware of the vastness of time and space, nuclear fusion in stars, the evolution of the solar system, and theories of life’s origin.

Communication, Volk shows, also has a deep history: from animal calls to speech among tribal associations and agrovillages to writing in geopolitical states. In Chapter 16, he introduces the concept of “alphakits”: 26 letters and 40 phonemes that have given rise to thousands of words and millions of linguistic artifacts and provided a foundation for cultural construction, preservation, and transfer. But the term applies to lower levels, too, he argues: A suite of quarks, elements, amino acids, and proteins constitutes early alphakits for evolution.

In a recent *Science* Editorial, Gordon McBean and Alberto Martinelli noted, “Despite decades of efforts toward better integration, much of society still presumes a stark divide between the disciplines, and most scientists continue to be trained, evaluated, and rewarded in disciplinary silos” (1). The “grand sequence” advocated in *Quarks to Culture* is, as they and others have argued, largely unrecognized. Volk, however, marshals evidence from a dozen disciplines filtered through a rigorous intellect with grace and without polemics.

With impressive learning, rigorous analysis, and artful writing, *Quarks to Culture* presents a unifying philosophy reminiscent of Alfred North Whitehead’s exploration of process. ■

REFERENCE

1. G. McBean, A. Martinelli, *Science* **358**, 975 (2017).

10.1126/science.aar8252

The reviewer is at the Department of English, University of Houston, Houston, TX 77204, USA. Email: bwood@uh.edu

SCIENCE LIVES

Enrico Fermi, flaws and all

A revealing biography falls short when it comes to the famous physicist's problematic treatment of women

By Megan Formato

With the title *The Last Man Who Knew Everything* and a first chapter entitled “Prodigy,” a reader could be forgiven for expecting David Schwartz’s new biography of Enrico Fermi to be a straightforward hagiography. Luckily, Schwartz’s ambitions are not as simple as providing yet another account of a great man of 20th-century physics. He has other, thornier questions in mind, some of which he credibly addresses and others that he handles less convincingly.

Moving from Fermi’s birth in Rome in 1901 to his death in Chicago in 1954, *The Last Man* grapples with Fermi’s legacy as a teacher and mentor, his contributions as a scientist in Mussolini’s Italy and later to the Manhattan Project in the United States, and his change of heart about the hydrogen bomb after World War II. (Fermi initially opposed the project but later worked on it.)

We learn how Fermi’s science was shaped by his family, by the many educational and scientific institutional contexts through which he moved, and by major political moments in both Italy and the United States.

The Last Man Who Knew Everything is at its best in the chapters devoted to the Manhattan Project. Following Fermi gives us a vantage point that is, at least initially, more focused on Columbia University and the University of Chicago than Los Alamos.

Shifting the focus to these locations allows us to see in great detail the institutional cultures and social worlds that the Fermis helped maintain. Here, Schwartz’s science communication skills also shine with patient, easy to follow, nontechnical descriptions of the construction of fission piles.

One of Schwartz’s sustained insights is that “[c]alculations of probabilities run like a bright thread throughout [Fermi’s] work.” It is particularly helpful to see his approach to the Manhattan Project through this statistical lens.

Fermi, we learn, was initially disinclined to keep the Manhattan Project a secret because he thought the development of such a weapon was statistically unlikely. His understanding of the probability of a fission weapon shifted during a conversation with the German physicist Werner Heisenberg in Michigan in 1939.



Enrico Fermi sits with his wife, Laura, an author and historian, in 1954.

Schwartz’s account of this exchange is one of the most illuminating in the book because it reveals a turning point for Fermi and therefore for the Manhattan Project.

Fermi and Heisenberg disagreed on what kind of agency and dissent is possible for scientists working under fascist governments. When Heisenberg returned to Germany, Fermi was convinced that he would work on a fission weapon for Hitler, a belief that instilled a sense of greater urgency in Fermi with regard to his own work.

The book is less compelling in early chapters when, in order to persuade the

The Last Man Who Knew Everything
The Life and Times of Enrico Fermi, Father of the Nuclear Age

David Schwartz
Basic Books, 2017. 477 pp.



reader that Fermi’s genius was recognized in its own time, Schwartz strays more frequently into cliché and hyperbole. It is hard to square the description of Fermi as “already a legend” or “the Voltaire of physics” during his time at Scuola Normale and university in Pisa alongside nearly simultaneous mentions that he was almost expelled for a prank and had difficulty on his exams.

More troubling is Schwartz’s handling of Fermi’s treatment of women. Describing a letter Fermi wrote to a male friend about a skit in which Fermi “ridiculed [women]—

‘barring one or two exceptions ugly enough to scare anybody’—by portraying them as incapable of reducing a simple fraction,” Schwartz acknowledges Fermi showed “a distinctly unattractive attitude toward his fellow female students.”

However, he quickly sets the skit aside, writing: “Perhaps it reflected merely the widely shared prejudices of his time and culture. More than likely, it was also a bravado with which he could mask his awkwardness around women, an ineptness he would eventually outgrow.”

Fermi’s lifelong teasing of his wife is similarly acknowledged throughout the book but then minimized as an “idiosyncrasy” that did not interfere with their marriage. By raising some of Fermi’s misogynistic behaviors only to repeatedly brush them aside, Schwartz props up the

tired, problematic trope of the eccentric male genius who is not held responsible for his destructive social behavior.

At its best, *The Last Man Who Knew Everything* resists drawing too clean a line between the personal and the scientific in order to explore how Fermi’s contributions as a scientist were contingent on “the specific circumstances” of his life. But by not letting Fermi’s mistreatment of women more thoroughly inform the biography, Schwartz falls short of his ambition to provide a complex portrait of the famous physicist. ■

The reviewer is at the Program in Writing and Rhetoric, Stanford University, Stanford, CA 94305, USA.
Email: mformato@stanford.edu

10.1126/science.aar2948

LETTERS

Edited by Jennifer Sills



The pitfalls of taking science to the public

Daily newspapers play a fundamental role in informing the public about research findings and discussions. Scientists, in turn, are increasingly expected to interact with the media. However, this system is devoid of peer review, and it is the editors, not authors, that often have the final say on how a piece is worded.

On 22 November 2017, *The Washington Post* published a perspective piece on the inevitability of large-scale extinctions (1). Stiff criticism followed on the comments page, social media, and blogs (2–5). We launched an effort to show that the views in that article do not represent the scientific majority (6). On 15 December, *The Washington Post* ran our response representing more than 3000 scientists from 88 countries, including many prominent scientists and Nobel laureates (7). Meanwhile, the perspective's author posted an essay on his laboratory blog expressing regret about how his words had been edited and interpreted (8). In a sense, the scientific record has been corrected. However, public debate is a matter of timing and scale. Our response ran after the peak of the original discussion, and the author's later comments ran on a local platform likely viewed by far fewer people than the perspective. Hence, the information that reached readers was fragmentary and skewed.

This event exemplifies why scientists,

before they approach the media, should understand the differences between publishing in scientific journals and publishing in news outlets. In the absence of peer review, it is essential to ask for advice about how text about scientific concepts could be misconstrued by nonspecialists. It is also helpful to suggest titles, which are often written by news staff who aim for accuracy but risk making sensationalistic statements. Scientists must also weigh the power of joint initiatives with the impact of speedy responses by individuals.

We urge scientists to continue engaging with the media, but beware these pitfalls. Ask for advice from colleagues or your communications office, request review and approval of any text before publication, and hurry up—without compromising the quality of your contribution.

Alexandre Antonelli^{1,2,3,4} and Allison Perrigo^{1,2*}

¹Gothenburg Global Biodiversity Centre, 405 30, Gothenburg, Sweden. ²Department of Biological and Environmental Sciences, University of Gothenburg, 405 30, Gothenburg, Sweden. ³Gothenburg Botanical Garden, Gothenburg, Sweden.

⁴Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02139, USA.

*Corresponding author.

E-mail: allison.perrigo@bioenv.gu.se

REFERENCES

1. R. A. Pyron, "We don't need to save endangered species: Extinction is part of evolution," *Washington Post* (2017).
2. S. Claramunt, "Extinction, macroevolution, and biodiversity conservation" (2017); <https://sites.google.com/site/sclaramuntuy/blog/extinctionmacroevolutionandbiodiversityconservation>.
3. C. Tucker, "Of course we need to save endangered species: a response," *The EEB and Flow* (2017); <http://evol-eco.blogspot.se/2017/11/of-course-we-need-to-save-endangered.html>.

<http://evol-eco.blogspot.se/2017/11/of-course-we-need-to-save-endangered.html>.

4. J. Atkins, "Saving species from ourselves," *PLOS Ecology Community* (2017); <http://blogs.plos.org/ecology/2017/12/01/saving-species-from-ourselves/>.
5. J. Schimel, "Do species matter: Responding to an op-ed by R. A. Pyron in the *Washington Post* as a piece of writing," *Writing Science* (2017); <https://schimelwritingscience.wordpress.com/2017/11/30/do-species-matter-responding-to-an-op-ed-by-r-a-pyron-in-the-washington-post-as-a-piece-of-writing/>.
6. A. Perrigo, A. Antonelli, "Don't give up on biodiversity!" (2017); www.protect-biodiversity.com.
7. A. Antonelli, A. Perrigo, "We must protect biodiversity," *Washington Post* (2017); www.washingtonpost.com/opinions/2017/12/15/53e6147c-e0f7-11e7-b2e9-8c636f076c76_story.html.
8. R. A. Pyron, "Biodiversity conservation is urgent and important, now and for the future" (2017); <http://colubroid.org>.

10.1126/science.aar8468

Vaccine mandates in France will save lives

In their Letter "France's risky vaccine mandates" (27 October 2017, p. 458), J. K. Ward *et al.* question the adoption of mandatory vaccination in France. Their prediction that such a step will encourage resistance to vaccination is unsupported by the facts and could prolong a dangerous situation in which French citizens have the right to allow their children to catch and transmit potentially fatal infections.

The French recommendation—which has now gone into effect (1)—was the product of two juries composed of both medical professionals and lay citizens (2), suggesting that Ward *et al.*'s concerns about acceptance by doctors and the public are unfounded. Moreover, evidence shows that mandates are effective. In California, immunization rates increased after so-called "philosophical exemptions" were eliminated (3).

Vaccine-hesitancy in French physicians has been found to be only moderate in prevalence (4). Even one vaccine-hesitant doctor is too many, but Ward *et al.* do



New French vaccine mandates are now in effect.

not offer a solution to the problem, such as better education by medical schools. Furthermore, a reference cited by Ward *et al.* does not, as they claim, show that mandating vaccines increases anti-vaccinationism, but rather that citing dangers of diseases is more effective than arguing for safety of vaccination (5).

Ward *et al.*'s reasoning could be extrapolated to argue against mandating car seats for young children. Like car seats, vaccination mandates will likely save lives.

**Stanley A. Plotkin,^{1,2} Paul Offit,^{3,4}
Pierre Bégué⁵**

¹University of Pennsylvania, Philadelphia, PA 19104, USA. ²Vaxconsult, Doylestown, PA 18902, USA. ³Children's Hospital of Philadelphia,

Philadelphia, PA 19104, USA. ⁴Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA. ⁵Académie Nationale de Médecine, Paris 75006, France.
*Corresponding author.
Email: stanley.plotkin@vaxconsult.com

REFERENCES

1. Association Française de Pédiatrie Ambulatoire, *Obligation Vaccinale: Ce Qu'il Faut Savoir* (2018); <https://afpa.org/obligation-vaccinale/> [in French].
2. Concertation Citoyenne sur la Vaccination, "Rapport sur la Vaccination" (2016); <http://concertation-vaccination.fr/la-restitution/>.
3. L. Sun, "California vaccination rate hits new high after tougher immunization law," *Washington Post* (2017).
4. P. Verger *et al.*, *Euro Surveill.* **27**, 30406 (2016).
5. Z. Horne *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10321 (2015).

10.1126/science.aar4422



Young Buddhist monks eat lunch in their monastery.

LIFE IN SCIENCE

Have your momos and eat them, too

Sitting at breakfast with a group of Tibetan Buddhist monks at a monastery in southern India, I am surprised to see a plate of chicken on the table. "Is vegetarianism not required at the monastery?" I ask. Their reply is unexpected: Although no meat is served in the monastery dining hall, individual monks can choose to eat meat elsewhere, including the monastery's guesthouse where I and my fellow scientists are staying.

On my previous visit to northern India, I was similarly surprised to find monks leaving the monastery to eat Tibetan meat buns ("momos") at nearby restaurants. At the time, I thought this reflected the attitudes of the monks themselves; they liked eating meat so much that they would leave the monastery to do it! But with this second trip and the additional context, I realize that the situational flexibility of meat eating is part of Tibetan Buddhism itself. When the monks are at the monastery, they eat no meat because the monastery is vegetarian. But that doesn't mean that they are unable to leave and eat meat on their own.

This experience shed light on my research, an exploration of ethical issues in neuroscience with the monastics as part of the Science for Monks program. I had asked individual monks for help solving ethical puzzles faced by scientists in the United States. Often their responses would avoid the puzzle altogether. When I asked whether it was right for people to augment their abilities with medication, they wondered why people would use enhancement technologies in the first place if they were neither free of negative side effects nor effective in the long term. What if people didn't debate whether to eat meat or not, but assessed how important it was to eat meat in each situation? The monks, happily eating their chicken for breakfast, seem perfectly content with this arrangement.

Laura Ellen Specker Sullivan

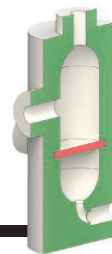
Center for Bioethics, Harvard Medical School, Boston, MA 02115, USA.
Email: laura_speckersullivan@hms.harvard.edu

10.1126/science.aar8472

RESEARCH

3D-printed interconnected plastic modules for chemical production

Kitson et al., p. 314



IN SCIENCE JOURNALS

Edited by **Stella Hurtley**



Migrants walking through the Little Ethiopia area of Los Angeles, California

SOCIAL SCIENCE

Data-driven refugee assignment

The continuing refugee crisis has made it necessary for governments to find ways to resettle individuals and families in host communities. Bansak *et al.* used a machine learning approach to develop an algorithm for geographically placing refugees to optimize their overall employment rate. The authors developed and tested the algorithm on segments of registry data from the United States and Switzerland. The algorithm improved the employment prospects of refugees in the United States by ~40% and in Switzerland by ~75%. —BJ

Science, this issue p. 325

VASCULAR BIOLOGY

Lymphatics limp along after MRSA

Lymphedema is associated with skin and soft tissue infections, and both can be recurring, causing continual suffering in affected patients. To better understand the relationship between bacterial infections and lymphedema, Jones *et al.* used intravital imaging. They examined the lymphatics of mice infected with MRSA (methicillin-resistant *Staphylococcus aureus*) and observed lymphatic muscle cell death, which led to prolonged dysfunction months after the bacteria had been cleared. In vitro experiments with human cells indicated bacterial toxins were responsible

for damaging the lymphatic muscle cells. The findings suggest that these bacterial toxins could be targeted in patients to interrupt this brutal cycle. —LP
Sci. Transl. Med. **10**, eaam7964 (2018).

NANOROBOTICS

Electrically driving a DNA arm

Most nanoelectromechanical systems are formed by etching inorganic materials such as silicon. Kopperger *et al.* improved the precision of such machines by synthesizing a 25-nm-long arm defined by a DNA six-helix bundle connected to a 55 nm-by-55 nm DNA origami plate via flexible single-stranded scaffold crossovers (see the Perspective by Hogberg). When placed in a cross-shaped

electrophoretic chamber, the arms could be driven at angular frequencies of up to 25 Hz and positioned to within 2.5 nm. The arm could be used to transport fluorophores and inorganic nanoparticles. —PDS

Science, this issue p. 296;
see also p. 279

SYNTHETIC BIOLOGY

Large-scale gene synthesis in tiny droplets

Gene synthesis technology is important for functional characterization of DNA sequences and for the development of synthetic biology. However, current methods are limited by their low scalability and high cost. Plesa *et al.* developed a gene synthesis method, DropSynth, which uses

barcoded beads to concentrate oligos and subsequently assemble them into synthetic genes within picoliter emulsion droplets. DropSynth allows generation of large libraries of thousands of genes and functional testing of all possible mutations of a particular sequence. —SYM

Science, this issue p. 343

VACCINES

Avoiding interferon avoidance

Interferon (IFN) expression is a mammal's first response to viral infection. Many viruses have thus evolved mechanisms to evade IFN. Du *et al.* developed a method to systematically ablate IFN evasion genes from live, attenuated influenza virus

(see the Perspective by Teijaro and Burton). A combination of mutants was assembled to construct a virus that triggered transient IFN responses in mice but that was unable to replicate effectively. The transient IFN responses led to robust antibody and memory responses that protected against subsequent challenge with different influenza viruses. This approach could be adapted to improve other RNA virus vaccines. —CA

Science, this issue p. 290;
see also p. 277

QUANTUM FLUIDS

Making dilute quantum droplets

In recent years, quantum fluids have been studied largely in gaseous form, such as the Bose-Einstein condensates (BECs) of alkali atoms and related species. Quantum liquids, other than liquid helium, have been comparatively more difficult to come by. Cabrera *et al.* combined two BECs and manipulated the atomic interactions to create droplets of a quantum liquid (see the Perspective by Ferrier-Barbut and Pfau). Because the interactions were not directional, the droplets had a roughly round shape. The simplicity of this dilute system makes it amenable to theoretical modeling, enabling a better understanding of quantum fluids. —JS

Science, this issue p. 301;
see also p. 274

STRUCTURAL BIOLOGY

Recognizing centromere by kinetochore

The kinetochore proteins CENP-N and CENP-C recognize the histone H3 variant CENP-A in the centromeric nucleosome. This ensures proper kinetochore assembly and accurate segregation of chromosomes. Chittori *et al.* describe the cryo-electron microscopy structure of the human CENP-A nucleosome–CENP-N complex. The interaction of CENP-N with CENP-A and the nucleosomal DNA together

ensure specific and stable centromeric nucleosome recognition. Mutational analyses using both human and *Xenopus* CENP-A and CENP-N proteins suggest that the proteins have coevolved to preserve the interacting surfaces. —SYM

Science, this issue p. 339

INDUCED SEISMICITY

Seismicity curbed by lowering volume

Determining why hydraulic fracturing (also known as fracking) triggered earthquakes in the Duvernay Formation in Canada is important for future hazard mitigation. Schultz *et al.* found that injection volume was the key operational parameter correlated with induced earthquakes in the Duvernay. However, geological factors also played a considerable role in determining whether a large injection volume would trigger earthquakes. These findings provide a framework that may lead to better forecasting of induced seismicity. —BG

Science, this issue p. 304

RESEARCH METHODS

Algorithms fail to improve predictions

In the United States, algorithms are commonly used to predict the likelihood that a criminal defendant will commit a crime, and these predictions influence pretrial, parole, and sentencing decisions. Commercial software, such as the widely used COMPAS system, promises to make these predictions more accurate than human judgments. Dressel and Farid show that COMPAS's impressive-sounding 137-feature black box is nearly equivalent to a trivial linear classifier using two features, and both approaches are no more accurate or fair than predictions made by people with little or no criminal justice expertise. —AC

Sci. Adv. 10.1126/sciadv.aao5580 (2018).

IN OTHER JOURNALS

Edited by **Caroline Ash**
and **Jesse Smith**



GEOPHYSICS

Going dry in the Pacific Northwest

Volcanic belts such as the Andes result from deep melting as water dragged down during subduction fluxes into the crust. Canales *et al.* show that the Juan de Fuca slab, which is subducting below the Pacific Northwest in North America, is much drier than other subducting slabs. The distribution of water in the slab may help determine the origins of seismic tremor and episodic slip that occur in this region. It also confirms a hypothesis that volcanism in the region is not the result of the influence of water, but rather is due to the decompression trigger melting more commonly seen along mid-ocean ridges. —BG

Nat. Geosci. 10, 864–870 (2017).

MOLECULAR BIOLOGY

Time-out for mRNAs in the nucleus

Cell cycle events are precisely orchestrated to ensure accurate

cell division. Yang *et al.* have discovered that sequestering mature mRNAs in the nucleus modulates cell cycle players. In dividing *Arabidopsis* cells, nuclear retention of CDC20 and CCS52B mRNAs prevents them from being released into the cytoplasm until the nuclear envelope breaks down at pro-metaphase. Released mRNAs are rapidly translated into proteins, ensuring their regulatory functions at the proper cell cycle stage. Similar nuclear sequestration strategies may be used for other mRNAs in different cellular contexts. —SYM

Mol. Cell. 10.1016/j.molcel.2017.11.008 (2017).

NEUROSCIENCE

Serious damage by soluble tau

Alterations in the metabolism of the neuronal microtubule-associated protein tau are central to several neurodegenerative diseases. In these diseases, tau usually loses solubility and forms aggregates that impair cell

PHOTO: ONDREJ PROSICKY/SHUTTERSTOCK.COM



CONSERVATION

How hunting affects brown bear populations

In many parts of the world, regulated hunting is used to control the size of predator populations such as wolves and brown bears. Bischof *et al.* explore how such regulated hunting affects the life history and demography of a brown bear population in Sweden that has been monitored continuously since 1985. The study shows that hunting was the leading cause of death for bears aged more than 3 years, resulting in reduced life expectancy; this contrasts with natural conditions, where mortality is reduced once bears reach adulthood. Hunting also substantially reduces the reproductive value—that is, the number of future offspring that female bears of a given age are expected to have. Thus, even if a carnivore population recovers numerically, regulated hunting transforms its makeup in multiple ways that need to be taken into account in management. —JFU

Nat. Ecol. Evol. **2**, 116 (2018).

Hunting is the leading cause of death for brown bears older than 3 years in Sweden.

function to trigger neuronal cell death and neurodegeneration. However, the *in vivo* neurotoxic potential of soluble tau is not yet fully understood. Bolós *et al.* stereotactically injected human soluble tau into the dentate gyrus of mice. Hippocampal granule neurons showed markedly reduced synapse numbers in the molecular layer. In addition, newborn granule cells showed reduced numbers of dendritic spines. Behaviorally, these animals exhibited an impaired capacity to perform pattern separation. Soluble tau thus causes long-term damage to the morphology and connectivity of newborn granule cells. —PRS

Transl. Psych. 10.1038/s41398-017-0013-6 (2017).

RNA FOLDING

Lighting up riboswitching

RNAs fold as they are synthesized, and this folding is required for function. Uhm *et al.* describe a single-molecule fluorescence energy transfer assay to monitor cotranscriptional RNA folding.

This approach revealed folding in the thiamine pyrophosphate (TPP) riboswitch that regulates translation of genes involved in the synthesis of thiamine, an essential vitamin. The riboswitch folds into the “off” conformation, in which translation is inhibited, even in the absence of the TPP ligand. If TPP is not bound to this off conformation, it can switch to the “on” conformation when transcription pauses near the translation start codon, and this allows translation to start. TPP binding stabilizes the off conformation and prevents the switch. The assay will allow investigation of other cases in which transcriptional speed and pausing affect RNA folding. —VV

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.1712983115 (2017).

BRAIN DEVELOPMENT

The value of scaffolds

The brain is built by groups of neurons that migrate and interdigitate to form layers and circuits. This process varies in different phyla of animals. García-Moreno *et al.* draw lessons from

the development of the chick brain to understand what makes the mammalian brain distinctive. In mammals, excitatory glutamatergic neurons born deep in the brain migrate radially to the cortex, whereas inhibitory GABAergic interneurons born elsewhere migrate tangentially across the

cortex. And, like the external scaffolds on a building under construction, some glutamatergic neurons migrate tangentially, instruct organization, then disappear. The developing chick brain, although it has tangentially migrating interneurons, lacks the tangentially migrating transient neurons. —PJH

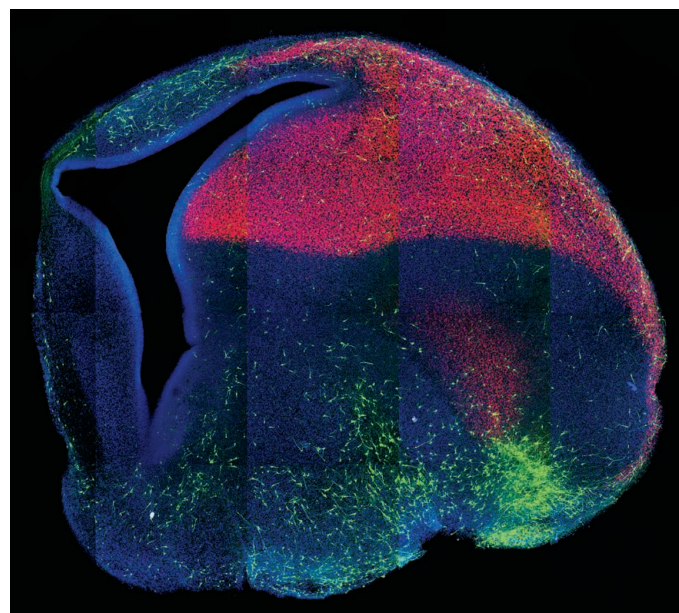
Cell Rep. **22**, 96 (2018).

SCIENCE CAREERS

Social skills to pay the bills

Employment requiring high math skills but low social skills, including many science and engineering jobs, has decreased in the United States as high social skills have become increasingly powerful predictors of employment and wage growth. Using surveys of occupations, skills, and wages, Deming shows that socially skilled people self-select into less structured jobs requiring a wide range of tasks, leading to wage gains. Increasing computerization may be a driver, replacing routine work and prioritizing social collaboration, but employment and wages have been especially strong in jobs demanding both high math and high social skills. —BW

Quart. J. Econ. **132**, 1593 (2017).



Neurons migrate to form surface layers of the chick brain.

ALSO IN SCIENCE JOURNALS

Edited by Stella Hurtley

CHEMICAL ENGINEERING

A plastic plan for organic synthesis

The infrastructure for chemical synthesis typically lies at either end of a spectrum: small-scale studies in ad hoc assemblies of glassware or large-scale production in capital-intensive custom reactors. Kitson *et al.* report a hybrid protocol that customizes a blueprint for synthesis of a target compound in a series of interconnected plastic modules, which can be assembled en masse by 3D printing (see the Perspective by Hornung). The approach, demonstrated for the commercial muscle relaxant baclofen, establishes a systematic workflow that is potentially amenable to automation: All that is necessary for synthesis and purification is the introduction of stock solutions and variation of temperature or pressure. —JSY

Science, this issue p. 314;
see also p. 273

MEMBRANE PROTEINS

Making your way through the side of a barrel

The mechanism of membrane insertion and assembly of β -barrel proteins is a central question of outer membrane biogenesis of mitochondria, chloroplasts, and Gram-negative bacteria. Höhr *et al.* developed assays to address this fundamental problem. They systematically mapped precursor proteins transported by the mitochondrial Omp85 channel (Sam50) to elucidate the entire membrane insertion pathway of a precursor in the native membrane environment. Their findings directly demonstrate translocation of precursor proteins through the lumen of the mitochondrial Omp85 channel, signal recognition by β -strand exchange between channel and precursor, and exit through the lateral gate into the membrane. —SMH

Science, this issue p. 289

BIOPHYSICS

Watching single molecules in motion

Structural techniques such as x-ray crystallography and electron microscopy give insight into how macromolecules function by providing snapshots of different conformational states. Function also depends on the path between those states, but to see that path involves watching single molecules move. This became possible with the advent of single-molecule Förster resonance energy transfer (smFRET), which was first implemented in 1996. Lerner *et al.* review how smFRET has been used to study macromolecules in action, providing mechanistic insights into processes such as DNA repair, transcription, and translation. They also describe current limitations of the approach and suggest how future developments may expand the applications of smFRET. —VV

Science, this issue p. 288

MAGNETIC MATERIALS

Boosting chiral nanoparticle responses

Optical nanomaterials that combine chirality and magnetism are useful for magneto-optics and as chiral catalysts. Although chiral inorganic nanostructures can exhibit high circular dichroism, modulating this optical activity has usually required irreversible chemical changes. Yeom *et al.* synthesized paramagnetic cobalt oxide (Co_3O_4) nanoparticles with L- and D-cysteine surface ligands. These ligands created chiral distortions of the crystal lattices, and this anisotropy led to much stronger chiroptical activity. The circular dichroism in the ultraviolet of nanoparticle gels could be modulated with magnetic fields of ~1.5 tesla. —PDS

Science, this issue p. 309

BIOGEOGRAPHY

A global map of soil bacteria

Soil bacteria play key roles in regulating terrestrial carbon dynamics, nutrient cycles, and plant productivity. However, the natural histories and distributions of these organisms remain largely undocumented. Delgado-Baquerizo *et al.* provide a survey of the dominant bacterial taxa found around the world. In soil collections from six continents, they found that only 2% of bacterial taxa account for nearly half of the soil bacterial communities across the globe. These dominant taxa could be clustered into ecological groups of co-occurring bacteria that share habitat preferences. The findings will allow for a more predictive understanding of soil bacterial diversity and distribution. —AMS

Science, this issue p. 320

MOLECULAR BIOLOGY

Substrate recognition by Dicer elucidated

The Dicer protein generates short RNAs from double-stranded RNA (dsRNA) substrates and is critical for RNA interference and antiviral defense. Sinha *et al.* report structures of a *Drosophila* Dicer protein that shed light on its two distinct mechanisms for recognizing and cleaving substrates: adenosine triphosphate (ATP)-independent, distributive cleavage of 3'-overhang dsRNAs and ATP-dependent, processive threading of blunt-end dsRNAs. This flexibility might provide invertebrates with the optimization capabilities needed for antiviral defense. —SYM

Science, this issue p. 329

CHEMICAL BIOLOGY

A naturally modified cellulose

Cellulose is the most abundant biopolymer on Earth and an

important component of bacterial biofilms. Thongsomboon *et al.* used solid-state nuclear magnetic resonance spectroscopy to identify a naturally derived, chemically modified cellulose, phosphoethanolamine cellulose (see the Perspective by Galperin and Shalaeva). They went on to identify the genetic basis and molecular signaling involved in introducing this modification in bacteria, which regulates biofilm matrix architecture and function. This discovery has implications for understanding bacterial biofilms and for the generation of new cellulosic materials. —SYM

Science, this issue p. 334;
see also p. 276

VASCULAR BIOLOGY

Processing microRNAs for blood vessels

Patients with hereditary hemorrhagic telangiectasia (HHT) are prone to hemorrhages and nose bleeds. This is usually (but not always) because of mutant proteins in a signaling pathway that regulates blood vessel formation. Jiang *et al.* found that zebrafish or mice deficient in the microRNA processing enzyme Drosha had vascular defects similar to those found in HHT patients. Rare mutations in *DROSHA* were overrepresented in HHT patients who lacked the typical disease-associated mutations. Two of these mutants showed reduced activity and could not rescue the vascular phenotypes of Drosha-deficient zebrafish. —WW

Sci. Signal. **10**, ean6831 (2018).

THYMUS

Regeneration circuits in the thymus

Chemotherapy and radiation treatments in cancer patients damage a number of tissues and organs, including the thymus. Prolonged thymic damage can lead to T cell deficiency

and increased susceptibility to opportunistic infections and malignancies. Wertheimer *et al.* examined thymic regeneration in mice after sublethal total body radiation. They found a critical role for bone morphogenetic protein 4 (BMP4) signaling in thymic regeneration. Endothelial cells provided a critical source of BMP4, which induces expression of the transcription factor FOXP1 in thymic epithelial cells to promote thymic regeneration. —AB

Sci. Immunol. **3**, eaal2736 (2018).

REVIEW SUMMARY

BIOPHYSICS

Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer

Eitan Lerner,* Thorben Cordes,* Antonino Ingargiola, Yazan Alhadid, SangYoon Chung, Xavier Michalet, Shimon Weiss†

BACKGROUND: Biomolecular mechanisms are typically inferred from static structural “snapshots” obtained by x-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM). In these approaches, mechanisms have to be validated using additional information from established biochemical and biophysical assays. However, linking conformational states to biochemical function requires the ability to resolve structural dynamics, as macromolecular structure can be intrinsically dynamic or altered upon ligand binding. Single-molecule Förster resonance energy transfer (smFRET) paved the way for studying such structural dynamics under biologically relevant conditions. Since its first implementation in 1996, smFRET experiments both confirmed previous hypotheses and discovered new fundamental biological mechanisms relevant for DNA maintenance, replication and transcription, translation, protein folding, enzymatic function, and membrane transport. We review the evolution of smFRET as a key tool for “dynamic structural biology” over

the past 22 years and highlight the prospects for its use in applications such as biosensing, high-throughput screening, and molecular diagnostics.

ADVANCES: FRET was first identified in the 1920s by Cario, Franck, and Perrin. In the late 1940s, Förster and Oppenheimer independently formulated a quantitative theory of the energy transfer between a pair of point dipoles. Stryer and Haugland verified this theory in the late 1960s and coined the term “spectroscopic ruler” for FRET. Simultaneously, Hirschfeld, and later Moerner and Orrit, pioneered optical single-molecule detection methods leading to the first demonstration of smFRET in 1996. This breakthrough made it possible to study heterogeneous systems, dynamic processes, and transient conformational changes on the nanometer scale. The smFRET technique was rapidly adopted by various research groups to provide mechanistic answers in diverse areas of biological research. In early pioneering applications of smFRET in biochemistry, Ha *et al.* visualized the conformational dynamics of the staphy-

lococcal nuclease enzyme; Deniz *et al.* obtained information on the structural dynamics of double-stranded DNA; and Zhuang *et al.* studied the conformation of individual RNA enzyme molecules and their folding dynamics in equilibrium. These pioneering studies were followed by others that used smFRET to unravel the inner workings of helicases and topoisomerases, DNA replication, DNA repair, transcription, translation, enzymatic reactions, molecular motors, membrane proteins, nucleic acids, protein and RNA folding, ribozyme catalysis, and many other molecular mechanisms.

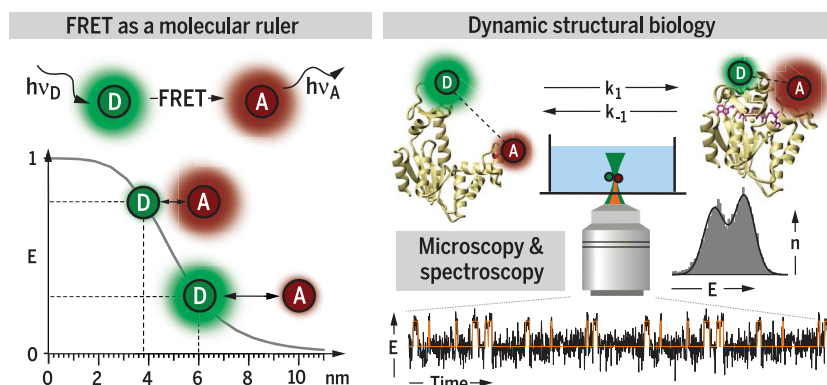
OUTLOOK: During the past two decades, smFRET has grown into a mature toolset with

ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aan1133>

capabilities to explore dynamic structural biology for both equilibrium and non-equilibrium reactions. The one-dimensional (“ruler”) character of the FRET approach, however,

only captures the complex three-dimensional structure of a system and needs to be complemented by other techniques that can provide additional information about the respective biochemical states of macromolecules. Approaches that explore smFRET combinations with other biophysical techniques (patch-clamp, optical, and magnetic tweezers; atomic force microscopy; microfluidics) or photophysical effects are hence gaining attention. Although smFRET is particularly useful for the observation of dynamic conformational changes and subpopulations, FRET efficiencies also carry very precise information on the actual distance between fluorophores attached to distinct moieties of a macromolecule. As shown by recent work from many laboratories (such as those of Seidel, Michaelis, Hugel, and Grubmüller), this quantitative information can be used to help define biological structures and in the future should find a place in the protein database of molecular structures. smFRET has so far mostly been used for in vitro experiments but can be used additionally to monitor conformational dynamics and heterogeneity in live cells. “In vivo smFRET” has recently emerged as a promising methodology, demonstrated by the groups of Sakon, Weninger, Schuler, and Kapanidis among others. We envision that further technological developments will expand smFRET applications beyond dynamic structural biology to allow fast nonequilibrium kinetic studies, high-throughput drug screening, and molecular diagnostics. Advancements of these applications will be impactful for systems that are highly heterogeneous and dynamic. ■



Dynamic structural biology using smFRET. Left: Principle of FRET as a molecular ruler. In a system with a pair of dyes, after the donor dye (D) is excited, it transfers the excitation energy to a nearby acceptor dye (A; top) with an efficiency (E) that depends on the sixth power of the distance between the dyes (bottom). Right: Use of FRET to study structural dynamics at the single-macromolecule level. The experimental setup (top), a combination of single-molecule fluorescence microscopy and spectroscopy, can be used to determine conformational states or dynamics in solution or on immobilized molecules. Here E is calculated per each single-molecule burst of photons, and bursts (n) are accumulated in E histograms (middle) or for different time bins to form a single-molecule E trajectory (bottom).

The list of author affiliations is available in the full article online.

*These authors contributed equally to this work.

†Corresponding author. Email: swiss@chem.ucla.edu

Cite this article as E. Lerner *et al.*, *Science* 359, eaan1133 (2018). DOI: 10.1126/science.aan1133

REVIEW

BIOPHYSICS

Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer

Eitan Lerner,^{1,*} Thorben Cordes,^{2,3*} Antonino Ingargiola,¹ Yazan Alhadid,¹ SangYoon Chung,¹ Xavier Michalet,¹ Shimon Weiss^{1,4,5†}

Classical structural biology can only provide static snapshots of biomacromolecules. Single-molecule Förster resonance energy transfer (smFRET) paved the way for studying dynamics in macromolecular structures under biologically relevant conditions. Since its first implementation in 1996, smFRET experiments have confirmed previously hypothesized mechanisms and provided new insights into many fundamental biological processes, such as DNA maintenance and repair, transcription, translation, and membrane transport. We review 22 years of contributions of smFRET to our understanding of basic mechanisms in biochemistry, molecular biology, and structural biology. Additionally, building on current state-of-the-art implementations of smFRET, we highlight possible future directions for smFRET in applications such as biosensing, high-throughput screening, and molecular diagnostics.

Until the mid-1990s, insights in structural biology came mainly from static macromolecular structures obtained by x-ray crystallography (1, 2). Nuclear magnetic resonance (NMR) spectroscopy allowed identification of many of the different structures associated with a single conformation of a biomolecule (1). However, a biomolecule can adopt many different conformations. Cryo-electron microscopy (cryo-EM) has recently complemented this toolkit, facilitating the determination of multiple conformations of macromolecular structures in the ensemble with near-atomic resolution (3, 4). Molecular mechanisms can be inferred from such static structural “snapshots” and validated using biochemical and biophysical assays [e.g., (5)]. Although structural snapshots can identify distinct conformational states that macromolecules explore at equilibrium (e.g., ligand-bound or unbound, folded or unfolded), they lack information on the interconversion dynamics between these states. Understanding the functional roles of these structures requires a full dynamic picture (6–8).

NMR, electron paramagnetic resonance (EPR) (9), and double electron-electron resonance (DEER) (10) spectroscopies, as well as fluorescence-based

techniques such as fluorescence anisotropy (11), ensemble Förster resonance energy transfer (FRET, Fig. 1) (12), or photo-induced electron transfer (13), can provide access to dynamic information about biomolecular interactions and macromolecular conformations. The interpretation of experimental results from these techniques is, however, highly model-dependent (14, 15). Even for two-state systems in equilibrium (e.g., transitions between open and closed conformations of a protein, or bound and unbound states of interacting molecules; Fig. 2, A and B, respectively), ensemble methods yield limited insight into structural and mechanistic details. This is because molecules in an ensemble undergo changes between conformational states asynchronously (Fig. 2C). This results in averaged-out signals (Fig. 2C), so that the underlying dynamical information can be retrieved by model fitting only, and only in the simplest cases (16, 17). One way of solving the problem of asynchronicity is by measuring one molecule at a time and retrieving the underlying conformational states and dynamics directly (Fig. 2, D and E).

Following the development of single-molecule detection techniques (18–24), the first demonstration of FRET at the single-molecule level was published in 1996 (25). It suggested that single-molecule FRET (smFRET) could be used to study dynamic processes and identify transient conformations and interactions between macromolecules labeled with a donor-acceptor dye pair. Schütz *et al.* used smFRET to monitor binding of ligands to streptavidin immobilized on phospholipid membranes (26), opening the way for similar experiments in live cells. In another pioneering application of smFRET, Ha *et al.* characterized the intricate conformational and substrate-binding dynamics of

the staphylococcal nuclease enzyme (27). Further smFRET studies of conformational changes in other enzymes and in RNA molecules followed, using either diffusing molecules (Figs. 2D and 3A) (28, 29) or immobilized molecules (Figs. 2E and 3C) (6, 27, 30, 31). In one example, Deniz *et al.* showed how information on distance-related distributions can be derived from smFRET measurements of double-stranded DNA (dsDNA; Fig. 3B) (28). Additionally, the combination of total internal reflection (TIR; Fig. 3, C and D) illumination with immobilized single molecules allowed Zhuang *et al.* to follow the conformation of individual RNA enzyme molecules and measure their folding dynamics at equilibrium (Fig. 3D) (6). In the two following decades, smFRET has matured into a toolkit to explore dynamic structural biology. This article reviews achievements in the use of smFRET to establish structure-function relationships and outlines challenges and prospects for the future.

A brief historical overview of single-molecule FRET in biochemistry and molecular biology

FRET was first identified in the 1920s by Cario, Franck, and Perrin. In the late 1940s, Förster and Oppenheimer independently formulated a quantitative theory of the energy transfer between a pair of point dipoles. Stryer and Haugland verified this theory in the late 1960s and coined the term “spectroscopic ruler” for FRET. Around the same time that the effect of heterogeneity on FRET was taken into account in ensemble measurements (32), Hirschfeld pioneered single-molecule fluorescence detection (33). The first observations of individual fluorescent molecules in the late 1980s and early 1990s (34–38) were followed by an explosion of studies, including imaging of complex biological systems such as molecular motors (39). Since its first demonstration in 1996, smFRET has been used to provide mechanistic answers in diverse areas of biological research. These studies unraveled molecular mechanisms of helicases and topoisomerases (40), DNA replication, DNA repair (41), transcription (42–44), translation (42, 45, 46), enzymatic function (47–49), molecular motors (50), membrane proteins (51), protein folding (52, 53), nucleic acids (54, 55), RNA folding (54, 56, 57), and ribozyme catalysis (58, 59). Because a short review cannot do justice to the large number and diversity of smFRET studies, we will discuss a few representative examples. The theory describing FRET is given in Box 1.

A good example of the power of smFRET to explore heterogeneous mixtures and distinguish subpopulations of conformers can be taken from the field of bacterial transcription. Here, the molecular mechanism of the long-known but poorly understood abortive transcription initiation was deciphered by two concerted single-molecule experiments (60, 61), one of which was based on smFRET (60). Both studies showed that RNA polymerase (RNAP) repeatedly and unsuccessfully attempts to reel the downstream DNA into its active site (using a mechanism called “DNA scrunching”) before clearing the promoter and

¹Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA. ²Molecular Microscopy Research Group, Zernike Institute for Advanced Materials, University of Groningen, 9747 AG Groningen, Netherlands. ³Physical and Synthetic Biology, Faculty of Biology, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany. ⁴Department of Physiology, University of California, Los Angeles, CA 90095, USA. ⁵California NanoSystems Institute, University of California, Los Angeles, CA 90095, USA.

*These authors contributed equally to this work.
†Corresponding author. Email: sweiss@chem.ucla.edu

proceeding to transcript elongation (Fig. 4A) (60). By using distinct labeling schemes, the FRET study ruled out other proposed mechanisms (“inchworming” and “transient excursion”; Fig. 4A). When the acceptor (A, red dot) labeled the promoter sequence and the donor (D, green dot) labeled the RNAP’s leading edge, no difference was observed between the FRET histograms of the RNAP-promoter open complex (RP_o) and the RNAP-promoter complex transcribing up to seven bases ($RP_{itc, \leq 7}$; Fig. 4A, left). This excluded the inchworming model. When the acceptor labeled the DNA upstream of the promoter sequence and the donor labeled the RNAP’s trailing edge, again no difference could be observed between the FRET histograms of RP_o and $RP_{itc, \leq 7}$ (Fig. 4A, center). This excluded the transient excursion model. In a third experiment, the acceptor and donor labeled the DNA downstream and upstream relative to the promoter sequence, respectively (Fig. 4A, right). In this case, smFRET showed an increase in the long-distance fraction (small apparent FRET efficiency, E^*) upon addition of nucleotides permitting transcription initiation. This data unambiguously supported the scrunching mechanism, where DNA is reeled into the active site by RNAP during the initial stages of transcription, resulting in an increase in the size of the transcription bubble.

Another good example of smFRET’s ability to disentangle conformational subpopulations is the study of the enzyme adenylate kinase (62, 63). Previous ensemble time-resolved FRET measurements had suggested a single conformational state characterized by a broad distance distribution for the enzyme in the absence of its substrates, adenosine monophosphate and Mg-adenosine triphosphate (64). smFRET measurements showed, however, that at least two distinct dynamically interconverting conformations were present in the absence of substrates: an apo conformation and an active-like conformation (Fig. 4B) (62, 63). These and similar studies (47, 48) have shed light on how enzymes exist in different precatalytic

conformations and how substrate binding can stabilize one of these conformations.

Another fruitful area of smFRET investigations is protein folding. The function of a protein is encoded in its three-dimensional (3D) structure. Although deduction of a protein’s tertiary structure (its native conformation) from its primary sequence has been revolutionized by computational techniques (65), smFRET experiments have provided many additional insights into the process of folding—whether into the correct structure or into incorrect structures (“misfolding”)—and characterization of possible folding intermediates (66, 67). Measurements involving a variety of denaturing agents have yielded evidence of a monotonic shift in the mean FRET value of the unfolded subpopulation as a function of denaturant concentration. These observations have been interpreted as a manifestation of rapid interconversion between the unfolded state and folding intermediates (29, 68), which would imply the existence of folding intermediates stabilized by non-native contacts (52, 53). Such studies have been expanded by many groups, and fast microfluidic mixers have enabled the extension of research from equilibrium to nonequilibrium regimes (69). The relevance of these in vitro folding studies to in vivo chaperone-assisted protein folding or to cotranslational protein folding is a topic of current investigation (70).

A related area of investigation to benefit from smFRET is the conformation of intrinsically disordered proteins (IDPs). IDPs are often stabilized in a folded state upon ligand binding (i.e., co-folding). For example, α -synuclein (α Syn), a major determinant in Parkinson’s disease, is an IDP that co-folds upon binding to membranes. Deniz and co-workers studied the conformational changes of α Syn upon co-folding with different ligands and characterized its associated rapid conformational dynamics (71). They found that α Syn gains different α -helical structures after binding to lipid-mimetic agents with varying surface curvature.

Similarly, smFRET helped to elucidate the conformational dynamics, folding mechanisms, and function of RNA molecules (54–59). Not all genes code for proteins. These RNAs become functional upon folding into specific structures. Many such RNA molecules serve as ribozymes (RNA-based enzymes) or regulate various cellular processes such as gene expression and ribosome translation (54). The complexity of folding scales with the structural complexity of the RNA. The folding of RNA molecules goes through multiple free energy local minima separated by barriers of various heights (72). Therefore, it is easy for ribozymes to become trapped in long-lived, nonfunctional states (73). Even hairpin ribozymes, previously presumed to be “simple,” exhibit multiple intermediates and multiple pathways during folding (74, 75).

These examples illustrate how smFRET can be used to study the conformational dynamics and function of biological macromolecules. Next, we consider different kinds of smFRET measurements, the type of data and analyses associated with them, and examples of the dynamics these methods are capable of exploring.

Conformational states and their dynamics

smFRET can be used to characterize distinct conformational states in macromolecules and the dynamics of their interconversion. However, transitions between states can only be measured if they occur over a time scale comparable to the technique’s temporal resolution. Transition time scales are proportional to the height of the activation barrier between states (Fig. 5A). Separation by a low barrier means rapid interconversion between states, which results in averaged-out smFRET data and indistinguishable states. Transitions occurring over time scales much longer than the typical observation time will, of course, not be detected either. The temporal resolution of a smFRET experiment depends on several parameters. One of the most important is whether the experiment involves freely diffusing molecules (Figs. 2D and 5B, left) or immobilized molecules (Figs. 2E and 5B, right). In both cases, FRET efficiency is calculated for each individual molecule over short, finite time intervals. For freely diffusing molecules, this observation period is set by the transit time of the diffusing molecule through the observation volume. These rare events generate a “burst” of fluorescence photons with a typical duration on the order of 1 ms. A given molecule may or may not be detected again subsequently, depending on its random diffusion path. For immobilized molecules, observation can last for several seconds or even minutes, generating time traces of fluorescence (or FRET efficiency, once processed) with a temporal resolution set by a combination of detector readout rate and signal level (a few milliseconds at best; Fig. 5B, right). Although improved organic dyes have been developed (76, 77), dye photobleaching remains the main constraint on the maximal observation time and temporal resolution (78). Analysis of burst (freely diffusing) or time-binned (immobilized)

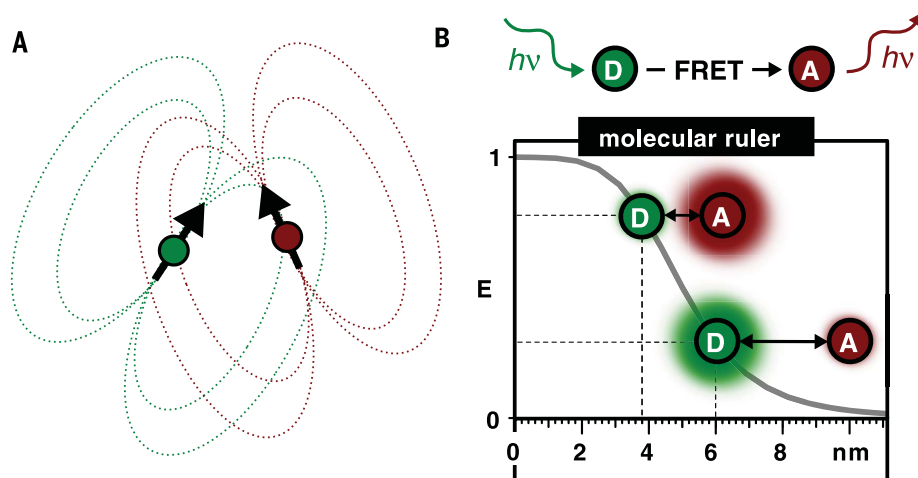


Fig. 1. The concept of FRET. (A and B) An electromagnetic transmitter-receiver (A) is a macroscopic analog for the molecular dipole-dipole coulombic interaction between donor and acceptor (D and A) fluorophores (B). The dependence of the efficiency of energy transfer from D to A on their distance provides a molecular ruler with a high dynamic range on the 3- to 9-nm scale.

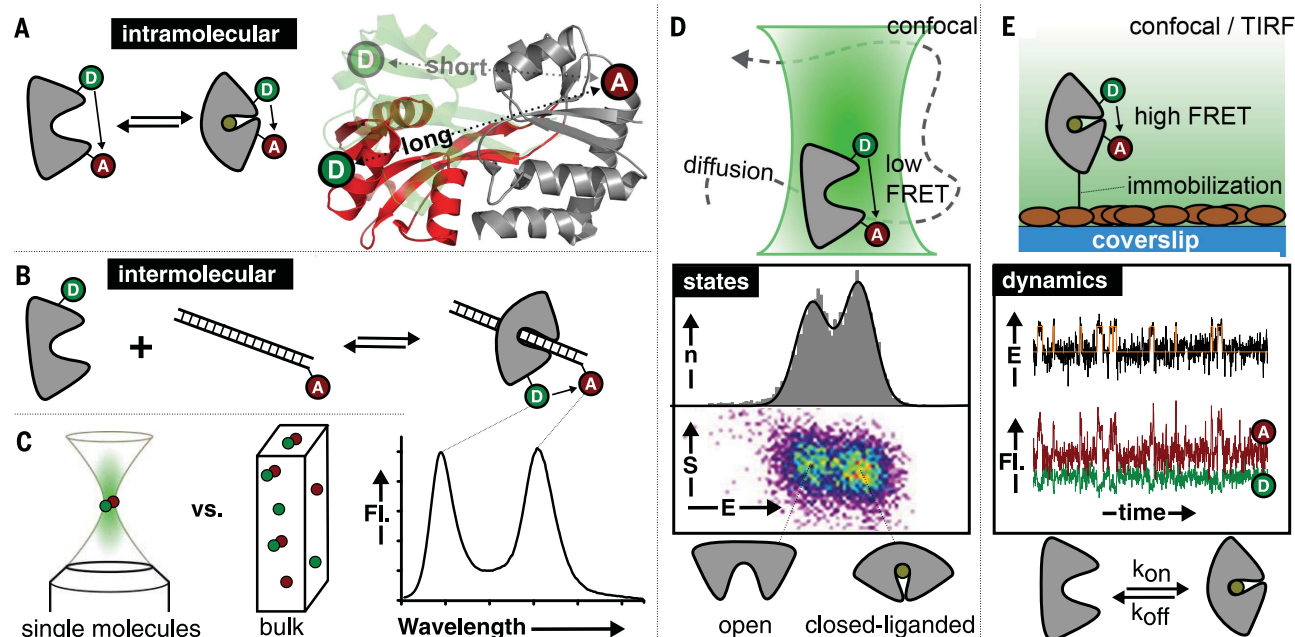


Fig. 2. Principle and use of FRET for elucidating biomolecular reaction mechanisms and structural dynamics. (A to C) Principle of intra-molecular (A) and intermolecular (B) FRET assays and their readout (C) in single-molecule and bulk fluorescence (Fl.) experiments. The bulk experiments always show an average value [i.e., donor (D) and acceptor (A) intensity of, e.g., hypothetical 50/50], whereas smFRET can determine (dynamically interconverting) states directly. The crystal structure overlay of substrate-binding domains of an ABC transporter in (A) shows open (red)

and closed (green) conformations. (D and E) smFRET with diffusing molecules (D) or immobilized molecules (E) including accessible biophysical parameters (i.e., conformational states and dynamical changes). For characterization of conformational states, histograms of FRET efficiency E with frequency n are used; dynamics are directly seen via temporal evolution of E obtained via ratio of acceptor (A) fluorescence to fluorescence from both donor (D) and acceptor after donor excitation. [(A) and (D) adapted, with permission, from (154)]

data allows identification of distinct conformational subpopulations, their FRET efficiency, and, in favorable cases, their interconversion rates.

Studying slow conformational dynamics (from 0.1 to 10 s) requires long observation times and is mostly done with immobilized molecules (Figs. 2E and 3C). Here, the different durations (dwell times) spent by a molecule in each state are analyzed, and energy transfer efficiencies are either directly extracted or obtained via hidden Markov modeling or Bayesian statistical analyses (79, 80) (Fig. 5B, right). Results from many individual molecules observed in parallel are pooled to obtain statistically meaningful information. This approach has been used to study the dynamics of nucleic acid-processing enzymes such as helicases (40), the complex molecular mechanism of translocation of the ribosome (81), and HIV reverse transcriptase initiation (82), among many others. Extraction of these dynamical parameters would be very difficult using ensemble techniques.

Faster conformational dynamics (10 μ s to 0.1 s) are typically best studied with diffusion-based smFRET (Figs. 2D and 3A). Here, dynamics can be extracted from fluctuations in FRET efficiency within single-molecule bursts or between consecutive bursts of the same molecules (moving in and out of the observation volume several times) with accessible time scales in the range of 0.1 to 10 ms. If diffusing molecules change conformation during transit through the observation volume, the time-averaged FRET efficiency within each burst is of little use, although it could hint at

the presence of faster dynamics (83, 84). Analytical methods to investigate such dynamics have been developed in recent years (85, 86). For instance, Torella *et al.* examined the short time scale variance of FRET efficiency within individual single-molecule bursts [burst variance analysis (BVA)] (86) (Fig. 5C, left). FRET variance exceeding that expected from photon-counting statistics (“shot noise”) was used to detect millisecond-time scale dynamics in complexes of the Klenow fragment of DNA polymerase. Using the same approach, Robb *et al.* showed that in transcription initiation, the transcription bubble (the DNA region opened up by RNAP) exhibits conformational dynamics on the submillisecond time scale (Fig. 5C, left) (87).

Single molecules freely diffusing in 3D may reenter the observation volume several times before diffusing away permanently. This results in a series of consecutive single-molecule bursts, between which the molecule may change its conformation. This opens the possibility of analyzing conformational dynamics by recurrence analysis of single particles (RASP; Fig. 5C, center) (88). Analyzing the succession of FRET efficiencies of consecutive bursts separated by variable recurrence times enabled quantification of the folding relaxation times of small proteins such as cold shock protein (Csp), spectrin R15, and the B domain of protein A (Bdpa), revealing time scales of 250 ms, 32 ms, and 0.7 ms, respectively (88).

Faster conformational changes (<0.1 ms) yield single-molecule bursts with averaged-out FRET values. Approaches that do not rely on the anal-

ysis of separate single-molecule bursts, but rather on photon statistics within bursts, are therefore called for. In addition, because such rapid conformational changes include multiple transitions within each single-molecule burst, the variance of the FRET efficiency becomes noisy (in a way that resembles shot noise). In this limit, techniques that resolve FRET dynamics through variance analysis (such as BVA) cannot resolve faster FRET dynamics. For this regime, fluorescence correlation spectroscopy (FCS) methods applied to smFRET (FRET-FCS) are the most straightforward to implement, even if demanding in terms of statistics (Fig. 5C, right). For instance, Nettels *et al.* performed diffusion-based smFRET measurements on Csp, acquiring data in order to compute correlation curves down to the picosecond time scale. Using this approach, they showed that the unfolded state of Csp undergoes structural reconfiguration within \sim 40 ns (89). The additional information attained from fluorescence lifetimes has also been used in the analysis of rapid FRET dynamics. Fluorescence lifetime analysis (using pulsed laser excitation) can also be used to unravel fast dynamics. Woźniak *et al.* used time-correlated single photon counting (TCSPC) to explore the bending dynamics of short dsDNA (90). Dolino *et al.* observed submillisecond dynamics in the ligand-binding domain of the N-methyl-D-aspartate receptor (91). Using alternating laser excitation on the nanosecond time scale (nsALEX; see Box 1), Laurence *et al.* analyzed fluorescence decays of specific FRET

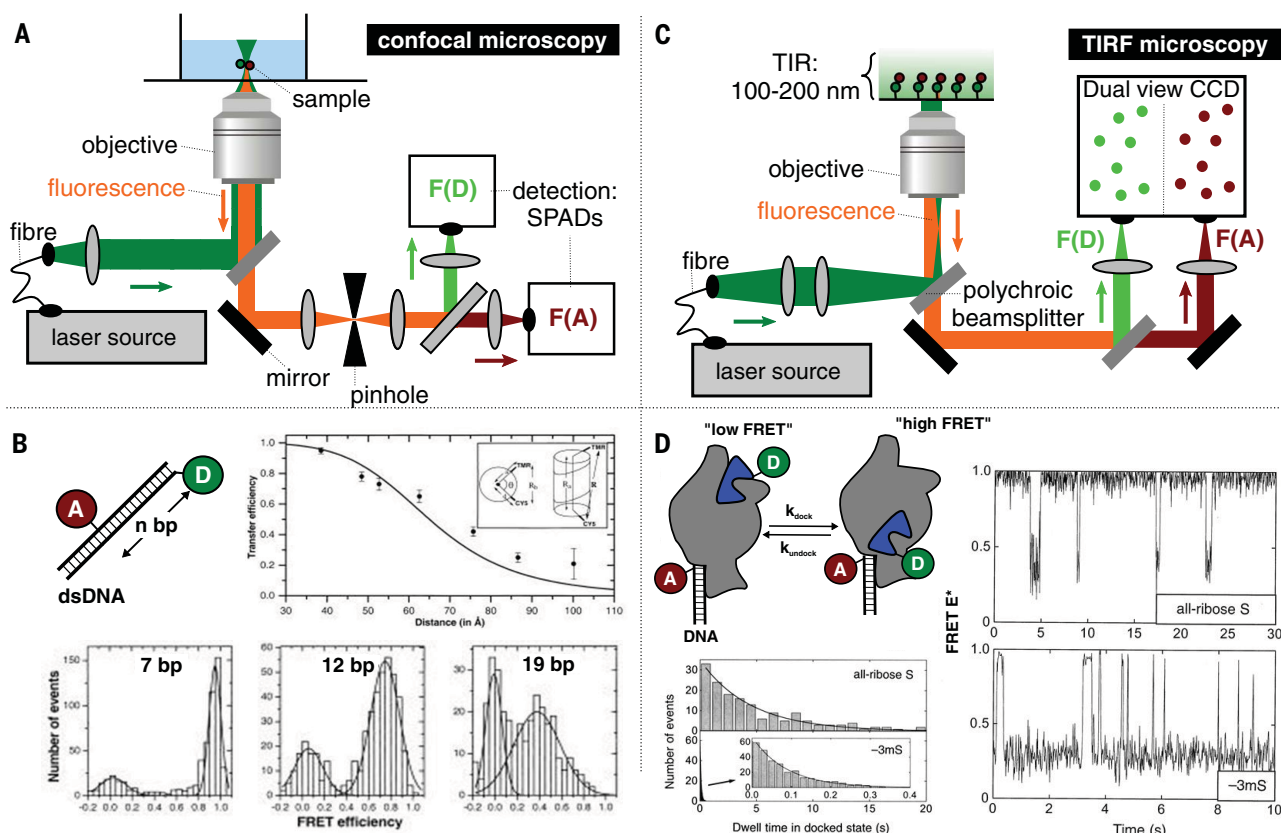


Fig. 3. Pioneering implementations of smFRET. (A) Schematic of a confocal microscope setup used for the acquisition of diffusion-based smFRET data; F(D) and F(A) indicate the donor and acceptor detection channels, respectively. (B) Example of data obtained with such a setup. The different histograms show the FRET efficiency distributions obtained for DNA samples differing by the distance between donor and acceptor labels; bp, base pairs. [Adapted,

with permission, from (28)] (C) Schematic of a total internal reflection fluorescence (TIRF) setup allowing the study of smFRET on surface-immobilized molecules. (D) Example of data obtained with such a setup, showing the real-time dynamics of RNA catalysis and folding. FRET trajectories were retrieved for individual RNA molecules (right) and histograms of dwell times reported on the time scale of the dynamics (lower left). [Adapted from (6)]

subpopulations to infer an effective distance distribution for the folded and unfolded chymotrypsin inhibitor 2 (CI2) (92).

Although powerful, these fast conformational dynamics methods usually do not provide information on the exact number of conformational states (and their mean FRET efficiencies) involved in the identified dynamics. Recently, Pirchi *et al.* reported an analytical method to extract the values of these parameters by performing a photon-by-photon hidden Markov modeling analysis of smFRET experiments (H2MM) (93), as previously suggested by Gopich and Szabo (94). They were able to extract rate constants (ranging from ~ 10 μ s to ~ 1 s) and the mean FRET efficiencies of the corresponding states. We are therefore on a path toward full characterization of fast conformational dynamics of macromolecules: the number of states, their FRET values, and the interconversion rate constants.

All of the methods mentioned above, although powerful, rely on a single reaction coordinate (the distance between a single donor-acceptor pair) and therefore provide a limited perspective on the underlying dynamics. We next discuss how multiple reaction coordinates can be simulta-

neously measured to untangle complex conformational dynamics.

Toward multiple reaction coordinates

A single smFRET measurement reports on a single distance within a macromolecular structure, projecting a complex 3D structure onto a single 1D reaction coordinate (Fig. 5A). In some macromolecules, domains or subunits may be approximated as rigid bodies linked by flexible linkers or interacting through well-defined binding interfaces. In other cases, allosteric ligand binding to one part of a macromolecule can cause conformational changes in other parts of the same macromolecule. In these cases, a single reaction coordinate may not be enough to report coordinated motions. Additionally, regardless of the presence or absence of allosteric binding and coordinated motion, some smFRET-derived single distances may be insensitive to conformational changes (e.g., structural change occurs tangential to the monitored distance or occurs in another part of the macromolecule). For all these reasons, it is generally desirable to study conformational changes with more than one set of positions for a pair of dyes.

An obvious solution is to label the macromolecule with more than two dyes. Multicolor smFRET techniques (95) can indeed provide a wealth of information (Fig. 6A). Ha and co-workers (96) and Person *et al.* (97) used three-color smFRET to study Holliday junctions, which spontaneously switch between two distinct conformations. They simultaneously determined three distances unambiguously and specified the correlated movement of the junction's hairpin structure. Multicolor smFRET techniques provide high information content but are challenging to implement. They require multiple orthogonal and efficient site-specific labeling chemistries, elaborate optics, and data analysis techniques. Some of these difficulties can be mitigated by using a dark quencher as one of the (three) dyes. Because the dark quencher accepts excitation energy through FRET but does not emit photons, there is no need for detection of emission, thereby simplifying data collection. Kapanidis and co-workers (98) used this approach to monitor the binding and unbinding of a DNA polymerase to its substrate in real time, without the need for three-color detection for simultaneous detection of protein binding and associated conformational changes.

It is, however, also possible to monitor more than one distance using multiple identical dyes in a two-color excitation and detection scheme, using some photophysical tricks. In biomolecular complexes with more than one donor and acceptor of the same kind, fluorophore interactions via FRET are highly complex, and the relation of FRET efficiency E to inter-dye distances R is generally nontrivial because of multiple energy transfer pathways. For immobilized molecules, this multiplicity can be removed by using chemically induced stochastic blinking of the acceptor fluorophores (99, 100), leaving only one active acceptor per molecule for a brief period of time. Using this “photoswitchable FRET” approach, Uphoff *et al.* measured the distances between DNA and two residues on the catabolite activator

protein (CAP), as well as the strand exchange dynamics in Holliday junctions (100). Because the acceptor blinks randomly, smFRET time traces exhibit different values over time, allowing measurement of multiple distances from a single donor to multiple acceptor fluorophores (Fig. 6B).

Another alternative to multicolor smFRET—which requires as many detection channels as there are different dyes—is using simple two-color smFRET with other independent observables (translational diffusion, fluorescence anisotropy, brightness, etc.) (101). For example, by using an alternating laser excitation (ALEX) scheme (see Box 1), Kapanidis *et al.* were able to simultaneously report FRET values E for each molecule as well as the “stoichiometry” S , defined as the ratio between fluorescence originating from donor excitation

and that originating from both donor and acceptor excitations (102). Changes in mean S values report changes in the ratio of donor and acceptor brightnesses. ALEX can therefore distinguish molecules with different numbers of donor and acceptor dyes, or molecules with altered dye fluorescence quantum yields. In a recent implementation of ALEX to simultaneously report two distances, smFRET was combined with protein-induced fluorescence enhancement (PIFE) (103, 104). One distance was between the dyes and within the FRET distance range (~3 to 9 nm); the other, between one dye and a bound protein, was in the shorter PIFE distance range (<3 nm). PIFE-FRET thus provided direct evidence for molecular coordination in the open transcription bubble (Fig. 6C) (104). Similarly, smFRET was combined

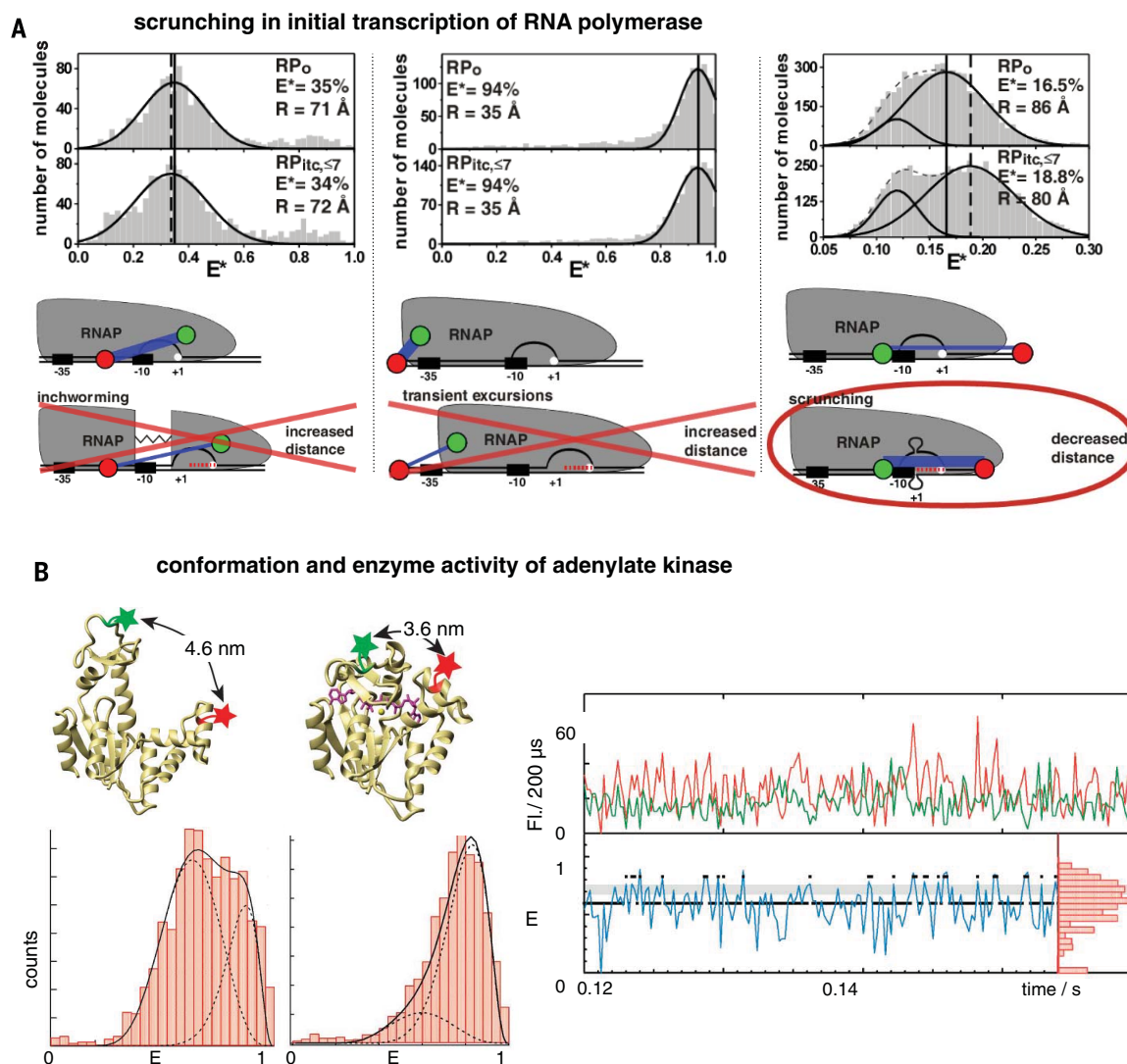


Fig. 4. Typical examples of smFRET studies. (A) Transcription initiation involves a DNA scrunching mechanism. The results of three experiments differing by the location of the donor and acceptor dyes are shown (see text). The cartoons indicate which model is or is not compatible with the results. [Adapted from (60)] **(B)** Intrinsic domain motions between conformations in adenylate kinase (AK). The experiment tracks the distance between substrate-binding domains

(donor and acceptor dyes as green and red stars, respectively) in the AK enzyme in apo form (left histogram) and when bound to the substrate-mimicking inhibitor ApsA (right histogram). FRET efficiency histograms (left) and single-molecule time traces (right) show that in apo conformation, AK dynamically switches between two conformations, one of which is similar to the substrate-bound state. [Adapted, with permission, from (62)]

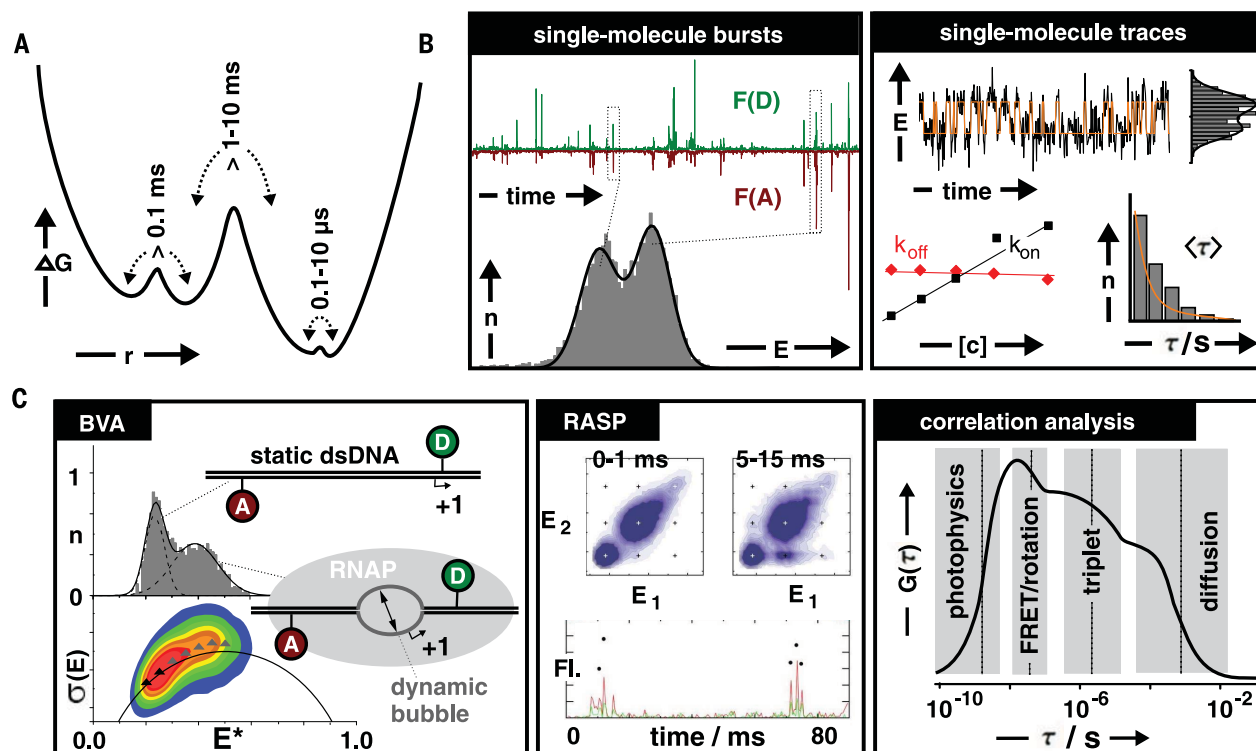


Fig. 5. Biomolecular dynamics accessible by smFRET. (A) Hypothetical energy landscape with Gibbs free energy projected onto a single reaction coordinate r showing different local minima (states) separated by energy barriers of different heights, giving rise to conformational transitions over different time scales. (B) smFRET data from diffusing molecules (bursts, left) and immobilized molecules (time traces, right) can be analyzed by various methods with differing temporal resolutions to study conformational transitions over different time scales. Conformational dynamics slower than ~ 0.1 s can be studied by analysis of single-molecule

traces and dwell times in each FRET-associated state. (C) Examples of data analysis techniques using details of burst properties and photon statistics: Burst variance analysis (BVA) identifies bursts with variance of the FRET efficiency larger than expected from shot noise; recurrence analysis (RASP) identifies whether the FRET efficiency has changed between consecutive bursts of the same molecule; and correlation techniques identify time scales (including <100 μ s) at which fluorescence-related processes occur, including changes in FRET efficiency. [Reproduced, with permission, from (87, 88)]

with photo-induced electron transfer (PET) (105), where the donor dye was quenched by a nearby tryptophan moiety. However, the steep dependence of PET on dye-quencher distance (angstroms) results in a binary output (contact/no contact) rather than a quantitative distance measurement. Finally, because both multicolor smFRET and ALEX achieve high information content, combining the two techniques doubles the information content (both FRET and brightness ratio for each dye pair permutation). Lee *et al.* used three-color ALEX to monitor the translocation of bacterial RNAP on DNA on two distinct reaction coordinates (106). Such high information content can be used for multiplexed sorting in molecular diagnostics. Yim *et al.* have shown the capability to sort and quantify multiple different biomarkers in a four-color ALEX experiment (107).

In each of these techniques, macromolecule labeling with fluorophores (or quenchers) is important. Although high-purity site-specific dye-labeled nucleic acids are now commercially available, preparation of site-specifically labeled proteins at multiple residues is far more challenging. Advances in this field [reviewed in (108, 109)] will allow the study of multiple reaction coordinates

and coordinated motions in single subunit proteins. As an alternative [already suggested in 1999 (110)], smFRET can be combined with other single-molecule techniques not involving fluorescence. These include patch-clamp (111) and single-molecule manipulation methods such as optical (112) and magnetic (113) tweezers, atomic force microscopy (114), and microfluidics and drag forces (115). These hybrid approaches are very powerful and simultaneously measure multiple orthogonal reaction coordinates. A detailed account of these methods is outside the scope of this review and can be found elsewhere (95).

Solving 3D structures with smFRET

If properly calibrated, FRET efficiency E carries information on the precise distance between donor and acceptor dyes. Can this information be extracted and used for 3D structure determination? X-ray crystallography, NMR, and cryo-EM are currently the gold standards for obtaining atomic-resolution 3D structures of complex macromolecules. However, crystallization conditions may preferentially stabilize one conformation over others (62); in extreme cases, crystallization may even induce a structure never observed in solu-

tion, as detected by comparison with structures solved by solution-based techniques (116). Thus, structural characterization of macromolecules in solution and at ambient temperature is desirable.

The ability to identify distinct conformational subpopulations can help in structure determination, because relevant subpopulations can be identified and selected for further processing. Structure determination requires the preparation and measurement of multiple donor-acceptor variants labeling different pairs of positions on the macromolecule. The nontrivial transformations from uncorrected FRET efficiency E^* to corrected FRET efficiency E , and then to inter-dye distance R and to inter-residue distance r , require additional preparations and measurements of control mutants, modeling and simulations, structural convergence procedures, and control and validation of refined structures. Several studies (117–119) have followed this route, reporting successful structure determination (120–122). The information retrieved from smFRET measurements of multiple distances can be used to directly triangulate a structure of a whole or part of a macromolecule, or can be used as experimental constraints for structural simulations (121). In

the latter approach, each iteration produces a structural snapshot. After assessing the dyes' accessible volumes via the "nanopositioning system" (NPS) approach (117), the computationally derived mean inter-dye distances are compared with the experimentally derived ones for all measured constructs, and the sum of all deviations (cost function) is computed. This process, performed on a large library of simulated structural snapshots, should result in a subset of candidate conformations selected by minimization of the cost function. Using this approach, we recently identified two conformations of the transcription bubble in the bacterial RNAP-promoter open (RPO) complex (123). The set of distances of one conformation agreed with the crystal structure of bacterial RPO (124), while the other did not. The latter conformation had characteristics of a scrunched transcription bubble, where a few bases from the duplex downstream to the bubble were reeled into the active site of RNAP and increased the size of the transcription bubble.

Although successful structural determinations by smFRET have been reported, single-particle cryo-EM has also gained the ability to resolve several (up to three) conformational states in the frozen ensemble (4, 125, 126). Nonetheless, single-particle cryo-EM fundamentally lacks what smFRET readily provides: the ability to detect the dynamics of transitions between conformations. We anticipate that FRET-derived macromolecular structures or distance constraints will also be accepted in the future as entries in the Protein Data Bank. However, different laboratories currently use different measurement and analysis techniques, different protocols, and different types of data files. Therefore, smFRET experiments—and more important, the control experiments and data analysis procedures required for obtaining exact distances—have to be standardized, as outlined below.

Standardizing smFRET measurements

Because of the challenges of smFRET data calibration, it is important to strive for reproducibility across laboratories by establishing standard protocols and data-sharing practices. Such a standardization effort, led by the Hugel and Seidel groups, was recently initiated through the wwPDB Hybrid/Integrative Methods Task Force (127). Equally important to this effort is a standard set of recommended practices that could be verified by peer review. These include (i) avoiding subjective selection of data sets (e.g., time traces in surface-immobilized experiments), (ii) requiring the donor-acceptor fraction of the labeled sample to be larger than 10%, (iii) using different excitation powers to assess photophysics effects, (iv) requiring fluorescence anisotropy measurements to characterize fluorophore rotational freedom, and (v) comparison with ensemble assays (denaturation curve, enzymatic assay, secondary structure content, thermal stability, ligand binding affinity, etc.) of the labeled macromolecule with its unlabeled counterpart to verify its activity and the relevance of the smFRET measurement.

We also recommend that every smFRET experiment, including experiments with surface-

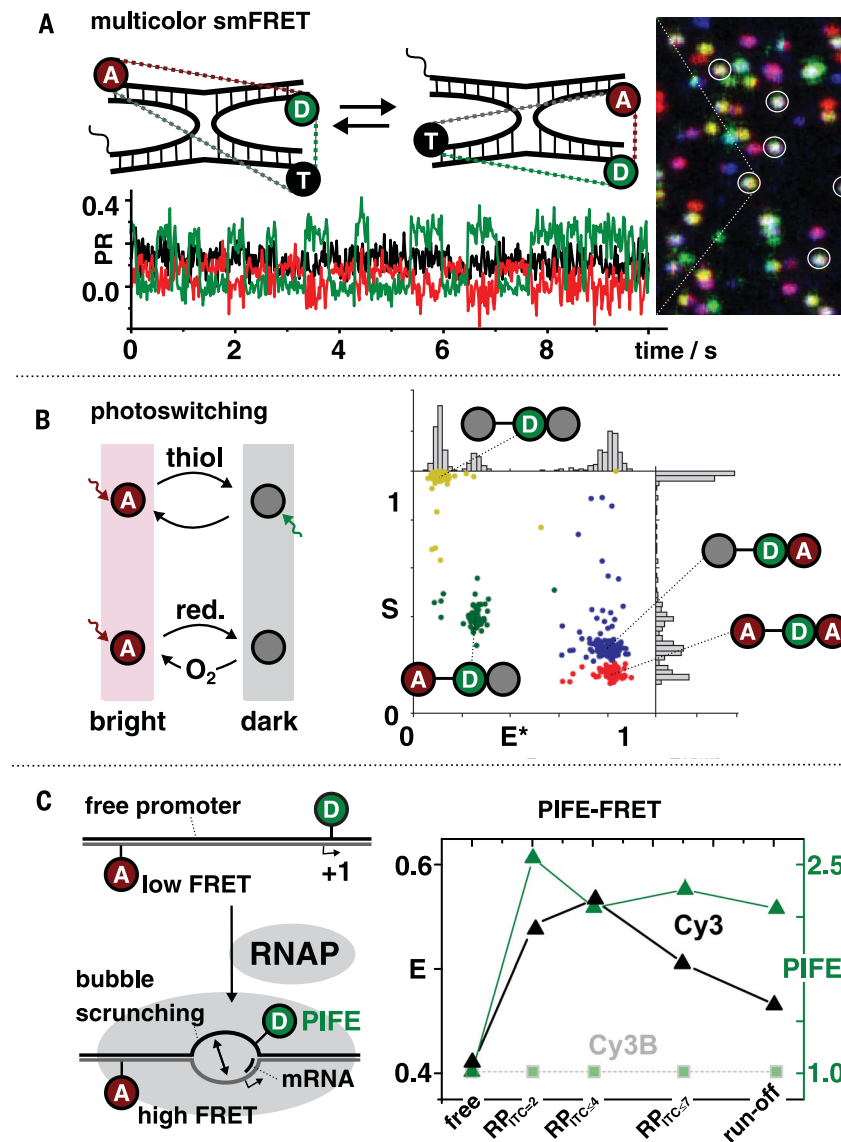


Fig. 6. smFRET-based approaches to study molecular coordination. (A) Multicolor smFRET studying coordinated movement of a Holliday junction via proximity ratio PR: donor-transmitter D-T (green trace), transmitter-acceptor T-A (black), and donor-acceptor D-A (red). [Adapted, with permission, from (97)] (B) Photoswitchable FRET relies on temporal separation of donor-acceptor interactions via photoswitching and isolation of molecular species with one distinct donor-acceptor pair at any given time point. [Adapted, with permission, from (100)] (C) PIFE-FRET uses a standard two-color assay with donor and acceptor (D-A) but adds information on protein binding via use of an environmentally sensitive donor (Cy3; Cy3B is used as the control dye that is insensitive to changes in the environment). [Adapted, with permission, from (104)]

immobilized molecules, should begin with a "control" solution-based smFRET assay, so as to determine (i) the quality of labeling, (ii) the number of states or biochemical species resolved as distinct FRET subpopulations in the sample, (iii) the mean FRET efficiencies of the resolved subpopulations, and (iv) interconversion rate constants between subpopulations. With this information, analysis of smFRET time trajectories from surface-immobilized molecules can be guided by, and compared to, a statistically robust diffusion-based analysis. Moreover, this two-step process will allow assessing whether biomolecule-surface interac-

tions are present and perturb the system under study—in particular, measured FRET efficiencies, population frequencies, and time constants.

Finally, to improve cross-checking and reproducibility, standardized data analysis protocols should be used and preferably based on open-source software. For instance, the FRETbursts open-source package provides a starting point for diffusion-based analysis (128), and similar packages are available for surface-immobilized smFRET (129, 130). The use of standardized file formats such as Photon-HDF5 (131) is a prerequisite to making the raw data freely available and

Box 1. FRET as a spectroscopic ruler for macromolecular distances.

FRET is a “spectroscopic ruler” with a distance-dependent efficiency E given by Förster theory (151) (Fig. 1). If R is the distance between two point dipoles [representing the center of the donor (D) and acceptor (A) fluorophores; Fig. 1A], E depends on the sixth power of the distance (Fig. 1B), assuming a “frozen” molecule (Eq. 1):

$$E = \frac{1}{1 + (R/R_0)^6} \quad (1)$$

$$R_0^6 = \frac{9 \ln(10)}{128\pi^2 N_A} \frac{\kappa^2 \phi_D}{n^4} \int f_D(\lambda) \epsilon_A(\lambda) \lambda^4 d\lambda \quad (2)$$

The Förster radius, R_0 , is the R at which $E = 50\%$. R_0 depends on parameters indicated in Eq. 2: N_A is Avogadro's number, n is the refractive index in the medium between the donor and acceptor, ϕ_D is the donor fluorescence quantum yield in the absence of acceptor, f_D is the donor emission spectrum with its area normalized to 1, ϵ_A is the spectrum of molar extinction coefficient of the acceptor, and κ^2 is the orientation factor of the dyes. The range of distances that can be accurately measured with FRET is $0.5R_0$ to $1.5R_0$ (for commonly used smFRET dye pairs, this translates into a dynamic range of ~3 to 9 nm). The parameters in Eq. 2 (R , ϕ_D , κ^2 , n) may dynamically change and therefore complicate the interpretation of smFRET distance measurements. Careful control experiments are therefore required. Note that E can be transformed into R if the dyes can be approximated by point dipoles. This approximation holds if the dye sizes are much smaller than R . Although smFRET has been demonstrated using quantum dots (152), they are too large to be approximated by point dipoles. Similarly, genetically encoded fluorescent proteins that are frequently used to monitor binding events and conformational changes in vivo (153) have chromophore groups that are bound deep inside their cores, complicating the transformation of E to R . Small and bright organic fluorophores are therefore the emitters of choice for smFRET measurements.

The average E can be measured experimentally using several approaches. The most straightforward way uses the donor mean fluorescence lifetimes (Eq. 3):

$$\langle E \rangle = 1 - \frac{\langle \tau_{DA} \rangle}{\langle \tau_{DO} \rangle} \quad (3)$$

where $\langle \tau_{DA} \rangle$ and $\langle \tau_{DO} \rangle$ are the donor mean fluorescence lifetimes in the presence or absence of an acceptor, respectively. Knowing the value of R_0 of the dye pair, it is possible to deduce the mean distance between dyes using Eq. 1. $\langle \tau_{DA} \rangle$ and $\langle \tau_{DO} \rangle$ can be retrieved in a single measurement via nanosecond alternating laser excitation (nsALEX) (92) [also known as pulsed-interleaved excitation (PIE) (103)], in which donor and acceptor are alternately excited with pulsed lasers and fluorescence photons are collected using time-correlated single photon counting (TCSPC).

A simpler approach extracts E from the donor and acceptor fluorescence intensities recorded using continuous-wave donor excitation, or with alternated donor and acceptor excitations [microsecond ALEX (103)] (Fig. 2, D and E; Eqs. 4 and 5):

$$\langle E \rangle = \frac{F_{\text{FRET}}}{\gamma F_D^D + F_{\text{FRET}}} \quad (4)$$

$$F_{\text{FRET}} = F_D^A - lk F_D^D - dir F_A^A \quad (5)$$

where F_D^D and F_A^A are the background-corrected fluorescence intensities of the donor and the acceptor, respectively, measured during donor excitation; in the case of ALEX, F_A^A is the background-corrected acceptor fluorescence intensity during acceptor excitation, lk is the donor fluorescence leakage into the acceptor detection channel, dir is the acceptor fluorescence when directly excited by the donor excitation laser, and γ is the ratio between acceptor and donor fluorescence quantum yields and detection efficiencies.

In ensemble FRET, the measured $\langle E \rangle$ reports on all molecules in all conformations; by contrast, in smFRET, $\langle E \rangle$ values (diffusing or immobilized formats, Fig. 5B) represent time-averaged FRET values over limited duration and/or limited number of molecules or events. During a single-molecule burst or time trace, the molecule might not visit all the states that define the system. The average of all $\langle E \rangle$ values for many single molecules and over a long enough observation will equal the ensemble-averaged $\langle E \rangle$. smFRET can distinguish between distinct subpopulations of $\langle E \rangle$ values, and each subpopulation may represent a distinct conformational state. However, if interconversion between conformational states takes place on time scales faster than the method's temporal resolution, $\langle E \rangle$ subpopulations may only represent time averages of these interconverting states.

preserved for the long term for independent validations and future reanalysis with new methods. Depositing the raw data, analysis tools, and pipelines in public repositories such as Dryad, Dataverse, Zenodo, Figshare, or Github (128) will allow different groups to cross-validate results and accelerate the development of new analysis tools.

So far, we have discussed past and present applications of smFRET in biophysics, biochemistry, molecular biology, and structural biology. Future technological advances striving to overcome the current limitations of smFRET measurements could further extend the power of smFRET. Several areas of improvement can be envisioned: temporal resolution, extension to more in vivo and in vitro experimental formats, simplification, and higher throughput compatible with biopharmaceutical applications.

What is next for smFRET?

smFRET has become the accepted method for dynamic structural biology but is still almost entirely used in the context of in vitro experiments. In vivo smFRET, which has recently emerged as a promising methodology requiring further development (43), may allow explorations of conformational dynamics and heterogeneity in the living cell—an approach so far limited to bulk “in-cell NMR” (132) and “in-cell FCS” (133). By removing the artificial constraints of in vitro experiments, in vivo smFRET promises to shed light on outstanding questions in biology by monitoring smFRET as a function of location, diffusivity, and interactions with other partners, thereby illuminating the long-sought link between conformational states and dynamics of biomolecules. Some of the challenges of in vivo smFRET measurements are the generally low signal-to-noise (S/N) and signal-to-background (S/B) ratios and poor photostability of fluorescence proteins. Organic fluorophores are the probes of choice, but their use in live cells requires specific delivery and tagging protocols, which generally introduce larger perturbations and uncertainties than conventional molecular biology techniques. Using microinjection in cultured cells, Sakon and Weninger were the first to track folding of individual SNARE proteins (134). Recently, Schuler and co-workers also used microinjection and smFRET to probe the submicrosecond dynamics of individual freely diffusing, intrinsically disordered proteins in different cellular compartments (Fig. 7, lower center) (70). The ability to distinguish between subpopulations while also detecting fast dynamics allows identification of different folding behaviors in the cytosol and the nucleus. While successfully used in these examples, microinjection relies on high-precision and low-throughput procedures. Techniques for internalization of labeled molecules, such as electroporation in bacteria and in yeast (135) or permeabilization using pore-forming reagents in mammalian cells (136), are more feasible. For instance, several studies have probed the conformations and localizations of doubly labeled oligonucleotides after microinjection in eukaryotic cells (137) or electroporation in bacteria (Fig. 7, lower left) (135). Simpler, robust delivery

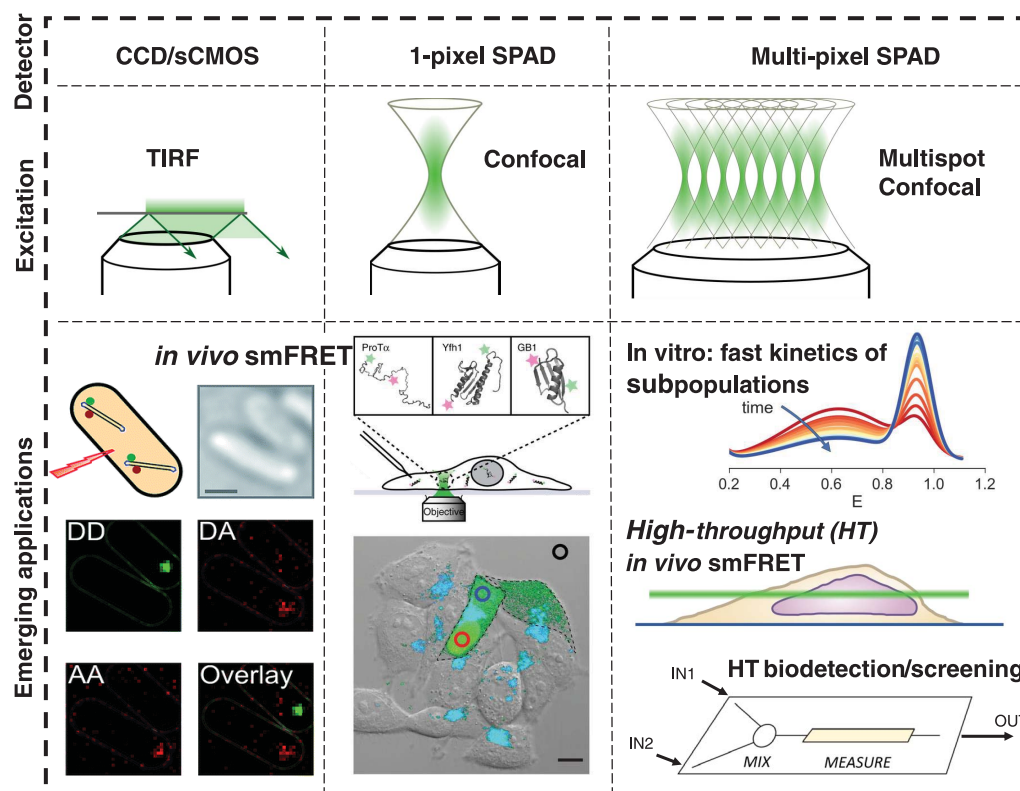


Fig. 7. Emerging applications and future directions of smFRET.

Top rows show current detector and excitation formats for smFRET; the bottom row shows emerging developments that go beyond existing capabilities. smFRET measurements have been demonstrated in live bacteria using TIRF with probes internalized via electroporation [left; adapted from (135)] and in eukaryotic cells using confocal

excitation and microinjected molecules [center; after (70)]. Multipixel SPADs (right) allow fast detection schemes and will allow retrieval of FRET trajectories of single molecules in vivo (scanning different z-layers via light-sheet microscopy) and in vitro (nonequilibrium kinetics via smFRET using mixers or continuous-flow microfluidic devices) [adapted from (143)].

strategies for better labeling yields and high cell viability are needed.

Commonly used illumination geometries such as TIR or confocal imaging are not always ideal. In TIR excitation mode, only a thin layer (~100 nm) of the cell above the cover glass is illuminated by an evanescent field (Fig. 7, upper left). Confocal excitation (Fig. 7, upper center) allows observation deeper into the cell while maintaining good S/N and S/B, but the diffraction-limited sampling volume requires raster scanning for image formation, which competes with continuous recording at each location. Traditional wide-field epi- or trans-illumination is unsuitable for single-molecule detection because of low S/N and S/B and high levels of photobleaching and phototoxicity. Light-sheet or single-plane illumination microscopy (SPIM) is more complex but enables 3D sectioning with high background rejection, limited phototoxicity, and bleaching, and has been successfully extended to single-molecule imaging (138). The use of SPIM for in vivo smFRET measurements could therefore provide new opportunities, as suggested by recent work using a simplified version (139). Combined with fast detectors [scientific-CMOS (sCMOS) cameras or single-photon avalanche diode (SPAD) arrays], probing fast biological events such as protein binding and conformational dynamics in live cells may become feasible (Fig. 7, right).

smFRET on immobilized molecules allows continuous monitoring of conformations or binding events with fairly good temporal resolution. A potential drawback is the introduction of artificial perturbations due to the surface proximity and immobilization chemistry. One way to bypass this problem is to entrap individual molecules in immobilized lipid vesicles (140, 141). However, this technique limits the ability to modulate the local environment (for instance, by buffer exchange). Another solution, the anti-Brownian electrokinetic (ABEL) trap (142), counteracts the Brownian diffusion of a single molecule in solution by active modulations of an external electric field, but this requires observing one molecule at a time and results in very low throughput. We note that in some cases, proteins and DNA molecules may gain different structures or activities under such conditions. To overcome the need for immobilizing or trapping molecules, we envision confinement within a thin chamber (<100 nm) limiting the diffusion along the *z* axis. Combined with fast detectors such as sCMOS cameras or SPAD arrays, this would enable tracking of multiple molecules for extended periods of time during their quasi-2D diffusion with reduced surface-interaction artifacts.

A major drawback of diffusion-based smFRET measurements is the long acquisition time (several minutes) needed to accumulate a large num-

ber of single-molecule bursts. Acquisition times on the order of a few seconds would enable an entirely new class of applications such as diffusion-based smFRET kinetic studies, high-throughput (HT) drug screening, and diagnostic assays (Fig. 7, lower right). Throughput can be multiplied by parallel acquisition from multiple excitation spots using SPAD arrays for detection (Fig. 7, right). This multispot approach provides a reduction in acquisition time proportional to the number of spots (143), which could potentially reach up to 1000 pixels for next-generation SPAD arrays suitable for single-molecule detection (144). Although the excitation geometry can be multispot as well, more scalable excitation schemes include zero-mode waveguides or light-sheet illumination. Such schemes will eliminate the tedious task of aligning the excitation pattern to the detector pixels. Additionally, as noted above, such 2D illuminations allow the use of fast sCMOS cameras (>100 Hz full frame, >1 kHz partial frame) (145), which may be sufficient for some applications such as high-throughput screening that currently relies on SPADs or live-cell smFRET imaging (Fig. 7, lower right) (135, 139).

Whereas ensemble kinetics can identify kinetic processes that are well separated in time, nonequilibrium smFRET kinetic studies can identify multiple conformations or binding states and their associated transitions. Non-equilibrium

smFRET kinetic studies rely on rapid exchange or mixing of reagents to initiate a perturbation in the system under study. In experiments on immobilized molecules, rapid exchange of conditions initiates the reaction, which is then recorded in as many time trajectories as there are molecules. Time-dependent FRET efficiency histograms that describe the reaction are computed by aligning all the smFRET trajectories (145). Non-equilibrium smFRET kinetic studies performed on diffusing molecules are more challenging because molecules randomly enter the observation volume, preventing continuous time traces to be acquired. Ingargiola *et al.* recently measured the kinetics of transcription initiation (promoter escape) using a multispot setup (143). They directly followed the kinetics by monitoring the conformation of the transcription bubble at the single-molecule level with 30-s temporal resolution, limited only by the number of single-molecule bursts detected across the eight spots (Fig. 7). A 48-spot system currently in development (146) should improve the temporal resolution accordingly. However, as the temporal resolution of the measurement is improved, faster mixing is required. A combination of a continuous-flow microfluidic mixer (147) together with multispot detection could provide the solution for fast non-equilibrium smFRET kinetic studies of diffusing molecules (Fig. 7, lower right).

Drug discovery using drug-ligand interactions measurements relies on high-throughput ensemble techniques to rapidly screen large libraries of small molecules for identification of interactions and quantification of affinities. Various screening methods differ in the range of affinities they can measure, their throughput, sample consumption, accuracy, measurement modality (kinetic, steady-state), possible requirement for immobilization, lowest binding stoichiometry, etc. Many of the techniques that allow high-throughput screening of more than 10^4 molecules per day report either quantitatively on low affinity ranges in bulk, or on interactions with immobilized small molecules measured by surface plasmon resonance (148). Diffusion-based smFRET assays could enable probing such interactions with minimal sample consumption. However, until recently, such measurements required long acquisition times. Additionally, such screening requires an automated system that can rapidly exchange conditions. Kim *et al.* used a microfluidic mixing device to automate titration from many input channels and perform serial smFRET measurements at different conditions (149). Multispot and multicolor smFRET in combination with an automated mixing device would allow highly multiplexed smFRET measurements, suitable for high-throughput screening (Fig. 7). As an example, a 1024-spot system would allow measurements lasting ~250 ms, translating to ~350,000 assays per day (assuming that the microfluidic chip enables dispensing of as many samples at this frequency). The same approach could be used to titrate binding components to produce affinity curves for each ligand down to picomolar concentrations and in varying conditions. Alternatively, a multispot setup could be used in conjunction with

a titer plate and scanning stage, where many different conditions in each well can be tested at a much higher rate than with a single-spot excitation. Lastly, using multispot smFRET acquisition in a stopped-flow format would allow measuring association and dissociation rate constants and extraction of molecular affinities (Fig. 7, right).

Likewise, molecular diagnostics could benefit from high-throughput smFRET capabilities. Such applications require highly specific and sensitive molecular recognition of low-abundance molecular markers (proteins, self-antibodies, microRNAs, freely circulating DNA, etc.) in a small volume of bodily fluids, ideally without any amplification (107, 150). Here again, fast acquisition coupled with automation of liquid handling is required. A combination of multispot smFRET with multicolor ALEX capabilities and a microfluidic chip could provide a powerful molecular diagnostics platform.

Conclusion

Two decades after its introduction, the promise of smFRET has largely materialized; several variants have now reached maturity to form a robust and mainstream toolkit available to biochemists, molecular biologists, and biophysicists. Commercial systems implementing smFRET have been introduced in recent years. We anticipate further development of such systems into turnkey and, eventually, fully automated devices based on open-source and validated data analysis algorithms, which will lower the barrier of entry to this powerful technology and further help to disseminate the method.

REFERENCES AND NOTES

1. A. T. Brünger, X-ray crystallography and NMR reveal complementary views of structure and dynamics. *Nat. Struct. Biol.* **4** (suppl.), 862–865 (1997). PMID: 9377160
2. A. H. Zewail, Diffraction, crystallography and microscopy beyond three dimensions: Structural dynamics in space and time. *Philos. Trans. R. Soc. A* **363**, 315–329 (2005). doi: 10.1098/rsta.2004.1513; PMID: 15664902
3. T. A. Bharat, C. J. Russo, J. Löwe, L. A. Passmore, S. H. Scheres, Advances in single-particle electron cryomicroscopy structure determination applied to sub-tomogram averaging. *Structure* **23**, 1743–1753 (2015). doi: 10.1016/j.str.2015.06.026; PMID: 26256537
4. Y. Cheng, Single-particle cryo-EM at crystallographic resolution. *Cell* **161**, 450–457 (2015). doi: 10.1016/j.cell.2015.03.049; PMID: 25910205
5. M. Hennig, A. Ruf, W. Huber, Combining biophysical screening and X-ray crystallography for fragment-based drug discovery. *Top. Curr. Chem.* **317**, 115–143 (2011). doi: 10.1007/128_2011_225; PMID: 21837555
6. X. Zhuang *et al.*, A single-molecule study of RNA catalysis and folding. *Science* **288**, 2048–2051 (2000). doi: 10.1126/science.288.5473.2048; PMID: 10856219
7. X. Michalet, S. Weiss, M. Jäger, Single-molecule fluorescence studies of protein folding and conformational dynamics. *Chem. Rev.* **106**, 1785–1813 (2006). doi: 10.1021/cr0404343; PMID: 16683755
8. M. Pirchi *et al.*, Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. *Nat. Commun.* **2**, 493 (2011). doi: 10.1038/ncomms1504; PMID: 21988909
9. H. S. McHaourab, P. R. Steed, K. Kazmier, Toward the fourth dimension of membrane protein structure: Insight into dynamics from spin-labeling EPR spectroscopy. *Structure* **19**, 1549–1561 (2011). doi: 10.1016/j.str.2011.10.009; PMID: 22078555
10. G. Jeschke, DEER distance measurements on proteins. *Annu. Rev. Phys. Chem.* **63**, 419–446 (2012). doi: 10.1146/annurev-physchem-032511-143716; PMID: 22404592
11. R. Rigler, M. Ehrenberg, Molecular interactions and structure as analysed by fluorescence relaxation

- spectroscopy. *Q. Rev. Biophys.* **6**, 139–199 (1973). doi: 10.1017/S003358350000113X; PMID: 4579675
12. E. Haas, Ensemble FRET methods in studies of intrinsically disordered proteins. *Methods Mol. Biol.* **895**, 467–498 (2012). doi: 10.1007/978-1-61779-927-3_28; PMID: 22760335
13. H. Yang *et al.*, Protein conformational dynamics probed by single-molecule electron transfer. *Science* **302**, 262–266 (2003). doi: 10.1126/science.1086911; PMID: 14551431
14. G. Lipari, A. Szabo, Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559 (1982). doi: 10.1021/ja00381a009
15. E. Meirovitch, Y. E. Shapiro, A. Polimeno, J. H. Freed, Structural dynamics of bio-macromolecules by NMR: The slowly relaxing local structure approach. *Prog. Nucl. Magn. Reson. Spectrosc.* **56**, 360–405 (2010). doi: 10.1016/j.pnmrs.2010.03.002; PMID: 20625480
16. E. Lerner, T. Orevi, E. Ben Ishay, D. Amir, E. Haas, Kinetics of fast changing intramolecular distance distributions obtained by combined analysis of FRET efficiency kinetics and time-resolved FRET equilibrium measurements. *Biophys. J.* **106**, 667–676 (2014). doi: 10.1016/j.bpj.2013.11.4500; PMID: 24507607
17. G. Rahamim, M. Chemerovski-Glikman, S. Rahimpour, D. Amir, E. Haas, Resolution of two sub-populations of conformers and their individual dynamics by time resolved ensemble level FRET measurements. *PLOS ONE* **10**, e0143732 (2015). doi: 10.1371/journal.pone.0143732; PMID: 26699718
18. X. S. Xie, R. C. Dunn, Probing single molecule dynamics. *Science* **265**, 361–364 (1994). doi: 10.1126/science.265.5170.361; PMID: 17838036
19. X. S. Xie, H. P. Lu, Single-molecule enzymology. *J. Biol. Chem.* **274**, 15967–15970 (1999). doi: 10.1074/jbc.274.23.15967; PMID: 10347141
20. F. Kulzer, M. Orrit, Single-molecule optics. *Annu. Rev. Phys. Chem.* **55**, 585–611 (2004). doi: 10.1146/annurev.physchem.54.011002.103816; PMID: 15117263
21. W. Min *et al.*, Fluctuating enzymes: Lessons from single-molecule studies. *Acc. Chem. Res.* **38**, 923–931 (2005). doi: 10.1021/ar040133f; PMID: 16359164
22. M. Orrit, T. Ha, V. Sandoghdar, Single-molecule optical spectroscopy. *Chem. Soc. Rev.* **43**, 973–976 (2014). doi: 10.1039/c4cs90001d; PMID: 24429724
23. M. Orrit, Single-molecule chemistry is more than superresolved fluorescence microscopy. *Angew. Chem. Int. Ed.* **54**, 8004–8005 (2015). doi: 10.1002/anie.201503674; PMID: 26074472
24. W. E. Moerner, Y. Shechtman, Q. Wang, Single-molecule spectroscopy and imaging over the decades. *Faraday Discuss.* **184**, 9–36 (2015). doi: 10.1039/C5FD00149H; PMID: 26616210
25. T. Ha *et al.*, Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6264–6268 (1996). doi: 10.1073/pnas.93.13.6264; PMID: 8692803
26. G. J. Schütz, W. Trabesinger, T. Schmidt, Direct observation of ligand colocalization on individual receptor molecules. *Biophys. J.* **74**, 2223–2226 (1998). doi: 10.1016/S0006-3495(98)77931-7; PMID: 9591649
27. T. Ha *et al.*, Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 893–898 (1999). doi: 10.1073/pnas.96.3.893; PMID: 9927664
28. A. A. Deniz *et al.*, Single-pair fluorescence resonance energy transfer on freely diffusing molecules: Observation of Förster distance dependence and subpopulations. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3670–3675 (1999). doi: 10.1073/pnas.96.7.3670; PMID: 10097095
29. A. A. Deniz *et al.*, Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5179–5184 (2000). doi: 10.1073/pnas.090104997; PMID: 10792044
30. Y. W. Jia *et al.*, Folding dynamics of single Gcn4 peptides by fluorescence resonant energy transfer confocal microscopy. *Chem. Phys.* **247**, 69–83 (1999). doi: 10.1016/S0301-0104(99)00127-5
31. T. Ha *et al.*, Ligand-induced conformational changes observed in single RNA molecules. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9077–9082 (1999). doi: 10.1073/pnas.96.16.9077; PMID: 10430898
32. E. Haas, I. Z. Steinberg, Intramolecular dynamics of chain molecules monitored by fluctuations in efficiency of

- excitation energy transfer. A theoretical study. *Biophys. J.* **46**, 429–437 (1984). doi: [10.1016/S0006-3495\(84\)80400-0](https://doi.org/10.1016/S0006-3495(84)80400-0); pmid: 6498263
33. T. Hirschfeld, Optical microscopic observation of single small molecules. *Appl. Opt.* **15**, 2965–2966 (1976). doi: [10.1364/AO.15.002965](https://doi.org/10.1364/AO.15.002965); pmid: 20168369
 34. W. E. Moerner, L. Kador, Optical detection and spectroscopy of single molecules in a solid. *Phys. Rev. Lett.* **62**, 2535–2538 (1989). doi: [10.1103/PhysRevLett.62.2535](https://doi.org/10.1103/PhysRevLett.62.2535); pmid: 10040013
 35. M. Orrit, J. Bernard, Single pentacene molecules detected by fluorescence excitation in a p-terphenyl crystal. *Phys. Rev. Lett.* **65**, 2716–2719 (1990). doi: [10.1103/PhysRevLett.65.2716](https://doi.org/10.1103/PhysRevLett.65.2716); pmid: 10042674
 36. E. Brooks Shera, N. K. Seitzinger, L. M. Davis, R. A. Keller, S. A. Soper, Detection of single fluorescent molecules. *Chem. Phys. Lett.* **174**, 553–557 (1990). doi: [10.1016/0009-2614\(90\)85485-U](https://doi.org/10.1016/0009-2614(90)85485-U)
 37. E. Betzig, R. J. Chichester, Single molecules observed by near-field scanning optical microscopy. *Science* **262**, 1422–1425 (1993). doi: [10.1126/science.262.5138.1422](https://doi.org/10.1126/science.262.5138.1422); pmid: 17736823
 38. R. Rigler, U. Mets, Diffusion of single molecules through a Gaussian laser-beam. *Proc. SPIE* **1921**, 239 (1993). doi: [10.1117/12.146154](https://doi.org/10.1117/12.146154)
 39. T. Funatsu, Y. Harada, M. Tokunaga, K. Saito, T. Yanagida, Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution. *Nature* **374**, 555–559 (1995). doi: [10.1038/374555a0](https://doi.org/10.1038/374555a0); pmid: 7700383
 40. J. G. Yodh, M. Schlierf, T. Ha, Insight into helicase mechanism and function revealed through single-molecule approaches. *Q. Rev. Biophys.* **43**, 185–217 (2010). doi: [10.1017/S0033583510000107](https://doi.org/10.1017/S0033583510000107); pmid: 20682090
 41. M. Stracy, S. Uphoff, F. Garza de Leon, A. N. Kapanidis, In vivo single-molecule imaging of bacterial DNA replication, transcription, and repair. *FEBS Lett.* **588**, 3585–3594 (2014). doi: [10.1016/j.febslet.2014.05.026](https://doi.org/10.1016/j.febslet.2014.05.026); pmid: 24859634
 42. A. Robinson, A. M. van Oijen, Bacterial replication, transcription and translation: Mechanistic insights from single-molecule biochemical studies. *Nat. Rev. Microbiol.* **11**, 303–315 (2013). doi: [10.1038/nrmicro2994](https://doi.org/10.1038/nrmicro2994); pmid: 23549067
 43. M. Sustarsic, A. N. Kapanidis, Taking the ruler to the jungle: Single-molecule FRET for understanding biomolecular structure and dynamics in live cells. *Curr. Opin. Struct. Biol.* **34**, 52–59 (2015). doi: [10.1016/j.sbi.2015.07.001](https://doi.org/10.1016/j.sbi.2015.07.001); pmid: 26295172
 44. Y. Alhadi et al., Studying transcription initiation by RNA polymerase with diffusion-based single-molecule fluorescence. *Protein Sci.* **26**, 1278–1290 (2017). doi: [10.1002/pro.3160](https://doi.org/10.1002/pro.3160); pmid: 28370550
 45. J. B. Munro, A. Vaiana, K. Y. Sanbonmatsu, S. C. Blanchard, A new view of protein synthesis: Mapping the free energy landscape of the ribosome using single-molecule FRET. *Biopolymers* **89**, 565–577 (2008). doi: [10.1002/bip.20961](https://doi.org/10.1002/bip.20961); pmid: 18286627
 46. S. C. Blanchard, Single-molecule observations of ribosome function. *Curr. Opin. Struct. Biol.* **19**, 103–109 (2009). doi: [10.1016/j.sbi.2009.01.002](https://doi.org/10.1016/j.sbi.2009.01.002); pmid: 19223173
 47. H. P. Lu, Revealing time bunching effect in single-molecule enzyme conformational dynamics. *Phys. Chem. Chem. Phys.* **13**, 6734–6749 (2011). doi: [10.1039/c0cp02860f](https://doi.org/10.1039/c0cp02860f); pmid: 21409227
 48. T. R. Weikl, F. Paul, Conformational selection in protein binding and function. *Protein Sci.* **23**, 1508–1518 (2014). doi: [10.1002/pro.2539](https://doi.org/10.1002/pro.2539); pmid: 25155241
 49. D. K. Sasmal, L. E. Pulido, S. Kasal, J. Huang, Single-molecule fluorescence resonance energy transfer in molecular biology. *Nanoscale* **8**, 19928–19944 (2016). doi: [10.1039/C6NR06794H](https://doi.org/10.1039/C6NR06794H); pmid: 27883140
 50. M. Börsch, T. M. Duncan, Spotlighting motors and controls of single F_0F_1 -ATP synthase. *Biochem. Soc. Trans.* **41**, 1219–1226 (2013). doi: [10.1042/BST20130101](https://doi.org/10.1042/BST20130101); pmid: 24059511
 51. J. Diao, Y. Ishitsuka, W. R. Bae, Single-molecule FRET study of SNARE-mediated membrane fusion. *Biosci. Rep.* **31**, 457–463 (2011). doi: [10.1042/BSR20110011](https://doi.org/10.1042/BSR20110011); pmid: 21919892
 52. B. Schuler, W. A. Eaton, Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **18**, 16–26 (2008). doi: [10.1016/j.sbi.2007.12.003](https://doi.org/10.1016/j.sbi.2007.12.003); pmid: 18221865
 53. Y. Gambin, A. A. Deniz, Multicolor single-molecule FRET to explore protein folding and binding. *Mol. Biosyst.* **6**, 1540–1547 (2010). doi: [10.1039/c003024d](https://doi.org/10.1039/c003024d); pmid: 20601974
 54. R. Zhao, D. Rueda, RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods* **49**, 112–117 (2009). doi: [10.1016/j.jmeth.2009.04.017](https://doi.org/10.1016/j.jmeth.2009.04.017); pmid: 19409995
 55. S. Preus, L. M. Wilhelmsson, Advances in quantitative FRET-based methods for studying nucleic acids. *ChemBioChem* **13**, 1990–2001 (2012). doi: [10.1002/cbic.201200400](https://doi.org/10.1002/cbic.201200400); pmid: 22936620
 56. X. Zhuang, Single-molecule RNA science. *Annu. Rev. Biophys. Biomol. Struct.* **34**, 399–414 (2005). doi: [10.1146/annurev.biophys.34.040204.144641](https://doi.org/10.1146/annurev.biophys.34.040204.144641); pmid: 15869396
 57. M. Helm, A. Y. Kobitski, G. U. Nienhaus, Single-molecule Förster resonance energy transfer studies of RNA structure, dynamics and function. *Biophys. Rev.* **1**, 161–176 (2009). doi: [10.1007/s12551-009-0018-3](https://doi.org/10.1007/s12551-009-0018-3); pmid: 28510027
 58. D. Klostermeier, Single-molecule FRET reveals nucleotide-driven conformational changes in molecular machines and their link to RNA unwinding and DNA supercoiling. *Biochem. Soc. Trans.* **39**, 611–616 (2011). doi: [10.1042/BST0390611](https://doi.org/10.1042/BST0390611); pmid: 21428949
 59. N. Bisaria, D. Herschlag, Probing the kinetic and thermodynamic consequences of the tetraloop/tetraloop receptor monovalent ion-binding site in P4-P6 RNA by smFRET. *Biochem. Soc. Trans.* **43**, 172–178 (2015). doi: [10.1042/BST20140268](https://doi.org/10.1042/BST20140268); pmid: 25849913
 60. A. N. Kapanidis et al., Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1144–1147 (2006). doi: [10.1126/science.1131399](https://doi.org/10.1126/science.1131399); pmid: 17110578
 61. A. Revyakin, C. Liu, R. H. Ebright, T. R. Strick, Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**, 1139–1143 (2006). doi: [10.1126/science.1131398](https://doi.org/10.1126/science.1131398); pmid: 17110577
 62. K. A. Henzler-Wildman et al., Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838–844 (2007). doi: [10.1038/nature06410](https://doi.org/10.1038/nature06410); pmid: 18026086
 63. J. A. Hanson et al., Illuminating the mechanistic roles of enzyme conformational dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18055–18060 (2007). doi: [10.1073/pnas.0708600104](https://doi.org/10.1073/pnas.0708600104); pmid: 17989222
 64. M. A. Sinev, E. V. Sineva, V. Ittah, E. Haas, Domain closure in adenylate kinase. *Biochemistry* **35**, 6425–6437 (1996). doi: [10.1021/bi952687j](https://doi.org/10.1021/bi952687j); pmid: 8639589
 65. P. S. Huang, S. E. Boyken, D. Baker, The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016). doi: [10.1038/nature19946](https://doi.org/10.1038/nature19946); pmid: 27629638
 66. B. Schuler, H. Hofmann, Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales. *Curr. Opin. Struct. Biol.* **23**, 36–47 (2013). doi: [10.1016/j.sbi.2012.10.008](https://doi.org/10.1016/j.sbi.2012.10.008); pmid: 23312353
 67. E. Shaw, P. St-Pierre, K. McCluskey, D. A. Lafontaine, J. C. Penedo, Using sm-FRET and denaturants to reveal folding landscapes. *Methods Enzymol.* **549**, 313–341 (2014). doi: [10.1016/B978-0-12-801122-5.00014-3](https://doi.org/10.1016/B978-0-12-801122-5.00014-3); pmid: 25432755
 68. V. A. Voelz et al., Slow unfolded-state structuring in Acyl-CoA binding protein folding revealed by simulation and experiment. *J. Am. Chem. Soc.* **134**, 12565–12577 (2012). doi: [10.1021/ja302528z](https://doi.org/10.1021/ja302528z); pmid: 22747188
 69. B. Schuler, E. A. Lipman, W. A. Eaton, Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419**, 743–747 (2002). doi: [10.1038/nature01060](https://doi.org/10.1038/nature01060); pmid: 12384704
 70. I. König et al., Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nat. Methods* **12**, 773–779 (2015). doi: [10.1038/nmeth.3475](https://doi.org/10.1038/nmeth.3475); pmid: 26147918
 71. A. C. Ferreón, Y. Gambin, E. A. Lemke, A. A. Deniz, Interplay of α -synuclein binding and conformational switching probed by single-molecule fluorescence. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5645–5650 (2009). doi: [10.1073/pnas.0809232106](https://doi.org/10.1073/pnas.0809232106); pmid: 19293380
 72. S. J. Chen, RNA folding: Conformational statistics, folding kinetics, and ion electrostatics. *Annu. Rev. Biophys.* **37**, 197–214 (2008). doi: [10.1146/annurev.biophys.37.032807.125957](https://doi.org/10.1146/annurev.biophys.37.032807.125957); pmid: 18573079
 73. D. K. Treiber, J. R. Williamson, Beyond kinetic traps in RNA folding. *Curr. Opin. Struct. Biol.* **11**, 309–314 (2001). doi: [10.1016/S0959-440X\(00\)00206-2](https://doi.org/10.1016/S0959-440X(00)00206-2); pmid: 11406379
 74. T. J. Wilson et al., RNA folding and the origins of catalytic activity in the hairpin ribozyme. *Blood Cells Mol. Dis.* **38**, 8–14 (2007). doi: [10.1016/j.bcmd.2006.10.004](https://doi.org/10.1016/j.bcmd.2006.10.004); pmid: 17150385
 75. S. Liu, G. Bokinsky, N. G. Walter, X. Zhuang, Dissecting the multistep reaction pathway of an RNA enzyme by single-molecule kinetic “fingerprinting”. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12634–12639 (2007). doi: [10.1073/pnas.0610597104](https://doi.org/10.1073/pnas.0610597104); pmid: 17496145
 76. Q. Zheng et al., Ultra-stable organic fluorophores for single-molecule research. *Chem. Soc. Rev.* **43**, 1044–1056 (2014). doi: [10.1039/C3CS60237K](https://doi.org/10.1039/C3CS60237K); pmid: 24177677
 77. J. H. van der Velde et al., A simple and versatile design concept for fluorophore derivatives with intramolecular photostabilization. *Nat. Commun.* **7**, 10144 (2016). doi: [10.1038/ncomms10144](https://doi.org/10.1038/ncomms10144); pmid: 26751640
 78. T. Ha, P. Tinnefeld, Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. *Annu. Rev. Phys. Chem.* **63**, 595–617 (2012). doi: [10.1146/annurev-physchem-032210-103340](https://doi.org/10.1146/annurev-physchem-032210-103340); pmid: 22404588
 79. R. Roy, S. Hohng, T. Ha, A practical guide to single-molecule FRET. *Nat. Methods* **5**, 507–516 (2008). doi: [10.1038/nmeth.1208](https://doi.org/10.1038/nmeth.1208); pmid: 18511918
 80. S. E. Kim, I. B. Lee, C. Hyeon, S. C. Hong, Deciphering kinetic information from single-molecule FRET data that show slow transitions. *J. Phys. Chem. B* **119**, 6974–6978 (2015). doi: [10.1021/acs.jpbc.5b03991](https://doi.org/10.1021/acs.jpbc.5b03991); pmid: 25989531
 81. J. Chen, A. Tsai, S. E. O’Leary, A. Petrov, J. D. Puglisi, Unraveling the dynamics of ribosome translocation. *Curr. Opin. Struct. Biol.* **22**, 804–814 (2012). doi: [10.1016/j.sbi.2012.09.004](https://doi.org/10.1016/j.sbi.2012.09.004); pmid: 23142574
 82. S. Liu, B. T. Harada, J. T. Miller, S. F. Le Grice, X. Zhuang, Initiation complex dynamics direct the transitions between distinct phases of early HIV reverse transcription. *Nat. Struct. Mol. Biol.* **17**, 1453–1460 (2010). doi: [10.1038/nmsb.1937](https://doi.org/10.1038/nmsb.1937); pmid: 21102446
 83. E. Nir et al., Shot-noise limited single-molecule FRET histograms: Comparison between theory and experiments. *J. Phys. Chem. B* **110**, 22103–22124 (2006). doi: [10.1021/jp063483n](https://doi.org/10.1021/jp063483n); pmid: 17078646
 84. S. Kalinin, E. Sisamak, S. W. Magennis, S. Felekyan, C. A. M. Seidel, On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. *J. Phys. Chem. B* **114**, 6197–6206 (2010). doi: [10.1021/jp100025v](https://doi.org/10.1021/jp100025v); pmid: 20397670
 85. T. E. Tomov et al., Disentangling subpopulations in single-molecule FRET and ALEX experiments with photon distribution analysis. *Biophys. J.* **102**, 1163–1173 (2012). doi: [10.1016/j.bpj.2011.11.4025](https://doi.org/10.1016/j.bpj.2011.11.4025); pmid: 22404939
 86. J. P. Torella, S. J. Holden, Y. Santoso, J. Hohlbein, A. N. Kapanidis, Identifying molecular dynamics in single-molecule FRET experiments with burst variance analysis. *Biophys. J.* **100**, 1568–1577 (2011). doi: [10.1016/j.bpj.2011.01.066](https://doi.org/10.1016/j.bpj.2011.01.066); pmid: 21402040
 87. N. C. Robb et al., The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: Implications for transcription start-site selection. *J. Mol. Biol.* **425**, 875–885 (2013). doi: [10.1016/j.jmb.2012.12.015](https://doi.org/10.1016/j.jmb.2012.12.015); pmid: 23274143
 88. A. Hoffmann et al., Quantifying heterogeneity and conformational dynamics from single molecule FRET of diffusing molecules: Recurrence analysis of single particles (RASAP). *Phys. Chem. Chem. Phys.* **13**, 1857–1871 (2011). doi: [10.1039/c0cp01911a](https://doi.org/10.1039/c0cp01911a); pmid: 21218223
 89. D. Nettels, A. Hoffmann, B. Schuler, Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. *J. Phys. Chem. B* **112**, 6137–6146 (2008). doi: [10.1021/jp076971j](https://doi.org/10.1021/jp076971j); pmid: 18410159
 90. A. K. Woźniak, G. F. Schröder, H. Grubmüller, C. A. Seidel, F. Oesterhelt, Single-molecule FRET measures bends and kinks in DNA. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18337–18342 (2008). doi: [10.1073/pnas.0800977105](https://doi.org/10.1073/pnas.0800977105); pmid: 19020079
 91. D. M. Dolino, S. Rezaei Adariani, S. A. Shaikh, V. Jayaraman, H. Sanabria, Conformational selection and submillisecond dynamics of the ligand-binding domain of the N-methyl-D-aspartate receptor. *J. Biol. Chem.* **291**, 16175–16185 (2016). doi: [10.1074/jbc.M116.721274](https://doi.org/10.1074/jbc.M116.721274); pmid: 27265681
 92. T. A. Laurence, X. Kong, M. Jäger, S. Weiss, Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17348–17353 (2005). doi: [10.1073/pnas.0508584102](https://doi.org/10.1073/pnas.0508584102); pmid: 16287971
 93. M. Pirchi et al., Photon-by-photon hidden Markov model analysis for microsecond single-molecule FRET kinetics. *J. Phys. Chem. B* **120**, 13065–13075 (2016). doi: [10.1021/acs.jpbc.6b10726](https://doi.org/10.1021/acs.jpbc.6b10726); pmid: 27977207
 94. I. V. Gopich, A. Szabo, Decoding the pattern of photon colors in single-molecule FRET. *J. Phys. Chem. B* **113**, 10965–10973 (2009). doi: [10.1021/jp903671p](https://doi.org/10.1021/jp903671p); pmid: 19588948

95. S. Hohng, S. Lee, J. Lee, M. H. Jo, Maximizing information content of single-molecule FRET experiments: Multi-color FRET and FRET combined with force or torque. *Chem. Soc. Rev.* **43**, 1007–1013 (2014). doi: [10.1039/C3CS60184F](#); pmid: [23970315](#)
96. S. Hohng, C. Joo, T. Ha, Single-molecule three-color FRET. *Biophys. J.* **87**, 1328–1337 (2004). doi: [10.1529/biophysj.104.043935](#); pmid: [15298935](#)
97. B. Person, I. H. Stein, C. Steinhauer, J. Vogelsang, P. Tinnefeld, Correlated movement and bending of nucleic acid structures visualized by multicolor single-molecule spectroscopy. *ChemPhysChem* **10**, 1455–1460 (2009). doi: [10.1002/cphc.200900109](#); pmid: [19499555](#)
98. L. Le Reste, J. Hohlbein, K. Gryte, A. N. Kapanidis, Characterization of dark quencher chromophores as nonfluorescent acceptors for single-molecule FRET. *Biophys. J.* **102**, 2658–2668 (2012). doi: [10.1016/j.bpj.2012.04.028](#); pmid: [22713582](#)
99. J. Vogelsang, T. Cordes, P. Tinnefeld, Single-molecule photophysics of oxazines on DNA and its application in a FRET switch. *Photochem. Photobiol. Sci.* **8**, 486–496 (2009). doi: [10.1039/b822318c](#); pmid: [19337662](#)
100. S. Uphoff et al., Monitoring multiple distances within a single molecule using switchable FRET. *Nat. Methods* **7**, 831–836 (2010). doi: [10.1038/nmeth.1502](#); pmid: [20818380](#)
101. E. Sisamak, A. Valeri, S. Kalinin, P. J. Rothwell, C. A. M. Seidel, Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods Enzymol.* **475**, 455–514 (2010). doi: [10.1016/S0076-6879\(10\)75018-7](#)
102. A. N. Kapanidis et al., Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8936–8941 (2004). doi: [10.1073/pnas.0401690101](#); pmid: [15175430](#)
103. J. Hohlbein, T. D. Craggs, T. Cordes, Alternating-laser excitation: Single-molecule FRET and beyond. *Chem. Soc. Rev.* **43**, 1156–1171 (2014). doi: [10.1039/C3CS60233H](#); pmid: [24037326](#)
104. E. Ploetz et al., Förster resonance energy transfer and protein-induced fluorescence enhancement as synergistic multi-scale molecular rulers. *Sci. Rep.* **6**, 33257 (2016). doi: [10.1038/srep33257](#); pmid: [27641327](#)
105. D. Haenni, F. Zosel, L. Reymond, D. Nettel, B. Schuler, Intramolecular distances and dynamics from the combined photon statistics of single-molecule FRET and photoinduced electron transfer. *J. Phys. Chem. B* **117**, 13015–13028 (2013). doi: [10.1021/jp402352j](#); pmid: [23718771](#)
106. N. K. Lee et al., Three-color alternating-laser excitation of single molecules: Monitoring multiple interactions and distances. *Biophys. J.* **92**, 303–312 (2007). doi: [10.1529/biophysj.106.093211](#); pmid: [17040983](#)
107. S. W. Yim et al., Four-color alternating-laser excitation single-molecule fluorescence spectroscopy for next-generation biodetection assays. *Clin. Chem.* **58**, 707–716 (2012). doi: [10.1373/clinchem.2011.176958](#); pmid: [22266381](#)
108. C. M. Haney, R. F. Wissner, E. J. Petersson, Multiply labeling proteins for studies of folding and stability. *Curr. Opin. Chem. Biol.* **28**, 123–130 (2015). doi: [10.1016/j.cbpa.2015.07.007](#); pmid: [26253346](#)
109. I. Nikić, E. A. Lemke, Genetic code expansion enabled site-specific dual-color protein labeling: Superresolution microscopy and beyond. *Curr. Opin. Chem. Biol.* **28**, 164–173 (2015). doi: [10.1016/j.cbpa.2015.07.021](#); pmid: [26302384](#)
110. S. Weiss, Fluorescence spectroscopy of single biomolecules. *Science* **283**, 1676–1683 (1999). doi: [10.1126/science.283.5408.1676](#); pmid: [10073925](#)
111. D. K. Sasmal, H. P. Lu, Single-molecule patch-clamp FRET microscopy studies of NMDA receptor ion channel dynamics in living cells: Revealing the multiple conformational states associated with a channel at its electrical off state. *J. Am. Chem. Soc.* **136**, 12998–13005 (2014). doi: [10.1021/ja506231j](#); pmid: [25148304](#)
112. M. Spies, Y. R. Chemla, Preface to “Single-Molecule Enzymology: Nanomechanical Manipulation and Hybrid Methods”. *Methods Enzymol.* **582**, i–xv (2017).
113. X. Long, J. W. Parks, M. D. Stone, Integrated magnetic tweezers and single-molecule FRET for investigating the mechanical properties of nucleic acid. *Methods* **105**, 16–25 (2016). doi: [10.1016/j.ymeth.2016.06.009](#); pmid: [27320203](#)
114. J. C. Cordova, D. K. Das, H. W. Manning, M. J. Lang, Combining single-molecule manipulation and single-molecule detection. *Curr. Opin. Struct. Biol.* **28**, 142–148 (2014). doi: [10.1016/j.sbi.2014.08.010](#); pmid: [25255052](#)
115. A. M. Streets, Y. Huang, Microfluidics for biological measurements with single-molecule resolution. *Curr. Opin. Biotechnol.* **25**, 69–77 (2014). doi: [10.1016/j.copbio.2013.08.013](#); pmid: [24484883](#)
116. K. Sikic, S. Tomic, O. Carugo, Systematic comparison of crystal and NMR protein structures deposited in the protein data bank. *Open Biochem. J.* **4**, 83–95 (2010). doi: [10.2174/1874091X01004010083](#); pmid: [21293729](#)
117. A. M. Schielek et al., A nano-positioning system for macromolecular structural analysis. *Nat. Methods* **5**, 965–971 (2008). doi: [10.1038/nmeth.1259](#); pmid: [18849988](#)
118. S. Kalinin et al., A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods* **9**, 1218–1225 (2012). doi: [10.1038/nmeth.2222](#); pmid: [23142871](#)
119. B. Hellenkamp, P. Wortmann, F. Kandzia, M. Zacharias, T. Hugel, Multidomain structure and correlated dynamics determined by self-consistent FRET networks. *Nat. Methods* **14**, 174–180 (2017). doi: [10.1038/nmeth.4081](#); pmid: [27918541](#)
120. T. D. Craggs, A. N. Kapanidis, Six steps closer to FRET-driven structural biology. *Nat. Methods* **9**, 1157–1158 (2012). doi: [10.1038/nmeth.2257](#); pmid: [23223168](#)
121. M. Dimura et al., Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. *Curr. Opin. Struct. Biol.* **40**, 163–185 (2016). doi: [10.1016/j.sbi.2016.11.012](#); pmid: [27939973](#)
122. T. O. Peulen, O. Opanasyuk, C. A. M. Seidel, Combining graphical and analytical methods with molecular simulations to analyze time-resolved FRET measurements of labeled macromolecules accurately. *J. Phys. Chem. B* **121**, 8211–8241 (2017). doi: [10.1021/acs.jpcb.7b03441](#); pmid: [28709377](#)
123. E. Lerner, A. Ingargiola, S. Weiss, Characterizing highly dynamic conformational states: The transcription bubble in RNAP-promoter open complex as an example. *J. Chem. Phys.* **148**, 123315 (2018). doi: [10.1063/1.5004606](#)
124. B. Bae, A. Feklistov, A. Lass-Napiorkowska, R. Landick, S. A. Darst, Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *eLife* **4**, e08504 (2015). doi: [10.7554/eLife.08504](#); pmid: [26349032](#)
125. C. D. K. Thompson, A. K. Sharma, J. Frank, R. L. Gonzalez Jr, D. Chowdhury, Quantitative connection between ensemble thermodynamics and single-molecule kinetics: A case study using cryogenic electron microscopy and single-molecule fluorescence resonance energy transfer investigations of the ribosome. *J. Phys. Chem. B* **119**, 10888–10901 (2015). doi: [10.1021/jp5128805](#); pmid: [25785884](#)
126. X. Wang, L. Vukovic, H. R. Koh, K. Schulten, S. Myong, Dynamic profiling of double-stranded RNA binding proteins. *Nucleic Acids Res.* **43**, 7566–7576 (2015). doi: [10.1093/nar/gkv726](#); pmid: [26184879](#)
127. A. Sali et al., Outcome of the first wwPDB Hybrid/Integrative Methods Task Force Workshop. *Structure* **23**, 1156–1167 (2015). doi: [10.1016/j.str.2015.05.013](#); pmid: [26095030](#)
128. A. Ingargiola, E. Lerner, S. Chung, S. Weiss, X. Michalet, FRETbursts: An open source toolkit for analysis of freely-diffusing single-molecule FRET. *PLOS ONE* **11**, e0160716 (2016). doi: [10.1371/journal.pone.0160716](#); pmid: [27532626](#)
129. M. Greenfield, D. S. Pavlichin, H. Mabuchi, D. Herschlag, Single Molecule Analysis Research Tool (SMART): An integrated approach for analyzing single molecule data. *PLOS ONE* **7**, e300024 (2012). doi: [10.1371/journal.pone.0030024](#); pmid: [22363412](#)
130. J. W. van de Meent, J. E. Bronson, C. H. Wiggins, R. L. Gonzalez Jr., Empirical Bayes methods enable advanced population-level analyses of single-molecule FRET experiments. *Biophys. J.* **106**, 1327–1337 (2014). doi: [10.1016/j.bpj.2013.12.055](#); pmid: [24655508](#)
131. A. Ingargiola, T. Laurence, R. Boutelle, S. Weiss, X. Michalet, Photon-HDF5: An open file format for timestamp-based single-molecule fluorescence experiments. *Biophys. J.* **110**, 26–33 (2016). doi: [10.1016/j.bpj.2015.11.013](#); pmid: [26745406](#)
132. H. Tochio, Watching protein structure at work in living cells using NMR spectroscopy. *Curr. Opin. Chem. Biol.* **16**, 609–613 (2012). doi: [10.1016/j.cbpa.2012.10.022](#); pmid: [23176973](#)
133. K. Bacía, S. A. Kim, P. Schuille, Fluorescence cross-correlation spectroscopy in living cells. *Nat. Methods* **3**, 83–89 (2006). doi: [10.1038/nmeth822](#); pmid: [16432516](#)
134. J. J. Sakon, K. R. Weninger, Detecting the conformation of individual proteins in live cells. *Nat. Methods* **7**, 203–205 (2010). doi: [10.1038/nmeth.1421](#); pmid: [20118931](#)
135. R. Crawford et al., Long-lived intracellular single-molecule fluorescence using electroporated molecules. *Biophys. J.* **105**, 2439–2450 (2013). doi: [10.1016/j.bpj.2013.09.057](#); pmid: [24314075](#)
136. K. W. Teng et al., Labeling proteins inside living cells using external fluorophores for microscopy. *eLife* **5**, e20378 (2016). doi: [10.7554/eLife.20378](#); pmid: [27935478](#)
137. T. Fessl et al., Towards characterization of DNA structure under physiological conditions in vivo at the single-molecule level using single-pair FRET. *Nucleic Acids Res.* **40**, e121 (2012). doi: [10.1093/nar/gks333](#); pmid: [22544706](#)
138. J. G. Ritter, R. Veith, A. Veenendaal, J. P. Siebrasse, U. Kubitschek, Light sheet microscopy for single molecule tracking in living tissue. *PLOS ONE* **5**, e11639 (2010). doi: [10.1371/journal.pone.0011639](#); pmid: [20668517](#)
139. L. Aigrain, M. Sustarsic, R. Crawford, A. Plochowitz, A. N. Kapanidis, Internalization and observation of fluorescent biomolecules in living microorganisms via electroporation. *J. Vis. Exp.* **2015**, e52208 (2015). doi: [10.3791/52208](#); pmid: [25741968](#)
140. E. Boukobza, A. Sonnenfeld, G. Haran, Immobilization in surface-tethered lipid vesicles as a new tool for single biomolecule spectroscopy. *J. Phys. Chem. B* **105**, 12165–12170 (2001). doi: [10.1021/jp012016x](#)
141. B. Okumus, T. J. Wilson, D. M. J. Lilley, T. Ha, Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophys. J.* **87**, 2798–2806 (2004). doi: [10.1529/biophysj.104.045971](#); pmid: [15454471](#)
142. A. E. Cohen, W. E. Moerner, Suppressing Brownian motion of individual biomolecules in solution. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4362–4365 (2006). doi: [10.1073/pnas.0509976103](#); pmid: [16537418](#)
143. A. Ingargiola et al., Multisite single-molecule FRET: High-throughput analysis of freely diffusing molecules. *PLOS ONE* **12**, e0175766 (2017). doi: [10.1371/journal.pone.0175766](#); pmid: [28419142](#)
144. X. Michalet et al., Silicon photon-counting avalanche diodes for single-molecule fluorescence spectroscopy. *IEEE J. Sel. Top. Quantum Electron.* **20**, 3804420 (2014). doi: [10.1109/JSTQE.2014.2341568](#); pmid: [25309114](#)
145. M. F. Juetz et al., Single-molecule imaging of non-equilibrium molecular ensembles on the millisecond timescale. *Nat. Methods* **13**, 341–344 (2016). doi: [10.1038/nmeth.3769](#); pmid: [26878382](#)
146. A. Ingargiola et al., 48-spot single-molecule FRET setup with periodic acceptor excitation. *bioRxiv* 156182 (2017). pmid: [156182](#)
147. B. Wunderlich et al., Microfluidic mixer designed for performing single-molecule kinetics with confocal detection on timescales from milliseconds to minutes. *Nat. Protoc.* **8**, 1459–1474 (2013). doi: [10.1038/nprot.2013.082](#); pmid: [23845960](#)
148. J. P. Renaud et al., Biophysics in drug discovery: Impact, challenges and opportunities. *Nat. Rev. Drug Discov.* **15**, 679–698 (2016). doi: [10.1038/nrd.2016.123](#); pmid: [27516170](#)
149. S. Kim et al., High-throughput single-molecule optofluidic analysis. *Nat. Methods* **8**, 242–245 (2011). doi: [10.1038/nmeth.1569](#); pmid: [21297618](#)
150. T. Kim, A. Reitmair, Non-coding RNAs: Functional aspects and diagnostic utility in oncology. *Int. J. Mol. Sci.* **14**, 4934–4968 (2013). doi: [10.3390/ijms14034934](#); pmid: [23455466](#)
151. T. Förster, Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **437**, 55–75 (1948). doi: [10.1002/andp.19484370105](#)
152. I. L. Medintz, H. Mattoussi, Quantum dot-based resonance energy transfer and its growing application in biology. *Phys. Chem. Chem. Phys.* **11**, 17–45 (2009). doi: [10.1039/B813919A](#); pmid: [19081907](#)
153. B. Hochreiter, A. P. Garcia, J. A. Schmid, Fluorescent proteins as genetically encoded FRET biosensors in life sciences. *Sensors* **15**, 26281–26314 (2015). doi: [10.3390/s151026281](#); pmid: [26501285](#)
154. G. Gouridis et al., Conformational dynamics in substrate-binding domains influences transport in the ABC importer GlnPQ. *Nat. Struct. Mol. Biol.* **22**, 57–64 (2015). doi: [10.1038/nsmb.2929](#); pmid: [25486304](#)

ACKNOWLEDGMENTS

We acknowledge contributions and discussions with past and present members of the Weiss lab. We thank M. Segal for editing the manuscript. Supported by NIH grants GM069709 and NSF grant MCB-1244175 (S.W.), NIH grant GM095904 (X.M. and S.W.), and European Research Council grant ERC-STG 638536–SM-IMPORT, grant of the Deutsche Forschungsgemeinschaft within GRK2062, the Center for NanoScience, Center for Integrated Protein Science Munich, and LMUexcellent (T.C.).

10.1126/science.aan1133

RESEARCH ARTICLE SUMMARY

MEMBRANE PROTEINS

Membrane protein insertion through a mitochondrial β -barrel gate

Alexandra I. C. Höhr, Caroline Lindau, Christophe Wirth, Jian Qiu, David A. Stroud, Stephan Kutik, Bernard Guiard, Carola Hunte, Thomas Becker, Nikolaus Pfanner,* Nils Wiedemann*

INTRODUCTION: The outer membranes of Gram-negative bacteria, mitochondria, and chloroplasts characteristically contain β -barrel membrane proteins. These proteins contain multiple amphipathic β strands that form a closed barrel. This arrangement exposes hydrophobic amino acid residues to the lipid phase of the membrane, with polar residues facing the lumen of the barrel. β -barrel proteins form outer membrane channels for protein import and export, and for metabolite and nutrient exchange.

An essential step in the biogenesis of β -barrel proteins is their insertion into the outer membrane. The β -barrel assembly machinery (BAM) of bacteria and the sorting and assembly machinery (SAM) of mitochondria are crucial for the membrane insertion of β -barrel precursors. The core subunits of these machineries, BamA and Sam50, are homologous 16-stranded β -barrel

proteins that belong to the outer membrane protein family 85 (Omp85). The β signal located in the last β strand of the precursor initiates protein insertion into the outer membrane; however, the molecular mechanism of β -barrel insertion has not been understood. Controversial models about the role of BAM and SAM have been discussed. These models either favor precursor translocation into the BamA or Sam50 barrel followed by lateral release through an opened β -barrel gate or suggest membrane thinning and precursor insertion at the BamA or Sam50 protein-lipid interface.

RATIONALE: Structural studies have suggested that BamA and Sam50 harbor a dynamic lateral gate formed between β strands 1 and 16. In addition, BamA and Sam50 have been proposed to induce a thinning of the lipid bilayer near

the lateral gate. To determine the translocation pathway during β -barrel membrane insertion, we probed the proximity of β -barrel precursors (Tom40, Por1, VDAC1) to Sam50 in intact mitochondria of the model organism baker's yeast, *Saccharomyces cerevisiae*. We engineered precursors and Sam50 variants with cysteine residues at defined positions and mapped the environment of precursors in transit by disulfide-bond scanning and cysteine-specific cross-linking.

RESULTS: Our findings indicated that during transport of β -barrel precursors by the SAM complex, the lateral gate of Sam50 between β strands 1 and 16 was open and contained accumulated

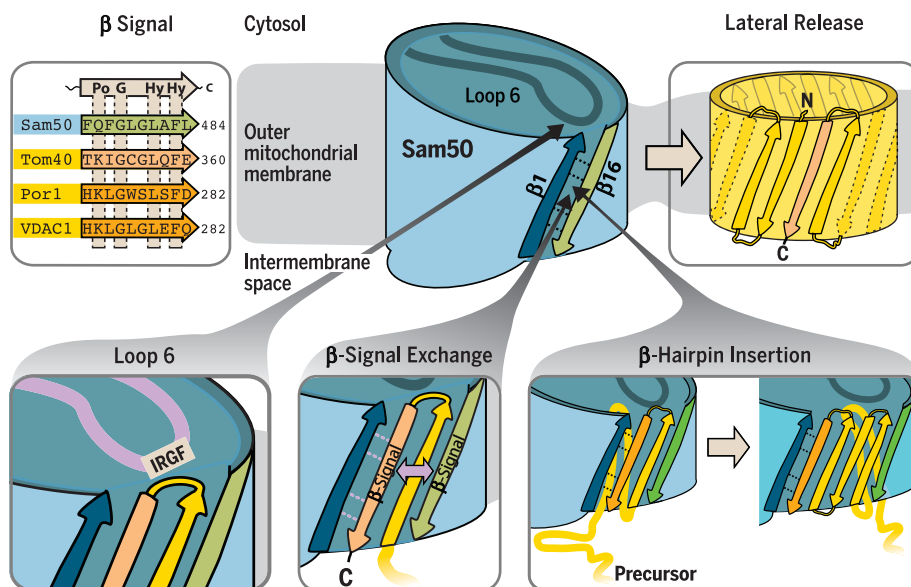
ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aah6834>

precursor. The β signal of the precursor specifically interacted with β strand 1 of Sam50 and thus replaced the endogenous β signal (β strand 16) of Sam50. Precursor transfer

to the lateral gate occurred via the channel lumen of Sam50 and required the conserved loop 6 located in the channel. β hairpin-like elements consisting of two antiparallel β strands of the precursor were translocated and inserted into the lateral gate. The precursor remained associated with the Sam50 gate until the folded full-length β -barrel protein was released into the outer membrane.

CONCLUSION: Our findings indicate that β -barrel precursors are inserted into the lumen of the Sam50 channel and are released into the mitochondrial outer membrane via the opened lateral gate of Sam50. The carboxy-terminal β signal of the precursor initiates opening of the gate by exchange with the endogenous Sam50 β signal. An increasing number of β hairpin-like loops of the precursor accumulate at the lateral gate. Upon folding at Sam50, the full-length β -barrel protein is laterally released into the outer membrane. Membrane thinning in the vicinity of the lateral gate likely facilitates insertion of the protein into the lipid bilayer. Thus, the membrane-insertion pathway of β -barrel proteins combines elements of both controversially discussed models: transport through the lumen of Sam50 and the lateral gate and subsequent insertion into the thinned membrane next to the gate. Owing to the conservation of both the β signal and Omp85 core machinery, we speculate that β -signal exchange, folding at the gate, and lateral release into the membrane represent a general mechanism for β -barrel protein biogenesis in mitochondria, chloroplasts, and Gram-negative bacteria. ■



β -Barrel protein insertion via the lateral gate of Sam50. β -Barrel precursors are transferred through the Sam50 interior to the lateral gate, which is formed by β strands 1 and 16. Upon gate opening, the β signal of the precursor substitutes for the endogenous Sam50 β signal. A conserved loop of Sam50 promotes β -signal binding to the gate and insertion of subsequent β hairpins. The folded β -barrel protein is released into the outer membrane. Po, polar amino acid residue; G, glycine; Hy, hydrophobic amino acid residue; C, C terminus; IRGF, binding motif.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: nikolaus.pfanner@biochemie.uni-freiburg.de (N.P.); nils.wiedemann@biochemie.uni-freiburg.de (N.W.)
Cite this article as Höhr et al., *Science* 359, eaah6834 (2018). DOI: [10.1126/science.aah6834](https://doi.org/10.1126/science.aah6834)

RESEARCH ARTICLE

MEMBRANE PROTEINS

Membrane protein insertion through a mitochondrial β -barrel gate

Alexandra I. C. Höhr,^{1,2} Caroline Lindau,^{1,2} Christophe Wirth,¹ Jian Qiu,^{1*} David A. Stroud,^{1†} Stephan Kutik,¹ Bernard Guiard,³ Carola Hunte,^{1,4} Thomas Becker,^{1,4} Nikolaus Pfanner,^{1,4‡} Nils Wiedemann^{1,4‡}

The biogenesis of mitochondria, chloroplasts, and Gram-negative bacteria requires the insertion of β -barrel proteins into the outer membranes. Homologous Omp85 proteins are essential for membrane insertion of β -barrel precursors. It is unknown if precursors are threaded through the Omp85-channel interior and exit laterally or if they are translocated into the membrane at the Omp85-lipid interface. We have mapped the interaction of a precursor in transit with the mitochondrial Omp85-channel Sam50 in the native membrane environment. The precursor is translocated into the channel interior, interacts with an internal loop, and inserts into the lateral gate by β -signal exchange. Transport through the Omp85-channel interior followed by release through the lateral gate into the lipid phase may represent a basic mechanism for membrane insertion of β -barrel proteins.

β -Barrel proteins are of central importance in the outer membranes of mitochondria, chloroplasts, and Gram-negative bacteria. In eukaryotic cells, β -barrel proteins are essential for the communication between the double membrane-bound organelles and the rest of the cell. β -Barrel channels mediate the translocation of a large number of metabolites and the import of organellar precursor proteins that are synthesized in the cytosol. The machineries for the biogenesis of β -barrel proteins have been identified in mitochondria and bacteria, termed sorting and assembly machinery (SAM) and β -barrel assembly machinery (BAM), respectively (1–6). The core component of the β -barrel insertion machinery is a member of the Omp85 superfamily, conserved from bacteria (BamA) to humans (Sam50, also called Tob55), whereas accessory BAM and SAM subunits are not conserved (1, 2, 4, 5, 7–11). The most C-terminal β strand of each precursor serves as a signal recognized by the Omp85 machinery (12, 13), and the assembly of a β -barrel protein was shown to occur from the C terminus (14). Upon closure of the barrel, the protein is released from the assembly machinery (15).

Members of the Omp85 superfamily form 16-stranded β barrels, including BamA and Sam50,

the filamentous haemagglutinin secretion protein FhaC, and the translocation and assembly module TamA (14, 16–19). In the case of FhaC, a substrate protein was shown to be translocated across the bacterial outer membrane through the interior of the β -barrel channel (20). The substrates of BamA, Sam50, and TamA, however, have to be inserted into the lipid phase to become integral outer membrane proteins. High-resolution structures of BamA and TamA and disulfide scanning revealed a flexible interaction of the first and last β strand, suggesting a lateral opening of a β -barrel gate toward the membrane and a distortion of the adjacent membrane lipids (16, 18, 21–27). Different models have been discussed for the BamA-, Sam50-, and TamA-mediated insertion of β -barrel precursors into the outer membrane (5, 15, 16, 18, 21–38). In the BamA- and Sam50-assisted models, the precursor is inserted at the protein-lipid interface; BamA or Sam50 creates a distortion and thinning of the membrane that favors spontaneous insertion of the precursor into the membrane. In the BamA- and Sam50-budding model, the precursor is threaded through the β -barrel interior of BamA or Sam50 and is laterally released through an opened lateral gate. The BamA structures, which were obtained in non-native environments and in the absence of precursor proteins (35), supported arguments for both models (16, 21–26), and thus the mechanism of β -barrel translocation by means of BAM or SAM is unknown.

Lateral gate of the Sam50 β barrel in the mitochondrial outer membrane

We developed a system to map the interaction of Sam50 with β -barrel precursors in transit in the native mitochondrial membrane environment. The β -barrel channel of Sam50 was modeled based on the BamA structures and cysteine and disulfide scanning of β strands 1 and 16 (Fig. 1, A

and B, and fig. S1, A to C) (39, 40). In the absence of precursor proteins, β strands 1 and 16 interacted, i.e., the putative lateral gate was closed (Fig. 1B and fig. S1C) (31). However, oxidation-induced disulfide formation between distinct cysteines also revealed a sliding of β strands 1 and 16, i.e., a dynamic behavior of the gate (27). To probe for the possible opening of the gate in the presence of substrate, we tested β -barrel precursors that contained the β -hairpin mitochondrial targeting signal (6) and imported them into isolated intact mitochondria, followed by position-specific SH cross-linking of β strands 1 and 16. The cross-linking reagent bismaleimido-hexane (BMH) showed a high efficiency for stably linking strands 1 and 16 in the absence of substrate (Fig. 1C, lane 2, and fig. S1C). A C-terminal fragment of the major mitochondrial β -barrel protein porin (Por1), also called VDAC, including the Por1 β signal, considerably disturbed the interaction of Sam50 β strands 1 and 16 (Fig. 1C, lane 4), indicating that the Por1 substrate interfered with gate closing.

β -Signal exchange in the lateral gate and release of the full-length β -barrel precursor

It has been speculated that the β signal may be specifically recognized by BamA and Sam50 by means of exchange of the endogenous BamA and Sam50 β signal (31, 33), yet experimental demonstration has been lacking (35). β Strand 16 of BamA and Sam50 functions as a β signal, and thus, in the exchange model, the β signal of the precursor, corresponding to the C-terminal β strand 19 of Por1, should interact with Sam50- β 1. To test this hypothesis, we synthesized a radiolabeled [³⁵S] Por1 substrate carrying a single cysteine residue at distinct positions of the β signal. After import into mitochondria containing Sam50 with a single cysteine residue at different positions in β strands 1 or 16, we probed the proximity of the β strands by disulfide formation. The Por1 β signal indeed specifically aligned with Sam50- β 1 such that residues predicted to point toward either the channel interior (black) or the lipid phase (gray) selectively interacted (Fig. 2A and fig. S2A).

We performed several control experiments and obtained the following results: (i) The Por1 β signal selectively interacted with Sam50- β 1, but not with Sam50- β 16 (Fig. 2A and fig. S2A). (ii) To test a different β signal, we imported a ³⁵S-labeled C-terminal precursor of the mitochondrial import channel Tom40 and observed a comparable pairing with Sam50- β 1 (fig. S2B). (iii) A precursor containing a mutant form of the Por1 β signal [replacement of a conserved hydrophobic residue (13, 41)] was strongly impaired in the interaction with Sam50- β 1 (Fig. 2B). These results show that the β signal of precursors specifically interacts with Sam50- β 1 (Fig. 2C). (iv) We analyzed substrates of different size, covering the range from 5 to 18 β strands, and observed disulfide formation between the Por1 β signal and Sam50- β 1 in each case (Figs. 2A and 3A and fig. S2A). (v) Comigration of the differently sized Por1 β -barrel precursors with the SAM complex, as observed by blue native gel analysis (1, 3, 8, 9, 13), showed that each substrate accumulated at the SAM complex

¹Institute of Biochemistry and Molecular Biology, Centre for Biochemistry and Molecular Cell Research (ZBMZ), Faculty of Medicine, University of Freiburg, 79104 Freiburg, Germany.

²Faculty of Biology, University of Freiburg, 79104 Freiburg, Germany.

³Centre de Génétique Moléculaire, CNRS, 91190 Gif-sur-Yvette, France. ⁴BIOS Centre for Biological Signalling Studies, University of Freiburg, 79104 Freiburg, Germany.

*Present address: Swiss Federal Institute of Technology (EPFL), 1015 Lausanne, Switzerland. †Present address: Department of Biochemistry and Molecular Biology and The Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, Victoria 3010, Australia.

‡Corresponding author. Email: nikolaus.pfanner@biochemie.uni-freiburg.de (N.P.); nils.wiedemann@biochemie.uni-freiburg.de (N.W.)

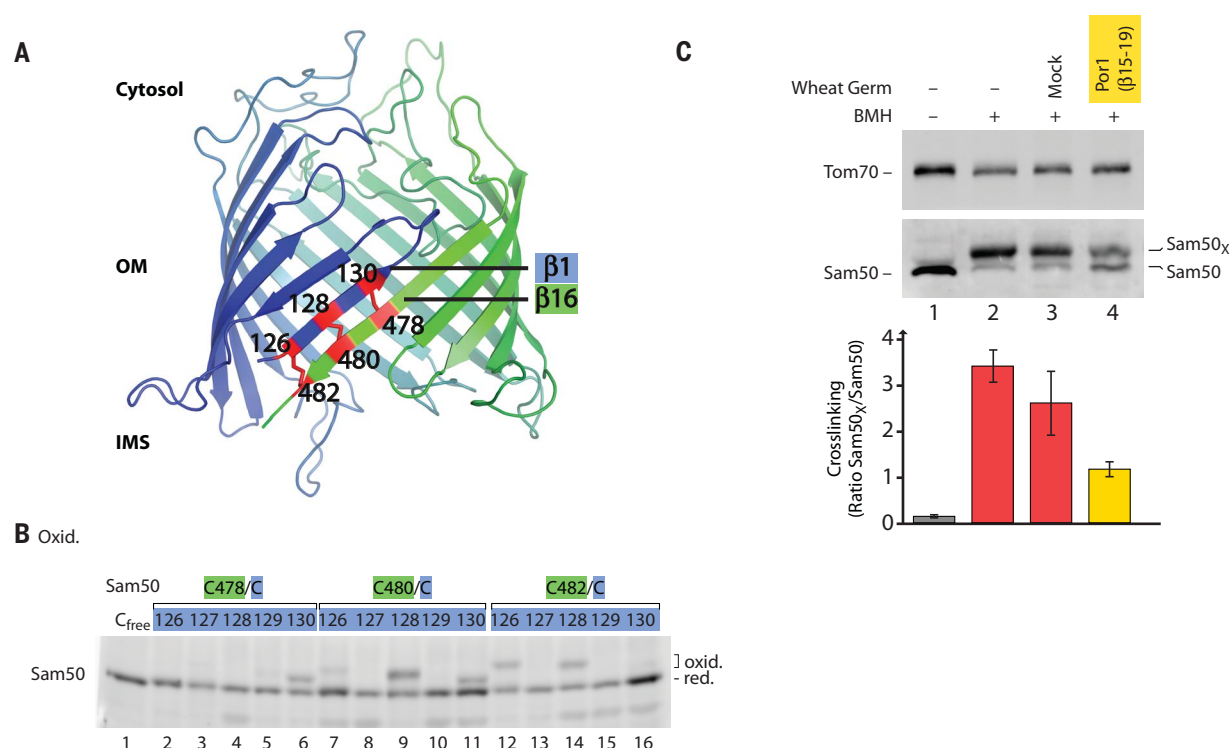


Fig. 1. Intramolecular interaction between the first and last β strands of Sam50.

(A) Model of the Sam50 β barrel. Engineered disulfide bonds between the first and last β strands of Sam50 are indicated in red, where numbers in black indicate positions of cysteine residue. Additional disulfide bonds are possible because of the dynamic interaction of β strands 1 and 16 (27). IMS, intermembrane space; OM, outer membrane. **(B)** Yeast strains expressing cysteine-free Sam50 (C_{free}) and Sam50 cysteine variants (containing exactly two cysteine residues, as indicated in green and blue) were treated in vivo with

the oxidant 4,4'-dipyridyl disulfide (4-DPS), followed by nonreducing SDS-polyacrylamide gel electrophoresis (PAGE), Western blotting, and immunodecoration. Oxid., oxidized; red., reduced. **(C)** Isolated Sam50_{C128/C480} mitochondria were incubated with a Por1-precursor construct (β strands 15 to 19 corresponding to amino acid residues 210 to 283) and controls, as indicated. Samples were treated with the cross-linker BMH and analyzed as in (B). Quantification of cross-linking efficiency (mean \pm SEM, $N = 3$ for samples 1, 2, and 4; mean with range, $N = 2$ for sample 3). Sam50_x, cross-linked Sam50.

(Fig. 3, B and C). (v) Only the full-length Por1 precursor, corresponding to 19 β strands, was released from the SAM complex and assembled into the mature porin complex (Fig. 3, B and C) (42–45).

Taken together, we conclude that the β signal of the precursor is bound by Sam50- $\beta 1$ through exchange with the endogenous Sam50 β signal ($\beta 16$) (Fig. 2C). Porin precursors of up to 18 β strands accumulate at the SAM complex, and only the full-size precursor is released into the lipid phase of the outer membrane.

β -Barrel precursors interact with both sides of the Sam50 gate

We asked if the substrate also interacted with β strand 16 of Sam50 and performed disulfide scanning between this β strand and the N-terminal region of the precursor, corresponding to β strand 14 of mature Por1. We tested five distinct amino acid positions corresponding to Por1- $\beta 14$ and observed disulfide formation with Sam50- $\beta 16$ in each case (Fig. 4, A and B). However, the interaction showed a considerably higher flexibility than that of the β signal of the precursor with Sam50- $\beta 1$ (Fig. 2 and fig. S2). A Por1 precursor with a mutant β signal strongly inhibited the interaction of the N-terminal precursor region with Sam50- $\beta 16$ (fig. S3). Because the β signal

itself did not interact with Sam50- $\beta 16$, this finding indicates that the specific binding of the β signal to Sam50- $\beta 1$ is a prerequisite for the accumulation of the N-terminal precursor region at Sam50- $\beta 16$. To provide further evidence that the precursor was intercalated between β strands 1 and 16 of Sam50, we studied if it interacted with both strands simultaneously. Por1 precursors containing two cysteine residues, one in the C-terminal β signal and one in the N-terminal region, were accumulated at Sam50, carrying a cysteine residue in $\beta 1$ as well as in $\beta 16$, and subjected to oxidation. In addition to the single disulfides formed (like in Figs. 2, A and B, and 4, A and B), we observed the formation of two disulfides simultaneously (Fig. 4C, lanes 3 and 7).

Our results indicate that β -barrel precursors are inserted into a Sam50 gate formed between β strands 1 and 16. The C-terminal β signal specifically exchanges with Sam50- $\beta 1$, whereas the N-terminal region of the precursor undergoes a flexible interaction with Sam50- $\beta 16$.

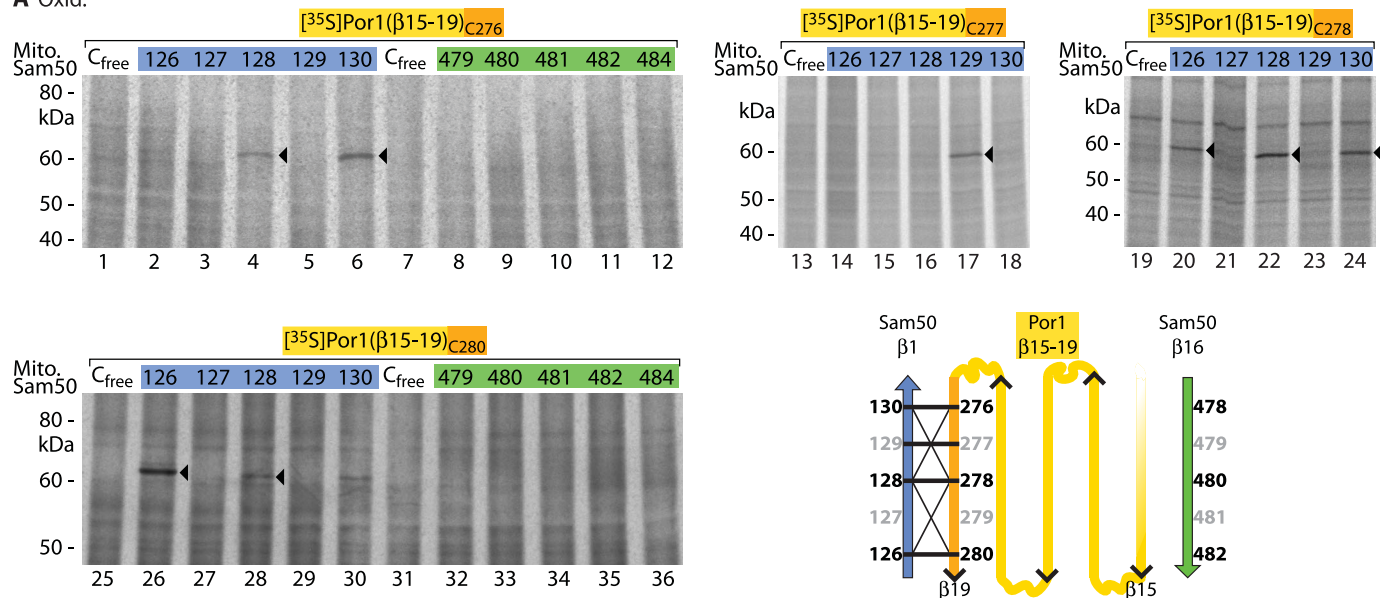
Translocation of β -barrel precursors into the Sam50 channel

The N-terminal region of the precursor (residues 204 to 207) was also found in close proximity to the first residue (residue 126) of Sam50- $\beta 1$ (Fig. 4,

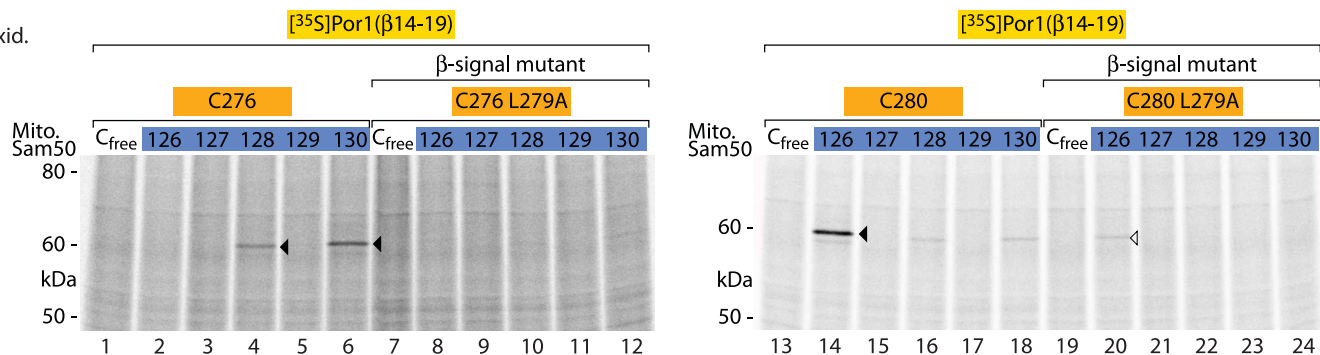
A and B). Sam50_{res126} is positioned at the intermembrane space opening of the Sam50 channel and predicted to point toward the channel interior (Fig. 1A). Por1_{res207}, which is located toward the cytosolic side of mature Por1 (42–44), was not only found in proximity to Sam50_{res126} but also to further residues of Sam50- $\beta 1$ that are predicted to face the channel interior (residues 128 and 130) (Fig. 4A and fig. S3). Disulfide formation between the N-terminal region of Por1 and Sam50- $\beta 1$ was impaired when the Por1 β signal was mutated (fig. S3). Thus, a functional C-terminal β signal is a prerequisite for the observed proximity of the N-terminal precursor region to Sam50- $\beta 1$ (pairing between Sam50- $\beta 1$ and the β signal involves hydrogen bonds of the polypeptide backbone, and thus, cysteine side chains are available for disulfide formation). These findings are compatible with a model that, upon binding of the β signal to Sam50- $\beta 1$, the N-terminal region of the precursor is passing at the interior of Sam50- $\beta 1$.

To obtain independent evidence that β -barrel precursors are using the interior of the Sam50 channel, we analyzed Sam50 β strand 15 and compared residues predicted to face either the channel interior (black) or the lipid phase (gray) (Fig. 5A). A [³⁵S]Por1 precursor with a single cysteine

A Oxid.



B Oxid.



C

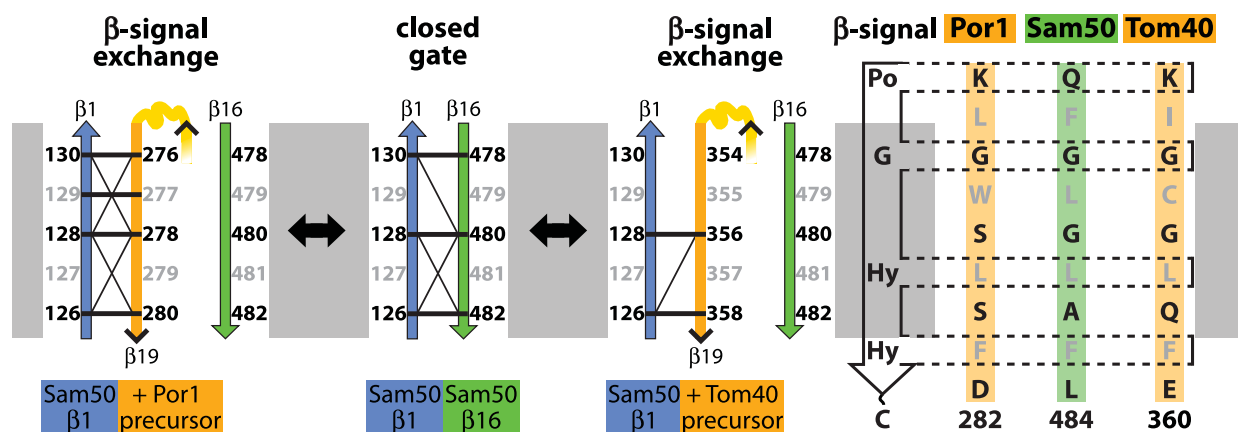


Fig. 2. Interaction of Sam50 β strand 1 with the C-terminal β signal of precursor proteins. (A) Radiolabeled Por1(β15 to 19) precursors containing one cysteine at the positions indicated were imported for 5 min into mitochondria isolated from yeast strains expressing Sam50 with the indicated cysteine residues, followed by oxidation with 4-DPS (lanes 1 to 12 and 19 to 36) or CuSO₄ (lanes 13 to 18). Samples were analyzed by nonreducing SDS-PAGE and autoradiography. Arrowheads, disulfide-bonded Sam50-Por1(β15 to 19) adducts. Schematic model, disulfide-bond formation of Sam50 β strand 1 with the β signal (β19) of the

porin precursor β15 to 19; thick and thin lines indicate strong and weak formation of Sam50-Por1 adducts, respectively. (B) [35S]Por1(β14 to 19)_{C276}, [35S]Por1(β14 to 19)_{C280}, and the corresponding β-signal mutants (L279A, leucine replaced with alanine) were incubated for 5 min with isolated mitochondria of Sam50 cysteine variants followed by oxidation with 4-DPS, nonreducing SDS-PAGE, and autoradiography. Arrowheads, cysteine-specific Sam50-precursor adducts. (C) Schematic model illustrating the β-signal exchange observed in Fig. 2 and fig. S2.

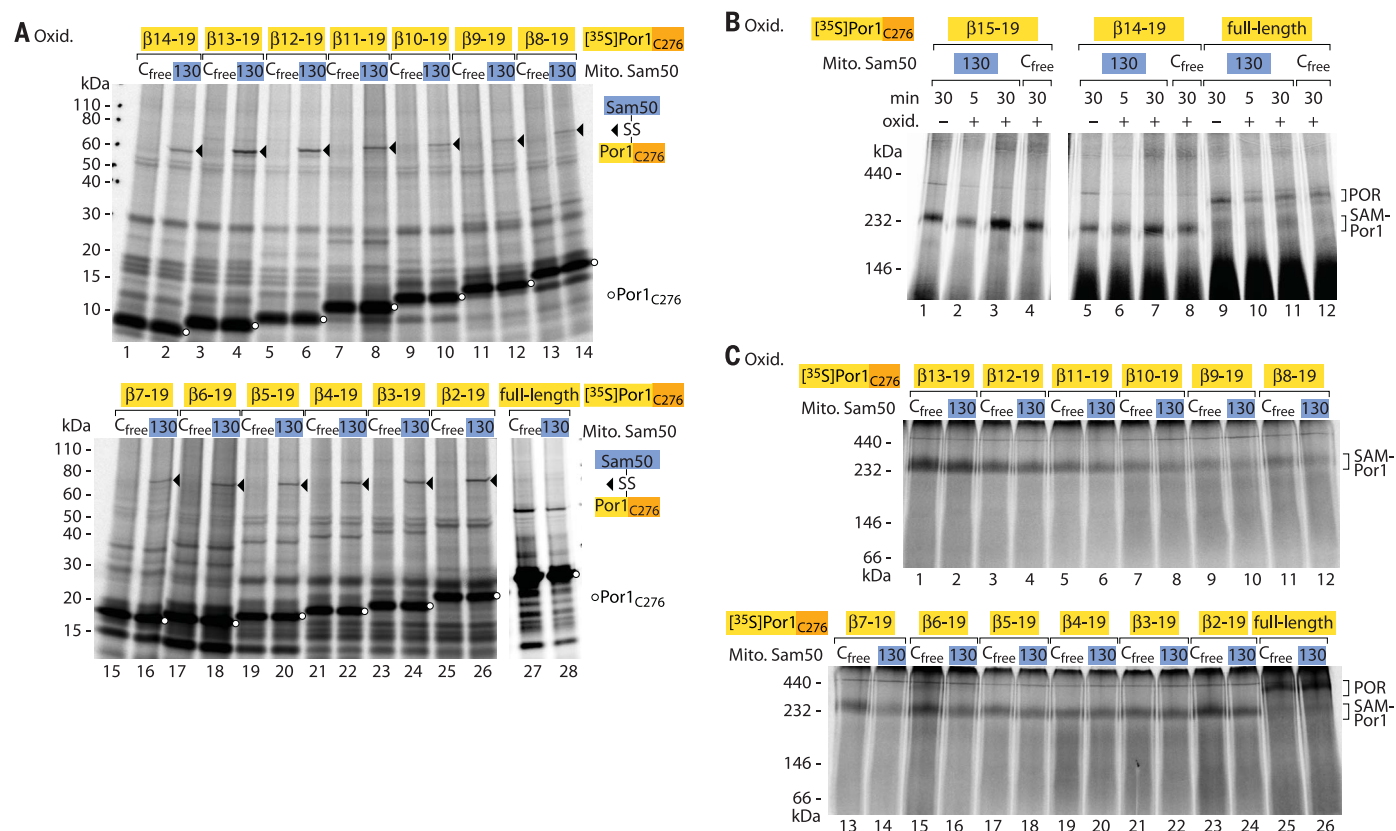


Fig. 3. Release of full-length β -barrel precursor from the SAM complex.

(A) [^{35}S]Por1 $_{\text{C276}}$ constructs of different lengths were incubated with isolated Sam50 $_{\text{Cfree}}$ or Sam50 $_{\text{C130}}$ mitochondria for 30 min, followed by oxidation with 4-DPS, analysis by nonreducing SDS-PAGE, and autoradiography. Lanes 27 and 28 were given a shorter exposure to compensate for the strong intensity of full-length Por1. Arrowheads,

disulfide-bonded Sam50-precursor adducts; circles, Por1 $_{\text{C276}}$ precursor.

(B and C) Samples as described for (A) were analyzed by blue native electrophoresis and autoradiography. Lanes 1 to 4 versus lanes 5 to 12 of (B) have a linear adjustment of brightness and contrast to compensate for the strong intensity of Por1(β 15 to 19) $_{\text{C276}}$. POR, assembled porin complex; SAM-Por1, Por1 precursor at SAM.

residue in the N-terminal region (residue 205) was imported into Sam50 containing a single cysteine at different positions of either β strand 15 or 16. In contrast to Sam50- β 16, we did not observe disulfide formation between the precursor and Sam50- β 15 upon oxidation (fig. S4), indicating that Por1 $_{\text{res205}}$ was not so close to Sam50- β 15 as to promote disulfide formation. By using SH-specific BMH, the precursor was cross-linked to Sam50- β 15 and β 16. Whereas the cross-linking occurred to various residues of Sam50- β 16 (comparable to the oxidation assay), only residues of Sam50- β 15 predicted to face the channel interior were cross-linked to the precursor (Fig. 5B). To probe further regions of the precursor, we used the short amine-to-sulphydryl cross-linking reagents *N*- α -maleimidoacet-oxy succinimide ester (AMAS) and succinimidyl iodoacetate (SIA) together with a cysteine-free Por1 precursor and Sam50 containing a single cysteine residue in β 15. Cysteine-specific cross-linking occurred only to Sam50- β 15 residues predicted to face the channel interior (Fig. 5C, arrowheads) (a larger non-specific band at 60 kDa was formed when no SH-group was available, i.e., also with cysteine-free Sam50). These results are fully compatible with the model that transfer of the Por1 pre-

cursor involves the interior of the Sam50 channel, but do not fit to a model in which the Por1 precursor is inserted at the protein-lipid interphase without getting access to the channel.

Sam50-loop 6 is required for β -signal binding

In addition to the β -barrel channel, Sam50 possesses two major characteristic elements, an N-terminal polypeptide transport-associated (POTRA) domain exposed to the intermembrane space and a highly conserved loop 6 that extends from the cytosolic side of the β barrel. Whereas bacterial BamA proteins contain several POTRA domains that interact with β -barrel precursors and are crucial for precursor transfer from the periplasm into the outer membrane (17, 46–49), Sam50 contains a single POTRA domain that is not essential for cell viability (13, 50, 51). Disulfide formation between the Por1 precursor and Sam50 β strands 1 and 16 was not blocked in mitochondria lacking the entire POTRA domain (fig. S5). Together with blue native gel analysis (13, 45), this result indicates that the single POTRA domain is not crucial for precursor transfer to Sam50. The second element, loop 6, extends from the outside, or cytosolic side, into the channel interior in

all Omp85 high-resolution structures analyzed (Fig. 6A) (16, 18, 21–25, 52). Deletion of Sam50-loop 6 was lethal to yeast cells. When wild-type Sam50 was depleted, expression of a Sam50 mutant form lacking the conserved segment of loop 6 did not rescue growth and led to strong defects in the import of ^{35}S -labeled β -barrel precursors such as Por1 and Tom40 into mitochondria (fig. S6, A and B). The steady-state amounts of β -barrel proteins and various Tom proteins were decreased (fig. S6C). As the translocase of the outer mitochondrial membrane (TOM complex) imports a large number of precursor proteins, this mutant did not permit a selective analysis of the function of loop 6. We thus generated point mutants of the conserved IRGF (Ile-Arg-Gly-Phe) motif of loop 6 (53, 54). Sam50 $_{\text{R366A}}$ yeast, in which the Sam50 arginine at position 366 is replaced with an alanine, exhibited a temperature-sensitive growth phenotype on nonfermentable medium (fig. S7A). Mitochondria isolated upon growth of the mutant cells at permissive temperature showed normal steady-state amounts of Sam, TOM and further control proteins (fig. S7, B and C). The import of ^{35}S -labeled β -barrel precursors such as Por1, Mdm10, and Tom40 was strongly inhibited (Fig. 6B), whereas the import of matrix-targeted and intermembrane

Fig. 4. Interaction of Sam50 with the N-terminal β strand of precursor proteins.

(A) [^{35}S]Por1(β 14 to 19) precursors containing a single cysteine residue, as indicated, were imported into mitochondria isolated from yeast strains expressing the specified Sam50 cysteine variants, followed by oxidation with 4-DPS, nonreducing SDS-PAGE, and autoradiography. Black and white arrowheads show cysteine-specific disulfide-bonded Por1(β 14 to 19) adducts to the C- and N-terminal β strand of Sam50, respectively. Right, schematic models. (B) [^{35}S]Por1(β 14 to 19) $_{\text{C}206}$ and [^{35}S]Por1(β 14 to 19) $_{\text{C}204}$ were treated as described in (A). (C) [^{35}S]Por1(β 14 to 19) single- and double-cysteine variants were incubated with isolated mitochondria from yeast strains expressing Sam50 $_{\text{Cfree}}$ or the double-cysteine variant Sam50 $_{\text{C}126/\text{C}480}$, followed by oxidation with 4-DPS. Samples were analyzed as described in (A).

space-targeted precursors, which depend on the TOM complex but not on SAM, was not or only mildly affected (fig. S7D). The import of [^{35}S]Tom40 can be dissected into distinct stages by blue native gel analysis (1, 3, 8, 9). Sam50 $_{\text{R}366\text{A}}$ mitochondria were impaired in the formation of SAM-bound intermediates (Fig. 6B). We conclude that loop 6 of Sam50 is required for a stable interaction of the precursor with SAM. It has been reported that both Sam50 and Sam35 are needed for binding of a β -barrel precursor to the SAM complex (13). To directly test the contribution of loop 6, we performed affinity purification from lysed mitochondria using a purified β signal-fusion protein, leading to the copurification of Sam50 and Sam35 from wild-type mitochondria; a mutant β signal did not pull down Sam50-Sam35 (Fig. 6C) (13). The interaction of Sam50-Sam35 with the β signal was strongly disturbed in Sam50 $_{\text{R}366\text{A}}$ mitochondria (Fig. 6C), demonstrating that loop 6 is required for stable precursor binding to Sam50-Sam35.

β Hairpin-like transport of precursor proteins by Sam50

To determine if a precursor in transit was in proximity to loop 6, [^{35}S]Por1 precursors with a single cysteine residue in the N-terminal region were imported into mitochondria containing Sam50 with a single cysteine residue in loop 6. By SH-specific cross-linking, the precursors were linked to residue 371 of loop 6 (Fig. 7A). A mutant β signal prevented cross-linking of the N-terminal precursor region to loop 6 (fig. S8A), whereas the β signal itself was not found in proximity to loop 6 (fig. S8B, lanes 1 to 6), supporting our conclusion that a functional β signal is a prerequisite for further translocation steps of the precursor. It has been suggested that β -barrel precursors transported by SAM and BAM may be partially folded, such that β hairpins consisting of two adjacent β strands are formed (35, 55). We used distinct approaches to assess this view: (i) By using precursors of different lengths, covering 5, 6, 7, or 8 β strands of mature

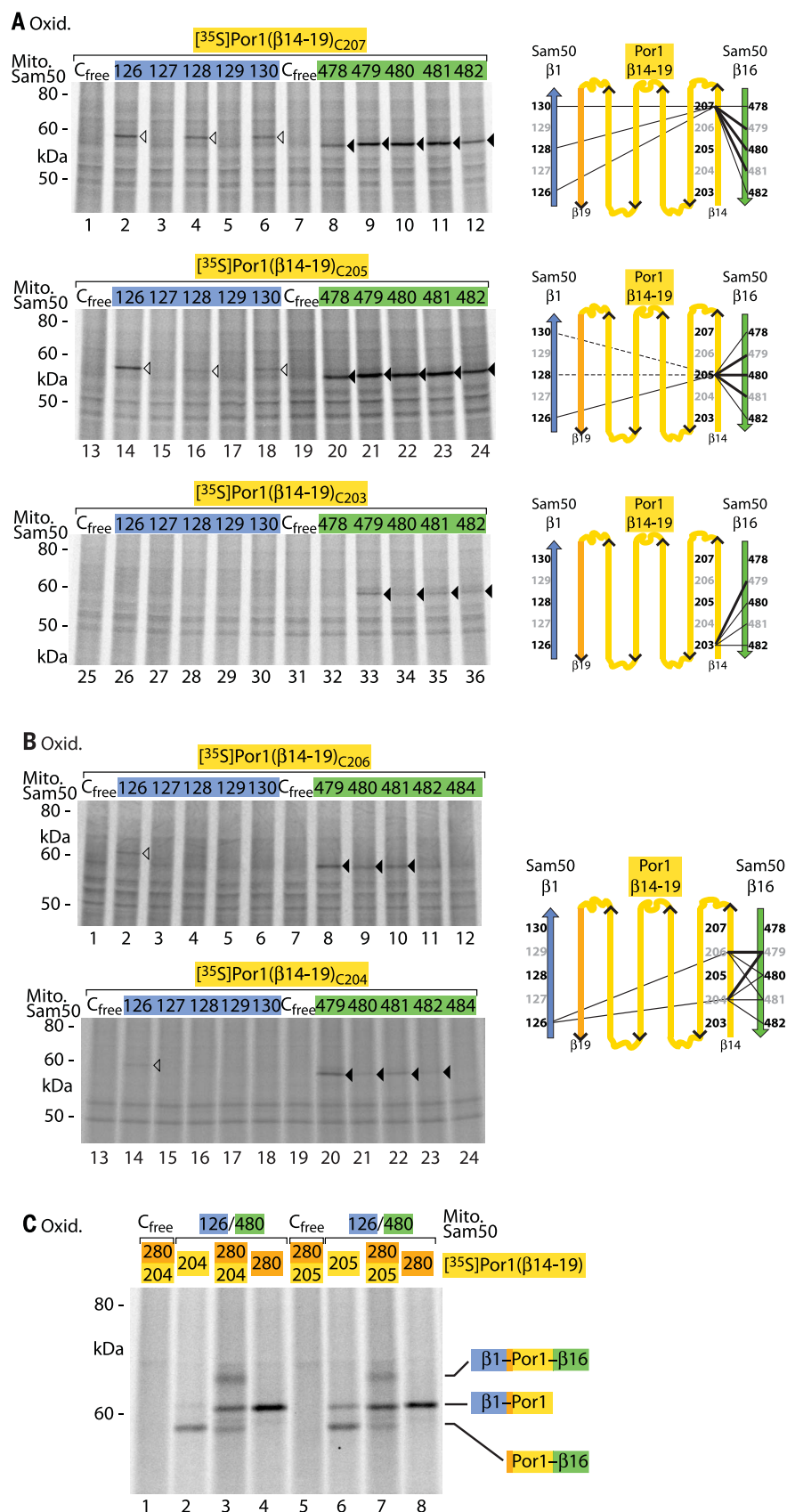
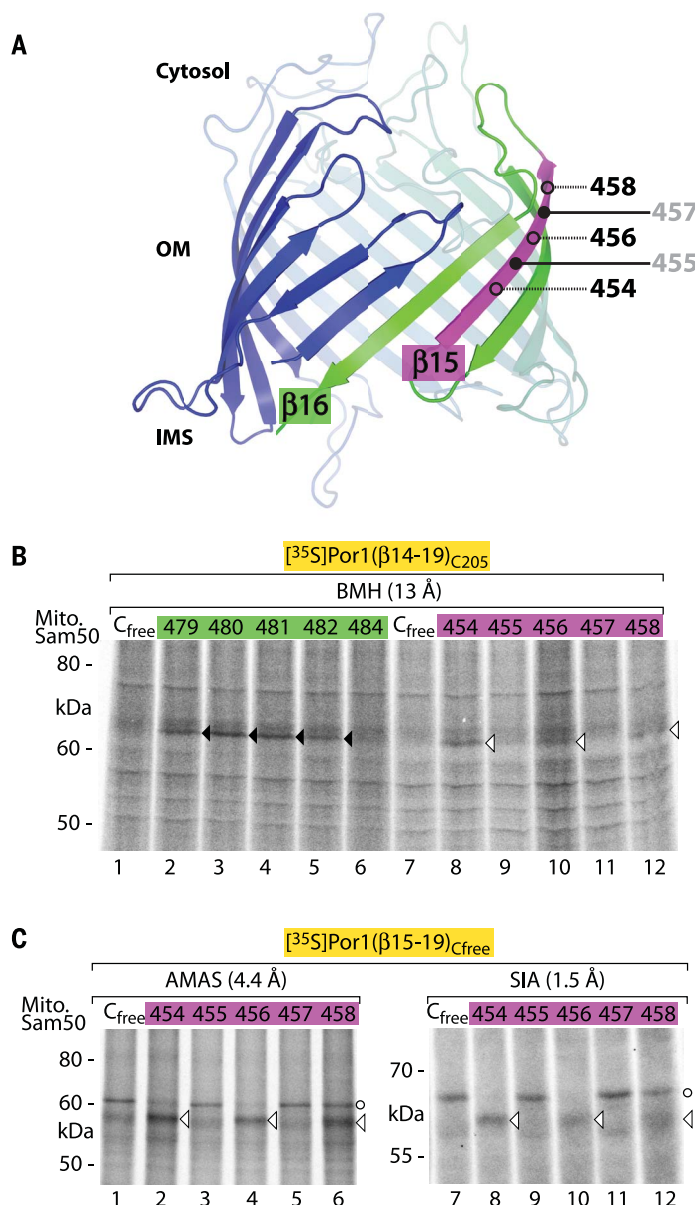


Fig. 5. Interaction of β -barrel precursor with Sam50 residues facing the channel interior.

(A) Model of the Sam50 β barrel. β Strand 15, purple; open and filled circles, residues facing the interior of the barrel (black) or the lipid phase (gray), respectively. **(B and C)** Radio-labeled Por1 precursor variants were imported into mitochondria of yeast strains expressing the indicated Sam50 variants. Samples were cross-linked with BMH, AMAS, or SIA and analyzed by non-reducing SDS-PAGE, and autoradiography. Black and white arrowheads, cysteine-specific precursor adducts to Sam50- β 16 or Sam50- β 15, respectively; circles, unspecific or cysteine-independent adducts (in the absence of free SH-groups, AMAS and SIA can react with other amino acids).



Por1, we found that only precursors corresponding to an even number of β strands were cross-linked to loop 6 (Fig. 7A and fig. S8B, lanes 7 to 30). (ii) We analyzed an internal precursor region that corresponds to a β hairpin in mature Por1 by inserting a pair of cysteine residues at the putative adjacent β strands and a tobacco etch virus (TEV) protease cleavage site at the predicted loop between the β strands. Upon import of the [35 S]precursor into mitochondria and lysis, TEV protease cleaved the precursor into two fragments (fig. S9A). When SH-specific cross-linking was performed before lysis, the fragments were not separated, demonstrating that the corresponding cysteines of the predicted adjacent β strands were indeed in close, hairpinlike proximity. (iii) We inserted single cysteine residues into precursor regions that correspond to cytosolic loops or intermembrane space-exposed turns of mature Por1 and

imported them into mitochondria containing a single cysteine in Sam50-loop 6 (summarized in Fig. 7B). The predicted most-C-terminal precursor loop was cross-linked to residue 369 of Sam50-loop 6, whereas the predicted most-N-terminal precursor loop was preferentially cross-linked to residue 371 (Fig. 7C and fig. S9B; precursors of different lengths and SH-specific cross-linkers with different spacer lengths yielded a comparable pattern). Cysteines inserted into the predicted precursor turns were not cross-linked to Sam50-loop 6 (Fig. 7B and fig. S9C). (iv) The specific pairing of the C-terminal β signal of the precursor with Sam50- β 1 (Fig. 2 and fig. S2) indicates that the β signal is likely in a β -strand conformation. These results suggest that β -barrel precursors interacting with Sam50 are not in a random conformation, but are partially folded and contain β hairpin-like elements.

Taken together, loop 6 of Sam50 is in proximity to the precursor in transit and plays a crucial role in β -barrel biogenesis. Thus, in contrast to the POTRA domain, the functional importance of loop 6 in precursor transfer has been conserved from the bacterial Omp85 proteins FhaC and BamA (53, 54, 56) to Sam50. The analysis of precursor interaction with Sam50 supports the view that precursor insertion involves β hairpin-like conformations.

Discussion

We conclude that the biogenesis of mitochondrial β -barrel precursors involves the gate formed by the first and last β strands of Sam50. The analysis in the native mitochondrial system provides strong evidence for both the exchange model of β -signal recognition and the lateral release model of precursor exit through the Sam50 β -barrel gate (31, 33, 35, 36). Our findings suggest the following translocation path of a mitochondrial β -barrel precursor through SAM (Fig. 8). The precursor enters the interior of the Sam50 channel from the intermembrane space side in close proximity to Sam50- β 1. The C-terminal β signal of the precursor is specifically bound to Sam50- β 1 by exchange with the endogenous Sam50 β signal (Sam50- β 16), leading to an opening of the lateral gate. The conserved loop 6 of Sam50 is involved in precursor transfer to the lateral gate. More and more N-terminal portions of the precursor are threaded through the gate in close proximity to Sam50- β 16. Upon translocation of the entire precursor polypeptide chain by Sam50, the full-length β barrel can be formed and released from the SAM complex (13).

When comparing mitochondrial and bacterial β -barrel biogenesis, the pathways start in different locations (eukaryotic versus bacterial cytosol) and converge at the central Sam50 or BamA β barrels. Three main stages can be distinguished: (i) Initial translocation into the intermembrane space and periplasm is mediated by nonrelated translocases—the TOM complex of the mitochondrial outer membrane and the Sec complex of the bacterial plasma membrane (5, 6). (ii) Subsequent precursor transfer to the outer membrane is performed in part by related machineries, including intermembrane space and periplasmic chaperones and POTRA domains (46–49, 57–59). The bacterial transfer machinery is considerably more complex than that of mitochondria, likely reflecting the large number of bacterial β -barrel substrates (60). Bacteria use multiple POTRA domains and several periplasm-exposed Bam proteins (5, 15), whereas mitochondria contain a single non-essential POTRA domain and no accessory intermembrane space-exposed proteins (13, 50). The two cytosol-exposed peripheral Sam proteins are involved in formation of a TOM-SAM super-complex (Sam37) and stabilization of the SAM-bound form of the precursor (Sam35) (9–11, 13, 39, 41). (iii) Finally, the membrane insertion process occurs by means of the highly conserved membrane-integral part of Sam50 and BamA. The β signal has been well conserved, and several examples were reported showing that the β signal is exchangeable

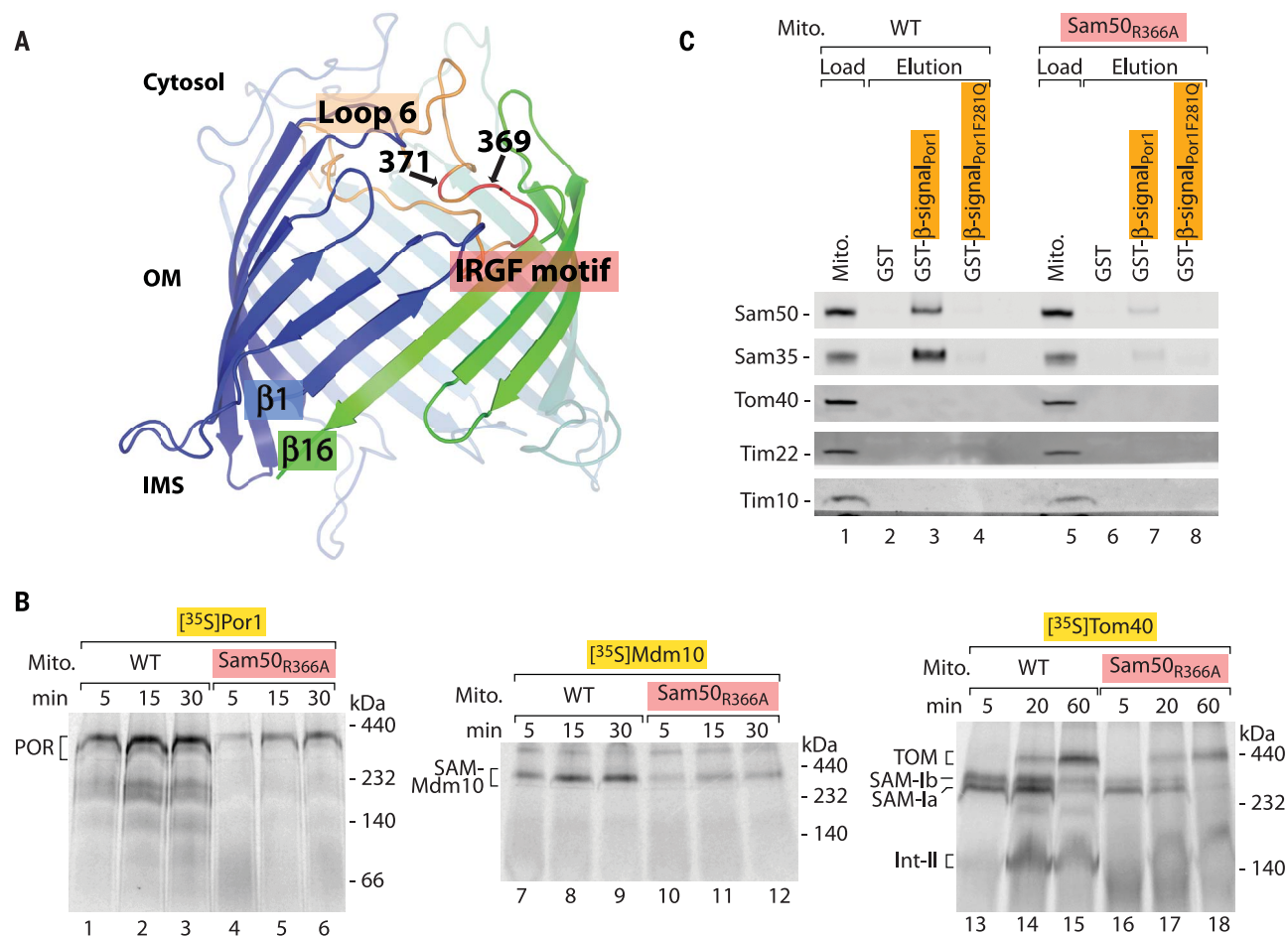


Fig. 6. Loop 6 of Sam50 is essential for β -barrel biogenesis. (A) Model of the Sam50 β barrel indicating loop 6 in peach and the conserved IRGF motif at the tip of loop 6 in red. The positions of residues 369 and 371 used for cross-linking in Fig. 7 are indicated. (B) Assembly of full-length β -barrel precursor proteins [³⁵S]Por1, [³⁵S]Mdm10, and [³⁵S]Tom40 in wild-type (WT) and Sam50_{R366A} mitochondria was analyzed by blue native electrophoresis

and autoradiography. SAM-Mdm10 indicates the SAM-Mdm10 complex; SAM-Ia, SAM-Ib, and Int-II are assembly intermediates of Tom40. (C) Immobilized glutathione-S-transferase (GST)-fusion proteins carrying the Por1 β signal were incubated with digitonin-lysed WT and Sam50_{R366A} mitochondria. The β signal was released by thrombin protease cleavage, and eluates were analyzed by SDS-PAGE and immunodecoration. Load 12.5%; elution 100%.

between bacteria, mitochondria, and chloroplasts (12, 13, 61), underscoring the conservation of basic mechanisms of β -barrel biogenesis. β -Barrel proteins are anchored in the lipid phase by a hydrophobic belt; the diminished hydrophobic area near the Sam50 and BamA lateral gates is thought to cause a membrane thinning (16, 21). In vitro studies on β -barrel membrane protein insertion demonstrate that membrane defects and BamA-mediated membrane distortion support membrane insertion (62–64). Sam50- and BamA-induced membrane thinning may contribute to β -barrel-membrane protein biogenesis in vivo by facilitating protein membrane insertion upon release from the SAM or BAM lateral gate. We propose that elements of both controversially discussed mechanisms, the budding model and assisted model, will be used in the lateral gate-sorting mechanism shown here.

The large diversity of bacterial β -barrel proteins and the involvement of multiple POTRA domains and accessory Bam proteins (5, 15, 51, 60)

raise the possibility that additional precursor-specific folding pathways may complement the central mechanism of β -signal exchange and sorting by means of the lateral gate elucidated here. For example, assembly of oligomeric β barrels in bacteria might be stalled at the BAM complex until all subunits are assembled (65), similar to the arrest of shortened precursor constructs of monomeric β barrels (Fig. 3). We envision that precursor insertion through the β -barrel channel and lateral gate demonstrated with mitochondrial Sam50 represents a basic mechanism that can also be employed by β -barrel assembly machineries of bacteria and chloroplasts.

Materials and methods

Site-directed mutagenesis

Mutagenesis was performed using the centromeric plasmid pFL39 (66) containing the wild-type open reading frame of *Saccharomyces cerevisiae* SAM50, TOM40 or POR1 and their corresponding native promoter and terminator

sequences (table S1). Primers listed in table S2, containing the specific mutational changes, were used for PCR with the high fidelity polymerases KOD (Sigma-Aldrich) or Q5 (NEB). After DpnI (NEB) template digestion (3 hours at 37°C), PCR products were transformed into competent XL-1 Blue *Escherichia coli* cells (Stratagene). Plasmids were isolated by using the QIAprep Spin Miniprep Kit (Qiagen). Successful mutagenesis was confirmed by sequencing.

Yeast strains and growth conditions

Because SAM50 is an essential gene, the plasmid shuffling method was used to exchange SAM50 wild-type with mutated versions of sam50 in a YPH499 background (67). The shuffling strain sam50 Δ contains a chromosomal deletion of SAM50 and expresses a wild-type copy of SAM50 on a YEp352 plasmid with a URA3 marker (7). After transformation of the centromeric TRP1 plasmid pFL39 containing a mutated sam50 allele, positive clones were selected on medium lacking

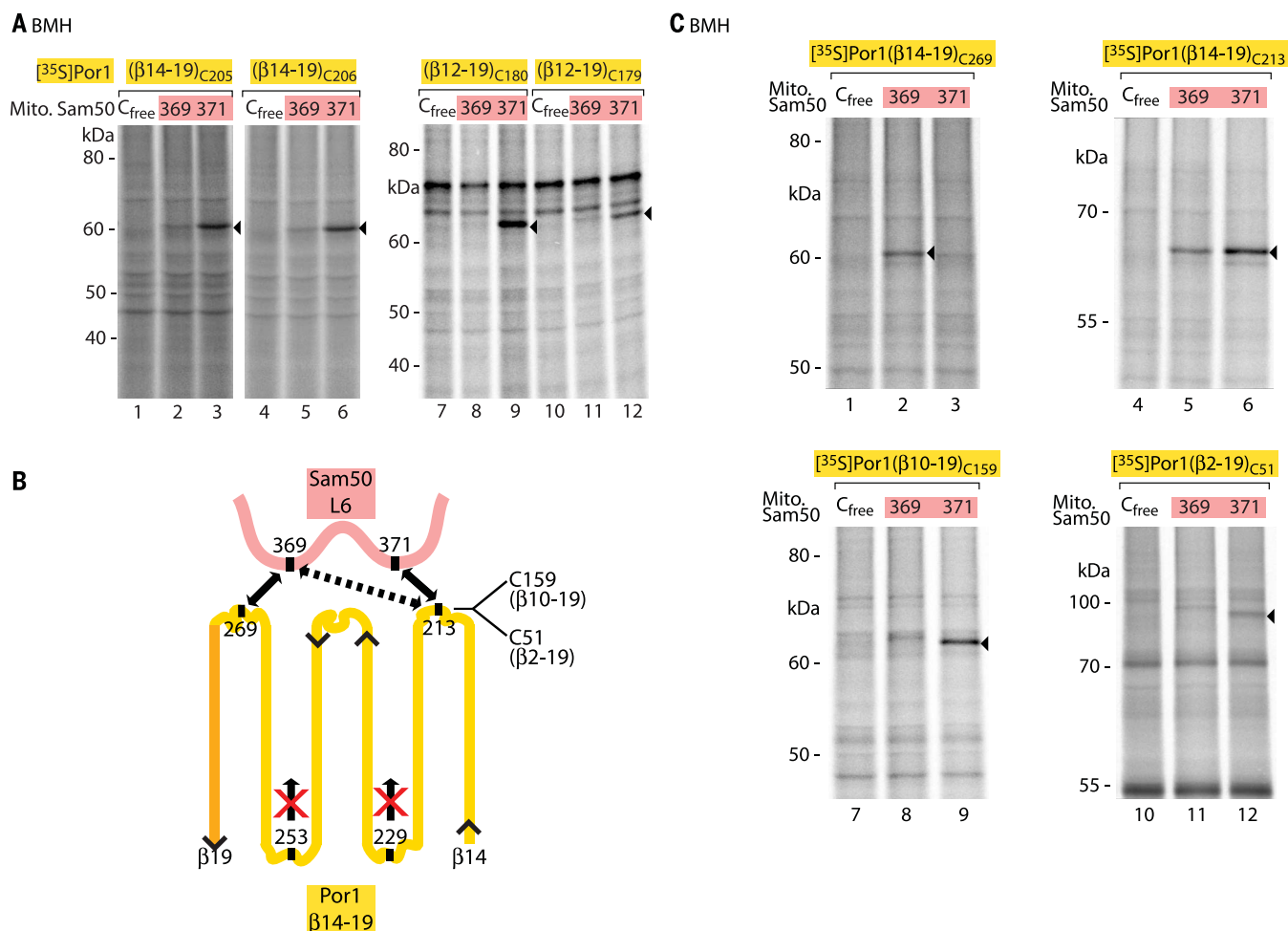


Fig. 7. β -Barrel precursors in transit are in close proximity to Sam50-loop 6. (A) [³⁵S]Por1(β 14 to 19) precursors carrying a cysteine in β strand 14 at position 205 or 206 and [³⁵S]Por1(β 12 to 19) precursors carrying a cysteine in β strand 12 at position 180 or 179 were imported for 5 min into isolated mitochondria containing Sam50 variants, followed by cross-linking with BMH. Samples were analyzed by nonreducing SDS-PAGE and autoradiography. Arrowheads, cysteine-specific Sam50-Por1

precursor cross-linking products. (B) Schematic model summarizing the crosslinking results of Fig. 7C and fig. S9, B and C. Black double-headed arrows, strong cross-links; dashed black double-headed arrow, weak cross-links; red X, no cross-links. (C) Radiolabeled Por1 constructs were imported into mitochondria of yeast strains expressing the indicated Sam50 variants for 5 min. Samples were treated with BMH and analyzed as described in (A).

tryptophan. By growth on plates containing 5-fluoroorotic acid (5-FOA) (Melford), cells that lost the *URA3* plasmid expressing wild-type *SAM50* were selected. Subsequently, yeast cells were grown on nonfermentable medium containing glycerol to rule out the loss of mitochondrial DNA. At each step, plates were incubated at 23°C to minimize possible temperature sensitive growth defects.

Yeast cells were cultured in liquid YPG medium (1% [w/v] yeast extract (Becton Dickinson), 2% [w/v] bacto peptone (Becton Dickinson), 3% [w/v] glycerol (Sigma), pH 5 - HCl (Roth)) at 23°C and shaking with 130 rpm. For growth tests, single yeast cells were picked and incubated overnight in 5 ml YPG. Cells corresponding to an OD₆₀₀ of 1 were taken from yeast strains indicated and resuspended in 1 ml autoclaved and distilled H₂O. The suspension was further diluted by factors of 1:10, 1:100, 1:1000 and 1:10,000. 3 or 5 μ l were dropped on solid YPG (1% [w/v] yeast extract,

2% [w/v] bacto peptone, 3% [w/v] glycerol, 2.5% [w/v] agar (Becton Dickinson)) and YPD (1% [w/v] yeast extract, 2% [w/v] bacto peptone, 2% [w/v] glucose (Roth), 2.5% [w/v] agar). Plates were incubated at indicated temperatures.

Yeast cells expressing Sam50 lacking loop 6 (*sam50* _{Δ loop6}) did not yield colonies after plasmid shuffling. Therefore, the plasmid encoding *SAM50* _{Δ loop6} was transformed into a YPH499 strain expressing *SAM50* under the control of a galactose promoter. After selection on galactose (Sigma-Aldrich) containing medium lacking tryptophan, the shutdown of *SAM50* wild-type was performed by growth in liquid SL-medium (0.3% [w/v] yeast nitrogen base w/o amino acids (Becton Dickinson), 0.077% [w/v] complete supplement mix (-TRP) (MP biomedical), 0.05% [w/v] NaCl (Roth), 0.05% [w/v] CaCl₂ (Roth), 0.06% [w/v] MgCl₂ (Roth), 0.1% [w/v] NH₄Cl (Roth), 0.1% [w/v] KH₂PO₄ (Roth), 0.6% [w/v] NaOH (Roth),

2.2% [v/v] lactic acid (Roth), 0.05% [w/v] glucose) (11, 13, 68). Yeast cells were diluted approximately every 20 hours with fresh medium. Yeast strains are listed in table S3.

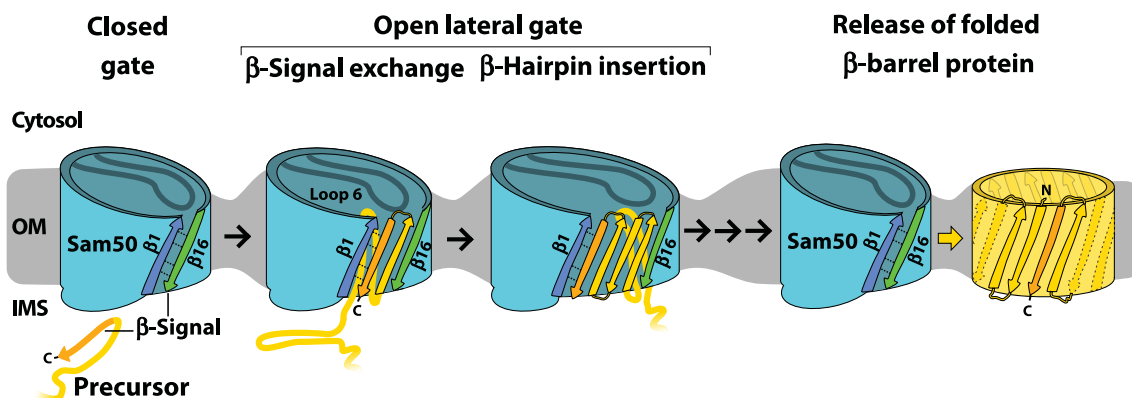
Isolation of mitochondria

Yeast cells were cultivated in YPG medium for 2 days as a preculture. The main culture was inoculated with the preculture and incubated for at least 15 hours with shaking at 130 rpm and 30°C. Yeast expressing *SAM50* _{Δ loop6} were grown in SL-Medium at 30°C for 42.5 hours to ensure proper shutdown of *SAM50* wild-type.

Yeast cells were harvested during log-phase by centrifugation at 1700 \times g (maximal relative centrifugal force; 4000 rpm, H-12000 Thermo-Fisher Scientific) for 10 min at room temperature. Yeast cells were washed twice with distilled H₂O, and incubated with 2 ml/g wet weight DTT buffer [100 mM Tris(hydrosymethyl)aminomethane

Fig. 8. Putative model for sorting of β -barrel precursors through the lateral gate of Sam50. β -Barrel precursors are translocated from the intermembrane space side into the lumen of Sam50. The C-terminal β signal of the precursor specifically binds to β strand 1 of Sam50 by replacing the endogenous β signal of Sam50 (β strand 16). This

induces an opening of the lateral gate of Sam50 between β strands 1 and 16. Further strands of the precursor are inserted into the lateral gate in β hairpin-like structures. Loop 6 of Sam50 promotes transfer of the precursor into



the lateral gate. The full-length precursor is released from the lateral gate into the lipid phase of the outer membrane. Thinning of the membrane in proximity to the lateral gate facilitates membrane insertion of the β -barrel protein.

(Tris)/H₂SO₄ (MP Biomedicals and Roth), pH 9.4, 10 mM dithiothreitol (DTT, Roth)] for 20 min with shaking at 130 rpm and 30°C. Yeast cells were reisolated by centrifugation for 5 min at 2700 × g (4000 rpm, SLA-3000 Sorvall) and incubated for 30 to 45 min in 6.5 ml/g wet weight Zymolyase buffer [16 mM K₂HPO₄ (Roth), 4 mM KH₂PO₄, pH 7.4, 1.2 M sorbitol (Roth), 3 mg/ml Zymolyase 20T (Seikagaku Kaygyo Co.)] with shaking at 130 rpm and 30°C. The resulting spheroplasts were pelleted by centrifugation for 5 min at 1500 × g (3000 rpm, SLA-3000 Sorvall) and washed with buffer (16 mM K₂HPO₄, 4 mM KH₂PO₄, pH 7.4, 1.2 M sorbitol). The pellet was resuspended in homogenization buffer (0.6 M sorbitol, 10 mM Tris/HCl, pH 7.4, 1 mM ethylenediaminetetraacetic acid (EDTA, Calbiochem), 0.4% [w/v] bovine serum albumin (Sigma), 1 mM phenylmethyl sulfonyl fluoride (PMSF, Sigma)). The spheroplasts were mechanically opened using a glass-Teflon potter by homogenizing the solution 17 times on ice. Mitochondria were isolated using a four-centrifugation step procedure. To remove cell debris, the solution was spun for 15 min, 1450 × g (3500 rpm, SS-34, Sorvall) at 4°C, followed by a high speed spinning step to pellet mitochondria at 18,500 × g (12,500 rpm, SS-34, Sorvall) for 15 min at 4°C. The mitochondrial pellet was resuspended in ice cold SEM buffer [250 mM sucrose (MP Biomedicals), 10 mM 3-(N-morpholino)propanesulfonic acid (MOPS, Sigma), pH 7.2, 1 mM EDTA] containing 1 mM PMSF and both centrifugation steps were repeated. The mitochondrial pellet was resuspended in ice cold SEM and the protein concentration was determined using the Bradford protein assay (69). The concentration was adjusted to 10 mg mitochondrial protein per 1 ml SEM. Mitochondria were aliquoted, snap-frozen in liquid nitrogen and stored at -80°C (70).

Oxidation assays with whole yeast cells

Yeast cells were grown overnight in YPG medium at 30°C. Cells corresponding to an OD₆₀₀ of 1 were taken and harvested by centrifugation for 10 min at 1500 × g (4000 rpm, FA-45-24-II, Eppendorf).

Cells were resuspended in 100 μ l buffer A [2 mM PMSF, 2 \times protease inhibitor w/o EDTA (Roche), 1 mM EDTA] and oxidized by adding 0.2 mM 4,4'-Dipyridyl disulfide (4-DPS, Sigma-Aldrich) (39). Cells were incubated on ice for 30 min followed by addition of 50 mM iodoacetamide (IA, Sigma) and further incubated for 15 min on ice. After addition of 60 mM NaOH, cells were centrifuged for 10 min at 1700 × g (4000 rpm, FA 45-30-II, Eppendorf) and 4°C, resuspended in Laemmli buffer and heated to 65°C for 10 min shaking vigorously.

Sam50 modeling

Potential templates were identified with the HHPRED server restricting the search sequence to the Sam50 β -barrel domain (residues 125 to 484) (71). The hidden-Markov model based homology search identified templates in the PDB with good p- and E-values. This included structures from FhaC [PDB code: 4QL0 (51)] and TamA [PDB code: 4C00 (18)] as well as several structures from BamA [PDB codes: 4K3B and 4K3C (16); 4C4V (21); 4N75 (22) and 5EKQ (23)], which exhibit considerable variations in the interaction between the β 1 and β 16 strands. A Sam50 model was calculated from each template using Modeler (72). The model obtained from the BamA structure with PDB code 4N75 (22) with optimized alignment fit best to the experimental results of disulfide bonds and cross-link formation (model S1).

Despite low sequence identity of 14%, the β -barrel model of Sam50 shows a very good agreement with the structure of BamA (PDB code: 4n75) with a core RMSD of 1.6 Å (C α -atoms of 310 out of 360 residues). Ramachandran analysis using RAMPAGE (73) showed similar geometrical quality of the model compared to the template (favored/allowed/outlier residues, model: 90.2% / 7.3% / 2.5% and template: 94.7% / 4.5% / 0.8%). Also, the distribution of charged and aromatic residues in respect to barrel inward and outward facing side chains agrees well between model and structure. In order to evaluate the position of loop 6, we superimposed the model

with five BamA structures (PDB codes: 4K3B, 4K3C, 4C4V, 4N75 and 5EKQ) as well as the TamA structure (PDB code: 4C00). They all show a highly similar overall structure for loop 6, with identical positions for the conserved IRGF motif including side chain orientations. IRGF faces the inside wall of the barrel (strands 13 to 16). Noteworthy is for instance the interaction between the guanidino group of the motif's arginine residue with an aromatic side chain of β barrel strand 13. The Sam50 model agrees overall with the structures of the loop and the position of IRGF side chains, for instance R366 is interacting with the aromatic ring of F413. Also, positions and orientations of residues 369 to 371 in the Sam50 model agree with those of the aforementioned structures. In addition, the side chain orientations of the Sam50 β signal (strand 16) toward either the β -barrel lumen or the lipid phase agree with the structure of the conserved β signal of mitochondrial VDAC/Porin (42–44).

For graphical presentations, cysteine residues were included in silico at relevant positions and disulfide bonds formed using coot (74) before figures were generated with Pymol (The PyMOL Molecular Graphics System, Version 1.6 Schrödinger, LLC). The Sam50 β -barrel models were oriented according to the localization of the N-terminal POTRA domain in the mitochondrial intermembrane space (13, 50).

In vitro transcription and translation

Plasmids containing the coding region of the gene of interest and carrying an upstream SP6 promoter binding region were incubated with TNT SP6 quick coupled kit (Promega), an in vitro eukaryotic translation system based on rabbit reticulocytes, in the presence of [³⁵S]methionine (PerkinElmer). The reaction was incubated for at least 90 min at 25°C with shaking at 300 rpm. Reactions were stopped upon addition of 20 mM unlabeled methionine (Roth). A clarifying step was performed at 125,000 × g (45,000 rpm, TLA-55, Beckman) for 30 min at 4°C. 0.3 M sucrose was added to the supernatant and the lysate was snap-frozen and stored at -80°C. Successful transcription/

translation was checked by SDS-PAGE and autoradiography.

Template DNA of cysteine mutants of Por1 and Tom40 constructs was generated by PCR using 2× REDTaq ReadyMix (Genaxxon). Forward primers contained a RTS™ wheat germ kit (5prime) specific 5'-CTTTAAGAAGGAGATA-TACC-3' sequence upstream of the start codon. The corresponding reverse primers contained downstream of the stop codon a 5'-TGATGAT-GAGAACCCCCCCC-3' wheat germ sequence. Cysteine mutagenesis was performed using a primer encoding the desired mutation. Successful mutations were confirmed by sequencing. For enhanced radiolabeling of the protein fragment, the methionine encoding sequence was added at the start codon and before the stop codon of the primers used for in vitro transcription. PCR products were analyzed by inspection of the DNA bands on 2% agarose (Biozym) gels. Products were purified using the QIAquick PCR Purification Kit (Qiagen). A consecutive PCR was performed according to the RTS™ wheat germ LinTempGenSet His₆-tag (5prime) manual using 2× REDTaq ReadyMix. PCR products were purified and concentrated using the MinElute PCR purification kit (Qiagen). Wheat germ lysate, an eukaryotic cell-free protein expression system based on wheat germ, was used as described in the RTS™ 100 wheat germ CECF Kit (5prime) with modification for radiolabeled lysates, including [³⁵S]methionine in the reaction solution and supplementation of unlabeled methionine with [³⁵S]methionine in the feeding solution. After incubation for 24 hours, lysates were clarified by centrifugation at 125,000 × g (45,000 rpm, TLA-55, Beckman) for 1 hour at 4°C. The supernatant was transferred to a fresh tube, snap-frozen in liquid nitrogen and stored at -80°C. Successful translation was checked by SDS-PAGE and autoradiography.

In vitro import into mitochondria and cross-linking and oxidation

Mitochondria were thawed on ice. For one import reaction, 50 µg mitochondria (protein amount) were resuspended in 100 µl import buffer (3% [w/v] bovine serum albumin, 250 mM sucrose, 80 mM KCl (Roth), 5 mM MgCl₂, 2 mM KH₂PO₄, 5 mM methionine (Sigma), 10 mM MOPS-KOH (Roth), pH 7.2) with 4 mM ATP (Roche), 4 mM NADH (Roth), 5 mM creatine phosphate (Roche) and 0.1 mM creatine kinase (Roche). To deplete the membrane potential ($-\Delta\psi$), AVO (8 µM antimycin A (Sigma), 1 µM valinomycin (Sigma), 20 µM oligomycin (Sigma), final concentrations) was added and NADH was omitted (75). When Tim9 was imported, bovine serum albumin was omitted. 4 to 10% [v/v] of rabbit reticulocyte lysate or wheat germ lysate containing the precursor proteins were incubated with mitochondria at 25°C with shaking at 300 rpm. Membrane potential dependent import reactions were stopped by addition of AVO, before the import reactions were transferred on ice. The mitochondria were pelleted for 10 min at 4°C and 18,500 × g (13,200 rpm, FA 45-30-II, Eppendorf), the supernatant was discarded and the pellet was washed

with 100 µl SEM. Mitochondria were resuspended in Laemmli buffer (0.2 M Tris, pH 6.8, 2% [w/v] dodecylsulfate-Na-salt (SDS, Serva), 4% [v/v] glycerol, 12.5% [w/v] bromophenol blue (Sigma), 1 mM PMSF, 50 mM iodoacetamide) and incubated for 10 min at 65°C shaking vigorously.

In case of experiments combining protein import and cross-linking or oxidation, mitochondria were incubated in 100 µl SEM including energy mix (4 mM ATP, 4 mM NADH, 5 mM creatine phosphate, 0.1 mM creatine kinase) and 2 to 6% [v/v] precursor-containing lysate was added. Import was conducted at 25°C for 5 to 30 min, shaking at 300 rpm, and the reactions were transferred on ice. To oxidize proteins, 0.36 mM 4-DPS or 2.5 mM CuSO₄ (Roth) was added to the reaction. For cross-linking experiments, 1 mM cross-linking reagent 1,6-bismaleimidoethane (BMH, Thermo-Fisher Scientific), bismaleimidoethane (BMOE, Thermo-Fisher Scientific), 1,3-propanediylbismethanethiosulfonate (M3M, Interchim), 1,1-methanediylbismethanethiosulfonate (MIM, Interchim), *N*-α-maleimidoacet-oxy succinimide ester (AMAS, Thermo-Fisher Scientific) or succinimidyl iodoacetate (SIA, Thermo-Fisher Scientific) were added from a 10 mM stock solution prepared in dimethyl sulfoxide (DMSO, Roth). Samples were gently mixed and incubated on ice for 30 min. Oxidation/cross-linking reactions were stopped by addition of 50 mM iodoacetamide and incubated on ice for 15 min. Reactions were laid on top of 500 µl S₅₀₀EM (500 mM sucrose, 10 mM MOPS, pH 7.2, 1 mM EDTA) and centrifuged for 15 min at 4°C and 20,800 × g (14,000 rpm, FA 45-30-II, Eppendorf) for purification. The pellet was resuspended in 100 µl SEM and processed as described above. Samples analyzed on blue native PAGE were resuspended in digitonin buffer (0.1 mM EDTA, 10% [v/v] glycerol, 50 mM NaCl, 1 mM PMSF, 20 mM Tris/HCl, pH 7.4, 1% [w/v] digitonin (Calbiochem)) and incubated on ice for 15 min before addition of blue native loading dye (0.5% [w/v] Coomassie blue G (Serva), 50 mM 6-aminocaproic acid, 10 mM Bis-Tris/HCl, pH 7).

β-Signal affinity assay

The method was performed as described in Kutik *et al.* (13). Briefly, *E. coli* cells expressing glutathione-S-transferase (GST), GST-β-signal_{Por1} and GST-β-signal_{Por1F281Q} were lysed and GST constructs were bound to glutathione sepharose 4B beads (GE Healthcare). Mitochondria were solubilized in GST buffer L (20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, Roth), 100 mM potassium acetate (KOAc, Roth), 10 mM magnesium acetate (Mg(OAc)₂, Roth), 10% [v/v] glycerol, 1% [w/v] digitonin (Calbiochem), 1 mM PMSF). After centrifugation for 10 min at 4°C and 20,800 × g (14,000 rpm, FA 45-30-II, Eppendorf), the supernatant was transferred to GST bound sepharose beads and incubated for 30 min at 4°C shaking end over end. Samples were centrifuged for 1 min at 4°C and 500 × g and washed at least seven times with GST buffer W (20 mM HEPES, 100 mM KOAc, 10 mM Mg(OAc)₂, 10% [v/v] glycerol, 0.5% [w/v] digitonin). To cleave bound

proteins from GST, samples were incubated overnight at 4°C, shaking at 800 rpm, in 200 µl GST buffer T (20 mM HEPES, 100 mM KOAc, 10 mM Mg(OAc)₂, 10% [v/v] glycerol, 0.5% [w/v] digitonin, 2.5 mM CaCl₂, 80 units thrombin protease (Calbiochem)). Columns were centrifuged for 1 min at 4°C and 500 × g. The flow-through was mixed with Laemmli buffer including 1% [v/v] β-mercaptoethanol (Roth) and heated to 95°C for 5 min. Samples were analyzed by SDS-PAGE.

TEV protease cleavage assay

In vitro import into mitochondria followed by cross-linking using BMOE was conducted as described above. After purification, samples were resuspended in solubilization buffer (20 mM Tris/HCl, pH 7.4, 0.1 mM EDTA, 50 mM NaCl) containing 6 M guanidinium hydrochloride (Roth). Samples were heated to 95°C and diluted 1:4 in solubilization buffer. TEV protease (Thermo-Fisher Scientific) was added and incubated for 30 min on ice. Samples were precipitated using 14% [w/v] trichloroacetic acid (TCA, Roth) and 0.0125% [w/v] sodium deoxycholate (Sigma). Samples were resuspended in Laemmli buffer containing 1 mM PMSF and 10 mM IA and heated to 65°C for 10 min shaking vigorously. Samples were separated by 4 to 12% NuPAGE gels (Thermo-Fisher Scientific) according to the manufacturer's protocol.

Swelling assay

To test the integrity of the mitochondrial outer membrane, 100 µg mitochondria (protein amount) were thawed on ice and resuspended in either 100 µl hypotonic swelling buffer (1 mM EDTA, 10 mM MOPS/KOH, pH 7.2) or isotonic SEM buffer. Mitochondria/mitoplasts were incubated on ice for 15 min before the indicated amount of proteinase K (Roche) was added. The samples were further incubated for 15 min on ice. Proteinase K was inactivated by addition of 2 mM PMSF and further incubated on ice for 10 min. Mitochondria/mitoplasts were pelleted and washed with SEM buffer including 1 mM PMSF. Samples were resuspended in Laemmli buffer, including 1% [v/v] β-mercaptoethanol and 1 mM PMSF, and separated by SDS-PAGE.

SDS-PAGE, NuPAGE, tris-tricine PAGE, blue native PAGE, and Western blotting

SDS-PAGE was performed using 10% polyacrylamide gels and SDS running buffer (25 mM Tris/HCl, pH 8.8, 191 mM glycine (MP Biomedicals), 1% [w/v] SDS). Gels were run at 30 to 35 mA for 3 to 5 hours. NuPAGE bis-tris pre-cast gels (10%, Thermo-Fisher Scientific) were used according to the manufacturer's instructions. Tris-Tricine PAGE gels consist of a 4 to 16% polyacrylamide gradient (48% [w/v] acrylamide (Roth), 1.5% [w/v] bisacrylamide (Serva)). Gels were run using anode buffer (0.2 M Tris/HCl, pH 8.9) and cathode buffer (0.1 M Tris, 0.1 M Tricine (Roth), 0.1% [w/v] SDS, pH 8.25) at 70 mA for 3 to 5 hours. For all above mentioned gel electrophoresis procedures, samples were resuspended in Laemmli buffer containing 1 mM PMSF, heated to 65°C for 10 min shaking

vigorously. When samples were cross-linked or oxidized, no DTT or β -mercaptoethanol was added but 50 mM iodoacetamide.

Native protein complexes were analyzed using blue native PAGE (76). After import of radio-labeled proteins, mitochondria were resuspended in cold digitonin buffer (0.1 mM EDTA, 10% [v/v] glycerol, 50 mM NaCl, 1 mM PMSF, 20 mM Tris/HCl, pH 7.4, 0.35 to 1% [w/v] digitonin) and incubated on ice for 15 min. Blue native loading dye (0.5% [w/v] Coomassie blue G (Serva), 50 mM 6-aminocaproic acid (Sigma), 10 mM Bis/Tris (Roth), pH 7) was added. Samples were centrifuged at 4°C for 15 min at 20,800 \times g (14,000 rpm, FA 45-30-II, Eppendorf) and the supernatant was loaded on a 6 to 16.5% discontinuous gradient gel. 8.5 cm gels were run in a cooled Hoefer SE600 vertical electrophoresis chamber using anode buffer (50 mM Bis/Tris/HCl, pH 7) and cathode buffer (50 mM tricine, pH 7, 15 mM Bis/Tris, 0.02% [w/v] Coomassie G) at 90 mA and 600 V for 90 min.

With the exception of blue native gels, gels containing radiolabeled samples were stained and fixed using staining buffer (30% [v/v] ethanol, 10% [v/v] acetic acid (Roth), 0.2% [w/v] Coomassie R250 (Roth)) followed by destaining with destain buffer (50% [v/v] methanol (Roth), 20% [v/v] acetic acid) until protein bands were clearly visible. Gels were dried onto Whatman paper (Macherey-Nagel) and exposed using PhosphorImager screens (GE Healthcare and Fuji), followed by autoradiographic detection (Storm PhosphorImager, GE Healthcare; FLA9000, Fujifilm).

When immunoblotting was performed, gels were incubated for 5 min in SDS running buffer after gel electrophoresis. Gel contents were transferred onto PVDF membranes (Immobilon-P, Millipore) using standard semi dry Western blotting (77) at 250 mA for 2 hours using blotting buffer (20 mM Tris, 150 mM glycine, 0.02% [w/v] SDS, 20% [v/v] methanol). PVDF membranes were stained with staining buffer, destained using destain buffer until visible bands confirmed equal loading, and completely destained using 100% methanol. Blocking was performed for 1 hour using 5% [w/v] fat-free dried milk powder (Frema Reform) in TBST (200 mM Tris/HCl, pH 7.5, 1.25 M CaCl₂, 0.1% [v/v] Tween20 (Sigma)) at room temperature. After washing in TBST, membranes were incubated with the designated primary antibodies listed in table S4, overnight at 4°C or for at least 1 hour at room temperature. After a second washing step in TBST, membranes were decorated with secondary antirabbit IgG antibody (Sigma), diluted 1:5000, that was coupled to horse radish peroxidase in 5% [w/v] fat-free dried milk powder in TBST for 1 hour. After washing a third time in TBST, membranes were incubated in ECL solution (GE Healthcare) and the chemiluminescence signal was detected by the LAS-4000 system (Fujifilm).

REFERENCES AND NOTES

- S. A. Paschen *et al.*, Evolutionary conservation of biogenesis of β -barrel membrane proteins. *Nature* **426**, 862–866 (2003). doi: [10.1038/nature02208](https://doi.org/10.1038/nature02208); pmid: [14685243](https://pubmed.ncbi.nlm.nih.gov/14685243/)
- R. Voulhoux, M. P. Bos, J. Geurtsen, M. Mols, J. Tommassen, Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science* **299**, 262–265 (2003). doi: [10.1126/science.1078973](https://doi.org/10.1126/science.1078973); pmid: [12522254](https://pubmed.ncbi.nlm.nih.gov/12522254/)
- N. Wiedemann *et al.*, Machinery for protein sorting and assembly in the mitochondrial outer membrane. *Nature* **424**, 565–571 (2003). doi: [10.1038/nature01753](https://doi.org/10.1038/nature01753); pmid: [12891361](https://pubmed.ncbi.nlm.nih.gov/12891361/)
- T. Wu *et al.*, Identification of a multicomponent complex required for outer membrane biogenesis in *Escherichia coli*. *Cell* **121**, 235–245 (2005). doi: [10.1016/j.cell.2005.02.015](https://doi.org/10.1016/j.cell.2005.02.015); pmid: [15851030](https://pubmed.ncbi.nlm.nih.gov/15851030/)
- C. L. Hagan, T. J. Silhavy, D. Kahne, β -Barrel membrane protein assembly by the Bam complex. *Annu. Rev. Biochem.* **80**, 189–210 (2011). doi: [10.1146/annurev-biochem-061408-144611](https://doi.org/10.1146/annurev-biochem-061408-144611); pmid: [21370981](https://pubmed.ncbi.nlm.nih.gov/21370981/)
- T. Jores *et al.*, Characterization of the targeting signal in mitochondrial β -barrel proteins. *Nat. Commun.* **7**, 12036 (2016). doi: [10.1038/ncomms12036](https://doi.org/10.1038/ncomms12036); pmid: [27345737](https://pubmed.ncbi.nlm.nih.gov/27345737/)
- V. Kozjak *et al.*, An essential role of Sam50 in the protein sorting and assembly machinery of the mitochondrial outer membrane. *J. Biol. Chem.* **278**, 48520–48523 (2003). doi: [10.1074/jbc.C300442200](https://doi.org/10.1074/jbc.C300442200); pmid: [14570913](https://pubmed.ncbi.nlm.nih.gov/14570913/)
- I. Gentle, K. Gabriel, P. Beech, R. Waller, T. Lithgow, The Omp85 family of proteins is essential for outer membrane biogenesis in mitochondria and bacteria. *J. Cell Biol.* **164**, 19–24 (2004). doi: [10.1083/jcb.200310092](https://doi.org/10.1083/jcb.200310092); pmid: [14699090](https://pubmed.ncbi.nlm.nih.gov/14699090/)
- D. Ishikawa, H. Yamamoto, Y. Tamura, K. Moritoh, T. Endo, Two novel proteins in the mitochondrial outer membrane mediate β -barrel protein assembly. *J. Cell Biol.* **166**, 621–627 (2004). doi: [10.1083/jcb.200405138](https://doi.org/10.1083/jcb.200405138); pmid: [15326197](https://pubmed.ncbi.nlm.nih.gov/15326197/)
- D. Milenkovic *et al.*, Sam35 of the mitochondrial protein sorting and assembly machinery is a peripheral outer membrane protein essential for cell viability. *J. Biol. Chem.* **279**, 22781–22785 (2004). doi: [10.1074/jbc.C400120200](https://doi.org/10.1074/jbc.C400120200); pmid: [15067005](https://pubmed.ncbi.nlm.nih.gov/15067005/)
- T. Waizenegger *et al.*, Tob38, a novel essential component in the biogenesis of β -barrel proteins of mitochondria. *EMBO Rep.* **5**, 704–709 (2004). doi: [10.1038/sj.embor.7400183](https://doi.org/10.1038/sj.embor.7400183); pmid: [15205677](https://pubmed.ncbi.nlm.nih.gov/15205677/)
- V. Robert *et al.*, Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLOS Biol.* **4**, e377 (2006). doi: [10.1371/journal.pbio.0040377](https://doi.org/10.1371/journal.pbio.0040377); pmid: [17090219](https://pubmed.ncbi.nlm.nih.gov/17090219/)
- S. Kutik *et al.*, Dissecting membrane insertion of mitochondrial β -barrel proteins. *Cell* **132**, 1011–1024 (2008). doi: [10.1016/j.cell.2008.01.028](https://doi.org/10.1016/j.cell.2008.01.028); pmid: [18358813](https://pubmed.ncbi.nlm.nih.gov/18358813/)
- C. Stubenrauch *et al.*, Effective assembly of fimbriae in *Escherichia coli* depends on the translocation assembly module nanomachine. *Nat. Microbiol.* **1**, 16064 (2016). doi: [10.1038/nmicrobiol.2016.64](https://doi.org/10.1038/nmicrobiol.2016.64); pmid: [27572967](https://pubmed.ncbi.nlm.nih.gov/27572967/)
- J. Lee *et al.*, Characterization of a stalled complex on the β -barrel assembly machine. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8717–8722 (2016). doi: [10.1073/pnas.1604100113](https://doi.org/10.1073/pnas.1604100113); pmid: [27439868](https://pubmed.ncbi.nlm.nih.gov/27439868/)
- N. Noinaj *et al.*, Structural insight into the biogenesis of β -barrel membrane proteins. *Nature* **501**, 385–390 (2013). doi: [10.1038/nature12521](https://doi.org/10.1038/nature12521); pmid: [23995689](https://pubmed.ncbi.nlm.nih.gov/23995689/)
- B. Clantin *et al.*, Structure of the membrane protein FhaC: A member of the Omp85-TpsB transporter superfamily. *Science* **317**, 957–961 (2007). doi: [10.1126/science.1143860](https://doi.org/10.1126/science.1143860); pmid: [17702945](https://pubmed.ncbi.nlm.nih.gov/17702945/)
- F. Gruss *et al.*, The structural basis of autotransporter translocation by TamA. *Nat. Struct. Mol. Biol.* **20**, 1318–1320 (2013). doi: [10.1038/nsmb.2689](https://doi.org/10.1038/nsmb.2689); pmid: [24056943](https://pubmed.ncbi.nlm.nih.gov/24056943/)
- J. Selkrig *et al.*, Discovery of an archetypal protein transport system in bacterial outer membranes. *Nat. Struct. Mol. Biol.* **19**, 506–510, S1 (2012). doi: [10.1038/nsmb.2261](https://doi.org/10.1038/nsmb.2261); pmid: [22466966](https://pubmed.ncbi.nlm.nih.gov/22466966/)
- C. Baud *et al.*, Translocation path of a substrate protein through its Omp85 transporter. *Nat. Commun.* **5**, 5271 (2014). doi: [10.1038/ncomms5271](https://doi.org/10.1038/ncomms5271); pmid: [25327833](https://pubmed.ncbi.nlm.nih.gov/25327833/)
- R. Albrecht *et al.*, Structure of BamA, an essential factor in outer membrane protein biogenesis. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 1779–1789 (2014). doi: [10.1107/S1399004714007482](https://doi.org/10.1107/S1399004714007482); pmid: [24914988](https://pubmed.ncbi.nlm.nih.gov/24914988/)
- D. Ni *et al.*, Structural and functional analysis of the β -barrel domain of BamA from *Escherichia coli*. *FASEB J.* **28**, 2677–2685 (2014). doi: [10.1096/fj.13-248450](https://doi.org/10.1096/fj.13-248450); pmid: [24619089](https://pubmed.ncbi.nlm.nih.gov/24619089/)
- J. Bakelar, S. K. Buchanan, N. Noinaj, The structure of the β -barrel assembly machinery complex. *Science* **351**, 180–186 (2016). doi: [10.1126/science.aad3460](https://doi.org/10.1126/science.aad3460); pmid: [26744406](https://pubmed.ncbi.nlm.nih.gov/26744406/)
- Y. Gu *et al.*, Structural basis of outer membrane protein insertion by the BAM complex. *Nature* **531**, 64–69 (2016). doi: [10.1038/nature17199](https://doi.org/10.1038/nature17199); pmid: [26901871](https://pubmed.ncbi.nlm.nih.gov/26901871/)
- L. Han *et al.*, Structure of the BAM complex and its implications for biogenesis of outer-membrane proteins. *Nat. Struct. Mol. Biol.* **23**, 192–196 (2016). doi: [10.1038/nsmb.3181](https://doi.org/10.1038/nsmb.3181); pmid: [26900875](https://pubmed.ncbi.nlm.nih.gov/26900875/)
- M. G. Iadanza *et al.*, Lateral opening in the intact β -barrel assembly machinery captured by cryo-EM. *Nat. Commun.* **7**, 12865 (2016). doi: [10.1038/ncomms12865](https://doi.org/10.1038/ncomms12865); pmid: [27686148](https://pubmed.ncbi.nlm.nih.gov/27686148/)
- P. A. Doerner, M. C. Sousa, Extreme dynamics in the BamA β -barrel seam. *Biochemistry* **56**, 3142–3149 (2017). doi: [10.1021/acs.biochem.7b00281](https://doi.org/10.1021/acs.biochem.7b00281); pmid: [28569500](https://pubmed.ncbi.nlm.nih.gov/28569500/)
- A. Klein *et al.*, Characterization of the insertase for β -barrel proteins of the outer mitochondrial membrane. *J. Cell Biol.* **199**, 599–611 (2012). doi: [10.1083/jcb.201207161](https://doi.org/10.1083/jcb.201207161); pmid: [23128244](https://pubmed.ncbi.nlm.nih.gov/23128244/)
- B. van den Berg, Lateral gates: β -barrels get in on the act. *Nat. Struct. Mol. Biol.* **20**, 1237–1239 (2013). doi: [10.1038/nsmb.2709](https://doi.org/10.1038/nsmb.2709); pmid: [24197163](https://pubmed.ncbi.nlm.nih.gov/24197163/)
- D. Gessmann *et al.*, Outer membrane β -barrel protein folding is physically controlled by periplasmic lipid head groups and BamA. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5878–5883 (2014). doi: [10.1073/pnas.1322473111](https://doi.org/10.1073/pnas.1322473111); pmid: [24715731](https://pubmed.ncbi.nlm.nih.gov/24715731/)
- N. Noinaj, A. J. Kuszak, C. Balusek, J. C. Gumbart, S. K. Buchanan, Lateral opening and exit pore formation are required for BamA function. *Structure* **22**, 1055–1062 (2014). doi: [10.1016/j.str.2014.05.008](https://doi.org/10.1016/j.str.2014.05.008); pmid: [24980798](https://pubmed.ncbi.nlm.nih.gov/24980798/)
- J. H. Kleinschmidt, Folding of β -barrel membrane proteins in lipid bilayers—Unassisted and assisted folding and insertion. *Biochim. Biophys. Acta* **1848**, 1927–1943 (2015). doi: [10.1016/j.bbame.2015.05.004](https://doi.org/10.1016/j.bbame.2015.05.004); pmid: [25983306](https://pubmed.ncbi.nlm.nih.gov/25983306/)
- N. Noinaj, S. E. Rollauer, S. K. Buchanan, The β -barrel membrane protein insertase machinery from Gram-negative bacteria. *Curr. Opin. Struct. Biol.* **31**, 35–42 (2015). doi: [10.1016/j.sbi.2015.02.012](https://doi.org/10.1016/j.sbi.2015.02.012); pmid: [25796031](https://pubmed.ncbi.nlm.nih.gov/25796031/)
- X. C. Zhang, L. Han, How does a β -barrel integral membrane protein insert into the membrane? *Protein Cell* **7**, 471–477 (2016). doi: [10.1007/s13238-016-0273-6](https://doi.org/10.1007/s13238-016-0273-6); pmid: [27236450](https://pubmed.ncbi.nlm.nih.gov/27236450/)
- N. Noinaj, J. C. Gumbart, S. K. Buchanan, The β -barrel assembly machinery in motion. *Nat. Rev. Microbiol.* **15**, 197–204 (2017). doi: [10.1038/nrmicro.2016.191](https://doi.org/10.1038/nrmicro.2016.191); pmid: [28216659](https://pubmed.ncbi.nlm.nih.gov/28216659/)
- R. S. Bamert *et al.*, Structural basis for substrate selection by the translocation and assembly module of the β -barrel assembly machinery. *Mol. Microbiol.* **106**, 142–156 (2017). doi: [10.1111/mmi.13757](https://doi.org/10.1111/mmi.13757); pmid: [28752534](https://pubmed.ncbi.nlm.nih.gov/28752534/)
- L. Estrada Mallarino *et al.*, TlOmp85, a β -barrel assembly protein, functions by barrel augmentation. *Biochemistry* **54**, 844–852 (2015). doi: [10.1021/bi5011305](https://doi.org/10.1021/bi5011305); pmid: [25537637](https://pubmed.ncbi.nlm.nih.gov/25537637/)
- Y. Gu, Y. Zeng, Z. Wang, C. Dong, BamA β 16C strand and periplasmic turns are critical for outer membrane protein insertion and assembly. *Biochem. J.* **474**, 3951–3961 (2017). doi: [10.1042/BJC20170636](https://doi.org/10.1042/BJC20170636); pmid: [28974626](https://pubmed.ncbi.nlm.nih.gov/28974626/)
- J. Qiu *et al.*, Coupling of mitochondrial import and export translocases by receptor-mediated supercomplex formation. *Cell* **154**, 596–608 (2013). doi: [10.1016/j.cell.2013.06.033](https://doi.org/10.1016/j.cell.2013.06.033); pmid: [23911324](https://pubmed.ncbi.nlm.nih.gov/23911324/)
- T. Shiota *et al.*, Molecular architecture of the active mitochondrial protein gate. *Science* **349**, 1544–1548 (2015). doi: [10.1126/science.aac6428](https://doi.org/10.1126/science.aac6428); pmid: [26404837](https://pubmed.ncbi.nlm.nih.gov/26404837/)
- L.-S. Wenz *et al.*, Sam37 is crucial for formation of the mitochondrial TOM-SAM supercomplex, thereby promoting β -barrel biogenesis. *J. Cell Biol.* **210**, 1047–1054 (2015). doi: [10.1083/jcb.201504119](https://doi.org/10.1083/jcb.201504119); pmid: [26416958](https://pubmed.ncbi.nlm.nih.gov/26416958/)
- S. Hiller *et al.*, Solution structure of the integral human membrane protein VDAC1 in detergent micelles. *Science* **321**, 1206–1210 (2008). doi: [10.1126/science.1161302](https://doi.org/10.1126/science.1161302); pmid: [18755977](https://pubmed.ncbi.nlm.nih.gov/18755977/)
- M. Bayrhuber *et al.*, Structure of the human voltage-dependent anion channel. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15370–15375 (2008). doi: [10.1073/pnas.0808115105](https://doi.org/10.1073/pnas.0808115105); pmid: [18832158](https://pubmed.ncbi.nlm.nih.gov/18832158/)
- R. Ujwal *et al.*, The crystal structure of mouse VDAC1 at 2.3 Å resolution reveals mechanistic insights into metabolite gating. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17742–17747 (2008). doi: [10.1073/pnas.0809634105](https://doi.org/10.1073/pnas.0809634105); pmid: [18988731](https://pubmed.ncbi.nlm.nih.gov/18988731/)
- D. A. Stroud *et al.*, Biogenesis of mitochondrial β -barrel proteins: The POTRA domain is involved in precursor release from the SAM complex. *Mol. Biol. Cell* **22**, 2823–2833 (2011). doi: [10.1091/mbc.E11-02-0148](https://doi.org/10.1091/mbc.E11-02-0148); pmid: [21680715](https://pubmed.ncbi.nlm.nih.gov/21680715/)
- S. Kim *et al.*, Structure and function of an essential component of the outer membrane protein assembly machine. *Science* **317**, 961–964 (2007). doi: [10.1126/science.1143993](https://doi.org/10.1126/science.1143993); pmid: [17702946](https://pubmed.ncbi.nlm.nih.gov/17702946/)
- P. Z. Gatzeva-Topalova, T. A. Walton, M. C. Sousa, Crystal structure of YaeT: Conformational flexibility and substrate recognition. *Structure* **16**, 1873–1881 (2008). doi: [10.1016/j.str.2008.09.014](https://doi.org/10.1016/j.str.2008.09.014); pmid: [19081063](https://pubmed.ncbi.nlm.nih.gov/19081063/)
- T. J. Knowles *et al.*, Fold and function of polypeptide transport-associated domains responsible for delivering unfolded

- proteins to membranes. *Mol. Microbiol.* **68**, 1216–1227 (2008). doi: [10.1111/j.1365-2958.2008.06225.x](https://doi.org/10.1111/j.1365-2958.2008.06225.x); pmid: [18430136](https://pubmed.ncbi.nlm.nih.gov/18430136/)
49. M. P. Bos, V. Robert, J. Tommassen, Functioning of outer membrane protein assembly factor Omp85 requires a single POTRA domain. *EMBO Rep.* **8**, 1149–1154 (2007). doi: [10.1038/sj.embor.7401092](https://doi.org/10.1038/sj.embor.7401092); pmid: [18007659](https://pubmed.ncbi.nlm.nih.gov/18007659/)
 50. S. J. Habib *et al.*, The N-terminal domain of Tob55 has a receptor-like function in the biogenesis of mitochondrial β -barrel proteins. *J. Cell Biol.* **176**, 77–88 (2007). doi: [10.1083/jcb.200602050](https://doi.org/10.1083/jcb.200602050); pmid: [17190789](https://pubmed.ncbi.nlm.nih.gov/17190789/)
 51. A.-K. Pfützner *et al.*, Mitochondrial-bacterial hybrids of BamA/Tob55 suggest variable requirements for the membrane integration of β -barrel proteins. *Sci. Rep.* **6**, 39053 (2016). doi: [10.1038/srep39053](https://doi.org/10.1038/srep39053); pmid: [27982054](https://pubmed.ncbi.nlm.nih.gov/27982054/)
 52. T. Maier *et al.*, Conserved Omp85 lid-lock structure and substrate recognition in FhaC. *Nat. Commun.* **6**, 7452 (2015). doi: [10.1038/ncomms8452](https://doi.org/10.1038/ncomms8452); pmid: [26058369](https://pubmed.ncbi.nlm.nih.gov/26058369/)
 53. A.-S. Delattre *et al.*, Functional importance of a conserved sequence motif in FhaC, a prototypic member of the TpsB/Omp85 superfamily. *FEBS J.* **277**, 4755–4765 (2010). doi: [10.1111/j.1742-4658.2010.07881.x](https://doi.org/10.1111/j.1742-4658.2010.07881.x); pmid: [20955520](https://pubmed.ncbi.nlm.nih.gov/20955520/)
 54. M. Leonard-Rivera, R. Misra, Conserved residues of the putative L6 loop of *Escherichia coli* BamA play a critical role in the assembly of β -barrel outer membrane proteins, including that of BamA itself. *J. Bacteriol.* **194**, 4662–4668 (2012). doi: [10.1128/JB.00825-12](https://doi.org/10.1128/JB.00825-12); pmid: [22753067](https://pubmed.ncbi.nlm.nih.gov/22753067/)
 55. A. I. C. Höhr, S. P. Straub, B. Warscheid, T. Becker, N. Wiedemann, Assembly of β -barrel proteins in the mitochondrial outer membrane. *Biochim. Biophys. Acta* **1853**, 74–88 (2015). doi: [10.1016/j.bbamcr.2014.10.006](https://doi.org/10.1016/j.bbamcr.2014.10.006); pmid: [25305573](https://pubmed.ncbi.nlm.nih.gov/25305573/)
 56. N. W. Rigel, D. P. Ricci, T. J. Silhavy, Conformation-specific labeling of BamA and suppressor analysis suggest a cyclic mechanism for β -barrel assembly in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5151–5156 (2013). doi: [10.1073/pnas.1302662110](https://doi.org/10.1073/pnas.1302662110); pmid: [23479609](https://pubmed.ncbi.nlm.nih.gov/23479609/)
 57. N. Wiedemann *et al.*, Biogenesis of the protein import channel Tom40 of the mitochondrial outer membrane: Intermembrane space components are involved in an early stage of the assembly pathway. *J. Biol. Chem.* **279**, 18188–18194 (2004). doi: [10.1074/jbc.M400050200](https://doi.org/10.1074/jbc.M400050200); pmid: [14978039](https://pubmed.ncbi.nlm.nih.gov/14978039/)
 58. S. C. Hoppins, F. E. Nargang, The Tim8-Tim13 complex of *Neurospora crassa* functions in the assembly of proteins into both mitochondrial membranes. *J. Biol. Chem.* **279**, 12396–12405 (2004). doi: [10.1074/jbc.M313037200](https://doi.org/10.1074/jbc.M313037200); pmid: [14722057](https://pubmed.ncbi.nlm.nih.gov/14722057/)
 59. J. G. Sklar, T. Wu, D. Kahne, T. J. Silhavy, Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. *Genes Dev.* **21**, 2473–2484 (2007). doi: [10.1101/gad.1581007](https://doi.org/10.1101/gad.1581007); pmid: [17908933](https://pubmed.ncbi.nlm.nih.gov/17908933/)
 60. J. W. Fairman, N. Noinaj, S. K. Buchanan, The structural biology of β -barrel membrane proteins: A summary of recent reports. *Curr. Opin. Struct. Biol.* **21**, 523–531 (2011). doi: [10.1016/j.sbi.2011.05.005](https://doi.org/10.1016/j.sbi.2011.05.005); pmid: [21719274](https://pubmed.ncbi.nlm.nih.gov/21719274/)
 61. T. Ulrich, D. Rapaport, Biogenesis of beta-barrel proteins in evolutionary context. *Int. J. Med. Microbiol.* **305**, 259–264 (2015). doi: [10.1016/j.ijmm.2014.12.009](https://doi.org/10.1016/j.ijmm.2014.12.009); pmid: [25596888](https://pubmed.ncbi.nlm.nih.gov/25596888/)
 62. G. J. Patel, J. H. Kleinschmidt, The lipid bilayer-inserted membrane protein BamA of *Escherichia coli* facilitates insertion and folding of outer membrane protein A from its complex with Skp. *Biochemistry* **52**, 3974–3986 (2013). doi: [10.1021/bi400103t](https://doi.org/10.1021/bi400103t); pmid: [23641708](https://pubmed.ncbi.nlm.nih.gov/23641708/)
 63. E. J. Danoff, K. G. Fleming, Membrane defects accelerate outer membrane β -barrel protein folding. *Biochemistry* **54**, 97–99 (2015). doi: [10.1021/bi501443p](https://doi.org/10.1021/bi501443p); pmid: [25513891](https://pubmed.ncbi.nlm.nih.gov/25513891/)
 64. B. Schiffrin *et al.*, Effects of periplasmic chaperones and membrane thickness on BamA-catalyzed outer-membrane protein folding. *J. Mol. Biol.* **429**, 3776–3792 (2017). doi: [10.1016/j.jmb.2017.09.008](https://doi.org/10.1016/j.jmb.2017.09.008); pmid: [28919234](https://pubmed.ncbi.nlm.nih.gov/28919234/)
 65. R. Sikdar, J. H. Peterson, D. E. Anderson, H. D. Bernstein, Folding of a bacterial integral outer membrane protein is initiated in the periplasm. *Nat. Commun.* **8**, 1309 (2017). doi: [10.1038/s41467-017-01246-4](https://doi.org/10.1038/s41467-017-01246-4); pmid: [29101319](https://pubmed.ncbi.nlm.nih.gov/29101319/)
 66. N. Bonneaud *et al.*, A family of low and high copy replicative, integrative and single-stranded *S. cerevisiae*/*E. coli* shuttle vectors. *Yeast* **7**, 609–615 (1991). doi: [10.1002/yea.320070609](https://doi.org/10.1002/yea.320070609); pmid: [1767589](https://pubmed.ncbi.nlm.nih.gov/1767589/)
 67. R. S. Sikorski, P. Hieter, A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989). pmid: [2659436](https://pubmed.ncbi.nlm.nih.gov/2659436/)
 68. M. S. Longtine *et al.*, Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961 (1998). doi: [10.1002/\(SICI\)1097-0061\(199807\)14:10<953::AID-YEA293>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1097-0061(199807)14:10<953::AID-YEA293>3.0.CO;2-U); pmid: [9717241](https://pubmed.ncbi.nlm.nih.gov/9717241/)
 69. M. M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976). doi: [10.1016/0003-2697\(76\)90527-3](https://doi.org/10.1016/0003-2697(76)90527-3); pmid: [942051](https://pubmed.ncbi.nlm.nih.gov/942051/)
 70. C. Meisinger, N. Pfanner, K. N. Truscott, Isolation of yeast mitochondria. *Methods Mol. Biol.* **313**, 33–39 (2006). pmid: [16118422](https://pubmed.ncbi.nlm.nih.gov/16118422/)
 71. A. Hildebrand, M. Remmert, A. Biegert, J. Söding, Fast and accurate automatic structure prediction with HHpred. *Proteins* **77** (Suppl 9), 128–132 (2009). doi: [10.1002/prot.22499](https://doi.org/10.1002/prot.22499); pmid: [19626712](https://pubmed.ncbi.nlm.nih.gov/19626712/)
 72. B. Webb, A. Sali, Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **47**, 1–32 (2014). pmid: [25199792](https://pubmed.ncbi.nlm.nih.gov/25199792/)
 73. S. C. Lovell *et al.*, Structure validation by $\text{C}\alpha$ geometry: ϕ , ψ and $\text{C}\beta$ deviation. *Proteins* **50**, 437–450 (2003). doi: [10.1002/prot.10286](https://doi.org/10.1002/prot.10286); pmid: [12557186](https://pubmed.ncbi.nlm.nih.gov/12557186/)
 74. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). doi: [10.1107/S0907444910007493](https://doi.org/10.1107/S0907444910007493); pmid: [20383002](https://pubmed.ncbi.nlm.nih.gov/20383002/)
 75. D. Stojanovski, N. Pfanner, N. Wiedemann, Import of proteins into mitochondria. *Methods Cell Biol.* **80**, 783–806 (2007). doi: [10.1016/S0091-679X\(06\)80036-1](https://doi.org/10.1016/S0091-679X(06)80036-1); pmid: [17445722](https://pubmed.ncbi.nlm.nih.gov/17445722/)
 76. H. Schagger, G. von Jagow, Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem.* **199**, 223–231 (1991). doi: [10.1016/0003-2697\(91\)90094-A](https://doi.org/10.1016/0003-2697(91)90094-A); pmid: [1812789](https://pubmed.ncbi.nlm.nih.gov/1812789/)
 77. W. N. Burnette, “Western blotting”: Electrophoretic transfer of proteins from sodium dodecyl sulfate–polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal. Biochem.* **112**, 195–203 (1981). doi: [10.1016/0003-2697\(81\)90281-5](https://doi.org/10.1016/0003-2697(81)90281-5); pmid: [6266278](https://pubmed.ncbi.nlm.nih.gov/6266278/)

ACKNOWLEDGMENTS

We thank C. Meisinger for discussion. This work was supported by the European Research Council Consolidator Grant no. 648235, the Deutsche Forschungsgemeinschaft grants PF 202/8-1 and BE 4679/2-1, the Sonderforschungsbereiche grants 746 and 1140, and the Excellence Initiative of the German federal and state governments grants EXC 294 BIOSS and GSC-4 Spemann Graduate School). Work included in this study has also been performed in partial fulfillment of the requirements for the doctoral theses of A.I.C.H. and C.L. and the diploma thesis of A.I.C.H. at the University of Freiburg. The data presented in this paper are tabulated in the main paper and the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6373/eaah6834/suppl/DC1
Figs. S1 to S9
Tables S1 to S4
Model S1
Reference (78)

31 July 2016; resubmitted 12 October 2017
Accepted 14 December 2017
10.1126/science.aah6834

RESEARCH ARTICLE

VACCINES

Genome-wide identification of interferon-sensitive mutations enables influenza vaccine design

Yushen Du,^{1,2*} Li Xin,³ Yuan Shi,¹ Tian-Hao Zhang,^{1,4} Nicholas C. Wu,⁴ Lei Dai,¹ Danyang Gong,¹ Gurpreet Brar,¹ Sara Shu,¹ Jiadi Luo,^{1,5,6} William Reiley,⁷ Yen-Wen Tseng,¹ Hongyan Bai,³ Ting-Ting Wu,¹ Jieru Wang,^{1,5} Yuelong Shu,^{3,8} Ren Sun^{1,2,4*}

In conventional attenuated viral vaccines, immunogenicity is often suboptimal. Here we present a systematic approach for vaccine development that eliminates interferon (IFN)-modulating functions genome-wide while maintaining virus replication fitness. We applied a quantitative high-throughput genomics system to influenza A virus that simultaneously measured the replication fitness and IFN sensitivity of mutations across the entire genome. By incorporating eight IFN-sensitive mutations, we generated a hyper-interferon-sensitive (HIS) virus as a vaccine candidate. HIS virus is highly attenuated in IFN-competent hosts but able to induce transient IFN responses, elicits robust humoral and cellular immune responses, and provides protection against homologous and heterologous viral challenges. Our approach, which attenuates the virus and promotes immune responses concurrently, is broadly applicable for vaccine development against other pathogens.

Most viruses adapt rapidly to diverse selection pressures, posing a challenge for deploying safe and effective vaccines. Influenza viruses, for example, are characterized by large genetic diversity across subtypes and rapid antigenic drift and shift, which present problems for traditional vaccine strategies. Attenuation or inactivation of viruses tends to reduce the strength and breadth of immune responses, resulting in ineffective protection against antigenic alterations (1–3). Previous pandemics and recent influenza outbreaks highlight the need to develop safe vaccines that elicit effective immune responses and confer broad protection.

The type I interferon (IFN) system is the major component of innate immune responses (4–6). The IFN response provides the first line of defense against viral infections by inducing the ex-

pression of hundreds of IFN-stimulated genes (ISGs), many of which have antiviral activities (7). The IFN response is also critical for dendritic cell maturation, development of B and T cells, and memory formation, bridging innate and adaptive immunity (8–12). Most viruses have evolved to efficiently suppress the production and function of IFN to allow replication in vivo. Thus, systematic elimination of IFN-modulating functions from the virus presents a potential approach for vaccine development (fig. S1) (13, 14). Removing the most well-characterized IFN modulator in influenza virus—namely, the NS1 protein—has shown promise in a vaccine candidate (delNS1) in phase 1/2 clinical trials (14, 15). Although studies have suggested that influenza proteins other than NS1 have IFN-modulating functions (16, 17), genome-wide identification and elimination of IFN-modulating functions without affecting viral replication fitness in vitro have remained challenging tasks.

To tackle this challenge, we developed a quantitative high-throughput genomics system, which combines saturation mutagenesis and next-generation sequencing, to comprehensively identify IFN-modulating functions in the entire viral genome (18). This system has enabled us to quantitatively measure the replication capacity of a large number of mutants in parallel under specific conditions (18, 19). We performed comparative profiling of the entire influenza genome with and without IFN selection, which led to the identification of IFN-modulating functions on multiple viral segments. By combining eight IFN-sensitive mutations across the viral genome, we generated

a hyper-interferon-sensitive (HIS) virus that is replication-competent in vitro but highly attenuated in IFN-competent hosts in vivo. The HIS virus showed desired properties as a safe and effective live attenuated influenza vaccine with robust humoral and cellular responses, and it provided broad protection against homologous and heterologous viral challenges in mice and ferrets.

Fitness profile of the influenza A viral genome at single-nucleotide resolution

The eight-plasmid reverse genetic system carrying the influenza A/WSN/33 (H1N1) virus genome was used for the construction of mutant plasmid libraries (20). The mutants were divided into 52 sublibraries, each of which contained single-nucleotide mutations in a small genome region of 240 base pairs that were generated by error-prone polymerase chain reaction (fig. S2) (21–23). Viral mutant libraries were reconstituted in human embryonic kidney 293T cells by cotransfecting the plasmid encoding the sublibrary of mutants with the other seven plasmids encoding wild-type (WT) viral proteins. To systematically identify IFN-modulating functions, all viral libraries were selected in A549 cells with or without exogenous IFN treatment (IFN- α_2 at inhibitory concentration 80) (19). Illumina sequencing was used to identify each mutant and to calculate the corresponding frequency within each sublibrary. The relative fitness (RF) score of a mutant virus was calculated as the ratio of the relative frequency in the selected virus library to that in the plasmid library (Fig. 1A and table S1). There were strong correlations between biological duplicates of transfection and of selection (fig. S3). We observed a clear separation of the distribution of fitness effects between synonymous mutations and nonsense mutations (fig. S4A), indicating effective selection on virus mutants. To further validate the accuracy of the fitness profiling, we randomly selected 26 missense mutations and characterized the corresponding mutant viruses individually. The replication capacity of each mutant was highly correlated with the RF scores from the fitness profiling (fig. S4B). Using synonymous mutations as a benchmark, 50.7% of missense mutations across the whole genome were deleterious, in accordance with previous findings that single mutations are poorly tolerated in the genomes of RNA viruses (fig. S5A) (24, 25).

Systematic identification of IFN-sensitive mutations

The RF scores of most mutants are correlated in the presence and absence of exogenous IFN treatment; however, we observed a set of mutations that were nearly neutral in the absence of IFN but highly deleterious under IFN selection (Fig. 1B and fig. S6). These putative IFN-sensitive mutations were widespread on multiple viral segments. Among all influenza A viral proteins, NS1 has been extensively studied for its interaction with the IFN pathway (19, 26, 27), which is validated both in our fitness profiling and individually constructed NS1 mutant viruses (fig. S7).

¹Department of Molecular and Medical Pharmacology, University of California, Los Angeles, CA 90095, USA. ²Cancer Institute, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, School of Medicine, Zhejiang University, Hangzhou 310058, China. ³National Institute for Viral Disease Control and Prevention, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Chinese Center for Disease Control and Prevention, Key Laboratory for Medical Virology and Viral Diseases, Ministry of Health of the People's Republic of China, Beijing 102206, China. ⁴Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA. ⁵Department of Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, PA 15224, USA. ⁶Department of Pathology, The Second Xiangya Hospital of Central South University, Changsha, Hunan 410005, China. ⁷Trudeau Institute, Saranac Lake, NY 12983, USA. ⁸School of Public Health (Shenzhen), Sun Yat-sen University, Guangdong 510275, China.
*Corresponding author. Email: lily.duyushen@gmail.com (Y.D.); rsun@mednet.ucla.edu (R.S.)

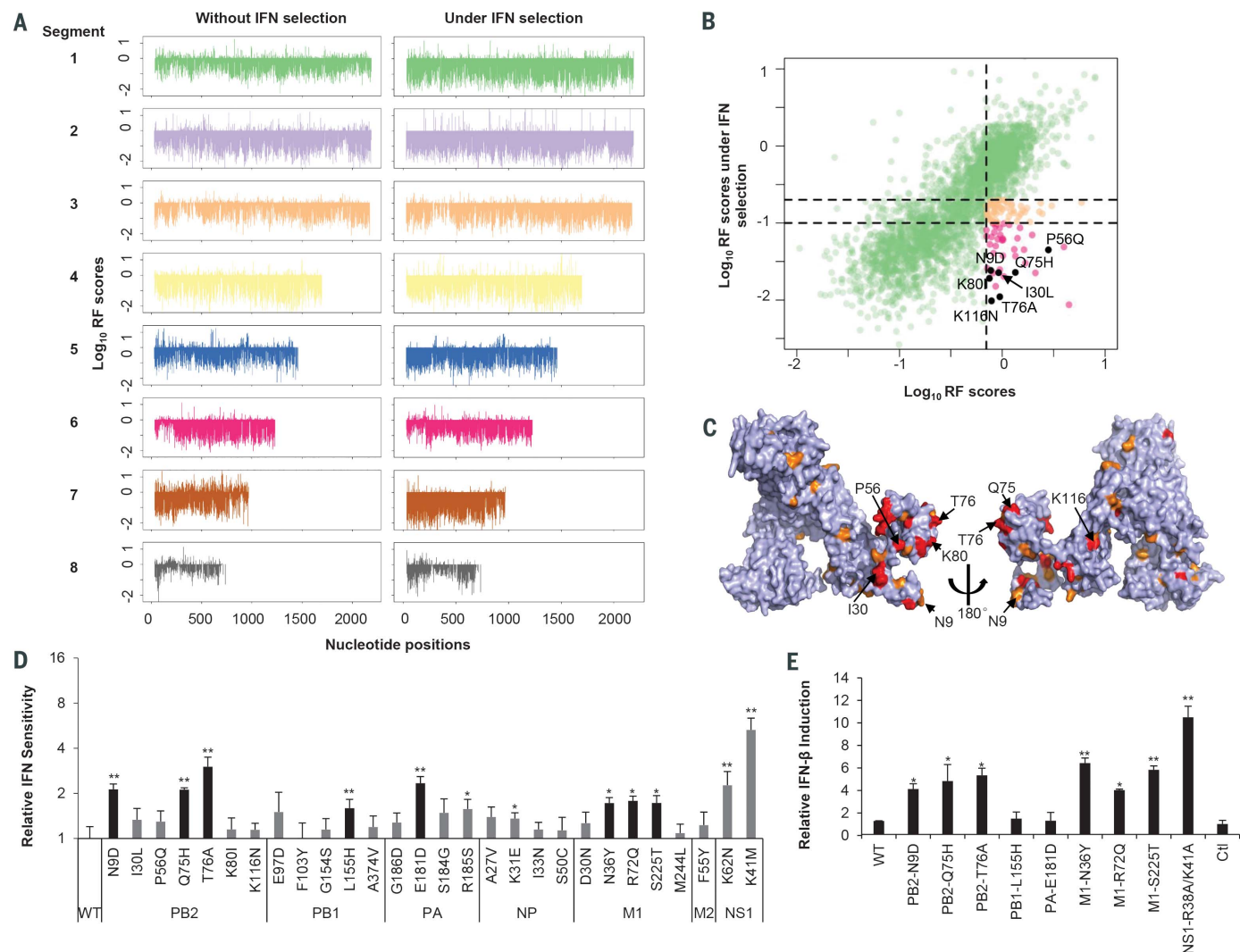


Fig. 1. Identification of IFN-sensitive mutations using quantitative high-throughput genomics. (A) Relative fitness (RF) scores for individual mutations in A549 cells with (right) and without (left) IFN selection across the influenza A/WSN/33 genome. (B and C) Identification of IFN-sensitive mutations with PB2 protein as an example [Protein Data Bank (PDB) ID, 4WSB] (38, 39). Red and orange represent strong and intermediate IFN sensitivity, respectively. (D) Validation of IFN sensitivity with individually reconstituted mutants ($n = 4$). The top eight mutations on nonsurface virion

proteins are shown in black. (E) Induction of IFN- β expression in A549 cells infected with WT virus or indicated mutants at 6 hours post-infection, with mock infection as control (Ctl) ($n = 3$). Error bars, SD. $*P < 0.05$, $**P < 0.01$ [two-tailed t test compared with WT (D) or with Ctl (E)]. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

To further explore IFN-modulating functions across the genome, we focused on IFN-sensitive mutations outside NS1, especially the solvent-exposed and structurally clustered residues in the polymerase complex (PB2, PB1, PA, and NP), as well as the M1 and M2 proteins (Fig. 1C and fig. S6). Twenty-six mutations were constructed individually, most of which were nearly neutral for viral replication with nearly intact polymerase activity (fig. S8). These included the previously characterized mutations PB2-N9D, which is known to counteract the inhibition of MAVS (mitochondrial antiviral signaling protein)-induced IFN- β production by PB2 (16), and M1-D30N, which has

been shown to induce IFN- β production (17). Several mutations significantly increased IFN sensitivity compared with WT, and the top eight were chosen for further characterization (Fig. 1D). Six of them (PB2-N9D, PB2-Q75H, PB2-T76A, M1-N36Y, M1-R72Q, and M1-S225T) elevated the expression of IFN- β and ISG54 (Fig. 1E and fig. S9) and stimulated nuclear translocation of IRF3 (fig. S10). We also observed that the IFN induction was MAVS-dependent and STING (stimulator of interferon genes)-independent (fig. S11). Moreover, these six mutants were not sensitive to IFN treatment in Vero cells, which are deficient in IFN production. However, the other two mutations (PB1-L155H and

PA-E181D) did not induce higher IFN production (Fig. 1E) and were still IFN-sensitive in Vero cells, suggesting that these mutants likely affect processes downstream of IFN production.

Combining mutations increases IFN sensitivity and IFN induction in vitro

To maximize IFN sensitivity and IFN induction, we combined three IFN-inducing mutations on PB2 (N9D, Q75H, and T76A), three on M1 (N36Y, R72Q, and S225T), and two previously reported ones on NS1 (R38A and K41A) to create the HIS virus. The growth of HIS virus in IFN-competent A549 cells showed significant attenuation compared

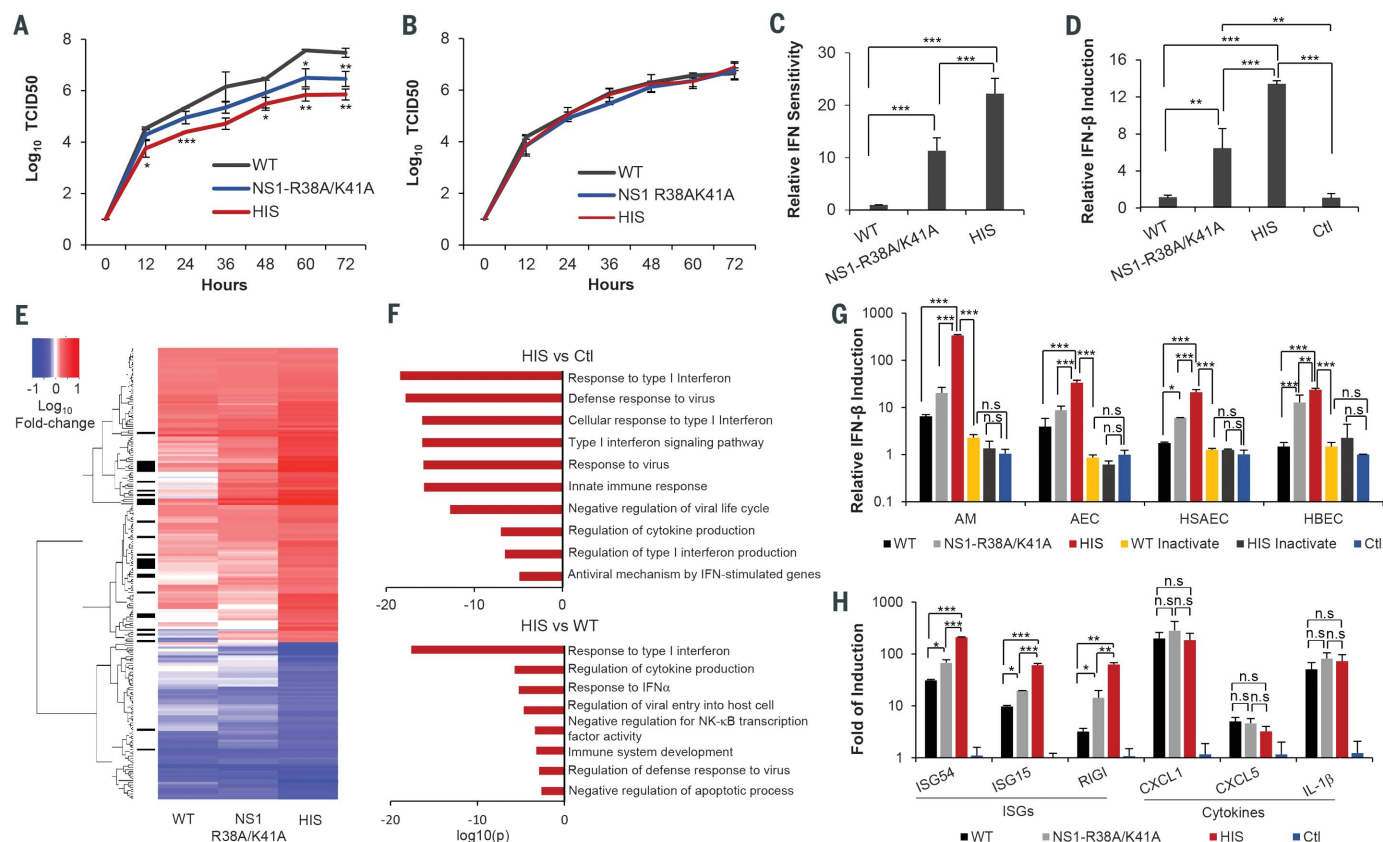


Fig. 2. The combination of mutations in HIS virus increases IFN sensitivity and IFN production. (A and B) Replication kinetics of WT, NS1 mutant, and HIS viruses in A549 (A) and Vero (B) cells. (C) IFN sensitivity of WT, NS1 mutant, and HIS viruses ($n = 4$). (D) Induction of IFN- β expression by indicated virus in A549 cells at 6 hours post-infection, with mock infection as Ctl ($n = 3$). (E) Global gene expression in A549 cells infected with indicated viruses was examined by RNA sequencing ($n = 2$). The heatmap shows the genes that were significantly differentially expressed in HIS-infected cells compared with mock-infected cells. IFN response genes are marked on the left with black bars. (F) GO enrichment analysis

of genes up-regulated in HIS-infected cells in comparison with mock- (top) or WT-infected (bottom) cells. (G) Induction of IFN- β expression by indicated viruses in primary human alveolar macrophages (AMs), human alveolar epithelial cells (AECs), human small airway epithelial cells (HSAECs), and human bronchial epithelial cells (HBECs) at 6 hours post-infection, with mock infection as Ctl ($n = 3$). (H) Induction of indicated ISGs and inflammatory cytokines in primary human AMs at 6 hours post-infection ($n = 3$). Error bars, SD. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ [one-way analysis of variance (ANOVA) with Bonferroni multiple comparisons test]; n.s., not significant.

with that of WT virus (1.4-log decrease at 36 hours and 1.8-log decrease at 60 hours) but was fully restored in IFN-deficient Vero cells (Fig. 2, A and B). The IFN sensitivity of HIS virus was significantly higher than that of the NS1-R38A/K41A mutant, indicating an independent effect of mutations on PB2 and M1 (Fig. 2C). Gene expression data from lung epithelial and macrophage cell lines (A549 and THP1) showed that HIS virus induced higher IFN production and responses (Fig. 2D and fig. S12, A to C). Using RNA sequencing, we evaluated the global gene expression changes in A549 cells infected with WT, NS1-R38A/K41A, or HIS virus. At 6 hours post-infection, the expression of 120 genes was significantly up-regulated (fold change > 2 and $P < 0.001$) in HIS-infected cells, of which 24% were IFN response genes (Fig. 2E, fig. S12D, and table S2). Gene Ontology (GO) enrichment analysis revealed that the pathways related to IFN production and response were the dominant ones activated by HIS virus, to a greater extent than by WT or

mock infection (Fig. 2F). Furthermore, HIS virus induced negative regulators of apoptosis process, such as TNFAIP3, an important inhibitor of TNF-mediated apoptosis. Slower cell death was observed with HIS infection than with WT infection (fig. S12E).

We further defined the phenotypes of HIS virus with a panel of human lung cells, including immortalized small airway epithelial cells, bronchial epithelial cells, primary alveolar epithelial cells, and primary alveolar macrophages (Fig. 2G). HIS virus induced the strongest up-regulation of IFN- β expression (~50-fold relative to WT) in the primary alveolar macrophages, an important target for influenza infection (Fig. 2G), and greater up-regulation of ISGs than WT virus (Fig. 2F). HIS virus did not enhance the expression of other inflammatory cytokines [CXCL1, CXCL5, or interleukin-1 β (IL-1 β)] in the infected macrophages, highlighting its specific effects on the IFN pathway (Fig. 2H). The phenotype of HIS virus is not limited to the WSN background: Introducing these

eight mutations into another H1N1 strain of influenza, A/PR8/34 (PR8-HIS), led to a similar phenotype (fig. S12, F and G). The up-regulation of the IFN pathway requires active viral infection, given that formalin-inactivated HIS virus lost the ability to induce higher IFN- β expression (Fig. 2G).

HIS virus is highly attenuated in IFN-competent mice and ferrets

We next measured the replication and pathogenesis of HIS virus in mice and ferrets, the most commonly used animal models for influenza virus. BALB/c mice were intranasally inoculated with WT or HIS virus at different doses. Whereas the median lethal dose of WT virus was 5×10^5 TCID₅₀ (50% tissue culture infective dose), and 1×10^3 TCID₅₀ caused obvious body weight loss in all animals, neither weight loss nor indicative clinical symptoms were observed in HIS-infected mice given 1×10^7 TCID₅₀, the highest dose that we have tested (Fig. 3, A and B). To compare the HIS

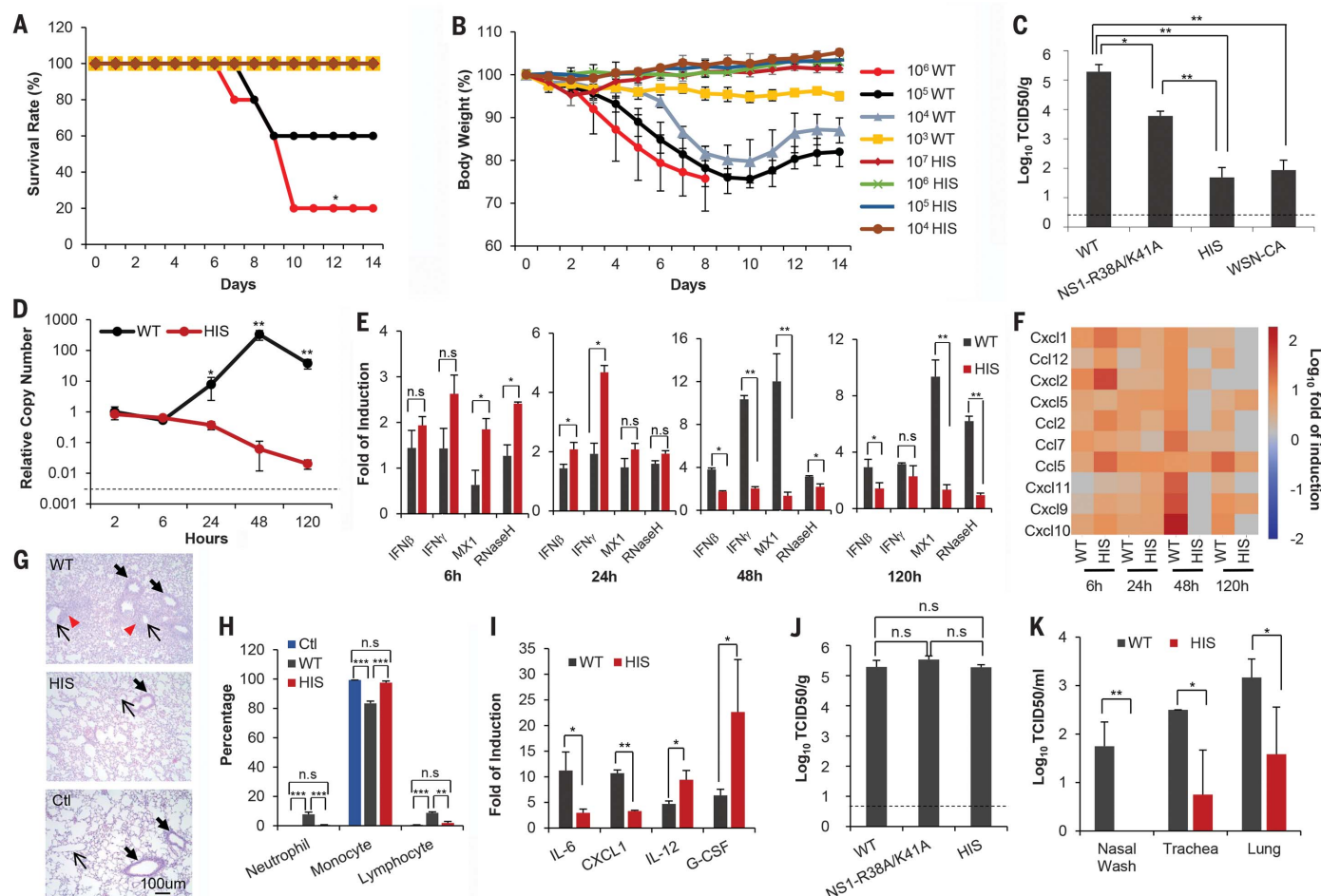


Fig. 3. HIS virus is replication-deficient in vivo and induces a transient IFN response. (A and B) Survival rate and percentage of body weight loss after intranasal infection ($n = 5$). (C and D) Viral titers at day 2 post-infection ($n = 4$) (C) and replication kinetics ($n = 3$) (D) of WT and HIS viruses in mouse lung tissues. (E) Induction of indicated ISGs in mouse lung tissues at 6, 24, 48, and 120 hours (h) post-infection ($n = 3$), shown as fold of induction over mock infection. RNase H, ribonuclease H. (F) Gene expression of indicated inflammatory cytokines in mouse lung tissues was examined by RNA sequencing ($n = 2$). (G) HE (hematoxylin and eosin) staining of lung tissues at

day 9 post-infection. Thick arrows, bronchioles; thin arrows, vessels; red triangles, inflammatory cell infiltration. (H) Percentage of neutrophils, monocytes, and lymphocytes in BAL cytopins at day 9 post-infection ($n = 3$). (I) Cytokines in BAL samples measured by Luminex multiplex assay ($n = 4$). (J) Replication of indicated viruses in lung tissues of IFNAR^{-/-} mice ($n = 4$). (K) Viral titer of WT and HIS viruses in ferret nasal wash, trachea, and lung tissues ($n = 3$). Dashed lines represent detection limits. Error bars, SD. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ [log-rank test for (A); ANOVA with Bonferroni multiple comparisons test for (C), (H), and (J); and two-tailed t test for (D), (E), (I), and (K)].

virus approach with the live attenuated vaccine strategy used in FluMist, we incorporated the five cold-adapted (CA) mutations from FluMist into the WSN background and generated a WSN-CA virus (28, 29). WSN-CA virus replicated well at 33°C but was highly attenuated at 39°C and induced IFN- β expression to a similar level as WT virus, which was significantly lower than that induced by HIS virus (fig. S13). By day 2 post-inoculation, replication of HIS virus in mouse lung tissues was significantly lower than that of WT virus (~ 3.6 -log decrease) or the NS1-R38A/K41A mutant (~ 2 -log decrease) and comparable to that of WSN-CA virus (Fig. 3C and fig. S14A). In contrast with the robust viral replication observed for WT infection, which peaked at 48 hours, no increase in viral copy number was detected in HIS-infected mice at any tested time point (Fig. 3D). PR8-HIS virus was also significantly attenuated com-

pared with WT PR8 virus in mouse lung tissues (fig. S14B). Although highly attenuated in replication, HIS virus showed transient yet significant up-regulation of IFN and ISGs at 6 and 24 hours post-infection, after which the response was diminished (Fig. 3E). In contrast, WT virus induced a robust pro-inflammatory response throughout the course of infection, exemplified by the high induction of CXCL10 at 48 and 120 hours post-infection (Fig. 3F). These results correlate well with histological analysis of infected lungs and cytopins of bronchoalveolar lavage (BAL) fluid (Fig. 3, G and H, and fig. S14, C to G). HIS-infected lungs showed infiltration of neutrophils and lymphocytes at day 2 post-infection; however, the infiltration was transient and cleared by day 9. Sustained inflammation and tissue damage was observed for WT-infected lungs, which became more severe by day 9 post-infection (Fig. 3H and

fig. S14, H and I). We also examined the cytokine response in the BAL samples at 48 hours post-infection by means of Luminex multiplex assay (Fig. 3I and fig. S14, J and K). WT infection showed significantly higher levels of IL-6 and CXCL1, consistent with the observed severe inflammation. In contrast, HIS virus induced higher amounts of IL-12 and G-CSF, which is important for granulocyte stimulation and T cell development. Furthermore, replication of HIS virus was fully restored to WT levels in IFNAR^{-/-} mice, indicating that the inability to counteract IFN response was the underlying mechanism for the highly attenuated replication of HIS virus in wild-type mice (Fig. 3J). In the ferret model, we also observed significant attenuation of HIS virus (Fig. 3K). By day 3 post-infection, HIS virus showed a ~ 2 -log decrease in trachea and a ~ 1.5 -log decrease in lung tissues compared with WT virus. Moreover, no infectious

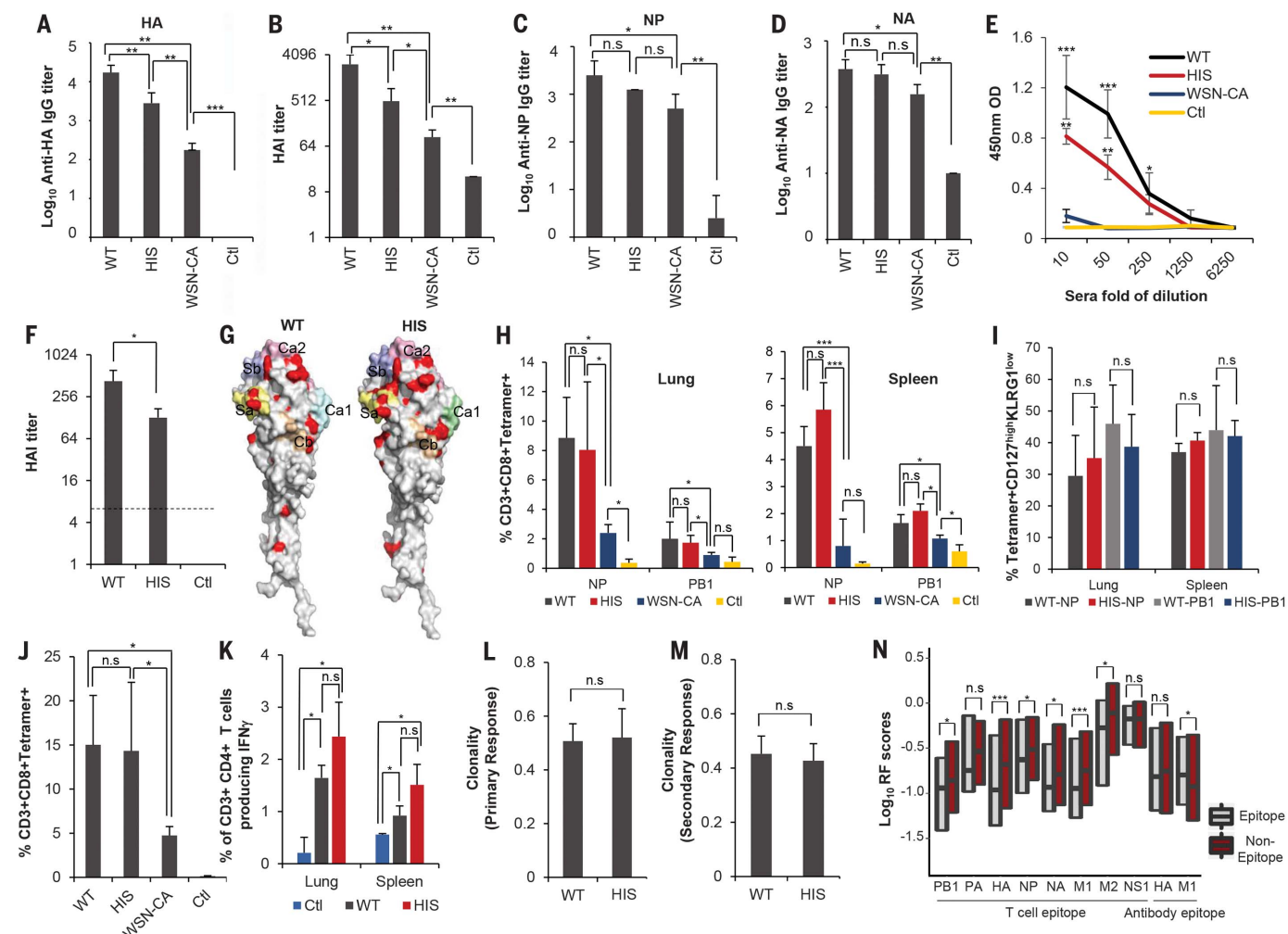


Fig. 4. HIS virus induces strong adaptive immune responses in mice and ferrets. (A to D) HA-binding IgG ($n = 7$), HA neutralizing antibody ($n = 7$), and NP- and NA-binding IgG ($n = 4$) in mouse sera at day 28 post-vaccination. HAI, hemagglutinin inhibition. (E) HA-binding IgA in BAL samples at day 28 post-vaccination ($n = 4$). The optical density (OD) in ELISA was 450 nm. (F) HA neutralizing antibody levels in ferret sera at day 22 post-vaccination ($n = 3$). The dashed line represents the detection limit. (G) Mutations not neutralized by mouse sera (red) were mapped onto the HA structure (PDB ID, 1RUZ; $n = 5$) (40). The other five colors represent five well-characterized neutralization epitopes. (H) Tetramer staining of antigen-specific CD8 T cells in mouse lung (left) and spleen (right) at day 10 post-vaccination ($n = 10$).

(I) Percentage of antigen-specific memory precursor effector cells in mouse lung and spleen ($n = 3$). (J) NP antigen-specific CD8 T cells during the secondary responses in lung tissues from mice vaccinated with indicated viruses ($n = 4$). (K) Intracellular IFN- γ staining of CD4 T cells induced by the indicated viruses ($n = 3$). (L and M) Clonality of TCR β sequences of NP antigen-specific CD8 T cells during the primary ($n = 5$) or secondary ($n = 4$) (M) responses. (N) Box plots show the fitness distribution of mutations on T cell epitopes or antibody epitopes. Error bars, SD. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ [ANOVA with Bonferroni multiple comparisons test for (A) to (D), (F), (H), (J), and (K); two-tailed t test for (I), (L), and (M); and Wilcoxon rank sum test for (N)].

viral particles were detected in nasal washes of HIS-infected ferrets, in contrast to the robust viral shedding observed during WT infection.

HIS virus induces strong and broad adaptive immune responses

We then examined the ability of the HIS virus to induce humoral and cellular responses. Mouse sera and BAL samples were collected at day 28 after single-dose (1×10^4 TCID₅₀) vaccination with WT, HIS, or WSN-CA virus. HIS virus induced robust antibody responses, as measured by ELISA (enzyme-linked immunosorbent assay) and hemagglutinin (HA) inhibition and neutralization antibody assays (Fig. 4, A to E, and fig. S15). The level of HA antibody responses elicited by HIS

virus was lower than for WT virus, yet significantly higher than for the WSN-CA, inactivated WT, and inactivated HIS viruses (Fig. 4, A and B, and fig. S15, C and D). Immunoglobulin G (IgG) antibodies against NP, NA, and M1 proteins, which have been shown to play an important role in limiting viral replication (30, 31), were also detected in the sera of HIS-vaccinated mice at a level comparable to that in WT-infected mice (Fig. 4, C and D, and fig. S15E). Furthermore, mucosal immune responses, indicated by secretory IgA antibodies against HA and NP proteins, were elicited by HIS vaccination (Fig. 4E and fig. S15F). Robust HA antibody responses were also observed in ferrets vaccinated with HIS virus (Fig. 4F and fig. S15G), which were sustained for at least 50 days post-

vaccination. To examine the epitope coverage of the neutralizing antibodies generated by HIS virus, we profiled the HA mutants in the presence or absence of mouse serum antibodies by using the high-throughput genomic approach (32). Mutations not neutralized by sera were observed in both head (Ca2 and Sa sites) and stem regions, with no significant difference in the number or the distribution of mutations between the WT and HIS viruses (Fig. 4G, fig. S16, and table S3). This suggests that the breadth and diversity of neutralizing antibodies induced by the HIS virus are comparable to those induced by the WT virus.

In addition to humoral responses, HIS virus elicited NP and PB1 antigen-specific CD8 T cell responses, similarly to WT virus and much more

strongly than the WSN-CA, inactivated WT, and inactivated HIS viruses (Fig. 4H and fig. S17, A to D). The CD8 T cells induced by the WT and HIS viruses had a similar capacity for IFN- γ production upon stimulation by viral epitope peptides (fig. S17E). We further examined the phenotypes of virus-specific T cells by quantifying the expression of KLRG1, CD127, CD44, CD62L, and CCR7. By day 21 post-infection, the NP and PB1 antigen-specific CD8 T cells induced by the WT and HIS viruses displayed similar levels of memory precursor effector cells with a CD127^{high}KLRG1^{low} phenotype and short-lived effector cells with a CD127^{low}KLRG1^{high} phenotype (Fig. 4I and fig. S17F). These virus-specific CD8 T cells also displayed a similar effector/memory phenotype, as measured by CD62L, CD44, and CCR7 expression (fig. S17, G and H). Consistently, after challenge infection at 1 month post-vaccination, HIS virus induced the secondary CD8 T cell responses similarly to WT but more strongly than WSN-CA virus (Fig. 4J and fig. S17I). Moreover, similar frequencies of influenza-specific CD4 T cells were elicited by the WT and HIS viruses (Fig. 4K). To examine the diversity of the primary and secondary T cell responses, we analyzed the T cell receptor repertoire by sequencing the β T cell receptor (TCR β) loci of NP-specific CD8 T cells in mice vaccinated with WT or HIS virus. The V β usage and clonality for both primary and secondary T cell responses were comparable between the WT and HIS viruses, documenting the diversity of T cell lineages induced by HIS vaccination (Fig. 4, L and M, and fig. S18).

We analyzed the potential impact of immune responses on the viral genome at the population level. Our whole-genome fitness profiling provides a data set for examining the genetic flexibility of viral sequences. We calculated the fitness cost of mutations in the previously identified B and T cell epitopes. Mutations on several T cell epitopes, but not on antibody epitopes, were generally correlated with lower fitness scores (Fig. 4N and table S4). Our results suggest that an escape from T cell selection will impose a higher fitness cost for the virus, and thus T cell responses will be effective against vaccine escape.

HIS virus protects against homologous and heterologous viral challenge

We examined whether HIS vaccination could offer protection against homologous and heterologous viral challenges. Immunized mice were challenged 28 days post-vaccination with 1×10^4 TCID₅₀ of WT virus. Vaccination by HIS virus reduced viral replication by ~ 3 log, with no sign of weight loss (Fig. 5 and fig. S19). Complete protection without detectable viral titers in the lung was achieved with one vaccination at a high dose (1×10^6 TCID₅₀) or two vaccinations at a low dose (1×10^4 TCID₅₀) (Fig. 5B and fig. S19B). Similar protective effects were observed in ferrets, which were challenged with 1×10^7 TCID₅₀ of WT virus at day 35 post-vaccination. Nasal washes were collected at days 1, 3, 4, 7, and 9 post-challenge, and no infectious viral particles were detected in nasal washes from HIS-vaccinated ferrets throughout this time period (Fig. 5C).

To test whether HIS vaccination provides protection against heterologous strains, we first challenged immunized mice with PR8 virus and examined viral titer at day 2 post-challenge. HIS vaccination reduced viral titer by ~ 3 log compared with mock vaccination and significantly more than WSN-CA vaccination (fig. S19C). We further challenged vaccinated mice with a lethal dose of three different influenza strains: H1N1 subtypes A/PR8/34 and A/Cal/04/09 and H3N2 subtype A/X-31. Protection by HIS vaccination was observed in all measures, including survival rate, percentage of body weight loss, and clinical scores (Fig. 5, D and E, and fig. S19D). Strong secondary antigen-specific T cell responses were observed in the challenged mice for all strains (fig. S20). HIS vaccination also protected ferrets from heterologous A/Cal/07/09 challenge, as shown by viral titer in nasal washes and percentage of body weight loss (Fig. 5F and fig. S19E).

Discussion

Conventional approaches to develop vaccines render the virus avirulent but also reduce immunogenicity. We developed a quantitative high-throughput genomics approach to systematically identify and eliminate immune-modulating functions

in the virus genome while maintaining replication fitness in vitro. This is a systems-based strategy to enhance viral immunogenicity while attenuating replication and pathogenesis. In this proof-of-principle study, we generated a HIS virus with a combination of eight IFN-sensitive mutations. These mutations also induced higher IFN production and response. We demonstrated that HIS virus is highly attenuated in vivo but is able to induce transient IFN responses, elicit robust and diverse humoral and cellular immunity, and provide protection against homologous and heterologous viral challenges in mice and ferrets.

Recent studies have suggested several strategies to design live attenuated vaccines (14, 15, 33–37). Our method is distinctive in the following aspects: (i) We systematically investigated the whole viral genome, and we eliminated immune-evasion functions at multiple loci to obtain a safe strain that has no detectable replication in vivo; (ii) we selected mutants that induce a higher IFN response, because a transient IFN response has been shown to be essential for adaptive immunity, including the strong and diverse T cell responses; (iii) HIS virus selectively induced a transient IFN response but no other tested inflammatory responses, which reduced potential pathogenesis or side effects for future

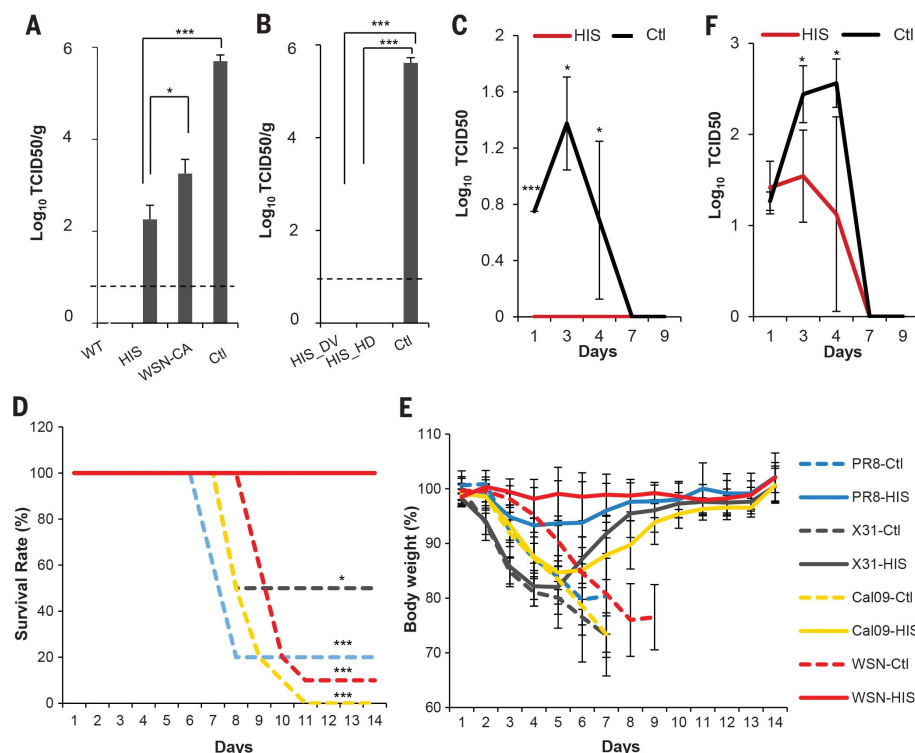


Fig. 5. HIS virus protects mice and ferrets from broad viral challenges. (A and B) Viral load in mouse lung tissues at day 2 post-challenge ($n = 4$). DV, double vaccinations with HIS virus at 1×10^4 TCID₅₀, 28 days apart; HD, high-dose vaccination with HIS virus at 1×10^6 TCID₅₀. Dashed lines represent detection limits. (C) Viral replication kinetics in ferret nasal wash after WSN virus challenge at day 35 post-vaccination ($n = 3$). (D and E) Survival rate and body weight loss of HIS-vaccinated mice after challenge with homologous and heterologous strains ($n = 10$). (F) Viral replication kinetics in ferret nasal wash after A/CALIFORNIA/07/09 virus challenge at 35 days post-vaccination ($n = 3$). Error bars, SD. * $P < 0.05$, *** $P < 0.001$ [ANOVA with Bonferroni multiple comparisons test for (A) and (B), two-tailed t test for (C) and (F), and log-rank test for (D)].

clinical usage. We have also applied this approach to a DNA virus and generated an effective vaccine candidate.

In general, this unbiased and quantitative high-throughput genomics system can be widely applied to other pathogens to define the impact of genome-wide mutations under certain selection conditions. Similar profiling of a viral genome can be performed with other immune components, such as cytokines, natural killer cells, or T cells, in vitro and in vivo. Inactivating additional immune evasion functions in the virus will further increase the safety and immunogenicity of its derivatives for prevention or therapy.

REFERENCES AND NOTES

1. M. T. Osterholm, N. S. Kelley, A. Sommer, E. A. Belongia, *Lancet Infect. Dis.* **12**, 36–44 (2012).
2. A. C. Tricco *et al.*, *BMC Med.* **11**, 153 (2013).
3. M. Darvishian, M. J. Bijlsma, E. Hak, E. R. van den Heuvel, *Lancet Infect. Dis.* **14**, 1228–1239 (2014).
4. A. García-Sastre, *Virus Res.* **162**, 12–18 (2011).
5. J. M. González-Navajas, J. Lee, M. David, E. Raz, *Nat. Rev. Immunol.* **12**, 125–135 (2012).
6. A. Iwasaki, P. S. Pillai, *Nat. Rev. Immunol.* **14**, 315–328 (2014).
7. J. W. Schoggins *et al.*, *Nature* **472**, 481–485 (2011).
8. J. P. Huber, J. D. Farrar, *Immunology* **132**, 466–474 (2011).
9. A. Le Bon *et al.*, *Immunity* **14**, 461–470 (2001).
10. R. M. Welsh, K. Bahl, H. D. Marshall, S. L. Urban, *PLOS Pathog.* **8**, e1002352 (2012).
11. A. Le Bon, D. F. Tough, *Curr. Opin. Immunol.* **14**, 432–436 (2002).
12. J. Crouse, U. Kalinke, A. Oxenius, *Nat. Rev. Immunol.* **15**, 231–242 (2015).
13. F. Krammer, P. Palese, *Nat. Rev. Drug Discov.* **14**, 167–182 (2015).
14. J. Talon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4309–4314 (2000).
15. C. Mössler *et al.*, *Vaccine* **31**, 6194–6200 (2013).
16. K. M. Graef *et al.*, *J. Virol.* **84**, 8433–8445 (2010).
17. M. Pérez-Cidoncha *et al.*, *J. Virol.* **88**, 4632–4646 (2014).
18. N. C. Wu *et al.*, *Sci. Rep.* **4**, 4942 (2014).
19. N. C. Wu *et al.*, *J. Virol.* **88**, 10157–10164 (2014).
20. E. Hoffmann, G. Neumann, Y. Kawaoka, G. Hobom, R. G. Webster, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6108–6113 (2000).
21. Y. Du *et al.*, *MBio* **7**, e01801-16 (2016).
22. N. C. Wu *et al.*, *PLOS Genet.* **11**, e1005310 (2015).
23. N. C. Wu *et al.*, *BMC Genomics* **17**, 46 (2016).
24. S. F. Elena, P. Carrasco, J.-A. Daròs, R. Sanjuán, *EMBO Rep.* **7**, 168–173 (2006).
25. E. Visher, S. E. Whitefield, J. T. McCrone, W. Fitzsimmons, A. S. Llaure, *PLOS Pathog.* **12**, e1005856 (2016).
26. B. G. Hale, R. A. Albrecht, A. García-Sastre, *Future Microbiol.* **5**, 23–41 (2010).
27. B. G. Hale, R. E. Randall, J. Ortin, D. Jackson, *J. Gen. Virol.* **89**, 2359–2376 (2008).
28. H. F. Maassab, M. L. Bryant, *Rev. Med. Virol.* **9**, 237–244 (1999).
29. H. Jin, H. Zhou, B. Lu, G. Kemble, *J. Virol.* **78**, 995–998 (2004).
30. M. J. Memoli *et al.*, *MBio* **7**, e00417–e16 (2016).
31. D. M. Carragher, D. A. Kaminski, A. Moquin, L. Hartson, T. D. Randall, *J. Immunol.* **181**, 4168–4176 (2008).
32. M. B. Doud, S. E. Hensley, J. D. Bloom, *PLOS Pathog.* **13**, e1006271 (2017).
33. S. Mueller *et al.*, *Nat. Biotechnol.* **28**, 723–726 (2010).
34. L. Si *et al.*, *Science* **354**, 1170–1173 (2016).
35. C.-Y. Wu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 280–285 (2017).
36. L. Wang *et al.*, *Cell Host Microbe* **21**, 334–343 (2017).
37. J. Steel *et al.*, *J. Virol.* **83**, 1742–1753 (2009).
38. A. Pflug, D. Guilleigay, S. Reich, S. Cusack, *Nature* **516**, 355–360 (2014).
39. S. Reich *et al.*, *Nature* **516**, 361–366 (2014).
40. S. J. Gamblin *et al.*, *Science* **303**, 1838–1842 (2004).

ACKNOWLEDGMENTS

We greatly appreciate the communication with A. te Velthuis and E. Fodor, who found that the same and similar mutations in PB2 produce mini viral RNAs (mvRNAs), which bind to RIG-I and induce IFN production. Using their method, we were able to detect these mvRNAs produced by our PB2 mutants. We thank Q. Zhou, A. York, and S. Bensinger for THP1 cells with specific gene knockouts and for discussions. We thank S. Park and S. Dubinett for the immortalized HSAECs and HBECs. Support was provided by the Whitcome Fellowship, the Burroughs Wellcome Fund, and the NIH (grants CA177322 and DE023591) to Y.D. and R.S.; the NIH (grant HL113655) to J.W.; the National Science Fund for Distinguished Young Scholars (grant 81525017) to Y. Shu; and National Science and Technology Major Project (grant 2015ZX09101044) to L.X. The sequencing data are deposited in the NIH Sequence Read Archive with accession numbers PRJNA383938, PRJNA254185, PRJNA318707, and PRJNA285135.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6373/290/suppl/DC1
Materials and Methods
Figs. S1 to S20
Tables S1 to S4
References (41–70)

1 June 2017; accepted 15 November 2017
10.1126/science.aan8806

NANOROBOTICS

A self-assembled nanoscale robotic arm controlled by electric fields

Enzo Kopperger,^{1*} Jonathan List,^{1*} Sushi Madhira,² Florian Rothfischer,¹
Don C. Lamb,^{2,3,4} Friedrich C. Simmel^{1,4,†}

The use of dynamic, self-assembled DNA nanostructures in the context of nanorobotics requires fast and reliable actuation mechanisms. We therefore created a 55-nanometer-by-55-nanometer DNA-based molecular platform with an integrated robotic arm of length 25 nanometers, which can be extended to more than 400 nanometers and actuated with externally applied electrical fields. Precise, computer-controlled switching of the arm between arbitrary positions on the platform can be achieved within milliseconds, as demonstrated with single-pair Förster resonance energy transfer experiments and fluorescence microscopy. The arm can be used for electrically driven transport of molecules or nanoparticles over tens of nanometers, which is useful for the control of photonic and plasmonic processes. Application of piconewton forces by the robot arm is demonstrated in force-induced DNA duplex melting experiments.

Nanoscale robotic systems will enable the programmable synthesis and assembly of molecular materials from the bottom up. Components of such systems were previously created with the tools of supramolecular chemistry (1–4) and bionanotechnology (5). In particular, DNA self-assembly (6, 7) has been used successfully to create nanoscale robotic walkers (8–13), assembly lines (14), movable molecular arms (15–18), and molecular mechanisms (19, 20). However, as a result of being driven by DNA hybridization reactions (8–10, 13–16, 18, 19), deoxyribozyme (11) or enzyme (12) action, changes in buffer composition, or using photoswitchable components (17), these systems were very slow, had a low assembly or operation yield, or were unable to exert appreciable forces against external loading. In one of the most successful methodologies (21, 22), DNA machines are driven through their operation cycle by hybridization with fuel and antifuel strands using toehold-mediated strand displacement reactions. Although this approach has the advantage of sequence addressability, DNA hybridization and strand-exchange reactions are slow, and structural switching often occurs with low yield. In our experiment, we deliberately abandoned sequence-specific switching and used electrical fields to move the components of a DNA machine with respect to each other. We thus gain many orders of magnitude in operation speed, almost perfect switch-

ing yield, and the capability of computer-controlled nanoscale motion and positioning.

A DNA-based molecular platform with an integrated robotic arm

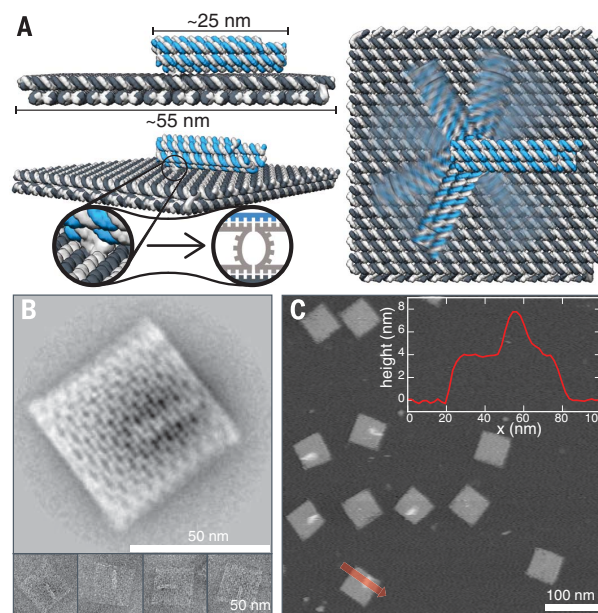
The actuator unit of our system is composed of a 55-nm-by-55-nm DNA origami plate with an integrated 25-nm-long arm defined by a DNA six-helix bundle (6HB) (Fig. 1A), allowing for a high-yield, one-pot folding procedure. For the rigid DNA plate, we used a crossed two-layer scaffold routing in which the top layer is rotated by 90° with respect to the bottom layer (supplementary materials and methods) (23). The 6HB,

functioning as the robot arm, is connected to the top layer of the base plate via a flexible joint created by two adjacent scaffold crossovers with three and four unpaired bases, respectively (see supplementary text section on the design of the joint) (23). Successful assembly of the structure with ~90% yield was verified using transmission electron microscopy (TEM) and atomic force microscopy (AFM) (Fig. 1, B and C, and fig. S1) (23). Consistent with our design, AFM indicates a height of 4 nm for the base plate and an additional 4 nm for the 6HB arm.

We first used single-molecule multicolor Förster resonance energy transfer (FRET) experiments to investigate diffusive motion of the arm with respect to the base plate (Fig. 2). For these experiments, we extended two staple strands on opposite sides of the plate with an identical short docking sequence, whereas a staple strand on the arm was extended with the complementary sequence. Transient binding of the arm results in stochastic switching between the two docking sites, which we observed with the help of three reporter dyes: a FRET donor at the tip of the arm and two different acceptor dyes at the docking sites (Fig. 2A). A typical trace of stochastically alternating FRET signals is shown in Fig. 2B. Upon donor excitation, a high donor fluorescence (blue) indicates a freely diffusing arm, whereas a high acceptor fluorescence (green or red) indicates docking at the respective site. Dwell times for the three states were extracted from fluorescence traces of more than 1000 robot-arm platforms via a hidden Markov model analysis (24) (fig. S2 and supplementary methods) (23). As expected, the dwell time in the bound states increases with docking duplex length (Fig. 2C, top). The dwell time spent in the unbound state also increases (Fig. 2C, bottom), indicating slower diffusion and/or a reduced hybridization rate for

Fig. 1. A molecular platform with an integrated rotatable positioning arm. (A) Sketch of the DNA origami structure in side (top left) and top (right) view. The close-up in the perspective view (bottom left) highlights the single-stranded scaffold crossovers that form the flexible joint. (B) TEM class-average (top) and single-particle (bottom) micrographs of the structure. (C) AFM image of particles on mica.

Only structures for which the actuator arms were buried below the plates could be imaged with high contrast. For imaging, the arms were fixed to the plates with a 10-bp duplex formed between two staple extensions on the plate and the tip of the arm. (Inset) Height profile of a platform measured along the direction indicated by the red arrow.



¹Physics Department E14, Technical University Munich, 85748 Garching, Germany. ²Department of Chemistry, Ludwig-Maximilians University Munich, 81377 Munich, Germany. ³Center for Integrated Protein Science Munich, Ludwig-Maximilians University Munich, 81377 Munich, Germany. ⁴Nanosystems Initiative Munich, 80539 Munich, Germany. *These authors contributed equally to this work. †Corresponding author. Email: simmel@tum.de

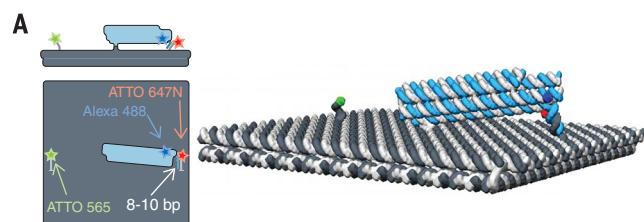


Fig. 2. Stochastic switching experiments. (A) For single-molecule multicolor FRET experiments, a donor fluorophore (Alexa Fluor 488) is attached to the six-helix bundle (6HB) arm and two acceptor fluorophores (ATTO 565 and ATTO 647N) are connected to staple-strand extensions on opposite sides of the plate. The pictograms on the left show hybridization of an extended staple of the arm to the staple extension of the base plate labeled with ATTO 647N. The length of the docking duplex was varied between 8 and 10 bp. A schematic three-dimensional representation is shown on the right. (B) Fluorescence traces obtained from the three fluorophores during donor excitation of the structures containing 9-bp docking duplexes. The change between green and red fluorescence indicates switching of the arm between corresponding docking sites. The zoomed-in view (bottom) reveals short periods of free diffusion between unbinding and rebinding events during which the donor (blue) fluorescence is dominant. a.u., arbitrary units. (C) Average dwell times for the bound and unbound states and their dependence on duplex length. Dwell times for the bound states (high acceptor signals shown in red or green; top panel) correspond to the times spent at the respective docking site. Dwell times for the unbound state (high donor signal shown in blue; bottom panel) represent the length of the traversal periods of the freely diffusing arm. (D) Average durations of the unbound states for various transitions and their dependence on duplex length. Corresponding to the start and end points of the traversal period (docking site or bound state shown in green or red before and after the unbound state), the unbound states can be classified as green→red and green→green or red→green and red→red traversals. In (C) and (D), the error bars denote the SD of the mean from three independent measurements.

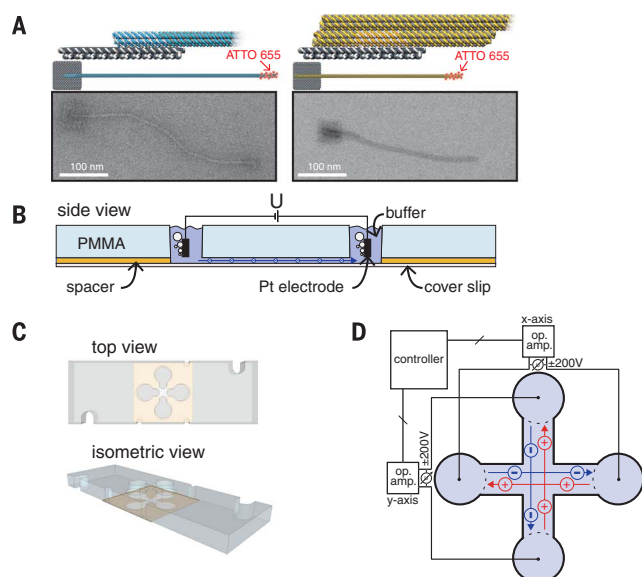
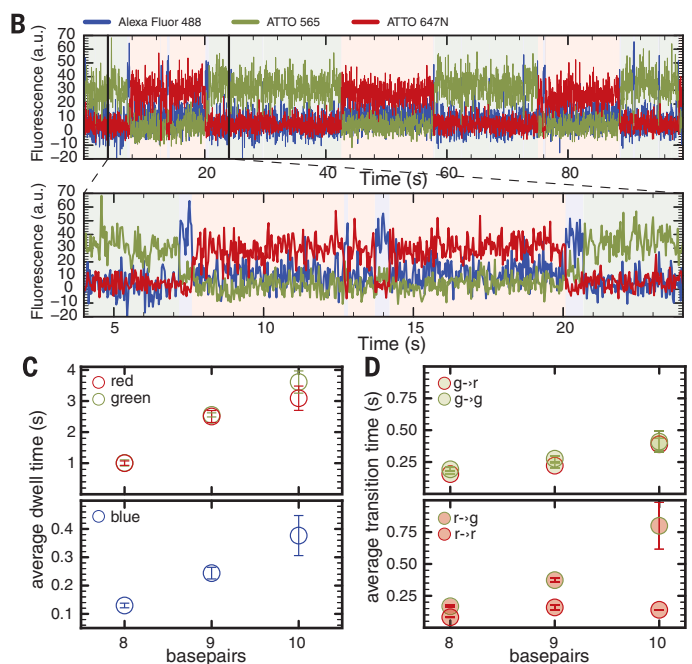
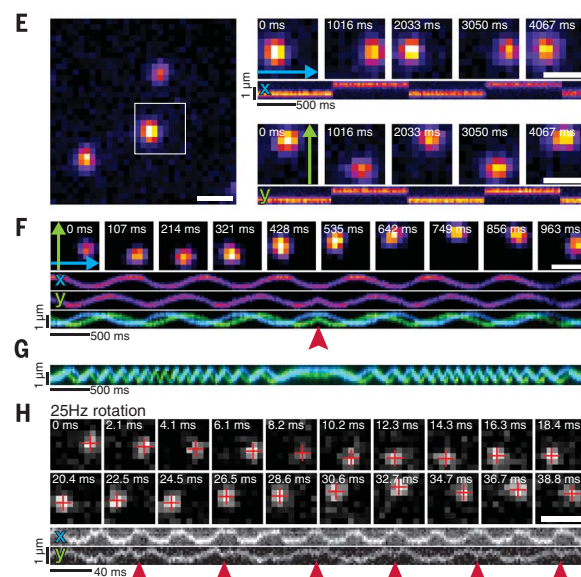


Fig. 3. External electric control of the robotic arm. (A) Two pointer extension designs for the robot arm and corresponding TEM images. The blue, linear extension pointer has a length of 411 nm (total length from center of rotation to tip: 436 nm). The orange pointer has a shape-complementary connection that withstands higher torque (total length: 354 nm; pivot point to tip: 332 nm; resulting arm extension: 308 nm). (B) Cross section and (C) top and isometric view of the cross-shaped electrophoretic sample chamber. PMMA, poly(methyl methacrylate); U, voltage. (D) Schematic depiction of the experimental setup with four electrodes. (E) Fluorescence microscopy images of three structures that are switched in the electric field. For the highlighted particle, movements are shown as snapshots and kymographs. The green and blue arrows indicate



the axes chosen for the kymographs. (Top) Switching left and right with 1 Hz. (Bottom) Switching up and down with 1 Hz. (F) (Top) One clockwise turn of 1-Hz rotation. (Bottom) Kymographs showing multiple turns of clockwise rotation followed by multiple counterclockwise turns, separately for the x and y axes and as a blue and green overlay. Reversal of the rotation direction is indicated by the red arrowhead (movie S1) (23). (G) Kymographs (x and y projections) obtained from a frequency sweep from 0 to 8 Hz and back, shown as an overlay of the kymographs along the x and y axes. (H) High-speed 360° clockwise and counterclockwise rotation with 25 Hz. For each frame, the center of the pointer tip is indicated by a red cross. Reversal of the rotation direction is indicated by red arrowheads (movie S2) (23). Unlabeled scale bars, 1 μ m.

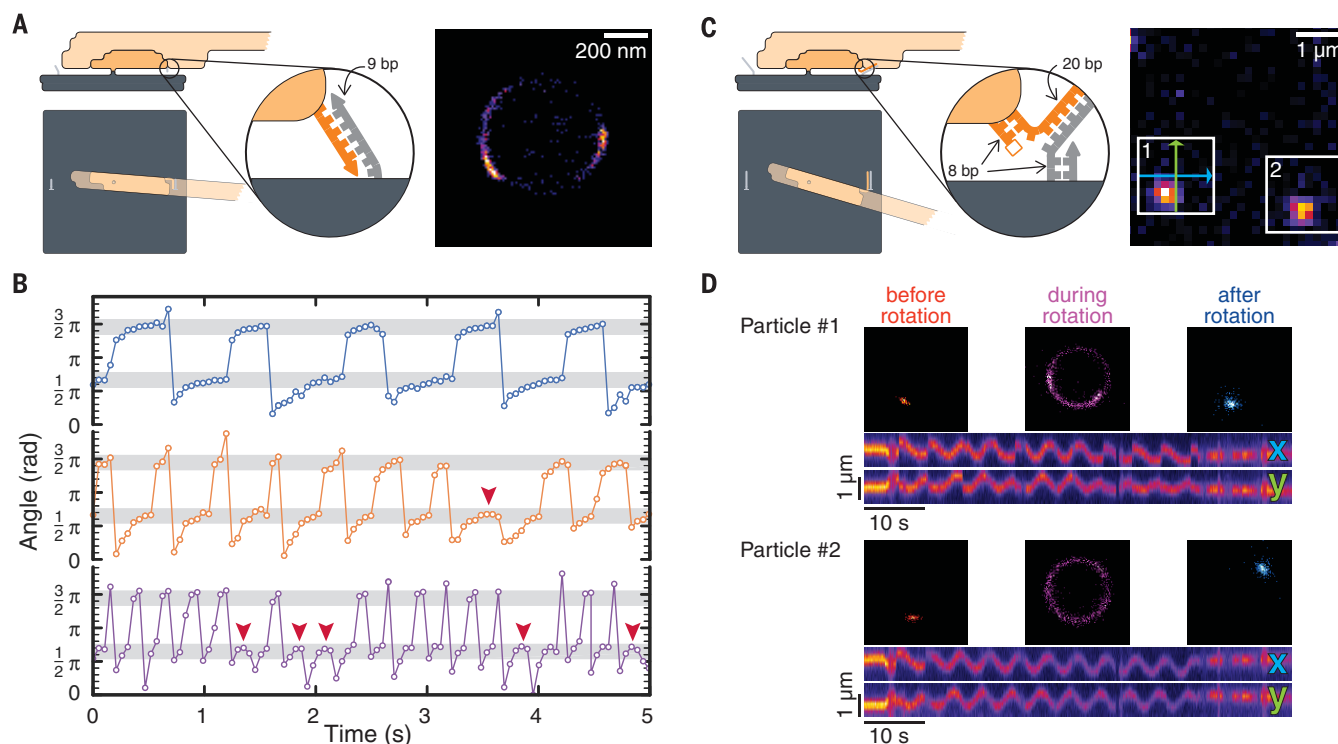


Fig. 4. Controlled hybridization and force-induced duplex dissociation. (A) Field-controlled switching of the extended robot arm between two 9-bp docking positions. (Left) Scheme of the setup. (Right) Single-molecule localization image of pointer positions acquired during electrical rotation at 1 Hz. The number of localizations is increased at angles corresponding to the two docking positions. (B) Angle plotted over time for 1-, 2-, and 4-Hz rotation with 110 V. The arm shows pronounced lagging for two angles (highlighted by gray bands). Higher frequencies result in a larger number of missed turns, which are indicated by the red arrowheads. (C) Unzipping of a

20-bp DNA duplex with the extended robot arm. (Left) Extensions from the platform and from the arm feature a short 8-bp strain-relief domain that prevents the staple strands from being pulled out of the structure. (Right) Experiments with two example particles are shown. Without an electric field, the arm is fixed at one of two docking positions on the base plate. (D) Rotation requires unzipping of the duplex, which is shown in the images (red, before rotation; violet, during rotation; and blue, after rotation) and kymographs at the bottom. Particle #1 rebinds to the starting position, whereas particle #2 rebinds to the position on the opposite side.

longer docking duplexes. Observed state transitions can be classified into transitions from one binding site to the other (green→red or red→green) or rebinding events to the same docking site (green→green or red→red). When the arm initially unbinds from the green docking site, it binds to either site with roughly the same transition time (Fig. 2D, top). Conversely, arms starting at the red docking site have a higher tendency to return to the same site (Fig. 2D, bottom). This bias is consistent with the expected orientation of the arm on the base plate, which is designed to point toward the red docking site (see Fig. 1A). The corresponding higher effective concentration of the arm in the vicinity of the red docking site results in faster rebinding transitions (16). Photophysical origins of the observed changes in the FRET signal (such as fluorescent dark states or environmental quenching of the fluorophores) were excluded by performing millisecond alternating laser excitation (25) experiments (fig. S3 and supplementary materials and methods) (23).

Modular extension with pointer structures

To facilitate direct observation of the arm's motion by diffraction-limited fluorescence microscopy,

we designed two versions of pointer structures that were multiply labeled with the fluorophore ATTO 655. Version one extended the arm linearly by 411 nm (Fig. 3A, blue). Version two extended the arm by 308 nm (Fig. 3A, orange) and was modularly plugged into the robot arm via a shape-complementary connector structure, creating a more stable connection between pointer and arm to allow for better torque transmission. Both pointers are based on a rigid 6HB with a persistence length $>1 \mu\text{m}$ (26). The two designs were motivated by the differing requirements for the experiments described below. For rotational diffusion experiments in the absence of docking sites, we found that the linear pointer interacted less with the base plate than the shape-complementary pointer (fig. S4) (23). However, when used to exert forces, the linear pointer displayed a reduced stability and tended to break at the connection site (supplementary text) (23). In the presence of docking sites, single-molecule localization images of both pointers were consistent with the positions of the docks on the platform, proving that the extensions point along the axis of the short arm (fig. S5) (23) and that the interactions with the docking sites dominated over unspecific sticking.

Electrical control of the robot arm

To realize dynamic external control of the robot arm, we applied electrical fields to the system—a natural choice for the manipulation of charged biomolecules (27). Electrical fields have been previously used only to stretch or orient substrate-immobilized DNA duplexes (28) but not to dynamically control the conformation of nanomechanical DNA devices. We created a cross-shaped electrophoretic chamber constituted by two perpendicular fluidic channels intersecting at the center of a microscopy cover slip, with two pairs of platinum electrodes inserted into the four buffer reservoirs (Fig. 3, B and C, and fig. S6) (23). DNA nanostructures immobilized at the center of the cross chamber experience a superposition of the fields generated by the electrode pairs. Hence, a voltage can be applied to arbitrarily control the pointing direction of the arm (Fig. 3D).

Electrical actuation of the arms results in a movement of the pointers, which we observed with an electron-multiplying charge-coupled device camera using total internal reflection fluorescence (TIRF) microscopy. In Fig. 3, we show switching of an arm in two perpendicular directions (Fig. 3E), as well as rotation with a constant frequency of 1 Hz (Fig. 3F) and with variable

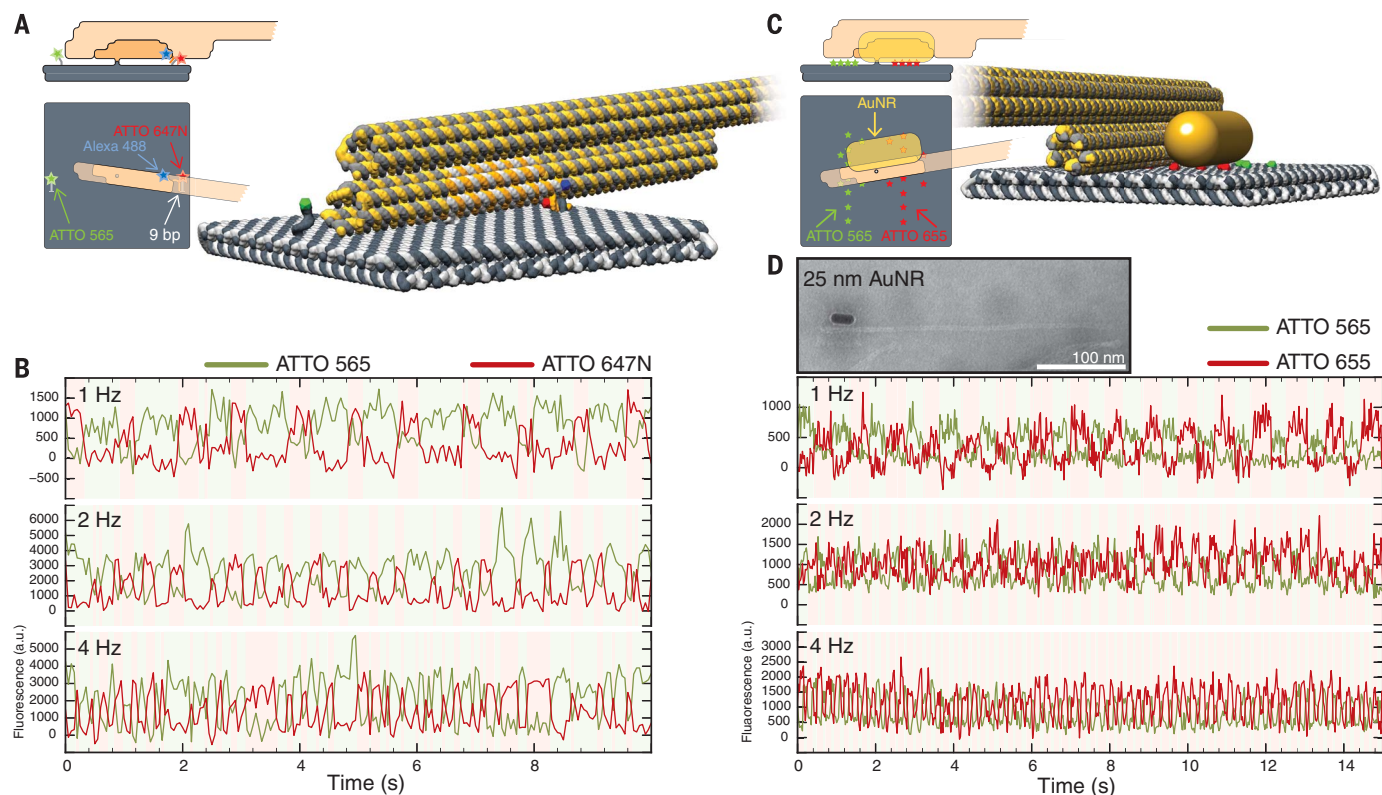


Fig. 5. Electrically controlled movement of molecules and nanoparticles.

(A) Configuration of the robot arm with shape-complementary extension for transport of the FRET donor Alexa Fluor 488 between two 9-nucleotide docking sites with the acceptors ATTO 565 and ATTO 647N. (B) Acceptor signals for continuous donor excitation for electrical rotation at 1 Hz (top), 2 Hz (middle), and 4 Hz (bottom). (C) For application of the robot arm in

switchable plasmonics, a 25-nm-long gold nanorod (AuNR) is attached to the side of the 6HB arm, and 11 ATTO 565 and ATTO 655 dyes are placed on opposite halves of the platform. (D) (Top) TEM micrograph of a structure with a 25-nm AuNR. (Bottom) Fluorescence traces for continuous excitation of the dyes while the robot arm is rotated at 1, 2, and 4 Hz (fig. S12) (23) for data obtained with a 50-nm AuNR.

frequency (Fig. 3G), ramping from 0 to 8 Hz and back to 0 Hz. Movie S1 (23) shows a range of movement patterns, underlining the capability of arbitrary angular control. To characterize faster arm movements, we used a complementary metal-oxide semiconductor camera to record TIRF microscopy videos with a 2-ms time resolution. An image series taken from a video in which the robot arm was rotated back and forth at a frequency $f = 25$ Hz is shown in Fig. 3H (see also supplementary movie S2) (23). Kymographs displaying the projected motion of the arm's pointer along the x and y axes show the expected sinusoidal characteristics. In a high-viscosity buffer solution containing 65% sucrose, motion of the arm was substantially slowed (fig. S7) (23).

Next, we assessed the angular positioning precision of the arm, which can be achieved in the absence of docking sites by the electrical field alone (fig. S8) (23). For large applied voltages (≥ 120 V in our setup), the angular standard deviation is ≈ 0.1 radians, which translates to a positioning precision of ≈ 2.5 nm on the plate.

Controlled interaction with docking positions on the platform

To investigate the interaction of the arm with binding sites on the platform during electrical manipulation, we performed latching experiments

with the same arrangement of docks as in Fig. 2 and an identical 9-base pair (bp) docking sequence (Fig. 4A). When rotated at frequencies of $f = 1, 2$, and 4 Hz, we observed temporary stalling of the pointer at the two angle positions that correspond to the two docking sites (Fig. 4B), indicating that the arm snaps into the binding positions during rotation. Whereas the signal followed the external control faithfully for $f = 1$ Hz, occasional skips occurred for $f = 2$ and 4 Hz. This behavior is caused by the statistical nature of single-molecule duplex dissociation, whose frequency increases exponentially with the application of a force (29) and, in dynamic experiments, also depends on the force rate (29). Apparently, the dissociation rate (~ 0.4 s $^{-1}$) (Fig. 2B) of the docking duplex is sufficiently enhanced by the electrical force to follow the 1-Hz rotation. For higher frequencies, the duplex does not always dissociate fast enough and the arm cannot follow the rotation of the electrical field. By contrast, at a slower rotation speed of $f = 0.1$ Hz, we were able to observe dynamic latching to four different docking sites (fig. S9) (23).

We next tested whether the robotic arm could wrest apart a 20-bp docking duplex, which is a stable structure at room temperature. Although the arm is firmly locked in place in the absence of an electrical field, it can be released from the

docking site by actuating the arm and rotated as shown in Fig. 4, C and D. Unzipping is expected to be most effective when the field is applied perpendicularly to the fixed arm. As the base plates are randomly oriented with respect to the sample chamber, the field is slowly rotated at a frequency of 0.2 Hz to guarantee that each structure has sufficient time to experience a strong enough unzipping force. When switching off the field during rotation at an arbitrary phase, the arm immediately localizes to an available docking site.

At the field strengths generated in our sample chamber, we do not expect field-induced melting of DNA duplexes as is observed, for instance, for DNA structures immobilized on electrode surfaces (30). Instead, the arm acts as a lever that mechanically transduces the electrical force acting on its large charge to the docking duplex. Force-induced unzipping of DNA duplexes has been previously achieved through the use of single-molecule manipulation techniques such as AFM (31) and optical tweezers (32) or within nanopores (33). These experiments have shown that DNA unzipping requires forces on the order of 10 to 20 pN, consistent with the typical binding free energy of DNA base pairs and their subnanometer spacing. A rough theoretical treatment (supplementary text) (23) suggests that forces that can be

generated by the robot arm are on this scale. The ability to separate stable duplexes by force facilitates the electrically controlled dissociation of the arm from one docking site and its subsequent placement at a different target position, which is then maintained at zero field (figs. S10 and S11) (23).

Electrically controlled movement of cargo

To show controlled movement of a cargo molecule attached to the arm, we used the three-color FRET system already employed in the stochastic switching experiments (Fig. 5A). In contrast to those experiments, the donor fluorophore is actively transported between two 9-nucleotide-long docking positions by rotating the arm with the help of the high torque extension at rotation frequencies of $f = 1, 2$, and 4 Hz, respectively. We observed alternating FRET traces (Fig. 5B) with the periodicity of the externally applied field. In agreement with the latching experiments (Fig. 4B), higher rotation frequencies correspond with an increase in the number of skips.

To demonstrate transport of inorganic nanoparticles by the robot arm, we attached a gold nanorod (AuNR) to one side of the 6HB arm and probed its plasmonic interaction with red and green fluorophores immobilized on the platform (Fig. 5C). As shown in Fig. 5D and fig. S12 (23), the AuNR alternately modulates the fluorescence of the fluorophores during rotation of the arm at the externally prescribed frequency. Electrical manipulation enables faster operation of switchable biohybrid plasmonic systems than previously achieved with the fuel-strand technique (34). More sophisticated systems involving multiple particles for the creation of switchable field enhancement or circular dichroism appear feasible (35).

Discussion

We have introduced electrical actuation as a viable strategy for fast, computer-controlled operation of biohybrid nanorobotic systems, which can exert forces at the molecular scale. Compared with nanoscale manipulation methods such as scanning probe techniques and optical or magnetic tweezers, electrical control is contact-free and can be implemented with low-cost instrumentation. The robotic movements achieved are at least five orders of magnitude faster than previously reported for the fastest DNA motor systems and are comparable to adenosine triphosphatase-driven biohybrids (5). The robot-arm system may be scaled up and integrated into larger hybrid

systems by a combination of lithographic and self-assembly techniques. For instance, the platforms can be easily connected to form long filaments with multiple DNA robot arms (fig. S13) (23) or to create extended lattices. The use of algorithmic self-assembly (36) will enable the creation of structures with different types of robot platforms with dedicated tasks. Lithographic patterning of the substrate (37–39) will further allow the fabrication of robot-arm arrays with defined platform orientations. By using nanostructured control electrodes, single robot arms could even be addressed individually, and their positioning state could act as a molecular mechanical memory. Combined with appropriate pick-up and release mechanisms (3, 40), it is conceivable that this technology can also be applied to DNA-templated synthesis (41). Electrically clocked synthesis of molecules with a large number of robot arms in parallel could then be the first step toward the realization of a genuine nanorobotic production factory.

REFERENCES AND NOTES

1. J. D. Badjic, V. Balzani, A. Credi, S. Silvi, J. F. Stoddart, *Science* **303**, 1845–1849 (2004).
2. T. Kudernac et al., *Nature* **479**, 208–211 (2011).
3. S. Kassem, A. T. L. Lee, D. A. Leigh, A. Markevicius, J. Solà, *Nat. Chem.* **8**, 138–143 (2016).
4. B. Lewandowski et al., *Science* **339**, 189–193 (2013).
5. R. K. Soong et al., *Science* **290**, 1555–1558 (2000).
6. P. W. Rothmund, *Nature* **440**, 297–302 (2006).
7. S. M. Douglas et al., *Nature* **459**, 414–418 (2009).
8. J. S. Shin, N. A. Pierce, *J. Am. Chem. Soc.* **126**, 10834–10835 (2004).
9. S. J. Green, J. Bath, A. J. Turberfield, *Phys. Rev. Lett.* **101**, 238101 (2008).
10. T. Ornabegho, R. Sha, N. C. Seeman, *Science* **324**, 67–71 (2009).
11. K. Lund et al., *Nature* **465**, 206–210 (2010).
12. S. F. J. Wickham et al., *Nat. Nanotechnol.* **6**, 166–169 (2011).
13. T. E. Tomov et al., *ACS Nano* **11**, 4002–4008 (2017).
14. H. Gu, J. Chao, S.-J. Xiao, N. C. Seeman, *Nature* **465**, 202–205 (2010).
15. Z.-G. Wang, J. Elbaz, I. Willner, *Angew. Chem. Int. Ed.* **51**, 4322–4326 (2012).
16. E. Kopperger, T. Pirzer, F. C. Simmel, *Nano Lett.* **15**, 2693–2699 (2015).
17. Y. Yang et al., *Chemistry* **23**, 3979–3985 (2017).
18. T. Tomaru, Y. Suzuki, I. Kawamata, S.-M. Nomura, S. Murata, *Chem. Commun.* **53**, 7716–7719 (2017).
19. A. E. Marras, L. Zhou, H.-J. Su, C. E. Castro, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 713–718 (2015).
20. P. Ketterer, E. M. Willner, H. Dietz, *Sci. Adv.* **2**, e1501209 (2016).
21. B. Yurke, A. J. Turberfield, A. P. Mills Jr., F. C. Simmel, J. L. Neumann, *Nature* **406**, 605–608 (2000).
22. D. Y. Zhang, G. Seelig, *Nat. Chem.* **3**, 103–113 (2011).
23. See supplementary materials.
24. S. A. McKinney, C. Joo, T. Ha, *Biophys. J.* **91**, 1941–1951 (2006).

25. A. N. Kapanidis et al., *Acc. Chem. Res.* **38**, 523–533 (2005).
26. D. J. Kauer, T. Kurth, T. Liedl, R. Seidel, *Nano Lett.* **11**, 5558–5563 (2011).
27. J. L. Vivoy, *Rev. Mod. Phys.* **72**, 813–872 (2000).
28. Y. Klapper, N. Sinha, T. W. S. Ng, D. Lubrich, *Small* **6**, 44–47 (2010).
29. O. K. Dudko, G. Hummer, A. Szabo, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15755–15760 (2008).
30. F. Wei et al., *Langmuir* **22**, 6280–6285 (2006).
31. B. Essevaz-Roulet, U. Bockelmann, F. Heslot, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11935–11940 (1997).
32. U. Bockelmann, P. Thomen, B. Essevaz-Roulet, V. Viasnoff, F. Heslot, *Biophys. J.* **82**, 1537–1553 (2002).
33. A. F. Sauer-Budge, J. A. Nyamwanda, D. K. Lubensky, D. Branton, *Phys. Rev. Lett.* **90**, 238101 (2003).
34. A. Kuzyk et al., *Nat. Mater.* **13**, 862–866 (2014).
35. T. Zhang, N. Gao, S. Li, M. J. Lang, Q. H. Xu, *J. Phys. Chem. Lett.* **6**, 2043–2049 (2015).
36. R. D. Barish, R. Schulman, P. W. K. Rothmund, E. Winfree, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6054–6059 (2009).
37. A. Gopinath, P. W. K. Rothmund, *ACS Nano* **8**, 12030–12040 (2014).
38. A. Gopinath, E. Miyazono, A. Faraon, P. W. K. Rothmund, *Nature* **535**, 401–405 (2016).
39. M. B. Scheible, G. Pardatscher, A. Kuzyk, F. C. Simmel, *Nano Lett.* **14**, 1627–1633 (2014).
40. S. K. Kufer, E. M. Puchner, H. Gump, T. Liedl, H. E. Gaub, *Science* **319**, 594–596 (2008).
41. X. Li, D. R. Liu, *Angew. Chem. Int. Ed.* **43**, 4848–4870 (2004).

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft through the Collaborative Research Centre (grants SFB1032 Project A2 to F.C.S. and SFB1032 Project B3 to D.C.L.). E.K. acknowledges support by the Technical University Munich (TUM) through the TUM International Graduate School of Science and Engineering. D.C.L. acknowledges the additional support of the Ludwig-Maximilians University through the Center for NanoScience and the Bioimaging Network. We thank J. Müller and F. Wilde for help with the design of the operational amplifier, F. Praetorius for providing the M13 scaffold strands, and S. Krishnan for help with TEM imaging. All data are reported in the main text and the supplementary materials. Contributions: E.K., J.L., and F.C.S. conceived the study. E.K. and J.L. designed the structure and built the electrical setup. E.K., J.L., S.M., and F.R. performed the experiments. D.C.L. guided the single-molecule fluorescence experiments. E.K., J.L., S.M., D.C.L., and F.C.S. interpreted and discussed the experiments. All authors commented on and wrote the paper. Competing interests: E.K., J.L., and F.C.S. have filed a patent covering electrically driven molecular machines (EPA 17 165 250.6). We declare no competing financial interests.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6373/296/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S13
DNA Sequences
Movies S1 and S2
References (42–61)

19 July 2017; accepted 29 November 2017
10.1126/science.aao4284

Cite as: C. R. Cabrera *et al.*, *Science*
10.1126/science.aao5686 (2017).

Quantum liquid droplets in a mixture of Bose-Einstein condensates

C. R. Cabrera,* L. Tanzi,* J. Sanz, B. Naylor, P. Thomas, P. Cheiney, L. Tarruell†

ICFO-Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain.

*These authors contributed equally to this work.

†Corresponding author. E-mail: leticia.tarruell@icfo.eu

Quantum droplets are small clusters of atoms self-bound by the balance of attractive and repulsive forces. Here we report on the observation of droplets solely stabilized by contact interactions in a mixture of two Bose-Einstein condensates. We demonstrate that they are several orders of magnitude more dilute than liquid helium by directly measuring their size and density via in situ imaging. We show that the droplets are stabilized against collapse by quantum fluctuations and that they require a minimum atom number to be stable. Below that number, quantum pressure drives a liquid-to-gas transition that we map out as a function of interaction strength. These ultra-dilute isotropic liquids remain weakly interacting and constitute an ideal platform to benchmark quantum many-body theories.

Quantum fluids can be liquids—of fixed volume—or gases, depending on the attractive or repulsive character of the inter-particle interactions and their interplay with quantum pressure. Liquid helium is the prime example of quantum fluids. For small particle numbers it forms self-bound liquid droplets: nanometer-sized, dense and strongly interacting clusters of helium atoms. Understanding the droplets' properties, which directly reflect their quantum nature, is challenging and requires a precise knowledge of the short-range details of the interatomic potential (1, 2). Very different quantum droplets, more than 2 orders of magnitude larger and 8 orders of magnitude more dilute, have recently been proposed in ultracold atomic gases (3, 4). Interestingly, these ultra-dilute systems enable a much simpler microscopic description, while remaining in the weakly interacting regime. They are thus amenable to well controlled theoretical studies.

The formation of quantum droplets requires a balance between attractive forces, which hold them together, and repulsive ones that stabilize them against collapse. In helium droplets, the repulsion is dominated by the short-range part of the interatomic potential (1, 2). In contrast, for ultracold atomic droplets several distinct stabilization mechanisms have been proposed, including three-body correlations (3) and quantum fluctuations (4). The latter can be revealed in systems with competing interactions, where mean-field forces of different origins almost completely cancel out and result in a small residual attraction. In such systems, beyond mean-field effects remain important even in the weakly interacting regime. To first order they lead to the Lee-Huang-Yang (LHY) repulsive energy (5), comparable in strength to the residual mean-field (MF) attraction. Recently, ultracold

atomic droplets stabilized by quantum fluctuations have been realized in single-component magnetic quantum gases with competing attractive dipolar and repulsive contact interactions (6–11). In this case, the anisotropic character of the magnetic dipole-dipole force leads to the formation of filament-like self-bound droplets with highly anisotropic properties (9, 12, 13). Given the generality of the stabilization mechanism, droplets should in fact also exist in simpler systems with pure isotropic contact interactions. Even though they were originally predicted in this setting (4), their experimental observation has so far remained elusive.

Here, we observe ultracold atomic droplets in a mixture of two Bose-Einstein condensates with competing contact interactions. Although a single-component attractive condensate with only contact interactions collapses (14, 15), quantum fluctuations stabilize a two-component mixture with inter-component attraction and intra-component repulsion (4). There, the repulsion in each component remains large and results in a non-negligible Lee-Huang-Yang energy. The beyond mean-field repulsion and the residual mean-field attraction scale differently with the total density n (in three dimensions the scaling of the Lee-Huang-Yang and mean-field energy densities are $\mathcal{E}_{\text{LHY}} \propto n^{5/2}$ vs. $\mathcal{E}_{\text{MF}} \propto n^2$, respectively). Hence, there is always a density for which these contributions balance each other and droplets are stabilized. Unlike their dipolar counterparts, these mixture droplets originate exclusively from s -wave contact interactions and are therefore isotropic. We demonstrate the self-bound character of mixture droplets and directly measure their ultra-low densities and micrometer-scaled sizes. Moreover, by comparison to a single-component condensate with only contact interactions, we confirm that their stability stems from quantum fluctuations.

We perform experiments with two ^{39}K Bose-Einstein condensates in states $|\uparrow\rangle \equiv |F, m_F\rangle = |1, -1\rangle$ and $|\downarrow\rangle \equiv |1, 0\rangle$, where F is the total angular momentum and m_F its projection. An external magnetic field allows us to control the interactions, which are parameterized by the intra- and inter-state scattering lengths $a_{\uparrow\uparrow}$, $a_{\downarrow\downarrow}$ and $a_{\uparrow\downarrow}$ (see Fig. 1A). These have been computed according to the model of ref. (16). The residual mean-field interaction is proportional to $\delta a = a_{\uparrow\downarrow} + \sqrt{a_{\uparrow\uparrow}a_{\downarrow\downarrow}}$. The condition $\delta a = 0$ separates the repulsive ($\delta a > 0$) and attractive ($\delta a < 0$) regimes. The experiment starts with a pure condensate in state $|\uparrow\rangle$ loaded in one plane of a vertical blue-detuned lattice potential (Fig. 1B). We choose a trapping frequency $\omega_z/2\pi = 635(5)$ Hz large enough to compensate for gravity, but small enough for the system to be in the three-dimensional regime. Indeed, the vertical harmonic oscillator length $a_{\text{ho}} = 0.639(3)$ μm exceeds the characteristic length of the most energetic Bogoliubov excitation branch by typically a factor of 3 (17). A vertical red-detuned optical dipole trap provides radial confinement in the horizontal plane. In order to prepare a balanced mixture of the two states, we apply a radio-frequency pulse at $B \approx 57.3$ G, which lies in the miscible regime ($\delta a \approx 7a_0$, where a_0 denotes the Bohr radius) (18). For all measurements, we verify independently the spin composition of the mixture via Stern-Gerlach separation during time-of-flight expansion (Fig. 1B). Subsequently, we slowly ramp down the magnetic field at a constant rate of 59 G/s and enter the attractive regime $\delta a < 0$ (17). We then switch off the vertical red-detuned optical dipole trap while keeping the lattice confinement, allowing the atoms to evolve freely in the horizontal plane. The integrated atomic density is imaged in situ at different evolution times. We use a high numerical aperture objective [<1 μm resolution, $1/e$ Gaussian width) along the vertical direction and a phase-contrast polarization scheme (19) which detects both states with almost equal sensitivity (Fig. 1) (17)].

Typical images of the mixture time evolution in the repulsive and attractive regimes are displayed in Fig. 1C. For $\delta a = 1.2(1) a_0 > 0$ (top row), the cloud expands progressively in the plane, as expected for a repulsive Bose gas in the absence of radial confinement (20). In contrast, in the attractive regime $\delta a = -3.2(1) a_0 < 0$ (central row), the dynamics of the system are remarkably different and the atoms reorganize in an isotropic self-bound liquid droplet. Its typical size remains constant for evolution times up to 25 ms. In an analogous experiment with a single-component attractive condensate $|\downarrow\rangle$ of scattering length $a = -2.06(2) a_0 < 0$, the system instead collapses (bottom row). In our experimental geometry, quantum pressure can never stabilize bright solitons because of the presence of weak anti-confinement in the horizontal

plane (17). At the mean-field level, the two-component attractive case has a description equivalent to the single-component one, provided that the scattering length a is replaced by $\sim \delta a/2$ and the density ratio between the two components is fixed at $n_{\uparrow}/n_{\downarrow} = \sqrt{a_{\downarrow\downarrow}/a_{\uparrow\uparrow}}$ (17). However, the role of the first beyond mean-field correction is very different in the two systems, explaining their very different behavior. In the single-component case, the Lee-Huang-Yang energy depends on a and in the weakly interacting regime constitutes a negligible correction to the mean-field term. Therefore, its contribution is most easily revealed in strongly interacting systems (21). In contrast, in the mixture the mean-field and Lee-Huang-Yang energy densities scale as $\mathcal{E}_{\text{MF}} \propto \delta a n^2$ and $\mathcal{E}_{\text{LHY}} \propto (\sqrt{a_{\uparrow\uparrow}a_{\downarrow\downarrow}} n)^{5/2}$, respectively. Because $\sqrt{a_{\uparrow\uparrow}a_{\downarrow\downarrow}} \gg |\delta a|$, for typical experimental parameters they balance at accessible atomic densities and stabilize liquid droplets (4). Therefore, the existence of liquid droplets is a striking manifestation of beyond mean-field effects in the weakly interacting regime.

To further characterize the mixture, we perform a quantitative analysis of the images fitting the integrated atomic density profiles with a two-dimensional Gaussian (17). We extract the atom number N and radial size σ_r and infer the peak density $n_0 = N/(\pi^{3/2}\sigma_r^2\sigma_z)$ by assuming a vertical size σ_z identical to the harmonic oscillator length a_{ho} . Figure 2A (top and central panel) shows the time evolution of N and σ_r measured for the interaction parameters of Fig. 1C. For $\delta a > 0$ (red circles) the gas quickly expands and its atom number does not vary. Instead, for $\delta a < 0$ (blue circles) the system is in the liquid regime and the radial size of the droplet remains constant at $\sigma_r \approx 6$ μm . Initially its atom number is $N = 24.5(7) \times 10^3$, corresponding to a peak density of $n_0 = 1.97(8) \times 10^{14}$ atoms/cm³. We attribute the subsequent decay of the droplet atom number (Fig. 2A, top panel) to three-body recombination. The observed timescale is compatible with the measured density and effective three-body loss rate (17). By directly measuring the density of our droplets we confirm that they are more than 8 orders of magnitude more dilute than liquid helium and remain very weakly interacting. Indeed, the interaction parameters of each component are extremely small ($n_{\uparrow}a_{\uparrow\uparrow}^3, n_{\downarrow}a_{\downarrow\downarrow}^3 \sim 10^{-5}$).

A closer view of the droplet size is displayed in the bottom panel of Fig. 2A. At $t \sim 25$ ms, σ_r starts to increase and the system behaves like the $\delta a > 0$ gas. Following refs. (4, 9, 22, 23), we attribute the dissociation of the droplet to the effect of quantum pressure, which acts as a repulsive force. As the atom number decreases, the relative weight between kinetic (\mathcal{E}_k) and interaction energies (\mathcal{E}_{MF} , \mathcal{E}_{LHY}) changes, for each term scales differently with N : $\mathcal{E}_k \propto N$, $\mathcal{E}_{\text{MF}} \propto N^2$ and $\mathcal{E}_{\text{LHY}} \propto$

$N^{5/2}$. Below a critical atom number, kinetic effects become sufficiently strong to drive a liquid-to-gas transition. To support this scenario, Fig. 2B depicts the radial size and atomic density as a function of atom number. For $\delta a < 0$ (blue circles) we observe that both size (top panel) and density (bottom panel) remain constant at large N . For decreasing atom number, we observe a point where the size diverges and the density drops abruptly. This indicates a liquid-to-gas transition, which takes place at the critical atom number N_c . Below this value, the attractive gas is still stabilized by quantum fluctuations but expands because of kinetic effects, similarly to the repulsive mixture ($\delta a > 0$, red circles).

The liquid-to-gas transition is expected to depend on δa , as sketched in the inset of Fig. 2B (top panel). We explore the phase diagram by tuning the interaction strengths with magnetic field (see Fig. 1A). Figure 3A displays the measured size as a function of the atom number for magnetic fields corresponding to δa between $-5.5(1) a_0$ and $-2.4(1) a_0$. The critical number N_c shows a strong dependence on the magnetic field. The top panel of Fig. 3B presents our experimental determination of the phase transition line. N_c increases when the attraction decreases, confirming that weakly bound droplets are more susceptible to kinetic effects and require a larger atom number to remain self-bound. Figure 3A also yields the droplet size as a function of atom number and magnetic field. In the bottom panel of Fig. 3B we display the measurements obtained at a fixed atom number $N = 1.5(1) \times 10^4$, always larger than N_c for our interaction regime. As expected, the droplet size decreases as the attraction increases.

We theoretically describe the system using a simple zero-temperature model based on an extended Gross-Pitaevskii equation that includes both the vertical harmonic confinement and an additional repulsive Lee-Huang-Yang term. The latter is obtained assuming the Bogoliubov spectrum of a three-dimensional homogeneous mixture (17). In Fig. 3B we compare the experimental results to the predicted critical atom number and droplet size (solid lines). We find qualitative agreement for the complete magnetic field range with no adjustable parameters. In the weakly attractive regime the agreement is even quantitative, similarly to the dipolar Erbium experiments of ref. (8). In contrast, when increasing the effective attraction, the droplets are more dilute than expected. In particular, their size exceeds the theoretical predictions by up to a factor of three. This is almost one order of magnitude larger than our imaging resolution, excluding finite-resolution effects. Furthermore, the critical atom number is a factor of two smaller than the theoretical value. Interestingly, a similar discrepancy was reported for dipolar Dysprosium droplets, with a critical atom number one order of magnitude smaller than expected (9). There, the deviation was attributed to an insufficient knowledge of the background scattering length. This explanation seems unlikely in

the case of potassium (17), where excellent interaction potentials are available (16, 24, 25).

Other physical mechanisms might be responsible for the diluteness of the observed droplets. Although our system is three-dimensional, the confinement along the vertical direction might affect the Lee-Huang-Yang energy, modifying its density and interaction dependence or introducing finite-size effects; however, a description of quantum fluctuations in the dimensional crossover between two and three dimensions is challenging. Interestingly, the almost perfect cancellation of the mean-field energy could reveal corrections other than the Lee-Huang-Yang term. Higher-order many-body terms might play a role, as proposed in ref. (3) for single-component systems. Taking them into account analytically requires knowledge of the three-body interaction parameters of the mixture, which are non-universal and difficult to estimate in our interaction regime. Alternatively, our results could be compared to *ab initio* quantum Monte Carlo simulations (26). Given the ultra-dilute character and simple microscopic description of our system, a direct comparison to different theoretical approaches could provide insights on yet unmeasured many-body effects.

Future research directions include studying the spectrum of collective modes of the droplets (8). Their unconventional nature not only provides a sensitive testbed for quantum many-body theories, but should also give access to zero-temperature quantum systems (4) not present in the dipolar case (27). Our experiments could also enable the exploration of low-dimensional systems, where the enhanced quantum fluctuations make droplets ubiquitous (28). Finally, a coherent coupling between the two components (29) is expected to yield effective three-body interactions (30) and provide control over the density dependence of the Lee-Huang-Yang term (31).

Note added in proof: After submission of this work, related experiments have been performed by the LENS group (32).

REFERENCES AND NOTES

1. F. Dalfovo, S. Stringari, Helium nanodroplets and trapped Bose-Einstein condensates as prototypes of finite quantum fluids. *J. Chem. Phys.* **115**, 10078 (2001). [doi:10.1063/1.1424926](https://doi.org/10.1063/1.1424926)
2. M. Barranco, R. Guardiola, S. Hernández, R. Mayol, J. Navarro, M. Pi, Helium nanodroplets: An overview. *J. Low Temp. Phys.* **142**, 1–81 (2006). [doi:10.1007/s10909-005-9267-0](https://doi.org/10.1007/s10909-005-9267-0)
3. A. Bulgac, Dilute quantum droplets. *Phys. Rev. Lett.* **89**, 050402 (2002). [doi:10.1103/PhysRevLett.89.050402](https://doi.org/10.1103/PhysRevLett.89.050402) [Medline](#)
4. D. S. Petrov, Quantum mechanical stabilization of a collapsing Bose-Bose mixture. *Phys. Rev. Lett.* **115**, 155302 (2015). [doi:10.1103/PhysRevLett.115.155302](https://doi.org/10.1103/PhysRevLett.115.155302) [Medline](#)
5. T. D. Lee, K. Huang, C. N. Yang, Eigenvalues and eigenfunctions of a Bose system of hard spheres and its low-temperature properties. *Phys. Rev. Lett.* **106**, 1135 (1957).

6. H. Kadau, M. Schmitt, M. Wenzel, C. Wink, T. Maier, I. Ferrier-Barbut, T. Pfau, Observing the Rosenzweig instability of a quantum ferrofluid. *Nature* **530**, 194–197 (2016). [doi:10.1038/nature16485](https://doi.org/10.1038/nature16485) [Medline](#)
7. I. Ferrier-Barbut, H. Kadau, M. Schmitt, M. Wenzel, T. Pfau, Observation of quantum droplets in a strongly dipolar Bose gas. *Phys. Rev. Lett.* **116**, 215301 (2016). [doi:10.1103/PhysRevLett.116.215301](https://doi.org/10.1103/PhysRevLett.116.215301) [Medline](#)
8. L. Chomaz, S. Baier, D. Petter, M. J. Mark, F. Wächtler, L. Santos, F. Ferlaino, Quantum-fluctuation-driven crossover from a dilute Bose-Einstein condensate to a macrodroplet in a dipolar quantum fluid. *Phys. Rev. X* **6**, 041039 (2016). [doi:10.1103/PhysRevX.6.041039](https://doi.org/10.1103/PhysRevX.6.041039)
9. M. Schmitt, M. Wenzel, F. Böttcher, I. Ferrier-Barbut, T. Pfau, Self-bound droplets of a dilute magnetic quantum liquid. *Nature* **539**, 259–262 (2016). [doi:10.1038/nature20126](https://doi.org/10.1038/nature20126) [Medline](#)
10. I. Ferrier-Barbut, M. Schmitt, M. Wenzel, H. Kadau, T. Pfau, Liquid quantum droplets of ultracold magnetic atoms. *J. Phys. B* **49**, 214004 (2016). [doi:10.1088/0953-4075/49/21/214004](https://doi.org/10.1088/0953-4075/49/21/214004)
11. M. Wenzel, F. Böttcher, T. Langen, I. Ferrier-Barbut, T. Pfau, Striped states in a many-body system of tilted dipoles. *arXiv:1706.09388*.
12. F. Wächtler, L. Santos, Quantum filaments in dipolar Bose-Einstein condensates. *Phys. Rev. A* **93**, 061603 (2016). [doi:10.1103/PhysRevA.93.061603](https://doi.org/10.1103/PhysRevA.93.061603)
13. D. Baillie, R. M. Wilson, R. N. Bisset, P. B. Blakie, Self-bound dipolar droplet: A localized matter wave in free space. *Phys. Rev. A* **94**, 021602 (2016). [doi:10.1103/PhysRevA.94.021602](https://doi.org/10.1103/PhysRevA.94.021602)
14. J. M. Gerton, D. Strekalov, I. Prodan, R. G. Hulet, Direct observation of growth and collapse of a Bose-Einstein condensate with attractive interactions. *Nature* **408**, 692–695 (2000). [doi:10.1038/35047030](https://doi.org/10.1038/35047030) [Medline](#)
15. E. A. Donley, N. R. Claussen, S. L. Cornish, J. L. Roberts, E. A. Cornell, C. E. Wieman, Dynamics of collapsing and exploding Bose-Einstein condensates. *Nature* **412**, 295–299 (2001). [doi:10.1038/35085500](https://doi.org/10.1038/35085500) [Medline](#)
16. S. Roy, M. Landini, A. Trenkwalder, G. Semeghini, G. Spagnolli, A. Simoni, M. Fattori, M. Inguscio, G. Modugno, Test of the universality of the three-body Efimov parameter at narrow Feshbach resonances. *Phys. Rev. Lett.* **111**, 053202 (2013). [doi:10.1103/PhysRevLett.111.053202](https://doi.org/10.1103/PhysRevLett.111.053202) [Medline](#)
17. See Supplementary Materials.
18. D. S. Hall, M. R. Matthews, J. R. Ensher, C. E. Wieman, E. A. Cornell, Dynamics of component separation in a binary mixture of Bose-Einstein condensates. *Phys. Rev. Lett.* **81**, 1539–1542 (1998). [doi:10.1103/PhysRevLett.81.1539](https://doi.org/10.1103/PhysRevLett.81.1539)
19. C. C. Bradley, C. A. Sackett, R. G. Hulet, Bose-Einstein condensation of Lithium: Observation of limited condensate number. *Phys. Rev. Lett.* **78**, 985–989 (1997). [doi:10.1103/PhysRevLett.78.985](https://doi.org/10.1103/PhysRevLett.78.985)
20. Here and in the following, error bars for the scattering lengths correspond to the 9 mG experimental uncertainty of the magnetic field and do not take into account the systematic uncertainties of the scattering length model.
21. F. Chevy, C. Salomon, Strongly correlated Bose gases. *J. Phys. B* **49**, 192001 (2016). [doi:10.1088/0953-4075/49/19/192001](https://doi.org/10.1088/0953-4075/49/19/192001)
22. R. N. Bisset, R. M. Wilson, D. Baillie, P. B. Blakie, Ground-state phase diagram of a dipolar condensate with quantum fluctuation. *Phys. Rev. A* **94**, 033619 (2016). [doi:10.1103/PhysRevA.94.033619](https://doi.org/10.1103/PhysRevA.94.033619)
23. F. Wächtler, L. Santos, Ground-state properties and elementary excitations of quantum droplets in dipolar Bose-Einstein condensates. *Phys. Rev. A* **94**, 043618 (2016). [doi:10.1103/PhysRevA.94.043618](https://doi.org/10.1103/PhysRevA.94.043618)
24. C. D'Errico, M. Zaccanti, M. Fattori, G. Roati, M. Inguscio, G. Modugno, A. Simoni, Feshbach resonances in ultracold³⁹. *New J. Phys.* **9**, 223 (2007). [doi:10.1088/1367-2630/9/7/223](https://doi.org/10.1088/1367-2630/9/7/223)
25. S. Falke, H. Knöckel, J. Friebe, M. Riedmann, E. Tiemann, C. Lisdat, Potassium ground-state scattering parameters and Born-Oppenheimer potentials from molecular spectroscopy. *Phys. Rev. A* **78**, 012503 (2008). [doi:10.1103/PhysRevA.78.012503](https://doi.org/10.1103/PhysRevA.78.012503)
26. V. Cikojević, K. Dželalija, P. Stipanović, L. Vranješ Markić, J. Boronat, Ultradilute quantum liquid drops. *arXiv:1708.01553*.
27. D. Baillie, R. M. Wilson, P. B. Blakie, Collective excitations of self-bound droplets of a dipolar quantum fluid. *arXiv:1703.07927*.
28. D. S. Petrov, G. E. Astrakharchik, Ultradilute low-dimensional liquids. *Phys. Rev. Lett.* **117**, 100401 (2016). [doi:10.1103/PhysRevLett.117.100401](https://doi.org/10.1103/PhysRevLett.117.100401) [Medline](#)
29. M. R. Matthews, B. P. Anderson, P. C. Haljan, D. S. Hall, M. J. Holland, J. E. Williams, C. E. Wieman, E. A. Cornell, Watching a superfluid untwist itself: Recurrence of Rabi oscillations in a Bose-Einstein condensate. *Phys. Rev. Lett.* **83**, 3358–3361 (1999). [doi:10.1103/PhysRevLett.83.3358](https://doi.org/10.1103/PhysRevLett.83.3358)
30. D. S. Petrov, Three-body interacting bosons in free space. *Phys. Rev. Lett.* **112**, 103201 (2014). [doi:10.1103/PhysRevLett.112.103201](https://doi.org/10.1103/PhysRevLett.112.103201) [Medline](#)
31. A. Cappellaro, T. Macrì, G. F. Bertacco, L. Salasnich, Equation of state and self-bound droplet in Rabi-coupled Bose mixtures. *Sci. Rep.* **7**, 13358 (2017). [doi:10.1038/s41598-017-13647-y](https://doi.org/10.1038/s41598-017-13647-y) [Medline](#)
32. G. Semeghini, G. Ferioli, L. Masi, C. Mazzinghi, L. Wolswijk, F. Minardi, M. Modugno, G. Modugno, M. Inguscio, M. Fattori, Self-bound quantum droplets in atomic mixtures. *arXiv:1710.10890 [cond-mat.quant-gas]* (30 October 2017).
33. K. Hueck, N. Luick, L. Sobirey, J. Siegl, T. Lompe, H. Moritz, Two-dimensional homogeneous Fermi gases. *arXiv:1704.06315*.
34. T. Weber, J. Herbig, M. Mark, H.-C. Nägerl, R. Grimm, Three-body recombination at large scattering lengths in an ultracold atomic gas. *Phys. Rev. Lett.* **91**, 123201 (2003). [doi:10.1103/PhysRevLett.91.123201](https://doi.org/10.1103/PhysRevLett.91.123201) [Medline](#)
35. Y. Kagan, B. V. Svistunov, G. V. Shlyapnikov, Effect of Bose condensation on inelastic processes in gases. *JETP Lett.* **42**, 209 (1985).
36. P. Cheiney, C. R. Cabrera, J. Sanz, B. Naylor, L. Tanzi, L. Tarruell, Bright soliton to quantum droplet transition in a mixture of Bose-Einstein condensates. *arXiv:1710.11079*.
37. M. Gajdacz, P. L. Pedersen, T. Mørch, A. J. Hilliard, J. Arlt, J. F. Sherson, Non-destructive Faraday imaging of dynamically controlled ultracold atoms. *Rev. Sci. Instrum.* **84**, 083105 (2013). [doi:10.1063/1.4818913](https://doi.org/10.1063/1.4818913) [Medline](#)
38. E. V. Goldstein, P. Meystre, Quasiparticle instabilities in multicomponent atomic condensates. *Phys. Rev. A* **55**, 2935–2940 (1997). [doi:10.1103/PhysRevA.55.2935](https://doi.org/10.1103/PhysRevA.55.2935)

ACKNOWLEDGMENTS

We acknowledge insightful discussions with G. Astrakharchik, J. Boronat, A. Celi, I. Ferrier-Barbut, M. A. García-March, M. Lewenstein, P. Massignan, D. Petrov, and A. Recati. We thank A. Simoni and M. Tomza for calculations of the potassium scattering lengths. We are grateful to M. Bosch, V. Brunaud, V. Lienhard, A. Muñoz, L. Saemisch, J. Sastre, and I. Uriaga for experimental assistance during the construction of the apparatus. We acknowledge funding from Fundació Privada Cellex, EU (MagQUPT-631633 and QUIC-641122), Spanish MINECO (StrongQSIM FIS2014-59546-P and Severo Ochoa SEV-2015-0522), DFG (FOR2414), Generalitat de Catalunya (SGR874 and CERCA program), and Fundación BBVA. CRC acknowledges support from CONACYT (402242/384738), JS from FPI (BES-2015-072186), PC from the Marie Skłodowska-Curie actions (TOPDOL-657439) and L. Tarruell from the Ramón y Cajal program (RYC-2015-17890). The experimental data presented in the figures is available at: <http://upcommons.upc.edu/handle/2117/111148> Other materials may be requested from the corresponding author.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aao5686/DC1
Supplementary Text
Figs. S1 and S2
References (33–38)

4 August 2017; accepted 28 November 2017
Published online 14 December 2017
[10.1126/science.aao5686](https://doi.org/10.1126/science.aao5686)

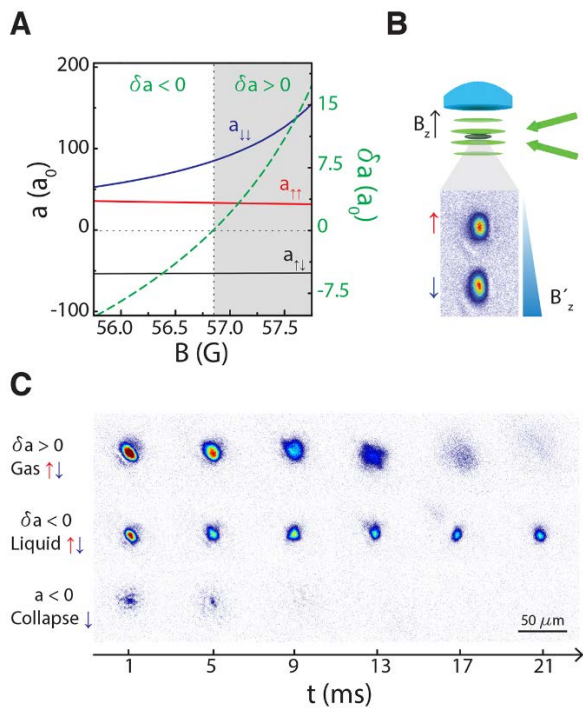


Fig. 1. Observation of quantum droplets. (A) Scattering lengths a (solid lines) and parameter $\delta a = a_{\uparrow\downarrow} + \sqrt{a_{\uparrow\uparrow}a_{\downarrow\downarrow}}$ (dashed line) vs. magnetic field B for a ^{39}K mixture in states $|\uparrow\rangle \equiv |F, m_F\rangle = |1, -1\rangle$ and $|\downarrow\rangle \equiv |1, 0\rangle$. The condition $\delta a = 0$ (dashed vertical line) separates the repulsive ($\delta a > 0$, grey area) and attractive ($\delta a < 0$, white area) regimes. (B) Schematic view of the experiment. Atoms are prepared in a plane of a blue-detuned optical lattice created by two beams intersecting at a small angle, and imaged in situ with a high numerical aperture objective [$\leq 0.97(4)$ μm measured resolution, $1/e$ Gaussian width). The spin composition of the system is verified independently via Stern-Gerlach separation by a magnetic field gradient during time-of-flight expansion. During the preparation sequence, a red-detuned optical dipole trap (not shown) provides radial confinement (17). (C) Typical in situ images taken at time t after removal of the radial confinement but in the presence of the lattice potential. Top row: expansion of a gaseous mixture [$B = 56.935(9)$ G and $\delta a = 1.2(1) a_0 > 0$]. Central row: formation of a self-bound mixture droplet [$B = 56.574(9)$ G and $\delta a = -3.2(1) a_0 < 0$]. Bottom row: collapse of a single-component $|\downarrow\rangle$ attractive condensate [$B = 42.281(9)$ G and $a = -2.06(2) a_0 < 0$]. In our geometry, quantum pressure cannot stabilize bright solitons. Therefore, the existence of self-bound liquid droplets is a direct manifestation of beyond mean-field effects.

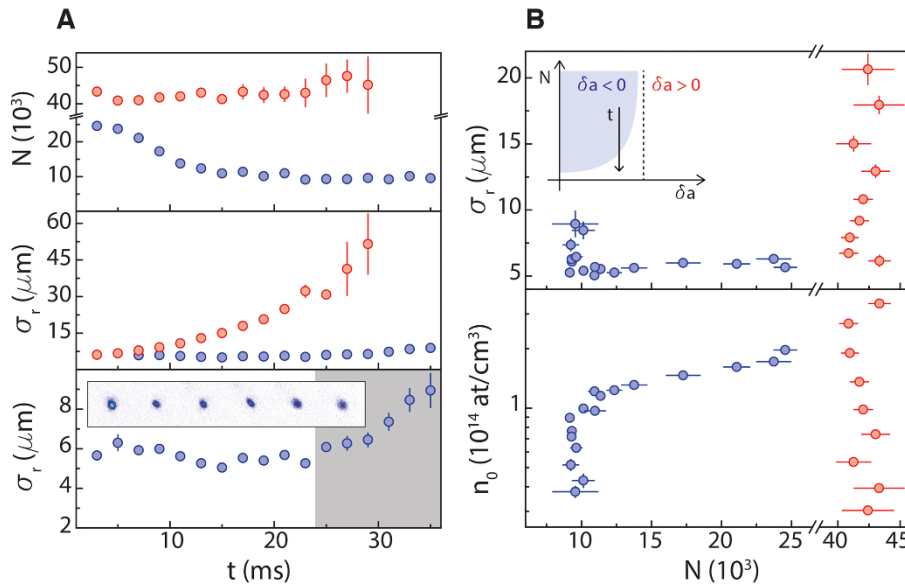


Fig. 2. Liquid-to-gas transition. (A) Atom number N and radial size σ_r of the mixture for different evolution times t . The measurements are taken in the repulsive [$\delta a = 1.2(1) a_0 > 0$, red circles) and attractive ($\delta a = -3.2(1) a_0$, blue circles) regimes. Top panel: while for $\delta a > 0$ the atom number in the gas remains constant, for $\delta a < 0$ it decreases on a timescale compatible with three-body recombination (17). Central panel: the radial size of the droplet remains constant at $\sigma_r \approx 6 \mu\text{m}$, demonstrating its self-bound nature. In contrast, the size of the gas increases continuously with time. Bottom panel: closer view of σ_r for $\delta a < 0$. For $t > 25$ ms the droplet dissociates and a liquid-to-gas transition takes place. The inset displays images taken at $t = 25-35$ ms in 2 ms increments, corresponding to the points in the grey area. (B) Radial size σ_r (top panel) and peak density n_0 (bottom panel) vs. N . For $\delta a < 0$ and large atom number both remain approximately constant, as expected for a liquid. For a critical atom number σ_r rises suddenly and n_0 plunges, signaling the liquid-to-gas transition. In the gas phase, the $\delta a < 0$ system behaves like the $\delta a > 0$ one. Inset (top panel): sketch of the phase diagram. In the liquid phase (blue region), observing the mixture at variable evolution times gives access to different values of N (black arrow). For all panels, error bars represent the standard deviation of 10 independent measurements. If not displayed, error bars are smaller than the size of the symbol. Additionally, N has a calibration uncertainty of 25% (17).

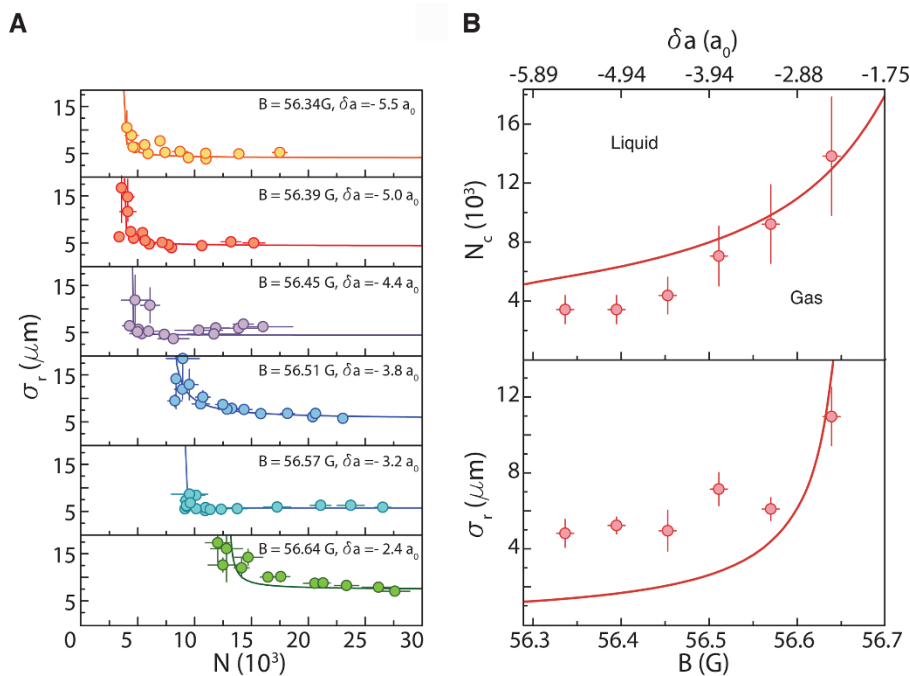


Fig. 3. Liquid-to-gas phase diagram.

(A) Radial size of the mixture σ_r as a function of atom number N for different magnetic fields B , from strong to weak attraction (top to bottom). The critical atom number N_c increases as attraction decreases. Solid lines display the phenomenological fit $\sigma_r(N) = \sigma_0 + A/(N - N_c)$ used to locate the liquid-to-gas phase transition. (B) N_c (top panel) and σ_r for fixed $N = 1.5(1) \times 10^4$ (bottom panel) as a function of B . The upper horizontal axis shows the corresponding values of δa . Solid lines are the predictions of an extended Gross-Pitaevskii model without fitting parameters (see text). Error bars for σ_r correspond to the standard deviation of 10 independent measurements. If not displayed, error bars are smaller than the size of the symbol. Error bars for B and N show the systematic uncertainty of the corresponding calibrations (17).

INDUCED SEISMICITY

Hydraulic fracturing volume is associated with induced earthquake productivity in the Duvernay play

R. Schultz,^{1*} G. Atkinson,² D. W. Eaton,³ Y. J. Gu,⁴ H. Kao⁵

A sharp increase in the frequency of earthquakes near Fox Creek, Alberta, began in December 2013 in response to hydraulic fracturing. Using a hydraulic fracturing database, we explore relationships between injection parameters and seismicity response. We show that induced earthquakes are associated with completions that used larger injection volumes (10^4 to 10^5 cubic meters) and that seismic productivity scales linearly with injection volume. Injection pressure and rate have an insignificant association with seismic response. Further findings suggest that geological factors play a prominent role in seismic productivity, as evidenced by spatial correlations. Together, volume and geological factors account for ~96% of the variability in the induced earthquake rate near Fox Creek. This result is quantified by a seismogenic index–modified frequency-magnitude distribution, providing a framework to forecast induced seismicity.

Subsurface injection of fluid may induce earthquakes (1) through anthropogenic alteration of crustal stresses (2). In the case of hydraulic fracturing (HF), high-pressure injection of fluid intended to increase the permeability of tight shales has been known to trigger earthquakes (3), some of which are large enough to be recorded or felt regionally (4–8). Within the Western Canada Sedimentary Basin, the recent increase in seismicity has been largely attributed to HF (9). Moreover, earthquakes in the Duvernay Formation (10) have been among the largest-magnitude events caused by HF completions globally (11, 12). These events have appreciably increased the seismic hazard in the area, and felt ground motions have resulted in the implementation of a traffic light protocol (TLP) (13).

Despite recent progress in characterizing the Duvernay-related earthquakes (14, 15), many scientifically critical questions have remained unresolved. For example, it is not clear why there was a large time delay (~3 years) between the first Duvernay play completion and the initiation of HF-related earthquakes (14), after which time such earthquakes became frequent occurrences. Moreover, it is not well understood why only a small subset of operations appear to be seismogenic (9). Within the Duvernay play specifically, only those completions located in the Kaybob region are seismogenic, whereas all HF completions in the Willesden Green and Edson regions have been seismically quiescent (fig. S1).

Geological factors have been suggested to contribute to the spatial distribution of these earthquakes (16, 17). However, to date, there has been little understanding or documentation of the relative contributions of surface injection pressure, rate, and volume to the seismogenic process.

To address these questions, we first examined the timing and location of earthquakes in relation to HF completions in the seismogenic Kaybob region of the Duvernay play (Fig. 1). We compiled a database of all (~300) horizontal HF well completions in the Kaybob Duvernay up to February 2016 (because of the ~1-year period of confidentiality) from public, hard-copy regulator records. Individual wells are aggregated into ~180 well “pads” on the basis of the proximity of multiple wells oriented along similar trajectories (fig. S2). A cursory examination of the time-averaged evolution of injected volumes per pad (Fig. 1B) indicates increasing pad design complexity and HF completion volumes, typical of maturing development in shale plays (18). The Kaybob Duvernay has been injected with more than 8.5×10^6 m³ of fracturing fluid (as of February 2016) to stimulate well productivity. The observation of a relative increase in pad volumes before the first recorded earthquakes suggests that injected volume may be a controlling factor in the Kaybob Duvernay earthquake activity.

Disposal is not likely to be a major factor in the induced seismicity in this area. The closest water-injection well to Crooked Lake that was actively injecting at depths similar to those of the Duvernay during the seismogenic period is ~35 km away. Disposal wells within a 50-km radius of Crooked Lake injected only $\sim 1.2 \times 10^6$ m³ of fluid during the Duvernay’s seismogenic period (much less than the total volume involved in the HF operations).

To examine the role of operational factors more closely, we associated clusters of seismicity with seismogenic pads on the basis of a

spatiotemporal association filter (SAF), which identifies the causally closest pad(s) in time and then space (supplementary materials). The validity of these associations can be demonstrated by comparison with prior case studies (14, 15) and direct communication with the regulator and the anonymous companies responsible. Based on the SAF, ~10% of the pads and ~15% of the wells in the Kaybob Duvernay are associated with seismicity. Of the completions associated with induced earthquakes, ~50% are single-well pads. Using this subset of associated pads, we then contrasted operational parameters at seismogenic pads, wells, and stages with those of their parent distributions, as derived from the entire Kaybob region of the Duvernay play (Fig. 2). Statistical distributions of operational parameters (fig. S3) were analyzed using the Kolmogorov-Smirnov (KS) test (19) to discern whether it is likely that the seismogenic distributions are sampled randomly from their parent distribution. In these tests, we used the standard significance level of 0.05. The computed *P* values (Fig. 2) allowed a rejection of the hypothesis that the seismogenic distributions are subsampled from the parent Kaybob Duvernay distribution. On the basis of finding nonrandomness, the Mann-Whitney (MW) *U* test (20) was applied next to determine whether the seismogenic subsets have significantly larger median values of key operational parameters than do their parent distributions. We performed both one- and two-tailed MW tests to first determine whether subset median values are different from the parent distribution (two-tailed) and then further assess whether the subset median values are larger than their parent distributions (one-tailed). This analysis (*P* < 0.05) demonstrated that seismogenic pads, wells, and stages are associated with larger injected volumes at a statistically significant level (Fig. 2). Analogously, an analysis of HF in the Horn River Basin found a relationship between induced earthquake productivity and volume (21). Examination of pad, well, and stage pressures suggests no significantly compelling association (fig. S3). Although injection rate has been suggested as a driving factor for disposal-induced seismicity in the central United States (22), we did not observe a meaningful association with injection rates. Potentially, this discrepancy could be due to differences in injection operations for disposal and HF; for example, the slowest Duvernay injection rates are nearly an order of magnitude faster than the critical rate threshold identified for seismogenic disposal (22). Thus, the effects of rate on HF-induced seismicity in the Kaybob region are either secondary, indiscernible, or negligible. The robustness of these findings was established using bootstrap (23) resampling sensitivity tests (supplementary materials and figs. S4, S5, and S6).

The results of the KS and MW tests show that volume is a controlling factor for HF-induced earthquakes in the Kaybob Duvernay. Potentially, this observation may be the result of greater injection volumes allowing for larger stimulated reservoir volumes and thus greater

¹Alberta Geological Survey, 4999 98 Avenue, Edmonton, Alberta, Canada. ²Western University, 1151 Richmond Street, London, Ontario, Canada. ³University of Calgary, 2500 University Drive, Calgary, Alberta, Canada. ⁴University of Alberta, 116 Street and 85 Avenue, Edmonton, Alberta, Canada. ⁵Natural Resources Canada, 9860 West Saanich Road, Sidney, British Columbia, Canada.

*Corresponding author. Email: ryan.schultz@aer.ca

likelihood of intersecting a critically stressed fault (24). However, this does not necessarily imply that it controls the maximum possible magnitude. This finding echoes similar conclusions from other induced cases (25). The relationship between volume and induced earthquakes has important implications for the management of HF-related seismic hazard. For example, controls on seismicity related to volume would be affected by the time-dependent nature of increasing pad completion volumes within a maturing play (Fig. 1B and fig. S7). On the other hand, the use of lower pad completion volumes

since the first TLP red light (Fig. 1B) could have resulted in an overall reduction in earthquake response from the Duvernay play. Whether this decrease in volume after the first TLP red light resulted in a net decrease in seismic hazard is more complicated, however, because hazard is dependent on multiple factors, some of which are independent of injection volume (26, 27). These points would require a physical model to validate the statistical earthquake-volume association.

To validate this relationship between injected volume and earthquakes (28), we first considered

the Gutenberg-Richter frequency-magnitude distribution (GR-FMD) (29, 30)

$$N_M = 10^a 10^{-bM} \quad (1)$$

In this formulation, N_M is the number of events greater than magnitude M , the a -value governs the rate at which earthquakes occur, and the b -value measures the proportion of relatively smaller events to larger ones. To consider the time-varying rate of induced earthquakes, modifications have been suggested on the basis of solutions to the diffusion equation that incorporate

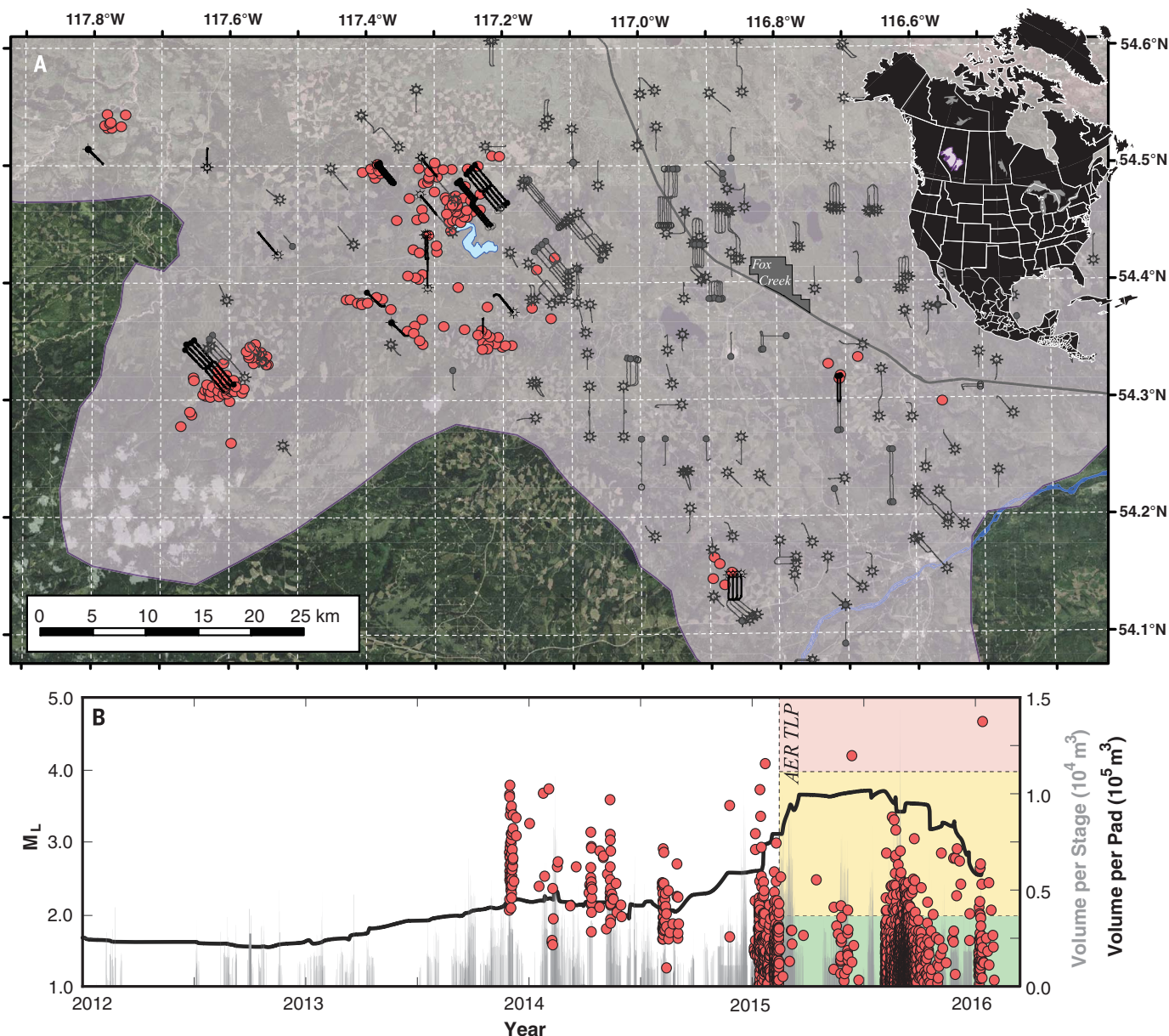


Fig. 1. HF-induced seismicity within the Duvernay play up to February 2016. (A) Spatial distribution of induced earthquakes (red circles) associated with HF wells (dark gray “tadpoles”; associated wells are bolded and black) in the Duvernay play (purple shaded area) near Fox Creek (dark gray) and Crooked Lake (blue). The inset map shows the location of the Duvernay

Formation in North America. **(B)** Timing and magnitude of induced earthquakes (red circles) alongside stage volumes (gray) and a 30-pad running average of pad volumes (black line). Colored areas indicate Alberta Energy Regulator (AER) traffic light protocol (TLP) cut-offs (13). Initial wells stimulated during March 2010 were included in the analysis but are not depicted.

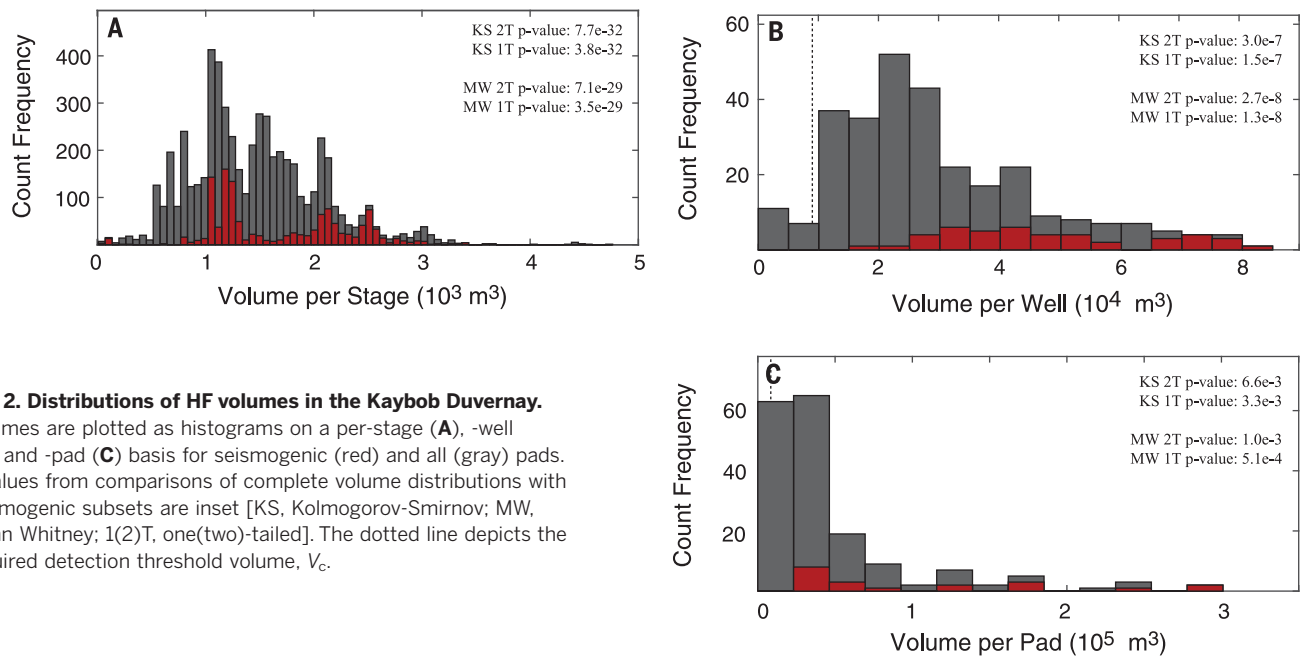


Fig. 2. Distributions of HF volumes in the Kaybob Duvernay.

Volumes are plotted as histograms on a per-stage (A), -well (B), and -pad (C) basis for seismogenic (red) and all (gray) pads. *P* values from comparisons of complete volume distributions with seismogenic subsets are inset [KS, Kolmogorov-Smirnov; MW, Mann Whitney; 1(2)T, one(two)-tailed]. The dotted line depicts the required detection threshold volume, V_c .

the nonstationary effects of injection volume $V(t)$ —e.g., $a = \Sigma + \log_{10}[V(t)]$ (25, 31–33). In this equation, Σ is the seismogenic index, an injection-invariant parameter that represents the seismotectonic response to increasing fluid pressure within the study area (25). Incorporation of Σ in the GR-FMD modifies Eq. 1 to (25, 31–33)

$$N_M = V(t) \cdot 10^{\Sigma} 10^{-bM} \quad (2)$$

We investigated the implications of Eq. 2 for the Kaybob Duvernay HF earthquakes by first compiling the time-dependent histories of completion volumes and the number of earthquakes above the detection threshold (fig. S8). We observed a strong correlation between these variables, especially if only the SAF-associated pads were considered (Fig. 3). The best-fit parameters of Eq. 2 are $b = 0.90 \pm 0.03$ and $\Sigma = -1.8 \pm 0.2$ for the detection threshold of local magnitude (M_L) 1.3 (supplementary materials and fig. S8). This regionally averaged Σ value near Fox Creek is very similar to Σ values computed for earthquakes in the Brazeau Cluster, which are related to wastewater disposal in the Cordell Field (-2.1 ± 0.2) (34), and to Σ values for numerous case studies worldwide (35). This is an important finding because it suggests that the regional seismic response to fluid injection in central Alberta is similar for both HF and disposal wells.

We repeated the fitting process for individual clusters associated with seismogenic wells and found Σ and b -values that vary from -2.5 to -0.5 and 0.7 to 1.7 , respectively (fig. S9). The relatively high variability of these parameters likely reflects the difficulty in making robust parameter determinations from small data subsets. More complete earthquake catalogs would enable a more robust analysis of seismic parameters for the

individual clusters, improving confidence and possibly reducing variability. Overall, we found seismic values for individual clusters that are roughly similar to Σ values previously determined for clusters on a local array (15). The determination of these values provides a powerful tool with which to forecast the expected number of future earthquakes at seismogenic wells and their likely magnitude distribution.

In light of these findings, we consider the hypothesis that the ~ 3 -year-delayed response in earthquake productivity was simply the result of the minimum injection volume (V_c) required to raise the seismicity rate to a sufficient level for observation (i.e., so that it produces an earthquake larger than the regional detection limit). Conservative upper-bound estimates using a detection limit of M_L 2.0 (the TLP requires operators to report all events of M_L 2.0 and greater) and 95% probability of exceedance suggest a V_c of $(8.0 \pm 0.2) \times 10^3 \text{ m}^3$ (supplementary materials). This number is comparable to the average total injected volume at pads at the time of their first corresponding earthquakes, $V_1 = (1.0 \pm 1.0) \times 10^4 \text{ m}^3$. Although V_c and V_1 roughly agree, it is likely that V_1 is systematically larger than V_c owing to insufficient resolution with which to discern aseismic stages or wells within a seismogenic pad. Corroborating these results, studies in the Horn River Basin found that seismic response appeared to “turn on” during months where HF injection volumes were greater than $2.0 \times 10^4 \text{ m}^3$ (21). Because fewer than 10 pads in the Kaybob Duvernay have a volume per pad less than V_c , we argue that it is unlikely that the spatial distribution of our catalog has been seriously biased by V_c . We justify this claim through comparison of our regional catalog (14, 36) with catalogs supplemented by local operator networks (15). In both cases, we observe

the same spatial distributions of earthquakes and the same associations with seismogenic wells and regions.

Although pad completion volumes have been increasing with time, it is interesting that more than 20 pads ($\sim 50\%$) that were completed before the first induced earthquake have volumes greater than $1.5 \times 10^4 \text{ m}^3$. This finding indicates that although volume appears to be a controlling factor for Kaybob Duvernay induced seismicity, other factors are also playing a nontrivial role. Prior work has suggested that the spatial distribution of all induced seismicity within central Alberta has been influenced by geological factors, with earthquakes preferentially occurring along underlying reef margins (16, 37) or within regions of relatively higher formation overpressure (17).

Following this rationale, we accommodate spatial factors using a modification $\Sigma = \Sigma' + \log_{10}[\delta(\bar{r})]$ that explicitly introduces a spatial variable \bar{r} so that Eq. 2 becomes

$$N_M = V(t) \cdot \delta(\bar{r}) \cdot 10^{\Sigma} 10^{-bM} \quad (3)$$

In our formulation, we define $\delta(\bar{r})$ to have binary values of 1 or 0 to indicate regions that do or do not experience earthquakes, respectively. In this limiting case, Σ' becomes equivalent to Σ [i.e., when $\delta(\bar{r}) = 1$]. Introducing this spatial term is a useful refinement, because numerous plays have been stimulated using similar per-well volumes (38, 39) with limited or undocumented seismic response. Even considering only the Duvernay, induced earthquakes have been restricted to one area, the Kaybob region (fig. S1). Furthermore, numerous high-volume Kaybob pads, wells, and stages are not associated with earthquakes (Fig. 2). In fact, the first seismogenic pads observed in the Duvernay (10) were also the first pads to complete in the most seismically susceptible region,

~10 km from Crooked Lake (fig. S7). The efficacy of introducing this spatial term for the Kaybob Duvernay is quantified by the significant enhancement of goodness-of-fit to the data when using the SAF (Fig. 3). In this sense, the SAF represents a rudimentary and empirical estimation of $\delta(\bar{r})$. Taken in the context of the original formulation, the spatial variability of Σ can be cast as $\Sigma = -1.8 + \log_{10}(\text{SAF})$. The incorporation of the SAF as an empirical estimate of $\delta(\bar{r})$ in a Σ -modified GR-FMD results in a model that accounts for ~96% of the seismic response variability within the Kaybob data set (Fig. 3B). This means that complexities related to other potential factors account for only the remaining 4% of variability or constitute part of the proposed $\delta(\bar{r})$. This is an important finding because these additional factors are numerous; they include well flow-back, effects of staged stimulation pressures,

well shut-in response, lagged seismicity relative to seismogenic pads, petroleum production, earthquake aftershock sequences, operator seismic mitigation strategies, varieties of fracturing fluids and proppants, SAF resolution limitations, interpad communication, poroelastic triggering effects, magnitude uncertainties, and spatiotemporal variability in b -values or Σ' . This observation is further confirmed by applying additional filtering to the data (figs. S10 and S11).

Although in reality, a fault is either reactivated or not, incomplete information about the subsurface often prevents definitive assessments a priori. Our spatial parameter is a useful concept in this regard. Statistically, we can extend $\delta(\bar{r})$ to represent the seismogenic activation potential, defined as the likelihood of a well inducing a detectable earthquake at a given location. We interpret this parameter as the probabilistic

intersection of all geological conditions required to cause an induced earthquake at a given location—i.e., the spatial variability of the geological susceptibility to induced seismicity. This interpretation is intuitive, owing to the derivation of $\delta(\bar{r})$ from Σ , which constrains the seismotectonic state of the point of injection. For a demonstration of this interpretation, we consider the effects of distance to underlying fossil reef margins (16, 37) and formation overpressure (17) as regional proxies for faulting and stress, respectively. The fractions of seismogenically associated pads as a function of these proxies are plotted (Fig. 4), confirming that regions of development that are closer to the reef margin or more highly overpressured have been more likely to induce earthquakes (fig. S12). For western Canada, an entire basin-wide average of this activation probability in regions that are also coincident with viable HF plays appears to be less than 0.3% (9).

Solely on the basis of the Σ modifications, it is possible to improve hazard estimates for future injections in real time during stimulation. Knowledge of fault size, hydraulic connectivity to proximal stimulation stages, and injected volumes that may be directed to reactivated faults can serve as input to calibrated models to estimate induced earthquake rates and magnitude distributions. Conversely, microseismic monitoring of HF completions may be used to scrutinize the number and rate of induced events resulting from individual stage stimulations as an indicator of hydraulic connectivity to nearby faults. For example, HF pads that are oriented subparallel to the north-south-oriented fault planes (12, 14, 15) are likely to be in extended hydraulic communication with seismogenic faults as compared with northeast-southwest-oriented pads. This rationale may explain why the three largest-magnitude clusters occurred at north-south-oriented pads: Greater volume pumped into these faults would allow for more numerous induced events and thus an increased likelihood of a larger event.

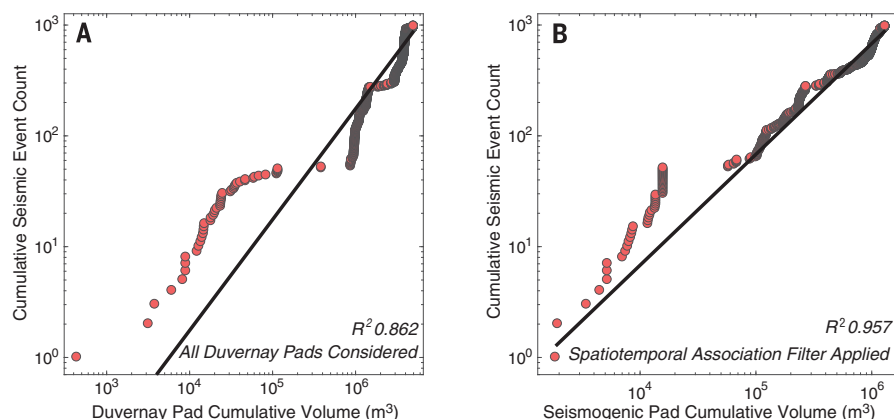


Fig. 3. Number of earthquakes above the detection threshold ($M_L 1.3$) versus cumulative injection volume. (A) Volumes from all HF operations within the Kaybob Duvernay play (red circles) are compared with the best fit of the data points (black line). **(B)** Analogous to (A), except only seismogenic HF pads are considered. Both panels are during a period where the detection threshold remains constant (July 2014 to February 2016). Well flow-back was not considered in computing volumes. In both panels, the goodness-of-fit of the data to the expected line is displayed with the R^2 (coefficient of determination) values.

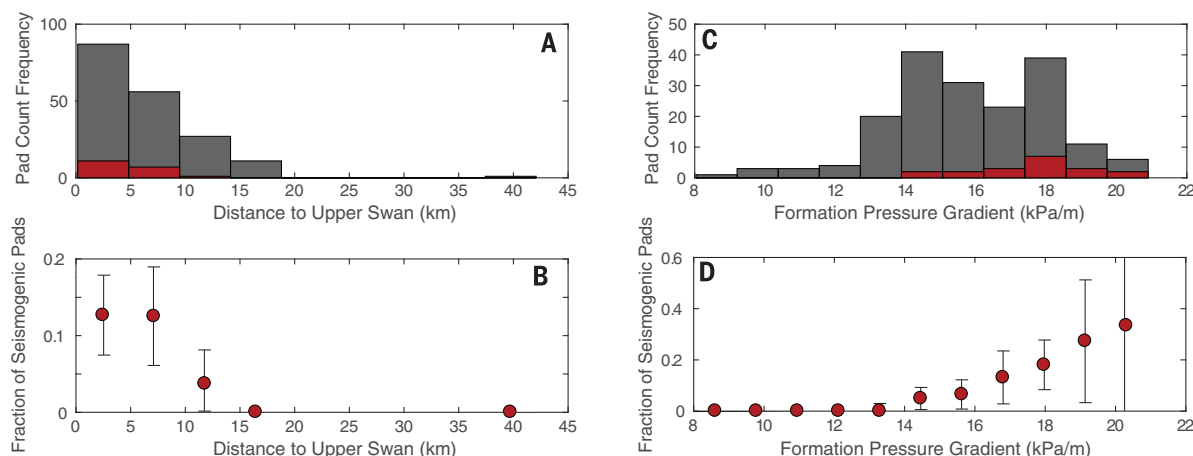


Fig. 4. Statistics of HF operations in relation to the Swan Hills Formation and Duvernay overpressure. (A) Distance of HF operations to the Swan Hills Upper Bank edge for seismogenic (red) and all (gray) pads, shown as a histogram. **(B)** Similar to (A), except each bin has been

normalized (red circles), and error bars have been added. **(C)** Pressure gradient for seismogenic (red) and all (gray) pads. **(D)** Fraction of seismogenic pads as a function of pressure gradient (red circles) with error bars from the standard deviation.

Extended to the regional scale, this study can be used to better inform earthquake rate models in induced seismic hazard forecasts. Coupled with an estimation of the seismogenic activation potential, our proposed framework (Eq. 3) would allow for the quantification of both induced earthquake rate and location models. Rate and location models are some of the most critical parameters in forecasting hazard related to induced earthquakes (26, 27). Although this study has focused on the Fox Creek area, the proposed framework can be applied to other jurisdictions to improve the management of induced seismic hazard.

We find that the most important operational parameter controlling induced earthquakes in the Duvernay play near Fox Creek is injected volume, which scales linearly with the total number of earthquakes in a Σ -modified Gutenberg-Richter formulation. Conversely, injection pressure and rate appear unrelated to induced seismicity response near Fox Creek. Furthermore, wells that exhibit seismicity appear to display a strong spatial bias related to geological factors, which has a pronounced effect on the resultant seismic response. Last, this study provides a framework with which to incorporate the seismogenic activation potential into seismic hazard analysis for induced earthquakes.

REFERENCES AND NOTES

- C. B. Raleigh, J. H. Healy, J. D. Bredehoeft, *Science* **191**, 1230–1237 (1976).
- P. Segall, S. Lu, *J. Geophys. Res. Solid Earth* **120**, 5082–5103 (2015).
- S. C. Maxwell *et al.*, “Fault activation during hydraulic fracturing,” in *SEG Technical Program Expanded Abstracts 2009* (Society of Exploration Geophysicists, 2009), pp. 1552–1556.
- H. Clarke, L. Eisner, P. Styles, P. Turner, *Geophys. Res. Lett.* **41**, 8308–8314 (2014).
- A. M. Farahbod, H. Kao, D. W. Walker, J. F. Cassidy, *Can. J. Earth Sci.* **52**, 112–122 (2015).
- R. J. Skoumal, M. R. Brudzinski, B. S. Currie, *Bull. Seismol. Soc. Am.* **105**, 189–197 (2015).
- R. Schultz *et al.*, *Bull. Seismol. Soc. Am.* **105**, 2871–2884 (2015).
- A. Babaie Mahani *et al.*, *Bull. Seismol. Soc. Am.* **107**, 101785/0120160175 (2017).
- G. M. Atkinson *et al.*, *Seismol. Res. Lett.* **87**, 631–647 (2016).
- R. Schultz, V. Stern, M. Novakovic, G. M. Atkinson, Y. J. Gu, *Geophys. Res. Lett.* **42**, 2750–2758 (2015).
- R. Wang, Y. J. Gu, R. Schultz, A. Kim, G. M. Atkinson, *Geophys. Res. Lett.* (2016).
- R. Wang, Y. J. Gu, R. Schultz, M. Zhang, A. Kim, *Geophys. J. Int.* **10.1093/gji/ggx204** (2017).
- Alberta Energy Regulator, *Subsurface Order No 2: Monitoring and Reporting of Seismicity in the Vicinity of Hydraulic Fracturing Operations in the Duvernay Zone, Fox Creek, Alberta* (AER Bulletin 2015-07, 2015); <https://aer.ca/documents/orders/subsurface-orders/SO2.pdf>.
- R. Schultz, R. Wang, Y. J. Gu, K. Haug, G. M. Atkinson, *J. Geophys. Res. Solid Earth* **122**, 492–505 (2017).
- X. Bao, D. W. Eaton, *Science* **354**, 1406–1409 (2016).
- R. Schultz *et al.*, *Geophys. Res. Lett.* **43**, 2534–2542 (2016).
- D. Eaton, “Dynamics of fault activation by hydraulic fracturing in overpressured shales,” in *79th EAGE Conference and Exhibition 2017 - Workshops* (European Association of Geoscientists & Engineers, 2017); available at <http://earthdoc.eage.org/publication/publicationdetails/?publication=89290>.
- G. E. King, “Thirty years of gas shale fracturing: What have we learned?” in *SPE Annual Technical Conference and Exhibition* (Society of Petroleum Engineers, 2010).
- F. J. Massey Jr., *J. Am. Stat. Assoc.* **46**, 68–78 (1951).
- J. D. Gibbons, S. Chakraborti, *Nonparametric Statistical Inference* (Chapman & Hall/CRC Press, Taylor & Francis Group, ed. 5, 2011).
- A. M. Farahbod, H. Kao, J. F. Cassidy, D. Walker, *Leading Edge* **34**, 658–663 (2015).
- M. Weingarten, S. Ge, J. W. Godt, B. A. Bekins, J. L. Rubinstein, *Science* **348**, 1336–1340 (2015).
- B. Efron, R. Tibshirani, *Stat. Sci.* **1**, 54–75 (1986).
- R. F. Westwood, S. M. Toon, N. J. Cassidy, *Fuel* **203**, 843–852 (2017).
- S. A. Shapiro, *Fluid-Induced Seismicity* (Cambridge Univ. Press, 2015).
- G. M. Atkinson, H. Ghofrani, K. Assatourians, *Seismol. Res. Lett.* **86**, 1009–1021 (2015).
- M. D. Petersen *et al.*, *Incorporating Induced Seismicity in the 2014 United States National Seismic Hazard Model—Results of 2014 Workshop and Sensitivity Studies* (Open-File Report 2015–1070, U.S. Geological Survey, 2015).
- A. McGarr, *J. Geophys. Res. Solid Earth* **119**, 1008–1019 (2014).
- M. Ishimoto, K. Iida, *Bull. Earthq. Res. Inst.* **17**, 391 (1939).
- W. Marzocchi, L. A. Sandri, *Ann. Geophys.* **46**, 1271–1282 (2003).
- S. A. Shapiro, C. Dinske, C. Langenbruch, F. Wenzel, *Leading Edge* **29**, 304–309 (2010).
- T. Hajati, C. Langenbruch, S. A. Shapiro, *Geophys. Res. Lett.* **42**, 10601–10606 (2015).
- C. Langenbruch, M. D. Zoback, *Sci. Adv.* **2**, e1601542 (2016).
- R. Schultz, V. Stern, Y. J. Gu, *J. Geophys. Res. Solid Earth* **119**, 3410–3423 (2014).
- C. Dinske, S. A. Shapiro, *J. Seismol.* **17**, 13–25 (2013).
- V. Stern, R. Schultz, L. Shen, Y. J. Gu, D. W. Eaton, *Alberta Earthquake Catalogue, Version 1.0, September 2006 Through December 2010* (Open-File Report 2013-15, Alberta Geological Survey, 2013).
- H. Ghofrani, G. M. Atkinson, *Geophys. Res. Lett.* **43**, 10164–10172 (2016).
- B. R. Scanlon, R. C. Reedy, J. P. Nicot, *Environ. Sci. Technol.* **48**, 12386–12393 (2014).
- H. Chen, K. E. Carter, *J. Environ. Manage.* **170**, 152–159 (2016).

ACKNOWLEDGMENTS

Waveform data are available through the Incorporated Research Institutions for Seismology, earthquake catalog data are available through the Alberta Geological Survey (14, 36), and tour reports of operator hydraulic fracturing parameters are available through the Alberta Energy Regulator. We thank K. Haug, G. Jean, S. Kolenosky, M. Speta, M. Araneta, and J. Chadi for their help in compiling the hydraulic fracturing database. We also thank H. Corlett and three anonymous reviewers for their comments. All authors declare no conflicts of interest.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6373/304/suppl/DC1
Materials and Methods
Figs. S1 to S12
References (40, 41)

5 June 2017; accepted 7 December 2017
10.1126/science.aao0159

MAGNETIC MATERIALS

Chiromagnetic nanoparticles and gels

Jihyeon Yeom,^{1,2} Uallisson S. Santos,³ Mahshid Chekini,^{2,4} Minjeong Cha,^{2,5}
 André F. de Moura,^{3*} Nicholas A. Kotov^{1,2,4,5,6*}

Chiral inorganic nanostructures have high circular dichroism, but real-time control of their optical activity has so far been achieved only by irreversible chemical changes. Field modulation is a far more desirable path to chiroptical devices. We hypothesized that magnetic field modulation can be attained for chiral nanostructures with large contributions of the magnetic transition dipole moments to polarization rotation. We found that dispersions and gels of paramagnetic Co₃O₄ nanoparticles with chiral distortions of the crystal lattices exhibited chiroptical activity in the visible range that was 10 times as strong as that of nonparamagnetic nanoparticles of comparable size. Transparency of the nanoparticle gels to circularly polarized light beams in the ultraviolet range was reversibly modulated by magnetic fields. These phenomena were also observed for other nanoscale metal oxides with lattice distortions from imprinted amino acids and other chiral ligands. The large family of chiral ceramic nanostructures and gels can be pivotal for new technologies and knowledge at the nexus of chirality and magnetism.

Optical materials that combine chirality and magnetism are essential for spintronics, magneto-optics, magnetochemistry, and chiral catalysts (1, 2) because they allow modulation of light beams, excited states, and chemical processes by means of a magnetic field. The junction of chirality and magnetism is central to skyrmions, spin catalysis, and the origin of homochirality in life on Earth (3–5), representing some of the newly emerged and long-standing problems of physics, chemistry, and biology. For all of these physicochemical phenomena, it is essential to increase the coupling of the photon's magnetic field with magnetic moments of electrons in chiral matter, which is expected to markedly enhance the chiroptical activity and first- and second-order magneto-optical phenomena, such as the Faraday effect, magnetic circular dichroism, and magneto-chiral dichroism (6–12). However, optical materials that combine a large magnetic moment and chiral asymmetry are rare (13). The common examples of such materials that can be dubbed chiromagnetic are typically based on complexes of transition metals (14, 15). But even for optical centers with rare-earth f orbitals accommodating multiple unpaired spins, the long-distance spin coupling enhancing the magnetic field effect on electronic transitions requires low temperatures of $T = 5$ to 7 K (14, 15).

Chiral inorganic nanoparticles (NPs) (16–19) and their assemblies (12, 20) provide a new toolbox for the design of materials combining chirality

and magnetism. Whereas the optical transitions in rare-earth coordination compounds involve localized molecular orbitals, the optical transitions in NPs may engage orbitals involving thousands of atoms, and so does their chirality (21). Unlike coordination compounds, the optical “center” responsible for chiroptical properties in inorganic NPs becomes orders of magnitude greater in volume compared with coordination compounds from f metals; magnetic coupling between atomic spins is also facilitated by shorter distances between magnetic atoms. Importantly, the chiral NPs may also show distorted crystal lattices (22, 23) and exhibit (super)paramagnetism (12, 24–26). This set of NP characteristics enables enhancement and spectral tuning of multiple chiroptical properties. (27, 28) In this study, we focus on the first-order effects of magnetic field on light absorption and circular dichroism (CD). The importance of magnetic properties of NPs for absorption of circularly polarized photons can be easily inferred from the quantum mechanical parameter known as rotational strength, R_{0a} , which can be calculated as

$$R_{0a} = \text{Im}[(\Psi_0|\hat{\mu}|\Psi_a) \cdot (\Psi_a|\hat{\mathbf{m}}|\Psi_0)] \\ = \text{Im}[\boldsymbol{\mu}_{0a} \cdot \mathbf{m}_{a0}] \quad (\text{Eq. 1})$$

where Ψ_0 and Ψ_a are the wave functions for the ground state (0) and excited state (a), $\hat{\mu}$ and $\hat{\mathbf{m}}$ are the corresponding electric and magnetic moment operators, and $\boldsymbol{\mu}_{0a}$ and \mathbf{m}_{a0} are the electric and magnetic transition dipole moments, respectively (6, 9). Equation 1 holds true for any and all quantum systems, whether chiral or not, whether bearing unpaired electrons or not, and whether in the presence of an external magnetic field or not (8, 29). As applied to chiral materials, Eq. 1 is usually simplified at the expense of the magnetic term (6–9, 30). For instance, in the case of plasmonic NPs, the magnetic moment term is reduced to a small constant, whereas the electric moment term is considered to be most essential

(6). However, in NPs with a large number of unpaired electrons and a chiral crystal lattice, the magnetic moment term $\langle \Psi_a|\hat{\mathbf{m}}|\Psi_0 \rangle$ should have a contribution comparable to that of the electric moment term $\langle \Psi_0|\hat{\mu}|\Psi_a \rangle$, which should lead to enhanced R_{0a} and potentially to practical realizations of NP chirality in magneto-optical devices operating at low fields and ambient temperatures.

Equally importantly, the large CD observed, for instance, for chiroplasmonic assemblies was difficult to translate to chiroptical devices, because all known approaches for real-time modulation of their optical activity are associated with irreversible chemical changes in the NP systems (31–37). At the same time, the helical assemblies of magnetic NPs are not known in the enantioselective form (38).

To address this set of fundamental and technological questions, we synthesized ~5-nm Co₃O₄ NPs using the L- and D-enantiomers of cysteine (Cys) as surface ligands. These NPs serve as the primary experimental model in this study and will be referred to as D-, L-, and DL-Cys Co₃O₄ when the corresponding Cys enantiomers or their equimolar mixture were used for NP synthesis. The choice of cobalt oxide as the inorganic core of the NPs was governed by its known magnetism and structural versatility, as well as the environmental robustness of cobalt-based ceramics. The chemical structure and atomic composition of the NPs was established by x-ray photoelectron spectroscopy (fig. S1, A to C) and atomic mapping (fig. S2) (39). Transmission electron microscopy (TEM) and scanning transmission electron microscopy (STEM) indicated the frequent presence of NPs with a seemingly amorphous inorganic phase (fig. S1, D to F). When the crystal structure was observed, the lattice plane distances could be adequately described by those in the cubic spinel phase. The crystalline domains are confined to the central part of the NPs, indicating that the seemingly amorphous shells originate from the crystal lattice distortions in the vicinity of the NP surface caused by molecular imprinting from attached enantiomers of amino acids. Instead of the typical antiferromagnetic behavior of Co₃O₄ nanostructures (40, 41) owing to exchange interactions in the spinel lattice, these NPs exhibit paramagnetic behavior even below the Néel temperature expected for Co₃O₄ (fig. S1, G to J); this observation confirms that crystal lattice distortions are characteristic of the vast majority of the NPs in the ensemble formed in this synthesis.

The brown transparent dispersions of these NPs yield CD spectra of high intensity, with up to eight positive and negative peaks in the ultraviolet (UV) and visible range (Fig. 1, A and C) corresponding to various transitions [intraparticle Co(II)→Co(III) (230, 280, and 350 nm) and surface states (including ligands)→Co(III) (450, 550, and 600 nm)], as demonstrated by the simplified time-dependent density functional theory (42, 43) calculation of a model NP (fig. S7 and table S1). The CD spectra for D- and L-Cys Co₃O₄ showed a nearly perfect mirror symmetry, whereas Co₃O₄ NPs made with equal amounts of D- and L-Cys were chiroptically silent (Fig. 1A). The spectral

¹Department of Macromolecular Science and Engineering, University of Michigan, Ann Arbor, MI 48109, USA.

²BioInterfaces Institute, University of Michigan, Ann Arbor, MI 48109, USA. ³Department of Chemistry, Federal University of São Carlos, 13.565-905, São Carlos, São Paulo, Brazil.

⁴Department of Chemical Engineering, University of Michigan, Ann Arbor, MI 48109, USA. ⁵Department of Materials Science and Engineering, University of Michigan, Ann Arbor, MI 48109, USA. ⁶Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA.

*Corresponding author. Email: moura@umfscar.br (A.F.d.M.); kotov@umich.edu (N.A.K.)

positions of the CD peaks were nearly perfectly aligned with those in absorption (Fig. 1C). Chiroptical anisotropy g -factors as high as 0.02 were obtained (Fig. 1B), which is ~ 10 times those obtained for other NPs of similar size, including plasmonic ones with known strong chiroptical activity (44, 45). The high value of the g -factor in the visible range can also be appreciated by the naked eye as the appearance of a distinct color when light passes through the NP-polyacrylamide gel and between crossed polarizers (Fig. 1D and scheme S1) (31); the green color corresponds to the 550-nm peak in the g -factor spectra in Fig. 1B.

The strong chiroptical activity was also observed for Co_3O_4 NPs synthesized and capped by L- and D-penicillamine, but CD peaks were at different positions (fig. S11) than those for Cys, especially for the visible range. Importantly, little change in the CD spectra was observed when Cys ligands were exchanged for penicillamine after the formation of NPs (fig. S12), indicating that chiroptical activity in the range of ~ 400 to 700 nm is associated with the inorganic core of NPs. Conservation of NP chirality after the ligand exchange is consistent with observations of chiral memory (46) and the high activation barrier for reconstruction of the Co_3O_4 lattice once NPs are formed. The kinetic stability of the chiral distortions also provides a pathway to other chiral nanostructures from Co_3O_4 by means of self-assembly using ligand-depleted NPs (9).

Raman scattering spectra further validated the chirality of the inorganic cores in the NPs. Characteristic bands at 380, 475, 516, 613, and 680 cm^{-1} observed for D- and L-Cys Co_3O_4 NPs are associated with Raman-active vibration modes of Co_3O_4 (Fig. 2C) (47). Raman optical activity spectra show peaks of opposite polarity at 377, 465, 531, and 719 cm^{-1} for the NPs carrying opposite enantiomers of Cys (Fig. 2D). Of particular importance is the strong peak at 380 cm^{-1} that corresponds to lattice phonons of Co_3O_4 . It occurs at frequencies higher than expected for Co_3O_4 with the cubic spinel crystal lattice (47) and shows a distinct antisymmetric relation of these vibrations in Co_3O_4 carrying L- and D-Cys (48). Both facts indicate the chirality transfer from amino acids to the crystal lattice of the inorganic core of the NPs, manifesting as crystal lattice distortions (49) propagating from the surface of the NPs into the core, which can also be seen in STEM images (fig. S1, E and F).

Computational study of atomic-scale dynamics in Co_3O_4 NPs having either L- or D-Cys on the surface were carried out to better understand the nature of the chirality transfer and distortions in the NP core. Relatively small NPs with Cys residues that were coordinated identically with experimental NPs were used in the simulations (fig. S6). In the course of full structural optimization, emblematic chiral geometries with mirror-image symmetries independently evolved for L- and D-Cys-bearing NPs. Specifically, three ligands on each corner of the tetrahedral model NP formed ringlike structures with either a clockwise or a counterclockwise sense of rotation with respect to the C_3 axis (Fig. 2, A and B). Using the Cahn-

Ingold-Prelog system, they can be classified as M (clockwise) and P (counterclockwise) enantiomers. Taking into account the hierarchical chirality in these structures that arises from superposition of the molecular chirality of the amino acids and their orientation on the surface, they can be denoted as M -D-Cys and P -L-Cys Co_3O_4 NPs.

The normal-mode analysis of the model NP indicated that peaks at 505, 529, 601, and 693 cm^{-1} correspond to the Raman active breathing mode of the inorganic core and ligand-core coupling (fig. S14). The experimental peaks in Fig. 2C match calculations for the M -D-Cys and P -L-Cys NPs well, with many of the bands representing the coupled vibrations of the surface atoms and ligands (movies S1 to S4).

In the ab initio molecular dynamics (AIMD) simulations, both enantiomers evolved independently, as demonstrated by the energy fluctuations of each model (fig. S19). Concomitantly, the degree of chirality increased by fourfold as determined by the Hausdorff chirality measure (50) for both the M -D- and P -L-Cys NPs (Fig. 2E), indicating that thermal fluctuations increase the

degree of distortion of the NPs. Notably, even when subjected to all of these vibrational distortions and bond reorganizations, the two model NPs followed nearly mirrored paths during the entire course of the AIMD simulations (Fig. 2F and movie S5).

MD simulations make it possible to follow the distortions being caused by the surface ligands in the ceramic crystal lattice. The NPs cores carrying L and D surface ligands present a pair of nearly mirrored structures after 2000 fs of structural relaxation (Fig. 3, B and C, and fig. S20, A and B). It is possible to analyze selected dihedral angles in these structures with respect to the value for the ideal crystallographic packing of Co_3O_4 cubic spinel (Fig. 3A). The three dihedral angles that we chose— $\phi_{35-15-8-18}$, $\phi_{35-12-11-22}$, and $\phi_{35-13-14-27}$ —share O atom number 35, located at the center of one of the faces of the NP model (Fig. 3, A to C, and fig. S21, C and D), and have values $\phi_{35-15-8-18} = \phi_{35-12-11-22} = \phi_{35-13-14-27} = 0$ in the undistorted cubic spinel of Co_3O_4 . The binding of the surface ligands led to noncoplanarity of these atoms that was already evident in the energy-minimization step, and these distortions increased

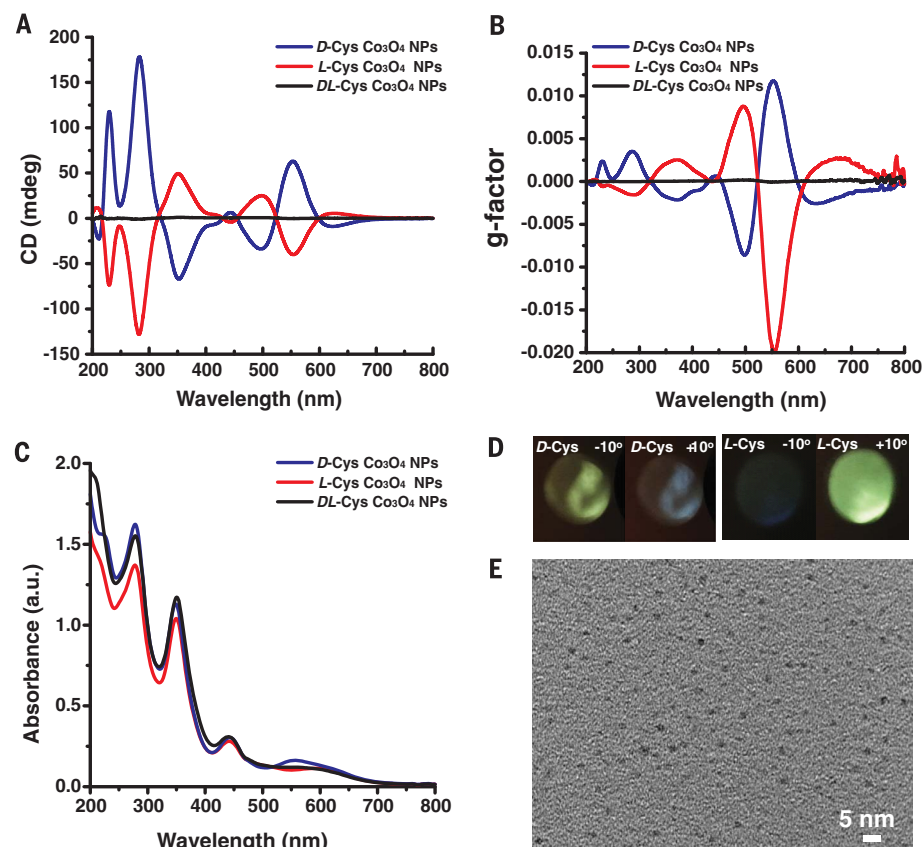


Fig. 1. Synthesized chiral Co_3O_4 NPs. (A) Circular dichroism (CD), (B) g -factor, defined as the ratio between the molar circular dichroism $\Delta\epsilon$ and the molar extinction coefficient ϵ ($g = \Delta\epsilon/\epsilon$), and (C) UV-visible absorption spectra of Co_3O_4 NPs stabilized by D-Cys, L-Cys, and DL-Cys. (D) Photographs of light transmitting through the NPs, with the rotation of the linear analyzer counterclockwise (-10°), and clockwise ($+10^\circ$). (E) TEM image of L-Cys-capped Co_3O_4 NPs. mdeg, millidegrees; a.u., arbitrary units.

during MD simulation. The mutual correlation of these angles stemming from the concerted movement of all the atoms in the nanoscale structures can be determined using Ramachandran-like plots. Similarly to proteins and other biomolecules, the pairwise probabilities of $\phi_{35-15-8-18}$, $\phi_{35-12-11-22}$, and $\phi_{35-13-14-27}$ acquiring specific values display unmistakable cross-correlation in the MD trajectories (Fig. 3, D to I). Importantly, the pattern of distortion is mirrored for D- and L-Cys NPs. These plots and MD simulations show the mechanism of chirality transfer in Co_3O_4 NPs that occurs during the growth of the NPs.

Chiral distortion of the original cubic spinel lattice changes the local magnetic fields within Co_3O_4 NPs because the overlap between atomic orbitals of Co^{3+} and O^{2-} and lattice symmetry are changed (51). Instantiating this point, the spin population of two Co(III) atoms changed from near zero to about two (tables S8 and S11) upon geometry optimization, owing to the distortions. The large spin and orbital angular momenta with corresponding operators \hat{S} and \hat{L} contribute to the chiroptical activity of NPs, according to Eq. 1, via magnetic transition dipole moments with the corresponding operator taking the form

$$\hat{\mathbf{m}} = \frac{-e\hbar}{2mc}(\mathbf{L} + g_S\mathbf{S}) \quad (\text{Eq. 2})$$

where g_S is the gyromagnetic ratio for spin angular momentum, m is electron mass, c is the speed of light, and \hbar is the reduced Planck constant. Unusually high g -factors can only result from large $|\mathbf{m}|$ (\mathbf{m} -allowed) and small $|\boldsymbol{\mu}|$ ($\boldsymbol{\mu}$ -forbidden) and either parallel ($\theta = 0^\circ$) or antiparallel ($\theta = 180^\circ$) orientations of the transient electrical and dipole moments (52). This is indeed the case for bands at 500, 550, and 650 nm (Fig. 1B) associated with the surface states of NPs, which provides direct experimental evidence for a strong contribution of magnetic transition dipole moments and paramagnetic enhancement of the optical activity of Co_3O_4 NPs. Naked-eye visualization of the large g -factors characteristic of Co_3O_4 NPs for an optical system with crossed polarizers indicates their importance for information technologies and photonics (Fig. 1D).

The key role of magnetically coupled unpaired electrons of Co atoms for the chiroptical activity of NPs was further confirmed by synthesizing mixed-metal NPs that included Cu^{2+} ions. Because the number of unpaired electrons in the NPs with Cu^{2+} was smaller than in those with neat Co_3O_4 , magnetic transition dipole moments should be lower according to Eq. 2. Indeed, the g -values gradually decreased as the amount of Cu^{2+} increased (fig. S22).

In addition to Co-Cu mixed oxides, the generality of the magnetic effects on the chiroptical activity of NPs was confirmed for chiral nickel(II) oxide NPs that also showed strong optical activity with g -factors up to 0.01 (fig. S23). These NPs were similar in size, shape, and magnetic properties (figs. S24 to S26) to the Co_3O_4 NPs in Fig. 1.

As one would expect from Eqs. 1 and 2, the rotatory optical activity of D- and L-Cys NPs

could be altered by the external magnetic field. To test magnetic field effects, the Co_3O_4 NPs were encapsulated in the transparent polyacrylamide gel (Fig. 4C) to avoid variation of optical properties owing to translational movement of the NPs. The NP gels showed the same optical and chiroptical bands at the same wavelengths in zero field as the NP dispersion (Fig. 4 and fig. S27). The magnetic field effect manifested as (i) a dramatic increase in UV transparency for circularly polarized light and (ii) its disappearance for racemic DL-Cys NPs (fig. S28), which is fundamentally different than the magnetic circular dichroism (MCD) that has been found for

Au, Ag, or Fe_3O_4 NPs (53–56). Importantly, the field-on/field-off ratios for 280-nm NP absorption peaks of left and right circularly polarized beams and their sum (Fig. 4, A and B) markedly exceed the ratios for even the giant Faraday rotation found in nanoscale plasmonic systems (19, 20, 57) and giant Zeeman splitting (58, 59). Also, the chiroptical effects of this magnitude were observed for D- and L-Cys NPs at room temperature, as opposed to the liquid helium temperatures often used for experimental observations of chiroptical effects in magnetic nanomaterials (14, 15) and the magneto-optical effects mentioned above.

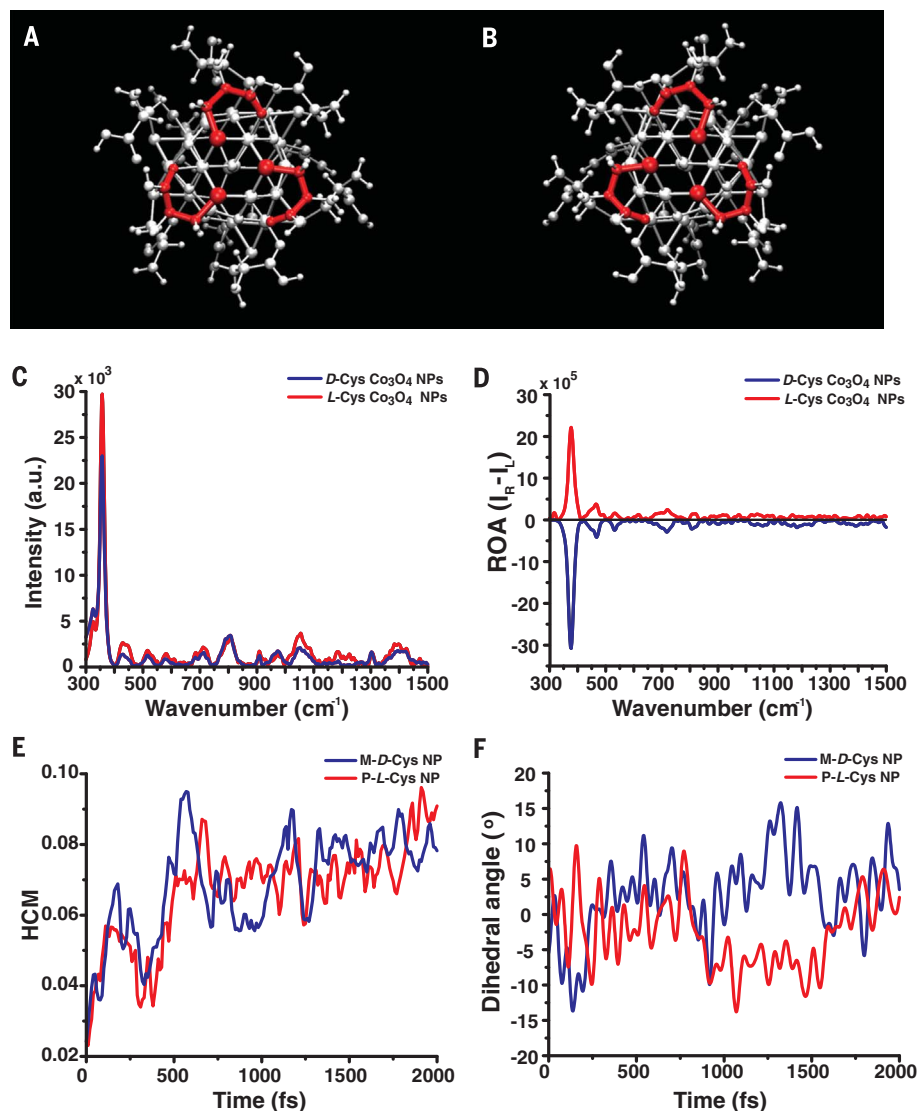


Fig. 2. Rotations of crystalline structures. (A and B) D-Cys (A) and L-Cys (B) Co_3O_4 NPs. Sulfur atoms are the larger red spheres forming one corner of the tetrahedra, and the remaining atoms depicted in red are C–C–C–O from the Cys ligands. (C and D) Raman (C) and Raman optical activity (ROA) (D) backscattering spectra with 532-nm excitation of D-Cys and L-Cys Co_3O_4 NPs in scattered circular polarization ROA mode. These spectra are courtesy of BioTools. (E) Hausdorff chirality measure (HCM) for the NP cores. (F) Dihedral angles between atoms 18, 7, 9, and 22 [O–Co(III)–Co(III)–O] of L-Cys and D-Cys Co_3O_4 NPs.

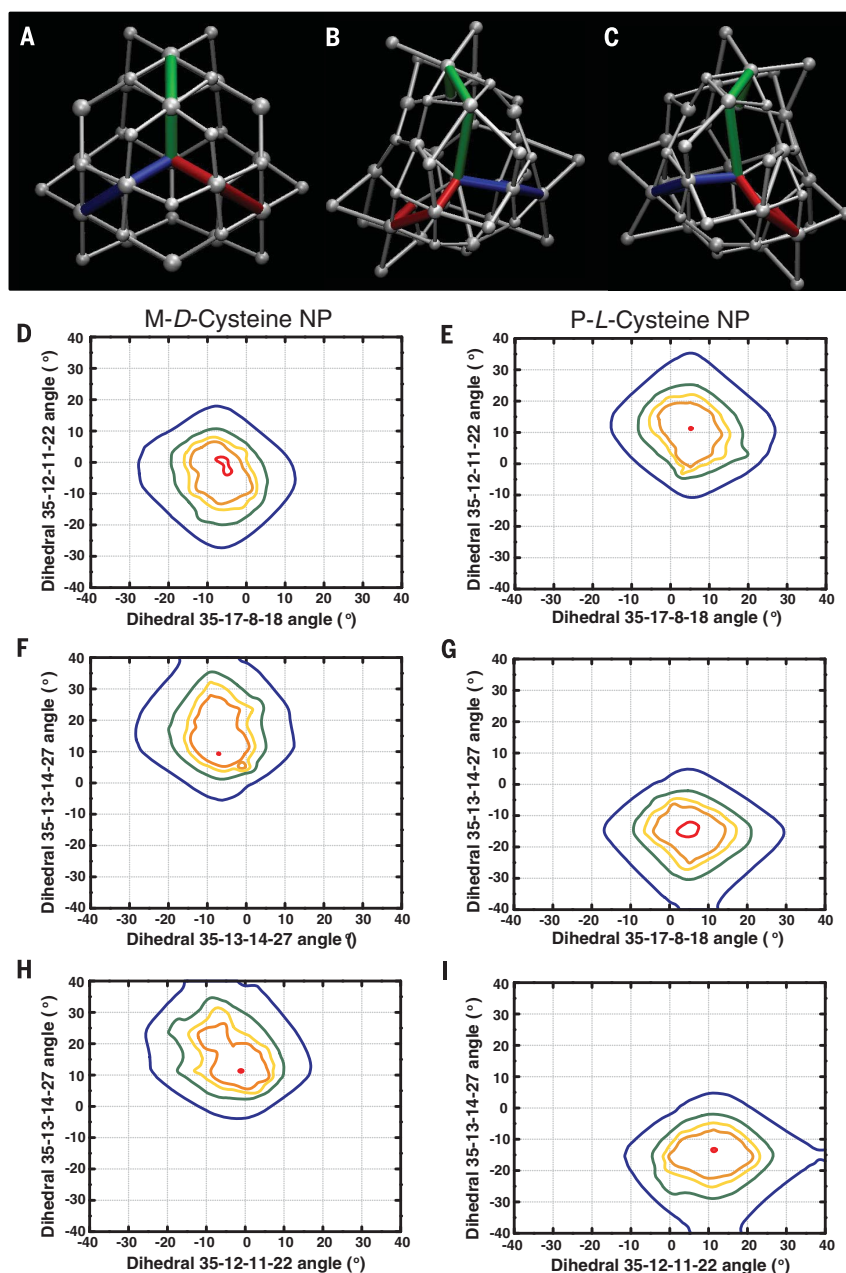


Fig. 3. Computed atomic geometry of chiral nanoparticles. (A to C) Graphical representation of the dihedral angles formed by four atoms: $\phi_{35-12-11-22}$ (red), $\phi_{35-17-8-18}$ (blue), and $\phi_{35-13-14-27}$ (green). A detailed description of the atomic types and numbering of each dihedral angle may be found in fig. S21. (A) NP with ideal crystallographic structure. (B) NP with M-D-Cys after 2000 fs of MD simulation. (C) NP with P-L-Cys after 2000 fs of MD simulation. The direction from the S atom to the carbonyl atom in the Cys molecules is taken as the basis for the geometry classification according to the Cahn-Ingold-Prelog rules. Ligands have been omitted for clarity. (D to I) Ramachandran plots for chiral NPs: two-dimensional probability maps for the relative orientation of adjoining octahedra pairs sharing the O atom number 35. Average probabilities were computed along the 2000-fs MD simulation for the NP functionalized with either M-D-Cys (left) or P-L-Cys (right). The isovalues depicted in the plots are probabilities of 0.0002 (blue), 0.0003 (green), 0.0004 (yellow), 0.0005 (orange), and 0.001 (red).

Although these experiments contain components of light-matter interactions from “natural” CD, the Faraday effect, and MCD (29), magneto-optical Kerr effects are unlikely to have a substantive

contribution (39). Magnetic field-induced transparency for circularly polarized light originates in the decrease of the absorption cross section of the NPs when the magnetic field is applied in non-

cubic crystal lattices (60), and not from plasmonic effects; the latter have a different mechanism and smaller magnitude (53–55, 61). The chiral distortions caused by the L- and D-Cys attachment to the surface of Co_3O_4 NPs make the dielectric permittivity and magnetic susceptibility tensors strongly anisotropic. The external magnetic field polarizes the paramagnetic NPs so that the spin and orbital magnetic moments become preferentially oriented in the direction of the field. The effect would be small for NPs with crystal lattices of high symmetry (e.g., cubic), and their experimental observation would require low temperatures to mitigate thermal broadening of the chiroptical peaks, because dielectric permittivity and magnetic susceptibility tensors are less anisotropic. In our observations, the transparency increase was minimal in DL-Cys NPs where chiral distortions were being partially compensated (Fig. 1A and fig. S28). The effect was also minimal for optical transitions involving surface states because the near-spherical symmetry of NPs averages the effect of external polarization on all the terms in Eq. 1. Consequently, the visible part of the CD spectra (>450 nm) of both D- and L-Cys NPs experiences little influence of the magnetic field, despite the high g -factor (Fig. 1B). However, the spectral region associated with $\text{Co(II)} \rightarrow \text{Co(III)}$ electronic transitions displays very strong effects because the external magnetic field forces both donor and acceptor states to become polarized simultaneously, and that is not averaged in the individual NPs or their ensembles. A simplified reason behind the reduced absorption cross section for the circularly polarized light could be the alignment of the transient magnetic moment for the $\text{Co(II)} \rightarrow \text{Co(III)}$ transition along the external magnetic field, which makes it orthogonal to the magnetic moment of incident photons, leading to a markedly reduced amplitude of the $\langle \Psi_a | \hat{\mathbf{m}} | \Psi_0 \rangle$ term.

The strong absorbance drop enables real-time optical modulation using a magnetic field, which was observed with excellent fidelity for 60 cycles (Fig. 4E); no agglomeration of NPs was observed. Cyclic switching of the gel transparency in the UV absorption band of NPs can also be converted into the modulation of photons in the visible range by using fluorescent targets (Fig. 4, D and F, and scheme S2).

Structural design of chiral NPs aimed at the maximization of the transient magnetic moment contribution to the NPs’ circular dichroism resulted in a large increase of the g -factor in the visible range of wavelengths and intense magnetic field-induced light modulation in the UV range. Data obtained in this study indicate that ceramic NPs with structural chirality and magnetism can be expanded to a large family of nanoscale materials with tunable chiroptical, magnetic, and other properties, enabled by the well-known tolerance of metal oxides to partial metal substitution. In addition to their technological relevance to magneto- and opto-electronic devices, the ceramic chiro-magnetic NPs based on metal oxides offer a versatile experimental system for different fields of science and fundamental problems unified by chiral properties of nanoscale matter.

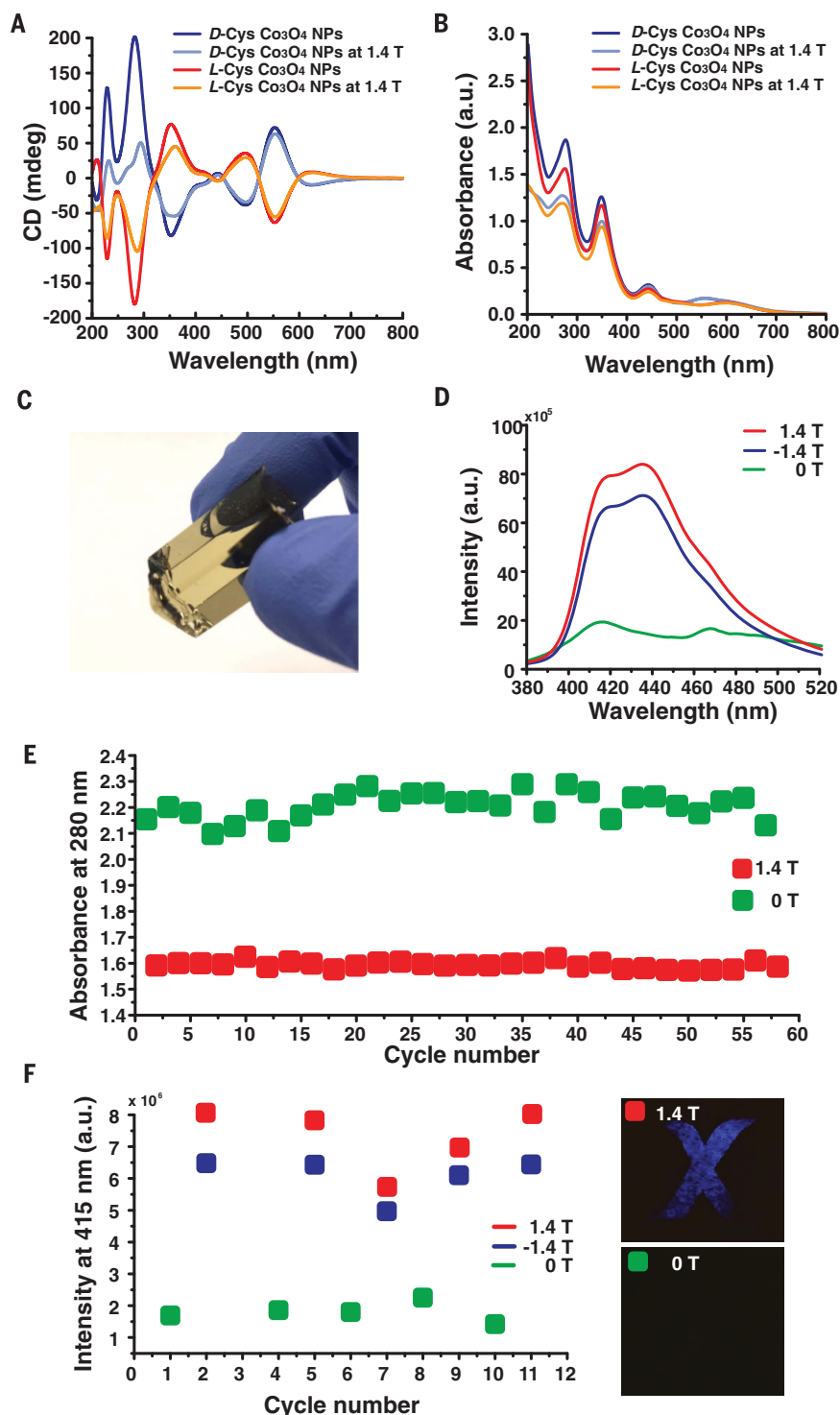


Fig. 4. Optical modulation. (A and B) CD and MCD (A) and corresponding absorbance spectra (B) of L-Cys and D-Cys Co_3O_4 NPs. (C) Photograph of the optically transparent gel made from L-Cys Co_3O_4 NPs. (D) Emission intensities of fluorescent paper plus the NP gel in front, with a magnetic field applied to the NP gel (red and blue) and without a magnetic field (green) (excitation, 280 nm). (E) Cycling performance of the NP gel's absorbance at 280 nm with and without magnetic fields. (F) Cycling profile of emission intensity at 415 nm with and without magnetic fields and corresponding photographs of blue-emitting light from the fluorescent paper.

REFERENCES AND NOTES

- S. Bordács et al., *Nat. Phys.* **8**, 734–738 (2012).
- L. D. Barron, *Nat. Mater.* **7**, 691–692 (2008).
- R. Naaman, D. H. Waldeck, *Annu. Rev. Phys. Chem.* **66**, 263–281 (2015).
- C. Pfeleiderer et al., *J. Phys. Condens. Matter* **22**, 164207 (2010).
- G. L. Rikken, E. Raupach, *Nature* **405**, 932–935 (2000).
- A. O. Govorov, Z. Fan, P. Hernandez, J. M. Slocik, R. R. Naik, *Nano Lett.* **10**, 1374–1382 (2010).
- T. Bruhn et al., *Angew. Chem. Int. Ed.* **53**, 14592–14595 (2014).
- W. Moffitt, A. Moscovitz, *J. Chem. Phys.* **30**, 648–660 (1959).
- K. Nakamishi, R. W. N. Berova, *Circular Dichroism: Principles and Applications* (Wiley-VCH, 2000).
- Y. Kitagawa, H. Segawa, K. Ishii, *Angew. Chem. Int. Ed.* **50**, 9133–9136 (2011).
- G. L. Rikken, E. Raupach, *Nature* **390**, 493–494 (1997).
- J. F. Hocheppied, P. Saintcavit, M. P. Pileni, *J. Magn. Magn. Mater.* **231**, 315–322 (2001).
- C. Train, M. Gruselle, M. Verdaguer, *Chem. Soc. Rev.* **40**, 3297–3312 (2011).
- R. Sessoli et al., *Nat. Phys.* **11**, 69–74 (2015).
- C. Train et al., *Nat. Mater.* **7**, 729–734 (2008).
- M. P. Moloney, J. Govan, A. Loudon, M. Mukhina, Y. K. Gun'ko, *Nat. Protoc.* **10**, 558–573 (2015).
- A. G. Mark, J. G. Gibbs, T.-C. Lee, P. Fischer, *Nat. Mater.* **12**, 802–807 (2013).
- K. M. McPeak et al., *Nano Lett.* **14**, 2934–2940 (2014).
- P. K. Jain, Y. Xiao, R. Walsworth, A. E. Cohen, *Nano Lett.* **9**, 1644–1650 (2009).
- V. Yannopapas, A. G. Vanakaras, *ACS Photonics* **2**, 1030–1038 (2015).
- J. Crassous, *Chem. Soc. Rev.* **38**, 830–845 (2009).
- M. V. Mukhina et al., *Nano Lett.* **15**, 2844–2851 (2015).
- C. E. Román-Velázquez, C. Noguez, I. L. Garzón, *J. Phys. Chem. B* **107**, 12035–12038 (2003).
- S. Sun et al., *J. Am. Chem. Soc.* **126**, 273–279 (2004).
- J.-H. Lee et al., *Nat. Nanotechnol.* **6**, 418–422 (2011).
- N. Lee, T. Hyeon, *Chem. Soc. Rev.* **41**, 2575–2589 (2012).
- J. Kumar, T. Kawai, T. Nakashima, *Chem. Commun. (Camb.)* **53**, 1269–1272 (2017).
- T. Goto et al., *Angew. Chem. Int. Ed.* **56**, 2989–2993 (2017).
- L. D. Barron, J. Vrbancich, *Mol. Phys.* **51**, 715–730 (1984).
- G. D. Fasman, *Circular Dichroism and the Conformational Analysis of Biomolecules* (Springer US, 1996).
- A. Kuzyk et al., *Nature* **483**, 311–314 (2012).
- W. Chen et al., *Nano Lett.* **9**, 2153–2159 (2009).
- A. Ben-Moshe, B. M. Maoz, A. O. Govorov, G. Markovich, *Chem. Soc. Rev.* **42**, 7028–7041 (2013).
- A. Guerrero-Martínez et al., *Angew. Chem. Int. Ed.* **50**, 5499–5503 (2011).
- J. K. Gansel et al., *Science* **325**, 1513–1515 (2009).
- A. Kuzyk et al., *Nat. Mater.* **13**, 862–866 (2014).
- Y. Kim et al., *Nat. Mater.* **15**, 461–468 (2016).
- G. Singh et al., *Science* **345**, 1149–1153 (2014).
- Supplementary materials.
- E. L. Salabaş, A. Rumpelcker, F. Kleitz, F. Radu, F. Schüth, *Nano Lett.* **6**, 2977–2981 (2006).
- S. Haffer, T. Walther, R. Köferstein, S. G. Ebbinghaus, M. Tiemann, *J. Phys. Chem. C* **117**, 24471–24478 (2013).
- A. B. Buda, K. Mislow, *J. Am. Chem. Soc.* **114**, 6006–6012 (1992).
- T. Risthaus, A. Hansen, S. Grimme, *Phys. Chem. Chem. Phys.* **16**, 14408–14419 (2014).
- N. Nishida, H. Yao, T. Ueda, A. Sasaki, K. Kimura, *Chem. Mater.* **19**, 2831–2841 (2007).
- Y. Zhou et al., *Angew. Chem. Int. Ed.* **50**, 11456–11459 (2011).
- T. Nakashima, Y. Kobayashi, T. Kawai, *J. Am. Chem. Soc.* **131**, 10342–10343 (2009).
- G. Wang et al., *J. Phys. Chem. C* **113**, 4357–4361 (2009).
- L. A. Nafie, *Chem. Phys.* **205**, 309–322 (1996).
- Y. Fujioka, J. Frantti, A. Puretzky, G. King, *Inorg. Chem.* **55**, 9436–9444 (2016).
- B. L. A. Harman, I. Sóvágó, *Inorg. Chim. Acta* **80**, 75–83 (1983).
- Y. Ikeda et al., *Phys. Rev. B* **75**, 054424 (2007).
- J. L. Alonso-Gómez, P. Rivera-Fuentes, N. Harada, N. Berova, F. Diederich, *Angew. Chem. Int. Ed.* **48**, 5545–5548 (2009).

53. B. Han *et al.*, *Nano Lett.* **17**, 6083–6089 (2017).
54. H. Yao, Y. Ishikawa, *J. Phys. Chem. C* **119**, 13224–13230 (2015).
55. H. Yao, T. Shiratsu, *Nanoscale* **8**, 11264–11274 (2016).
56. F. Pineider *et al.*, *Nano Lett.* **13**, 4785–4789 (2013).
57. D. A. Kuzmin, I. V. Bychkov, V. G. Shavrov, V. V. Temnov, *Nano Lett.* **16**, 4391–4395 (2016).
58. J. H. Yu *et al.*, *Nat. Mater.* **9**, 47–53 (2010).
59. N. Norberg, D. Gamelin, *J. Appl. Phys.* **99**, 08M104 (2006).
60. J. Li *et al.*, *Opt. Mater. Express* **2**, 1760 (2012).
61. I. Zubritskaya *et al.*, *Nano Lett.* 10.1021/acs.nanolett.7b04139 (2017).

ACKNOWLEDGMENTS

All data are reported in the main text and supplementary materials. The central part of this work was supported by the

NSF Ceramics program through the project “Ceramics quasicrystals” (grant no. 1411014). Partial support of this work was also provided by the NSF-funded project “Energy- and cost-efficient manufacturing employing nano-particles” (grant no. 1463474) and the Air Force Office of Scientific Research–funded project “Nanocomposite ion conductors from branched aramid nanofibers” (grant no. FA9550-16-1-0265). We also thank the Brazilian funding agencies Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; grants 2012/15147-4 and 2013/07296-2) for financial support. The authors acknowledge the National Laboratory for Scientific Computing (LNCC/MCTI, Brazil) for providing the high-performance computing resources of the SDumont supercomputer (<http://sdumont.lncc.br>), which have contributed to the results reported in this paper. A.F.d.M. thanks

Ministério da Educação/Programa de Educação Tutorial for a fellowship. U.S.S. thanks CNPq for a scholarship. M.Che. acknowledges the support of the Swiss National Science Foundation Early Postdoc Mobility Fellowship (grant no. P2GEP2-165061).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6373/309/suppl/DC1
Materials and Methods
Figs. S1 to S33
Tables S1 to S15
Schemes S1 to S3
References (62–93)
Movies S1 to S5

8 September 2017; accepted 4 December 2017
10.1126/science.aao7172

CHEMICAL ENGINEERING

Digitization of multistep organic synthesis in reactionware for on-demand pharmaceuticals

Philip J. Kitson, Guillaume Marie, Jean-Patrick Francoia, Sergey S. Zaleskiy, Ralph C. Sigerson, Jennifer S. Mathieson, Leroy Cronin*

Chemical manufacturing is often done at large facilities that require a sizable capital investment and then produce key compounds for a finite period. We present an approach to the manufacturing of fine chemicals and pharmaceuticals in a self-contained plastic reactionware device. The device was designed and constructed by using a chemical to computer-automated design (ChemCAD) approach that enables the translation of traditional bench-scale synthesis into a platform-independent digital code. This in turn guides production of a three-dimensional printed device that encloses the entire synthetic route internally via simple operations. We demonstrate the approach for the γ -aminobutyric acid receptor agonist, (\pm)-baclofen, establishing a concept that paves the way for the local manufacture of drugs outside of specialist facilities.

The manufacture of active pharmaceutical ingredients (APIs) is vital for modern health care, yet critical drugs are regularly manufactured for a finite period in a limited number of sites. The manufacture of chemical products—whether bulk, fine, or specialty chemicals, such as APIs—is currently based on a model whereby a central plant is exclusively designed for the manufacture of the product, or range of products, sold by that particular company (1). This model holds whether the manufacturer is a large pharmaceutical company or, as is increasingly the case, a contract research organization operating large chemical manufacturing plants to order from the pharmaceutical industry. This process leads to safety issues around both the storage and transport of such materials, as well as the issues inherent in the large-scale manufacture of chemicals (2). In addition, these large-scale plants are often at the mercy of complicated and global supply chains of raw materials, the failure of which at any point will reduce or halt the capacity of the plant to produce materials and deliver them effectively (3, 4). Also, when a given complex intermediate or API goes out of production, the plants are often repurposed and the manufacturing capacity is lost. The reinstatement of the process would require, in the best case, substantial capital investment to reconfigure a chemical plant for its synthesis. To alleviate this issue, we propose a concept whereby the large-scale manufacturing process of complex fine chemicals, such as APIs, is augmented by distributed, point-of-use manufacturing in self-contained cartridges, requiring limited user interaction to produce the desired products on demand. To achieve this, we developed a methodology for the translation of bench-scale synthesis procedures into a step-by-step

workflow that could be used to create digital designs for custom reactionware that can be fabricated by using three-dimensional (3D) printing technologies. In this way, we aim to move beyond the preserve of industrial manufacturing and prototyping applications (5), to revolutionize the relationship between the design, manufacture, and operation of functional devices (6–11) and exploit the increasing use of 3D printing in the automation of the chemical sciences (12–15). This methodology, which is in stark contrast to both large- and medium-scale traditional chemical manufacture, and also to the use of continuous-flow and microreactor approaches (1, 16, 17), allows for the distribution of simple chemical precursors and solvents rather than the complex products themselves. These precursors could then continue to benefit from the economies of scale brought by traditional manufacturing processes while complex products with short shelf lives, or lower and more distributed demand, can be produced locally. This has added benefits in terms of manufacture of the final products as the synthesis of smaller quantities is inherently safer than large-scale processes and poses less risk to both operators and infrastructure. Further, the translation of these synthetic approaches into a digitally defined format, where the reactor design and, eventually, an automated synthesis procedure are encoded, could allow the digitization of all chemical products into a very low-cost manufacturing format. This could allow large numbers of discontinued APIs to be made available as they can be brought back into production on a small scale by the fabrication and use of the appropriate cartridges (18, 19).

As a proof of principle, we present a process by which the traditional laboratory-scale synthesis of a commercially available API can be translated into the design of an integrated cartridge. To do this, all the reaction steps and intrasynthesis purification processes are encoded into the 3D

architecture of the cartridge so that the chemical reactions, work-ups, and purification are done with minimal user intervention and exposure automatically. We have demonstrated this process in the full synthesis of the anticonvulsant medication (\pm)-baclofen (see below).

This method for translating traditional laboratory syntheses into a form that can be encapsulated in a single cartridge is split into three layers of consideration, which were iteratively reevaluated during the cartridge development process. The first is the “conceptual layer,” where the chemical reactions and processes necessary are identified and developed. The second is the “digital layer,” in which these processes are translated into digital 3D models of the final cartridge devices. Finally, a “physical layer,” where the digital models are realized as either a modular implementation or a monolithic implementation, is used to generate the finalized cartridge, which can be used to effect the designed synthesis (Fig. 1). These physical systems can then be tested for efficacy as a final implementation, before iterating the process to develop reliable cartridge syntheses.

First, the fundamental chemistry required for the transformations is considered and optimized to minimize the necessary interstep purification for the completion of the full synthesis. This approach is similar to that taken to develop telescoped (i.e., consecutive transformations in a single reactor or sequence of reactors without isolation and purification of intermediates) and “one-pot” syntheses (20, 21), often used in process chemistry, both of which aim to maximize the efficiency of the synthetic route. Although here it is not necessary to produce genuinely telescoped syntheses, as modules for interstep purification can be built into the cartridge design, the synthesis of the desired compound, including all reagents and starting materials for all the necessary steps, must be considered as a unified process. The choice of synthetic route to any target compound will be dictated by a number of factors, including the relative availability and cost of starting materials, reagents, and solvents, as well as the compatibility of reaction and purification sequences with the reactor modules produced. In any wide-scale application of our approach, a cost analysis of any proposed synthetic route will have to be performed to ensure its viability for the product. Once the chemistry for the synthesis is developed, a sequence can be produced where the physical processes and reaction parameters—such as heating, cooling, phase separations, reaction volumes, and times—can be identified.

Vital to the success of these modules is the compatibility of the cartridge material with the chemistry being performed. Whereas traditional laboratory syntheses take place mostly in glassware, we use polypropylene (PP) as a basic structural material for the fabrication of the cartridges. We have found that this polyolefinic material, while demonstrating a robust range of chemical compatibility for traditional synthetic organic reactions, is also a suitable substrate for 3D printing applications (22–24). This gives the best balance of chemical resistance and material

WestCHEM, School of Chemistry, The University of Glasgow, Glasgow G12 8QQ, UK.

*Corresponding author. Email: lee.cronin@glasgow.ac.uk

properties for 3D printing. Therefore, the first step in the design process is testing the reactions necessary for compatibility in the reactor materials. Future iterations of the concept could expand on the materials and fabrication processes available for the reaction modules to further develop the range of chemistries feasible in this system, using, for example, perfluorinated polymers to increase the chemical resistance of the module.

To demonstrate the feasibility of incorporating these PP reactors into the production of APIs, we tested a number of reactions that lead to such targets in test modules fabricated from PP (Fig. 2). We tested reactions for the synthesis of three APIs: the central nervous system inhibitor (\pm)-baclofen (25), the anticonvulsant lamotrigine (26), and the gastroprotective agent zolimidine (27). As can be seen, all of the reactions tested were observed to work, but with slightly lower efficiency in PP reactors than in traditional glass

reactors, owing to physical loss of material on the relatively rough PP surface hampering product recovery. Surface roughness is inherent in the 3D printed process; however, use of other, as yet undeveloped, materials or different manufacturing techniques could reduce this issue. The zolimidine reactions, particularly the copper-catalyzed iodination reaction, experienced a pronounced reduction in efficiency, compared to (\pm)-baclofen or lamotrigine. We surmised that this was due to side reactions of the iodine with the polypropylene. These issues highlight that the process of translation from glassware must take into account both the physical and chemical properties and limitations of the reactor substrate (23). For this reason, the first two syntheses were selected for further development into reaction cartridges to explore the concept.

Once the processes needed for the reaction sequence are identified, the combined continuous

protocol is mapped onto the 3D digital designs for the target-specific cartridge. The sequence of processes is split into a series of modules, with each representing a telescoped series of processes that can take place in a single chamber of the 3D printed system. Each process module is then created as a digital model that can be manipulated to provide the correct physical dimensions necessary for the reaction scheme. The 3D models of the cartridges used in this study were created with OpenSCAD software, an open-source framework for CSG (constructive solid geometry) modeling that allows a highly flexible and configurable approach to create versatile libraries of components as reusable pieces of code. Once defined, these pieces of code can be manipulated by the software, allowing the generation of complex reactor geometries with minimal human inputs. For example, in this study, we designed a module library consisting of interchangeable top

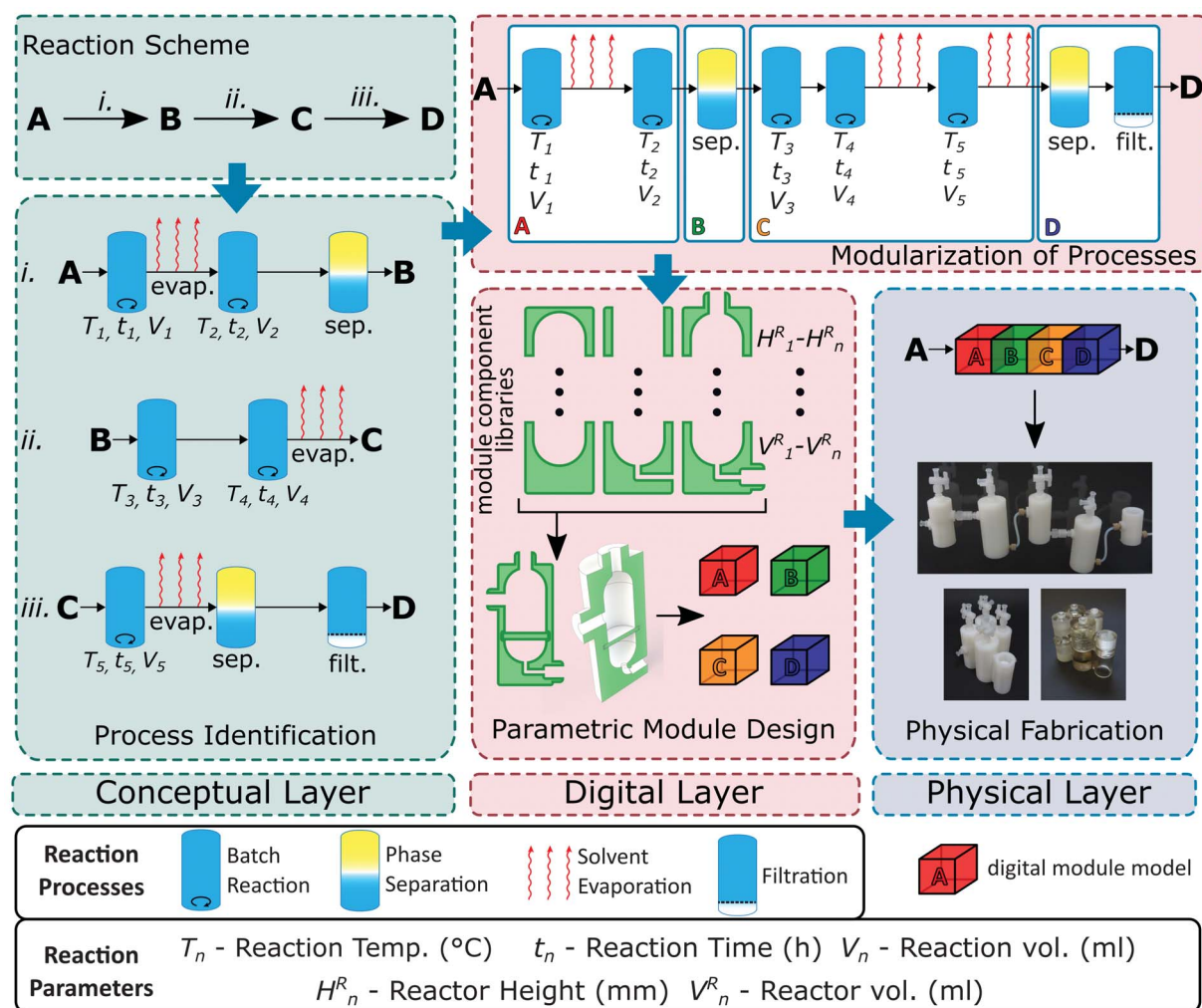


Fig. 1. Schematic representation of the translation of a multistep synthesis from conception through to implementation as a reaction cartridge. Reactions necessary for the synthesis are identified ($A \rightarrow B \rightarrow C \rightarrow D$, top left panel) and the specific chemical and physical processes and reaction parameters necessary for each reaction are laid out (conditions $i - iii$, left panel). These processes are then translated into bespoke reaction

modules designed to accomplish one or more of the chemical processes identified in the previous step (top right panel). The modules are then designed as 3D CAD models (lower center panel), with libraries of module components to accommodate the required reaction parameters. These digital models can then be fabricated to produce either a modular or monolithic implementation (lower right panel) of the process.

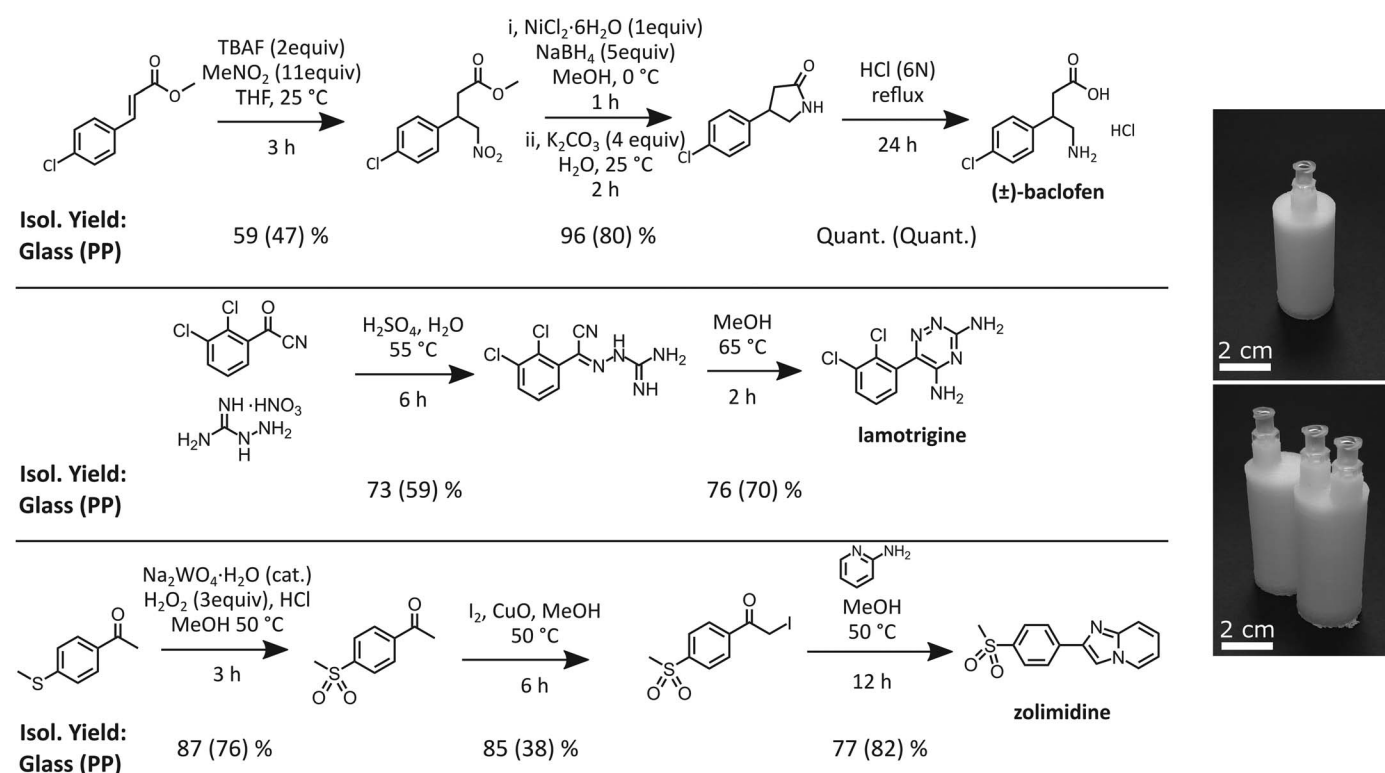


Fig. 2. Comparison of glass reactors with plastic reactionware for the optimized synthetic routes to (±)-baclofen (top), lamotrigine (middle), and zolimidine (bottom) with reaction yields for each step (reaction yields in PP vessels given in parentheses). Single (top right) or double (bottom right) chambered polypropylene reaction test cartridges were used. PP, polypropylene; TBAF, tetrabutylammonium fluoride; THF, tetrahydrofuran.

and bottom components with varying features that can be easily combined to produce reaction vessels with different shapes and features. From a single line of code, an entire module can be created, with 18 different shapes available (i.e., three different tops and six different bottoms can be selected; Fig. 3). The modules were designed around simple chambers where each reaction or process could be performed in as close a manner as possible to the way it would be carried out with traditional batch chemical techniques, easing the transition between published synthesis in glassware and “cartridge” synthesis. Typically, a standard module would have an opening on the top of the wall of the chamber for transfer of reaction mixtures from previous modules and an opening at the bottom of the chamber for expelling material from the module subsequent to the completion of the desired process. The transfer of material between modules is facilitated by a further opening in the roof of the compartment, which can be used to apply pressure that forces the reaction medium out of the chamber via the outlet at the bottom. The opening at the top otherwise equalizes pressure throughout the device to prevent the premature transfer of material, and also allows for application of vacuum to remove and exchange solvents. These modules can then be combined in sequences by use of further components of our module library such as siphon tubes for the transfer of material from one reaction module to another.

Once a reaction chamber is created, new features can be introduced by subtracting or adding shapes to the module. For example, a filtration device can be made from a module with a top input, a round bottom with a port, and a glass filter. To achieve this feature, a cylindrical model conforming to the dimensions of the physical filter to be inserted is created and subsequently subtracted from the model of a reaction chamber, producing a void space in the model into which the filter fits (see supplementary materials). Phase separation modules were achieved in a similar manner by using hydrophobic frit inserts that effectively separate organic and aqueous phases for product extractions. In keeping with our desire to design synthesis cartridges that can be produced outside traditional manufacturing regimes, we have exploited our group’s development of 3D printed reactors—reactionware—for synthetic chemical applications as a method of prototyping the physical reactors (28, 29). Three-dimensional printing-based fabrication approaches have the added advantage of being intimately linked to the design process.

Fabrication of the modular system was carried out on low-cost (~\$2000) 3D printers, Ultimaker 2 and 2+, although many other fused deposition modeling (FDM) printers could print the 3D modules produced through this approach. If it is necessary to incorporate nonprinted materials during 3D printing of the final module, a pre-programmed pause in the printing process is

instigated at a point just above the designed void, and the component is inserted in this space before the resumption of printing. Upon completion of printing, the inlet and outlet ports were tapped with a ¼ inch unified national fine (UNF) thread to allow ease of integration with the external infrastructure for performing the reaction sequences. Using standard ports allowed us to attach either standard fluidic tubing connectors such as those found in traditional flow synthesis setups, or widely used Luer lock adapters. These Luer lock connectors are easily reconfigurable, facilitating feedback into the design process.

The API chosen to accomplish a complete end-to-end synthesis was the central nervous system depressant and antispastic medication (±)-baclofen (30, 31) [*RS*-β-(4-chlorophenyl)-γ-aminobutyric acid] (4) (Fig. 4), a derivative of γ-aminobutyric acid (GABA) that modulates the action of this central inhibitory neurotransmitter (25). This target was chosen as an example to demonstrate that even relatively short syntheses require a disproportionately larger set of chemical processing steps to effect the full synthesis; in the future, we envision that the synthesis of larger numbers of compounds and compound classes will greatly expand the scope of this approach. (±)-Baclofen has found a number of applications since its first reported synthesis and is currently being investigated beyond its traditional use, as a high-dose treatment for alcoholism (32). Many syntheses of (±)-baclofen have been published since it was first

reported, often proceeding through the formation, and subsequent hydrolysis, of β -(4-chlorophenyl)- γ -butyrolactam (**3**). We have modified such a traditional synthesis of (\pm)-baclofen starting from the commercially available material methyl 4-chloro-cinamate (**1**), and proceeding via the Michael addition of nitromethane to form 4-nitro-3-(4-chlorophenyl)butanoic acid (**2**), followed by nickel-catalyzed reductive lactamization and subsequent acid hydrolysis to produce the final product in its commercially available racemic form as a hydrochloride salt. This three-reaction-step sequence contains 12 individual processing steps that must be incorporated into the reactionware device to complete the synthesis (Fig. 4). This sequence was designed to be particularly amenable to translation into the modular or monolithic system as at each stage, the reactions are either sufficiently clean, or reaction impurities that would impinge on subsequent process in the synthesis could be readily removed by phase partition. The final product is purified through a methanol-diethyl ether crystallization, which yields a crystalline solid that can be retrieved directly from the cartridge device. An animation of the entire process, showing the passage of reagents, processes, and work-ups, is shown in movie S1.

Each of these processes was translated into operations that could be successfully embodied in one or more reaction or purification modules. The specific reaction modules used for the synthesis of (\pm)-baclofen were (a) a combined Michael addition, evaporation and ether extraction module; (b) a combined solvent exchange and reduction module; (c) a phase separation and filtration module; (d) a combined solvent exchange and hydrolysis module; and (e) a filtration module. Individual modules were fabricated for a “plug-and-play” approach to the reaction process development by using Luer lock fittings to connect individual modules and Luer taper-compatible valves to interface with pressure or vacuum systems. This design allowed testing of each individual process in isolation before the modules were combined to build up the full synthesis. Finally, the module designs were “digitally stitched together” by using the developed CAD libraries for internal fluidic pathways to create the design for a monolithic synthesis cartridge. Once fabricated, the individual modules and the monolithic cartridges were evacuated and filled with a nitrogen atmosphere to ensure an inert environment for the subsequent chemistry.

The first chamber, (a), consists of a lower volume (4.9 ml) where the initial reaction can take place and is separated from the upper outlet by a hydrophobic frit. Reactor modules (b) and (d) consist of a single unbroken reaction chamber

(31.8 ml) with sufficient volume to accommodate the reaction volumes and extraction solvents from the previous processes before concentration under reduced pressure. Extraction module (c) consists of a chamber of sufficient volume (4.7 ml) to contain the aqueous phase from the previous chamber, which has a drain at the bottom covered by a hydrophobic frit that prevents both solid material and aqueous solution from passing into the next chamber or module. The final module is a filtration module for separating and retrieving the final product. This single module can be either open to the atmosphere or enclosed as required. During the fabrication process, chambers or modules that required stirring were equipped with a PTFE (polytetrafluoroethylene)-coated magnetic stirring bead (length 10 mm) to enable mixing of the contents. Each module or chamber of the monolith was equipped with a $\frac{1}{4}$ inch UNF threaded port carrying a female Luer lock adapter, which was used to introduce an inert (dry, N_2) atmosphere, or reduced pressure, into the system. The modular system was designed such that there

was a single fluidic path through the reactor; flow from one chamber into the next was induced either by pressure from excess solvent, in the case of the phase separation processes, or the introduction of nitrogen pressure difference between the relevant chambers to push the reaction mixture through an embedded channel running from the bottom of one chamber to the top of the next.

Starting materials were prepared as simple solutions and transferred to the cartridge via standard Luer syringes. The cooling and heating required for the reaction sequence were achieved by the immersion of the reaction cartridge or module in an ice or sand bath, respectively, and the temperature required for the reactions can be achieved automatically on a stirring-hotplate. The exact sequence of operations, positioning of the module in the heating or cooling bath, and time intervals necessary for completing the synthesis are outlined in the supplementary materials (figs. S12 and S13 and table S3).

Performing the synthesis starting from 200 mg of **1** in the manner described yielded 98 mg

(39% yield over three reaction steps and 12 processing steps from **1** with $\geq 95\%$ purity as determined by high-performance liquid chromatography) of (\pm)-baclofen hydrochloride salt, which is more than 1 day's maximum dosage of the drug. Better efficiency of reaction can be achieved with lower concentrations of starting materials (using a similar cartridge at half concentration, i.e., 100-mg scale, gave a 44% yield over three steps of similar purity). Increasing the volume of the reactor as well increases the quantity of (\pm)-baclofen obtained [a 300-mg scale synthesis yielded 133 mg (35%) (\pm)-baclofen]. The integration of the reaction processing steps into the design of the modules greatly simplifies the operations required to perform the reaction sequence compared to traditional bench synthesis and simultaneously reduces the level of technical skills required to perform the process down to simple operations that do not require the specific skills of a trained synthetic chemist. Although the total time for the reaction sequence is around 40 hours in this case, including all intermediate operations, the workflow is constrained by the geometry of the device, so all human interaction is limited to simple interventions at specific time periods, and it should be possible to shorten the interaction time further. The use of such bespoke, single-use cartridges would greatly reduce the time spent on glassware preparation, liquid handling, and other ancillary tasks associated with the majority of chemical syntheses at this scale. Also, by using the geometry of the reactor to constrain the operation of the synthesis, we reduce the human decision making

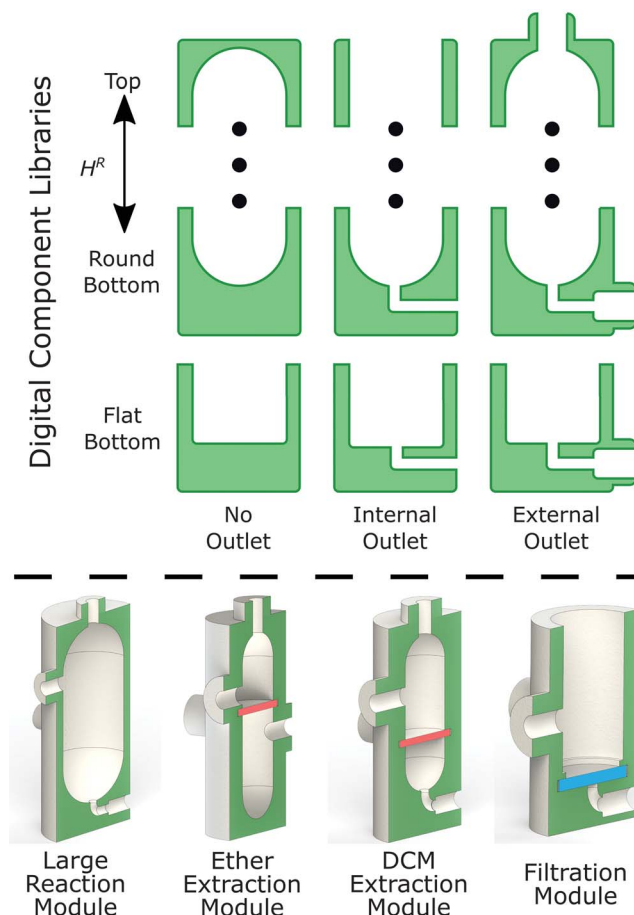


Fig. 3. Parameterized approach to the design of individual process modules. Digital libraries of module components (top) can be easily assembled to produce a wide range of module geometries dictated by the specific process and reaction parameters (e.g., solvent volumes, number of inputs and outputs, etc.) (bottom). Hydrophobic filters for phase separation are shown in red, and fritted glass filters are shown in blue. DCM, dichloromethane. H^R , reactor height.

involved in the synthesis processes, making the sequence more reproducible. Given sufficient facilities, several instances of the synthesis cartridge could be used at once, achieving scalability by numbering-up arrays of cartridges, and using these in parallel to increase the output. As a result of the ability to parameterize and encode multistep organic synthesis reactions with work-

ups embedded, we envisage that a digital programmable universal heater-stirrer-solvent-reagent plug-and-play device can be constructed into which only the cartridge, specific to a given synthesis, can be plugged in.

The (±)-baclofen synthesis necessitated liquid handling and separation of reaction chambers to effect the full reaction sequence. In some cases,

however, syntheses can be conducted in single reaction cartridges, depending on the nature and quality of the interstep purification required. For example, the synthesis of lamotrigine (Fig. 2) can be achieved in a single cartridge as the intermediate material is insoluble in the reaction solvent at low temperatures. In a single, closed, filtration module, the initial reaction product could be

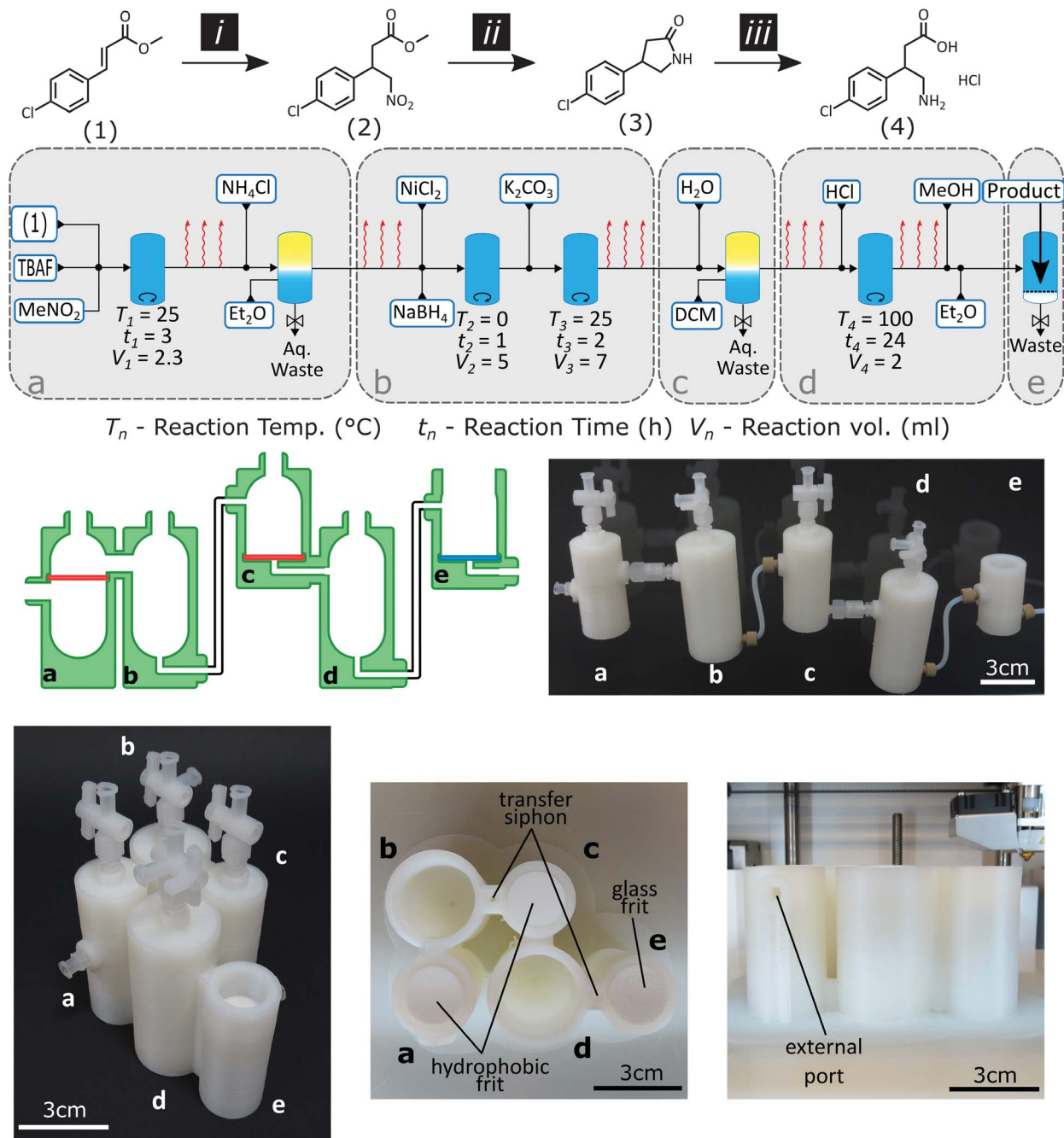


Fig. 4. Synthesis of (±)-baclofen in a series reaction cartridges. (Top) Conceptual synthetic procedure for the synthesis of (±)-baclofen under the conditions described in Fig. 2, showing the necessary processing sequence to effect this synthetic pathway. These processes were then split into modules (a) to (e) (indicated by gray boxes in the process sequences), which we translated into a digital design (middle left) and finally fabricated as

either a modular (middle right) or monolithic (bottom left) implementation. A partially fabricated monolithic cartridge is also shown indicating the placement of non-3D printed components and internal fluidic pathways (bottom center and right). Both modular and monolithic cartridges are shown with Luer taper-compatible valving for interfacing with external fluidic inputs and pressure or vacuum lines.

washed and processed in situ before introduction of the solvent for the subsequent cyclization step. This stands in contrast to the traditional procedure, which requires the solid product of the first step to be removed from the initial reactor to be filtered, dried, and then reintroduced to a reactor for the second step of the synthesis. Performing the synthesis of lamotrigine on a 250-mg scale of starting material yields 112 mg (46% over two reaction steps) of the final product, giving an off-white crystalline powder.

The digital approach to the design of the system that we have adopted allows the blueprints for these cartridges to be stored electronically for implementation as and when required. The distribution model for fine and specialty chemicals, such as the APIs implied by this approach, would lead to a decentralizing of logistical approaches to chemical manufacture. Here, any location with access to a sufficiently diverse market of chemical precursors and suitable cartridge fabrication facilities could be used to produce chemical products, which could previously be achieved only in a fully equipped synthesis laboratory with highly trained staff. This approach not only holds promise for eventually delivering on-demand personalized medicines manufactured at, or near, the point of use, but also has short-term potential applications in the synthesis of APIs that are currently out of production. An immediate impact of digitization is that the cost for synthesis at the bench scale (milligrams) could decrease markedly owing to savings in labor and infrastructure with only a one-off digitization cost (and allow operators to make 5 to 10 different products at the same time). Once the saving meets the digitization cost, the efforts of the expert chemist will shift from bespoke on-demand chemical manufacturing to chemical digitization (see supplementary materials for an economic analysis). Our methodology will have the most rapid impact for chemicals that are currently produced on demand in small batches and that occupy a gap in the market where the demand for a product is sufficient for it to be commercially viable but insufficient to justify plant-scale production. This gap lies between the high cost of bench-scale versus reactor-scale synthesis, and thus the digitization benefit of compounds in this zone is high.

The regulatory framework necessary to produce complex materials in this fashion will need thorough attention; indeed, our approach would

require a completely new system for the regulation of API manufacture. This system would have to be developed alongside the evolution of this approach as a method for pharmaceutical synthesis, which we have presented here in proof-of-concept form; however, we can envision a situation in which regulatory agencies certify specific cartridge or module designs as soon as a digitized process is fully established (including the embedded quality-control protocols), independent of the physical location of person who uses the cartridge. This approach has multiple benefits. First, the framework can adopt well-established methods of digital object certification from the information technology universe (e.g., digital signing with asymmetric ciphers). Second, no explicit certification would be needed for each new “facility” (which might be a hospital or a private house) that would need the drug. Third, existing methods for protecting and manipulating digital content provide much more efficient models for distribution and regulation compared to the retail and patent system, respectively. These regulatory issues surrounding the commercial or clinical application of this approach are not trivial, and care must be taken to ensure that end-user safety is not compromised. However, we believe that the benefits in terms of efficiency of delivery, robustness of supply, and range of materials available could lead to the digitization of chemical synthesis.

REFERENCES AND NOTES

1. D. M. Roberge, L. Ducry, N. Bieler, P. Cretton, B. Zimmermann, *Chem. Eng. Technol.* **28**, 318–323 (2005).
2. D. A. Crowl, J. F. Louvar, *Chemical Process Safety: Fundamentals with Applications* (Pearson Education, 2001).
3. M. A. Ehlen, A. C. Sun, M. A. Pepple, E. D. Eidson, B. S. Jones, *Comput. Chem. Eng.* **60**, 102–111 (2014).
4. G. E. Applequist, J. F. Pekny, G. V. Reklaitis, *Comput. Chem. Eng.* **24**, 2211–2222 (2000).
5. D. Dimitrov, K. Schreve, N. de Beer, *Rapid Prototyping J.* **12**, 136–147 (2006).
6. F. Rengier et al., *Int. J. Comput. Assist. Radiol. Surg.* **5**, 335–341 (2010).
7. X. M. Li et al., *Int. J. Polym. Sci.* **2014**, 829145 (2014).
8. S. J. Hollister, *Nat. Mater.* **4**, 518–524 (2005).
9. S. Hong et al., *Adv. Mater.* **27**, 4035–4040 (2015).
10. M. Bogers, R. Hadar, A. Bilberg, *Technol. Forecast. Soc. Change* **102**, 225–239 (2016).
11. R. Bogue, R. Bogue, *Assem. Autom.* **33**, 307–311 (2013).
12. S. V. Ley, D. E. Fitzpatrick, R. J. Ingham, R. M. Myers, *Angew. Chem. Int. Ed.* **54**, 3449–3464 (2015).
13. B. C. Gross, J. L. Erkal, S. Y. Lockwood, C. Chen, D. M. Spence, *Anal. Chem.* **86**, 3240–3253 (2014).
14. B. Gross, S. Y. Lockwood, D. M. Spence, *Anal. Chem.* **89**, 57–70 (2017).
15. O. Okafor et al., *React. Chem. Eng.* **2**, 129–136 (2017).
16. N. Kockmann, M. Gottsponer, B. Zimmermann, D. M. Roberge, *Chemistry* **14**, 7470–7477 (2008).
17. A. Adamo et al., *Science* **352**, 61–67 (2016).
18. J. M. Pearce, in *Open-Source Lab* (Elsevier, Boston, 2014), pp. 1–11.
19. M. Coakley, D. E. Hurt 3rd, *J. Lab. Autom.* **21**, 489–495 (2016).
20. C. Vaxelaire, P. Winter, M. Christmann, *Angew. Chem. Int. Ed.* **50**, 3605–3607 (2011).
21. Y. Hayashi, *Chem. Sci.* **7**, 866–880 (2016).
22. P. J. Kitson, S. Glatzel, L. Cronin, *Beilstein J. Org. Chem.* **12**, 2776–2783 (2016).
23. P. J. Kitson, R. J. Marshall, D. Long, R. S. Forgan, L. Cronin, *Angew. Chem. Int. Ed.* **53**, 12723–12728 (2014).
24. P. J. Kitson, M. D. Symes, V. Dragone, L. Cronin, *Chem. Sci. (Camb.)* **4**, 3099–3103 (2013).
25. M. Da Prada, H. H. Keller, *Life Sci.* **19**, 1253–1263 (1976).
26. A. Fitton, K. L. Goa, *Drugs* **50**, 691–713 (1995).
27. L. Almirante et al., *J. Med. Chem.* **8**, 305–312 (1965).
28. M. D. Symes et al., *Nat. Chem.* **4**, 349–354 (2012).
29. P. J. Kitson et al., *Nat. Protoc.* **11**, 920–936 (2016).
30. P. Camps, D. Muñoz-Torrero, L. Sánchez, *Tetrahedron Asymmetry* **15**, 2039–2044 (2004).
31. A. Mann et al., *J. Med. Chem.* **34**, 1307–1313 (1991).
32. G. Addolorato et al., *Alcohol Alcohol.* **37**, 504–508 (2002).

ACKNOWLEDGMENTS

We acknowledge the help of S. Marshall in compiling our analysis of the economic impact of our methodology. Supplementary materials include a PDF document detailing the materials and methods used in this article; the STL and OpenSCAD files used to generate all 3D printed objects mentioned; a schematic movie illustrating the process of baclofen synthesis in the monolithic cartridge; and Python code that can be used to automate the stirrer-hotplate operations for the baclofen synthesis. We gratefully acknowledge financial support from the Engineering and Physical Sciences Research Council (grant nos. EP/H024107/1, EP/J015156/1, EP/K021966/1, EP/L015668/1, EP/L023652/1) and European Research Council (project 670467 SMART-POM). This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA). The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. L.C. is the founder and director of CroninGroupPLC and is listed as an inventor on a patent application filed by The University of Glasgow (GB 1800299.8). L.C. conceived the initial concept and the design approach; P.J.K. designed the reactionware with help from J.P.F. and S.Z.; G.M. and R.C.S. telescoped the methods, porting them from glass to plastic; and P.J.K. developed the monolithic cartridges with help from S.Z. J.S.M. helped evaluate the purity of the products, and P.J.K. coordinated the team with help from L.C.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6373/314/suppl/DC1
Materials and Methods
Figs. S1 to S16
Tables S1 to S4
Movie S1
OpenSCAD libraries.zip
STL files.zip
Python Code

10 July 2017; accepted 1 November 2017
10.1126/science.aao3466

BIOGEOGRAPHY

A global atlas of the dominant bacteria found in soil

Manuel Delgado-Baquerizo,^{1,2*} Angela M. Oliverio,^{1,3} Tess E. Brewer,^{1,4} Alberto Benavent-González,⁵ David J. Eldridge,⁶ Richard D. Bardgett,⁷ Fernando T. Maestre,² Brajesh K. Singh,^{8,9} Noah Fierer^{1,3*}

The immense diversity of soil bacterial communities has stymied efforts to characterize individual taxa and document their global distributions. We analyzed soils from 237 locations across six continents and found that only 2% of bacterial phylotypes (~500 phylotypes) consistently accounted for almost half of the soil bacterial communities worldwide. Despite the overwhelming diversity of bacterial communities, relatively few bacterial taxa are abundant in soils globally. We clustered these dominant taxa into ecological groups to build the first global atlas of soil bacterial taxa. Our study narrows down the immense number of bacterial taxa to a “most wanted” list that will be fruitful targets for genomic and cultivation-based efforts aimed at improving our understanding of soil microbes and their contributions to ecosystem functioning.

Although soil bacteria have been studied for more than a century, most of the diversity of soil bacteria remains undescribed. This is unsurprising given that soil bacteria rank among the most abundant and diverse group of organisms on Earth (1–4), challenging our capacity to understand their specific contributions to ecosystem processes, including nutrient and carbon cycling, plant production, and greenhouse gas emissions (1–3). Put simply, characterizing the ecological attributes (environmental preferences and functional traits) of the thousands of bacterial taxa found in soil is unfeasible. Most soil bacteria do not match those found in preexisting 16S ribosomal RNA (rRNA) gene databases (5), we have genomic information for relatively few of them (5–7), and the majority of soil bacteria have not been successfully cultivated in vitro (6, 7). For these reasons, we lack a predictive understanding of the ecological attributes of most soil individual bacterial taxa, with their environmental preferences, traits,

and metabolic capabilities remaining largely unknown.

Previous work has shown that only a small fraction of soil bacteria is typically shared between any pair of unique soil samples (4, 8, 9). However, we also know that, as with most “macrobial” communities (10), not all bacterial taxa are equally abundant in soil. There are often subsets of soil bacterial taxa that are far more abundant than others. For example, the genus *Bradyrhizobium* has been found to be dominant in forest soils from North America (11). Similarly, a lineage within the class *Spartobacteria* was found to be highly abundant in undisturbed grassland soils (12). Perhaps more important, many individual taxa that are highly abundant in individual soil samples may also be abundant across distinct soil samples, even when those soil samples are from sites located far apart (e.g., *Candidatus Udaeobacter copiosus*) (13). Therefore, a critical and logical next step to advance our understanding of soil bacterial communities is to identify the dominant bacterial phylotypes that are abundant and ubiquitous across soils, and determine their ecological attributes.

From the large body of literature using marker gene sequencing to characterize soil bacterial communities, we know which major phyla tend to be more abundant in soil (14) and we have a growing understanding of how various factors, including soil properties (e.g., pH) (15), climate (9, 16), vegetation type (17), and nutrient availability (18), structure the composition of soil bacterial communities worldwide. What is currently missing is a detailed ecological understanding of common soil bacterial species, which we refer to as phylotypes (as bacterial species definitions can be problematic) (19). Understanding the ecological attributes of dominant phylotypes will increase our ability to successfully cultivate them in vitro and allow us to build a more predictive understanding of how soil bacterial communities vary across space, time, and in response

to anthropogenic changes. For example, if we could identify those dominant phylotypes with strong preferences for a given set of environmental conditions (e.g., low or high pH), we could then use this information to predict their distributions and enrich for these dominant phylotypes in vitro. Ultimately, a better understanding of dominant soil bacterial taxa will improve our ability to actively manage soil bacterial communities to promote their functional capabilities.

Here we conducted a global analysis of the bacterial communities found in surface soils from 237 locations across six continents and 18 countries (fig. S1) to (i) identify the most dominant (i.e., most abundant and ubiquitous) soil bacterial phylotypes worldwide; (ii) determine which of these dominant phylotypes tend to co-occur and share similar environmental preferences; (iii) map the abundances of these ecological clusters of dominant soil bacteria across the globe; and (iv) assess the genomic attributes that differentiate phylotypes with distinct environmental preferences. The soils included in this study were selected to span a wide range of vegetation types, edaphic characteristics, and bioclimatic regions (arid, temperate, tropical, continental and polar) (20).

We first identified the most dominant bacterial phylotypes by 16S rRNA gene amplicon sequencing (20). Dominant phylotypes (taxa that share ≥97% sequence similarity across the amplified 16S rRNA gene region) include those that are highly abundant (top 10% most common phylotypes sorted by their percentage of 16S rRNA reads) (21) and ubiquitous (found in more than half of the 237 soil samples evaluated) (20). Not surprisingly, our global data set comprised bacterial communities that were highly variable with respect to their diversity and overall composition (fig. S2). For example, observed phylotype richness ranged from 774 to 2869 phylotypes per sample, and there was a large amount of variability in the relative abundances of major phyla across the studied sites (fig. S2). Also, as expected, only a small fraction of phylotypes was found to be shared across soil samples, and most phylotypes were relatively rare (fig. S3). Based on our criteria, only 2% of the bacterial phylotypes (511 out of 25,224 phylotypes) were dominant (Fig. 1A and table S1). However, this small number of phylotypes accounted for, on average, 41% of 16S rRNA gene sequences across all samples (Fig. 1A), although they collectively accounted for more than half of the bacterial communities in some environments (e.g., forests from arid environments; Fig. 1B). In other words, most soil bacterial phylotypes are rare and relatively few are abundant, but many of these are found across a wide range of soils.

Notably, 85% of the dominant phylotypes identified from our data set were also found to be dominant in the bacterial communities recovered from 123 global soils that were analyzed using a shotgun metagenomic approach (20) (table S1). This cross-validation indicates that our list of dominant phylotypes is not biased by polymerase chain reaction amplification or by our choice

¹Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA.

²Departamento de Biología y Geología, Física y Química Inorgánica, Escuela Superior de Ciencias Experimentales y Tecnología, Universidad Rey Juan Carlos, c/ Tulipán s/n, 28933 Móstoles, Spain. ³Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309, USA. ⁴Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309, USA. ⁵Departamento de Biología Vegetal II, Fac. Farmacia, Universidad Complutense de Madrid, 28040 Madrid, Spain. ⁶Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales 2052, Australia. ⁷School of Earth and Environmental Sciences, Michael Smith Building, The University of Manchester, Oxford Road, Manchester M13 9PT, UK. ⁸Hawkesbury Institute for the Environment, Western Sydney University, Penrith, 2751, New South Wales, Australia. ⁹Global Centre for Land-Based Innovation, Western Sydney University, Penrith South DC, NSW 2751, Australia.

*Corresponding author. Email: m.delgadobaquerizo@gmail.com (M.D.-B.); noah.fierer@colorado.edu (N.F.)

of primers, as most of the identified dominant phylotypes were shared between two independent sets of soils analyzed using two different approaches (amplicon versus shotgun metagenomic sequencing). In addition, we compared the results from our sample set with those soils analyzed via amplicon sequencing as part of the Earth Microbiome Project (EMP) (22). The majority of the dominant phylotypes in the EMP data set (80%)—identified using the same criteria explained above—were included within our list of dominant taxa (>97% similarity) (20). Also, the top 511 phylotypes, comparable to our top 511 dominant taxa, accounted for 0.5% of all bacterial phylotypes and 41% of all 16S rRNA gene reads in the EMP data set. Despite important methodological differences between the two data sets (20), this concordance between the results from EMP and our study reinforces our conclusion that a relatively small subset of bacterial phylotypes dominate soils across the globe.

On average, the dominant bacterial phylotypes identified from our data set were highly abundant in soils across multiple continents, eco-

system types, and bioclimatic regions (Fig. 1B). The only exception was soil from tropical forests, where the dominant phylotypes accounted for only ~20% of 16S rRNA gene sequences, which is likely a product of soils from tropical forests being under-represented in our database and/or tropical forest bacterial communities being very distinct from those found in other ecosystem types (fig. S4). Together, our results suggest that soil bacterial communities, like plant communities (10), are typically dominated by a relatively small subset of phylotypes. As such, we focus all downstream analyses on the 511 phylotypes found to be the most abundant and ubiquitous in soils from across the globe.

The identified dominant phylotypes accurately predicted overall patterns in β -diversity for the “subdominant” component of the bacterial communities surveyed (98% of phylotypes; figs. S2 and S5 and Fig. 1C). That is, patterns in the distribution of the dominant bacterial phylotypes across the globe closely mirrored those observed for the remaining 98% of bacterial phylotypes. The most abundant and ubiquitous of these 511 phylotypes

included Alphaproteobacteria (*Bradyrhizobium* sp., *Sphingomonas* sp., *Rhodoplanes* sp., *Devosia* sp., and *Kaistobacter* sp.), Betaproteobacteria (*Methylobium* sp. and *Ramlibacter* sp.), Actinobacteria (*Streptomyces* sp., *Salinibacterium* sp., and *Mycobacterium* sp.), Acidobacteria (*Candidatus Solibacter* sp. and order iii1-15), and Planctomycetes (order WD2101) (see table S1 for a complete list). Notably, less than 18% of the 511 phylotypes that we identified had a match to an available reference genome at the >97% 16S rRNA sequence similarity level, the level commonly used for delineating different bacterial species (23) (Fig. 2 and table S1). Approximately 42% of the dominant 511 phylotypes had no genome match even at the >90% 16S rRNA sequence similarity level, indicating that we do not have genomic information for taxa even within the same genus or family (Fig. 2A and table S1). Further, only 45% of the identified 511 dominant phylotypes are related to cultivated isolates and <30% of the phylotypes have representative type strains at the >97% sequence similarity level (Fig. 2B and table S1), which emphasizes the limited amount

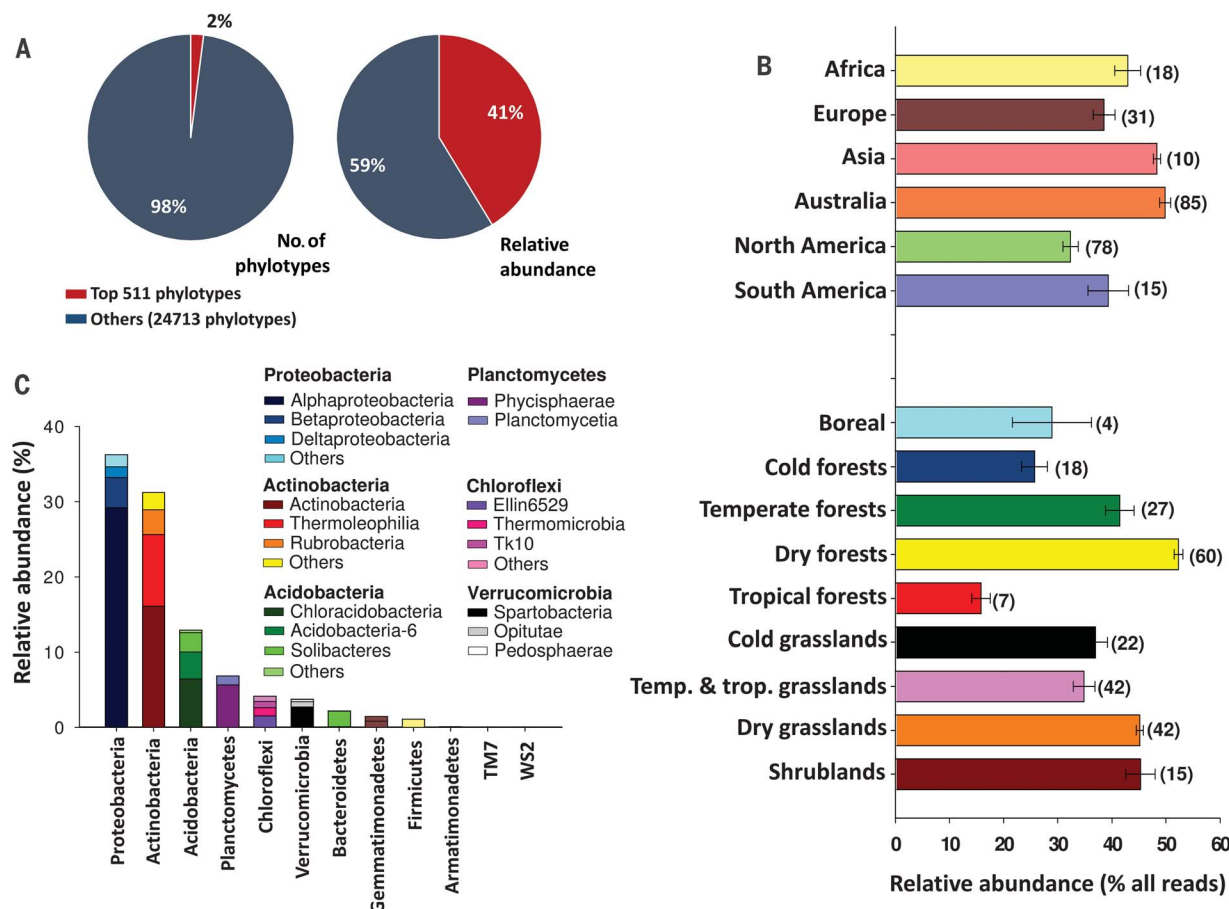


Fig. 1. Abundance and composition of dominant soil bacterial phylotypes across the globe. (A) Percentage of phylotypes and relative abundance of 16S rRNA genes representing the dominant versus the remaining bacterial phylotypes. (B) Relative abundance (mean \pm SE) of dominant phylotypes across continents and ecosystem types. Ecosystem type classification followed the Köppen climate classification and the major vegetation types found in our database. Grasslands include both

tropical and temperate grasslands. Shrublands include polar, temperate, and tropical shrublands. The number of samples in each category is indicated in parentheses. (C) The taxonomic composition of the dominant phylotypes. The phylotypes assigned to the least abundant phyla are not shown (including *Armatimonadetes* = 0.08%, *TM7* = 0.05%, and *WS2* = 0.03%). Details on the top 511 dominant phylotypes are shown in table S1.

of phenotypic information that we have available for these dominant phylotypes. Not surprisingly, phylotypes closely related to previously cultivated taxa tended to come from a few well-studied taxonomic groups, mostly *Proteobacteria*

and *Actinobacteria*, with only a few representatives available from other phyla (Figs. 1C and 2B and table S1), highlighting the well-known taxonomic biases of many preexisting culture collections (6).

After identifying the dominant 511 phylotypes, we used random forest modeling (24) to identify habitat preferences for each phylotype (20). Our statistical models included 15 environmental factors: climate (aridity index, minimum and maximum temperature, precipitation seasonality, and mean diurnal temperature range), ultraviolet (UV) radiation, net primary productivity, soil abiotic properties (soil texture; pH; total C, N, and P concentrations; and C:N ratio), and dominant ecosystem type (forests and grasslands) (20). We found that 53% (270) of the dominant 511 phylotypes had predictable habitat preferences [models explaining >30% of the variation; see (20) and table S1], with soil pH, climatic factors (aridity index, maximum temperature, and precipitation seasonality), and plant productivity consistently being the best predictors of their abundances across the globe (fig. S6). These findings are in line with previous research demonstrating that climatic factors and soil pH are often highly correlated with observed differences in overall soil bacterial community composition (4, 8, 9, 15, 16), but additionally, we found a strong link between microbial community composition and plant productivity (fig. S7). We were unable to identify a strong ecological preference for the remaining 241 of the 511 phylotypes, which included representatives from a wide range of phyla and subphyla (fig. S8). Our inability to predict the distributions of these 241 phylotypes could be related to the absence of key, but hard to measure, environmental predictors (e.g., soil C availability) or the fact that our models did not take into account specific associations between the bacteria and plants, fungi, or animals (e.g., pathogen-host or predator-prey interactions), which may be driving their distribution patterns. Alternatively, we may not have been able to identify the habitat preferences of these phylotypes because of low variability in their abundances across the samples (figs. S9 and S10). Indeed, the relative abundance of the group including all 241 undetermined phylotypes showed a much lower coefficient of variation than the relative abundance of those phylotypes for which we could identify their habitat preferences, as explained below (fig. S9). This result suggests that the undetermined phylotypes, those with no clearly identifiable habitat preferences, represent a “core” group of dominant phylotypes that are ubiquitous across global soils with proportional abundances that are relatively invariant.

We then used semipartial correlations (Spearman) and clustering analyses (20) to identify groups of phylotypes with shared habitat preferences, restricting our analyses to those 270 phylotypes with predictable distribution patterns. We found that the phylotypes group into five reasonably well-defined ecological clusters sharing environmental preferences for (i) high pH; (ii) low pH; (iii) drylands; (iv) low plant productivity; and (v) dry-forest environments (Figs. 2B and 3A, fig. S11, and table S1). These five clusters of phylotypes included 200 out of the 270 phylotypes for which we could identify their habitat preferences (table S1). Each of the ecological clusters

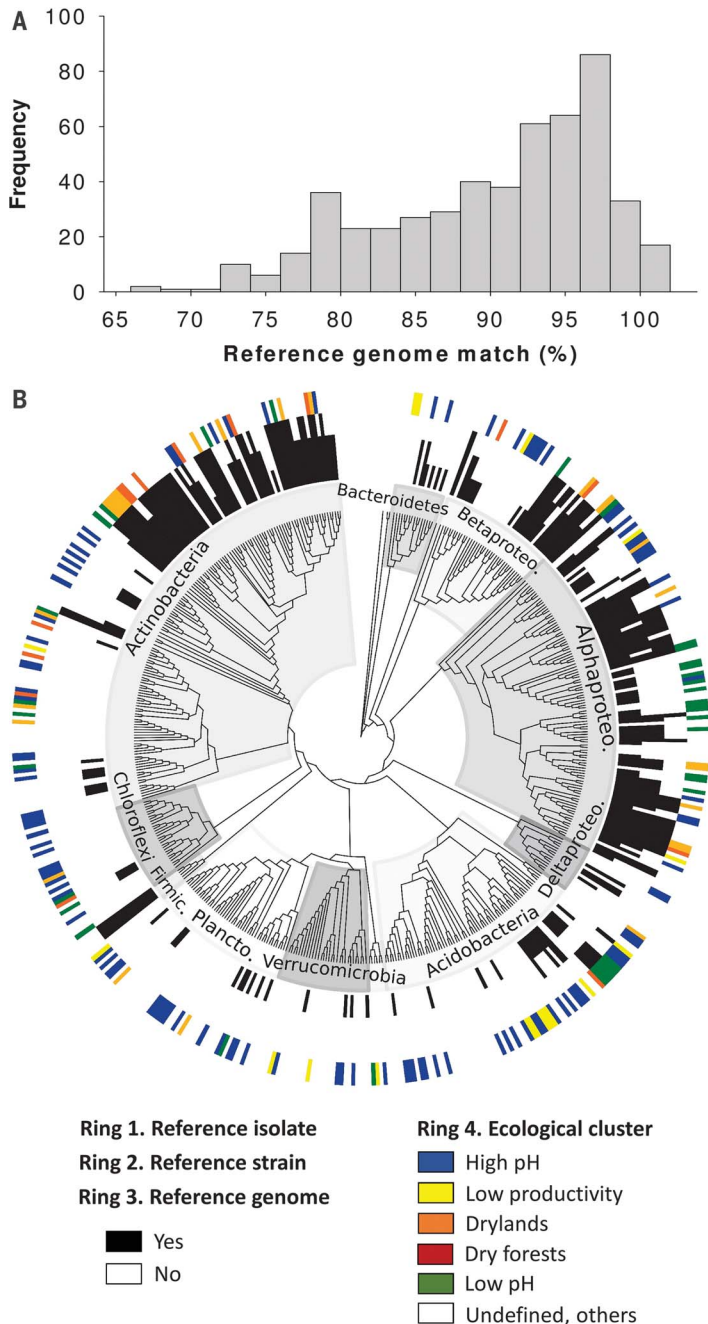


Fig. 2. Phylogenetic tree including the taxonomic information on dominant soil bacterial phylotypes. (A) Histogram showing the percentage 16S rRNA gene sequence similarity between the 511 dominant phylotypes and the most closely related available reference genome for each phylotype. (B) Phylogenetic distribution of the 511 dominant phylotypes. Black shading on the innermost and middle rings indicates, for each phylotype, whether there is a representative isolate and a genome match at the $\geq 97\%$ 16S rRNA gene sequence similarity level. The coloring on the outermost ring highlights the distribution of environmental preferences for all phylotypes ($n = 511$). For the few phylotypes where taxonomic assignment did not correspond to tree topology, no manual corrections were made. Betaproteo., Betaproteobacteria; Alphaproteo., Alphaproteobacteria; Deltaproteo., Deltaproteobacteria; Plancto., Planctomycetes; Firmic., Firmicutes.

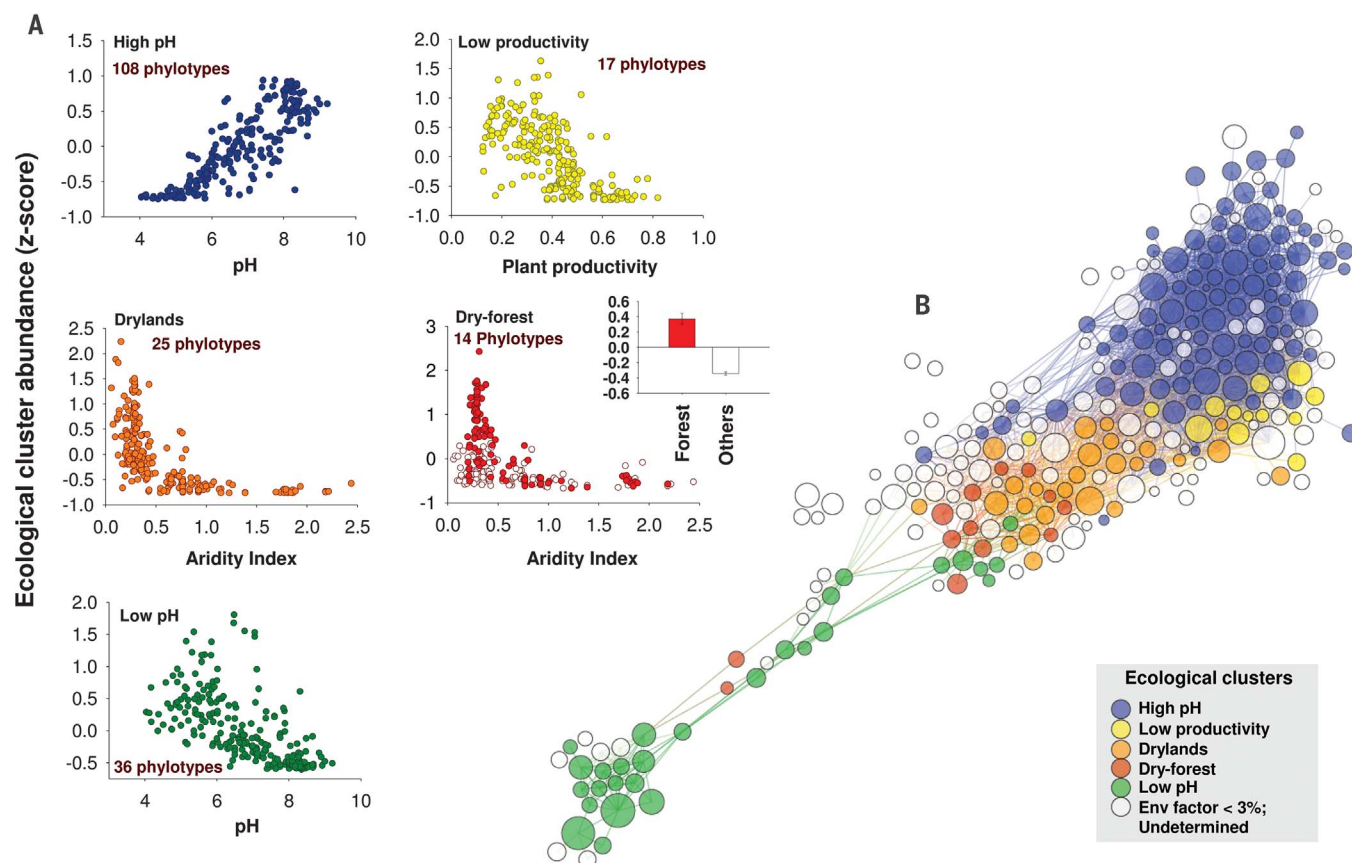


Fig. 3. Identified habitat preferences for dominant soil bacterial phylotypes. (A) Relationships between the relative abundance of the phylotypes assigned to each ecological cluster and their major environmental predictors (statistical analyses and identity of phylotypes within each cluster are presented in table S1). (B) Network

diagram with nodes (bacterial phylotypes) colored by each of the five major ecological clusters that were identified, highlighting that the phylotypes within each ecological cluster tend to co-occur more than expected by chance (statistical analyses presented in fig. S12).

identified included phylotypes from multiple phyla, suggesting that habitat preferences are not linked to phylogeny at coarse levels of resolution (fig. S8). The remaining 70 phylotypes were classified into three minor clusters, including a small cluster consisting of six phylotypes (high pH–forest preference; table S1 and fig. S11) and two clusters that included phylotypes with preferences including warm-forests, sites with low seasonal variation in precipitation, mesic environments, and soils of low phosphorus content (table S1 and fig. S11). These results suggest that the dominant bacterial phylotypes can be clustered into predictable ecological groups that share similar habitat preferences. To cross-validate the ecological clusters, we used correlation network analyses (20, 25) to investigate whether bacterial phylotypes sharing similar habitat and environmental preferences tend to co-occur (Fig. 3B). Indeed, our network analyses indicated that bacterial phylotypes sharing a particular habitat preference (e.g., low pH) tend to co-occur with other phylotypes belonging to the same cluster more than we would expect by chance ($P < 0.001$ for all clusters; Fig. 3B and fig. S12).

We next sought to determine if we could identify genomic attributes that delineate bacteria as-

signed to the individual ecological clusters. These analyses were restricted to the relatively small subset of bacterial phylotypes for which genomic data were available (>97% 16S rRNA sequence similarity to a reference genome). An insufficient number of representative unique genomes were available from phylotypes in four of the five major clusters identified (fig. S13). However, we had genomic data for 10 unique genomes out of 25 phylotypes assigned to the “drylands” cluster, including representatives of the *Proteobacteria* and *Actinobacteria* phyla (fig. S13). We then identified functional genes that were overrepresented in this “drylands” cluster as compared to the genomes available for the other dominant taxa. A total of 72 genomes were included in this analysis, with 10 of these genomes belonging to the dryland cluster (20). We found that the genomes within this dryland cluster had significantly higher relative abundances of 18 genes (fig. S14) compared to genomes representative of phylotypes assigned to other ecological clusters. Notably, *Mnh* and *Mrp* genes, which encode membrane transport proteins responsible for the proton-mediated efflux of monovalent cations (e.g., Na^+ , K^+), were overrepresented in the “drylands” cluster (fig. S14). These genes have frequently been

linked to increased bacterial tolerance to alkaline or saline conditions and, more generally, a greater capacity to tolerate external changes in the osmotic environment (26). These adaptations are likely to be important for bacteria living in arid soils, which are often saline, have high pH values, and experience prolonged periods of low moisture availability (27). Given the low number of reference genomes available, these findings are not conclusive and are simply a “proof of concept.” Nevertheless, our results highlight that it is possible to identify genomic attributes that differentiate soil bacteria with distinct environmental preferences. They also emphasize the importance of acquiring new genomes to further understand the ecological attributes of dominant soil bacterial taxa. As such, our results pave the way for leveraging genomic data to predict the spatial distributions of soil bacterial taxa, efforts that will be improved as the collections of reference genomes from these microorganisms increase in size.

Together, our results suggest that there are predictable clusters of co-occurring dominant bacterial phylotypes in soils from across the globe. This finding indicates that commonly available environmental information could be

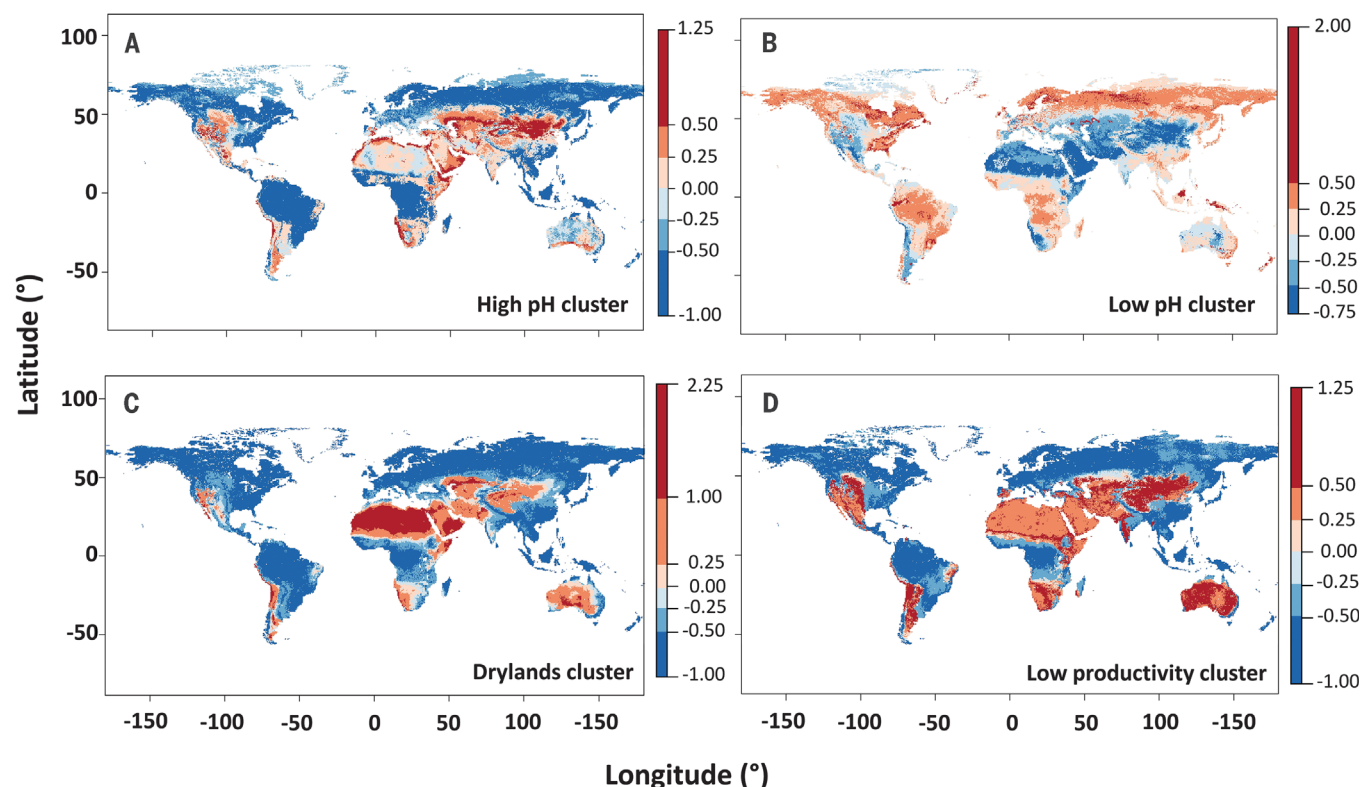


Fig. 4. A global atlas of the dominant bacteria found in soil.

(A to D) Predicted global distribution of the relative abundances of the four major ecological clusters of bacterial phylotypes sharing habitat preferences for high pH, low pH, drylands, and low plant productivity. R^2 (percentage of variation explained by the models) as follows:

(i) high-pH cluster, $R^2 = 0.53$, $P < 0.001$; (ii) low-pH cluster, $R^2 = 0.36$, $P < 0.001$; (iii) drylands cluster, $R^2 = 0.64$, $P < 0.001$; and (iv) low-productivity cluster, $R^2 = 0.40$, $P < 0.001$. The scale bar represents the standardized abundance (z-score) of each ecological cluster. An independent cross-validation for these maps is available in (20).

used to build predictive maps of the global distributions of these bacterial clusters at a global scale. We did so for the four major ecological clusters (i.e., low pH, high pH, drylands, and low productivity, Fig. 4) (20) using the prediction-oriented regression model Cubist (28) and information on 12 environmental variables for which we could acquire globally distributed information (20). Our models confirm that pH, aridity levels, and net primary productivity are major drivers of the low-pH, high-pH, dryland, and low-productivity clusters observed, respectively (Appendix S1). Notably, our maps (which accounted for 36 to 64% of the spatial variation in these clusters, Fig. 4) provide estimates of the regions where we would expect the groups of dominant soil bacterial phylotypes to be most abundant (Fig. 4). As expected, the dryland and low-productivity clusters were relatively abundant across the globe, and the low- and high-pH clusters were particularly abundant in areas known for their low- or high-pH soils, respectively.

This global inventory of dominant soil bacterial phylotypes represents a small subset of phylotypes that account for almost half of the 16S rRNA sequences recovered from soils. We show that we can predict the environmental preferences for more than half of these dominant phylotypes, making it possible to predict how future envi-

ronmental change will affect the spatial distribution of these taxa. Following Grime's mass ratio hypothesis (10), we would expect that identifying the physiological attributes of these dominant taxa will be critical for improving our understanding of the microbial controls on some key soil processes, including those that regulate soil C and nutrient cycling (1–3, 29). Also, given the strong links between the distribution of bacterial phylotypes and their functional attributes across the globe (8, 12), and the observed associations between dominant and subdominant phylotypes (fig. S5), we expect that these dominant bacteria will be critical drivers, or indicators, of key soil processes worldwide. We also found that habitat preferences were not predictable from phylum-level identity alone, given that all of the ecological clusters included phylotypes from multiple phyla. This suggests that phylotypes from diverse taxa share some phenotypic traits (e.g., osmoregulatory capabilities) or life-history strategies (29, 30) that allow them to survive under particular environmental conditions. By narrowing down the number of phylotypes to be targeted in future studies from tens of thousands to a few hundred, our study paves the way for a more predictive understanding of soil bacterial communities, which is critical for accurately forecasting the ecological consequences of ongoing global environmental change.

REFERENCES AND NOTES

1. J. M. Tiedje, S. Asuming-Brempong, K. Nüsslein, T. L. Marsh, S. J. Flynn, *Appl. Soil Ecol.* **13**, 109–122 (1999).
2. R. D. Bardgett, W. H. van der Putten, *Nature* **515**, 505–511 (2014).
3. P. L. E. Bodelier, *Front. Microbiol.* **2**, 80 (2011).
4. K. S. Ramirez et al., *Proc. R. Soc. B* **281**, 20141988 (2014).
5. M. Land et al., *Funct. Integr. Genomics* **15**, 141–161 (2015).
6. P. D. Schloss, R. A. Girard, T. Martin, J. Edwards, J. C. Thrash, *MBio* **7**, e00201–e00216 (2016).
7. C. Lok, *Nature* **522**, 270–273 (2015).
8. N. Fierer et al., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21390–21395 (2012).
9. F. T. Maestre et al., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15684–15689 (2015).
10. J. P. Grime, *J. Ecol.* **86**, 902–910 (1998).
11. D. VanInsberghe et al., *ISME J.* **9**, 2435–2441 (2015).
12. N. Fierer et al., *Science* **342**, 621–624 (2013).
13. T. E. Brewer, K. M. Handley, P. Carini, J. A. Gilbert, N. Fierer, *Nat. Microbiol.* **2**, 16198 (2016).
14. P. H. Janssen, *Appl. Environ. Microbiol.* **72**, 1719–1728 (2006).
15. C. L. Lauber, M. Hamady, R. Knight, N. Fierer, *Appl. Environ. Microbiol.* **75**, 5111–5120 (2009).
16. J. Zhou et al., *Nat. Commun.* **7**, 12083 (2016).
17. S. M. Prober et al., *Ecol. Lett.* **18**, 85–95 (2015).
18. J. W. Leff et al., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10967–10972 (2015).
19. K. T. Konstantinidis, A. Ramette, J. M. Tiedje, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 1929–1940 (2006).
20. Materials and methods are available as supplementary materials.
21. S. Soliveres et al., *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150269 (2016).
22. C. Thompson, *Nature* **551**, 33 (2017).
23. E. Stackebrandt, B. M. Goebel, *Int. J. Syst. Bacteriol.* **44**, 846–849 (1994).
24. L. Breiman, *Mach. Learn.* **45**, 5–32 (2001).

25. A. B. de Menezes *et al.*, *Environ. Microbiol.* **17**, 2677–2689 (2015).
26. T. H. Swartz, S. Ikewada, O. Ishikawa, M. Ito, T. A. Krulwich, *Extremophiles* **9**, 345–354 (2005).
27. W. G. Whitford, *Ecology of Desert Systems* (Academic Press, San Diego, CA, 2002).
28. J. R. Quinlan, *C4.5: Programs for Machine Learning* (Morgan Kaufmann, San Mateo, CA, 1993).
29. A. Barberán, H. Caceres Velazquez, S. Jones, N. Fierer, *MSphere* **2**, e00237–e17 (2017).

ACKNOWLEDGMENTS

We thank V. Ochoa and B. Gozalo for help with soil analyses. We also thank S. Karunaratne for advice on mapping analyses, and T. Hartshorn and B. Turner for providing soil samples from Montana and Panama. M.D.-B. acknowledges support from the Marie Skłodowska-Curie Actions of the Horizon 2020 Framework Programme H2020-MSCA-IF-2016 under Research Executive

Agency grant agreement no. 702057. N.F. was supported by grants from the U.S. National Science Foundation (EAR 1331828 and DEB 1554415). B.K.S. and M.D.-B. were supported by the Australian Research Council project (DP170104634). D.J.E. was supported by the Hermon Slade Foundation, and R.D.B. was supported by the UK Department of Environment, Food and Rural Affairs project no. BD5003 and a Biotechnology and Biological Sciences Research Council International Exchange Grant (BB/L026406/1). A.B.-G. was supported by a Formación de Personal Investigador (FPI) PhD fellowship (BES-2013-062945) and two research grants (CTM2015-64728-C2-1-R and CTM2012-38222-C02-01) from the Spanish Ministerio de Economía y Competitividad. The work of F.T.M. and the global drylands database were supported by the European Research Council [ERC Grant Agreements 242658 (BIOCOM) and 647038 (BIODESERT)] and by the Spanish Ministry of Economy and Competitiveness (BIOMOD project, ref. CGL2013-44661-R). M.D.-B. and N.F. designed the study. Field sampling was conducted by N.F., M.D.-B., F.T.M., A.B.-G., D.J.E., and R.D.B. Lab analyses

were done by F.T.M. and B.K.S. Bioinformatic analyses were done by T.E.B. Statistical modeling and mapping were done by M.D.-B. Network analyses were done by A.M.O. The manuscript was written by M.D.-B. and N.F. with contributions from all coauthors. All data used in this study are publicly available in Figshare (<https://figshare.com/s/82a2d3f5d38ace925492>; DOI: 10.6084/m9.figshare.5611321).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6373/320/suppl/DC1

Material and Methods

Table S1

Appendix S1

Figs. S1 to S14

References (30–58)

13 September 2017; accepted 13 December 2017

10.1126/science.aap9516

SOCIAL SCIENCE

Improving refugee integration through data-driven algorithmic assignment

Kirk Bansak,^{1,2*} Jeremy Ferwerda,^{2,3*} Jens Hainmueller,^{1,2,4*†} Andrea Dillon,² Dominik Hangartner,^{2,5,6} Duncan Lawrence,² Jeremy Weinstein^{1,2}

Developed democracies are settling an increased number of refugees, many of whom face challenges integrating into host societies. We developed a flexible data-driven algorithm that assigns refugees across resettlement locations to improve integration outcomes. The algorithm uses a combination of supervised machine learning and optimal matching to discover and leverage synergies between refugee characteristics and resettlement sites. The algorithm was tested on historical registry data from two countries with different assignment regimes and refugee populations, the United States and Switzerland. Our approach led to gains of roughly 40 to 70%, on average, in refugees' employment outcomes relative to current assignment practices. This approach can provide governments with a practical and cost-efficient policy tool that can be immediately implemented within existing institutional structures.

Refugees are among the world's most vulnerable populations (1, 2). After experiencing war, violence, and years of living in overcrowded refugee camps, refugees arrive in a new country with few resources and must acclimate to an unfamiliar local language, economy, and culture. Refugees frequently remain economically marginalized, with low levels of employment in the years following their arrival (3–5).

The assignment of refugees to different resettlement locations within a host country is one of the first policy decisions made during the resettlement process (6). It is also one of the most consequential in maximizing refugees' economic integration and self-sufficiency as a first step toward a more comprehensive integration into society (7–9). Three sets of factors affect refugee integration: geographical context, personal characteristics, and synergies between geography and personal characteristics (Fig. 1 and fig. S1). For instance, some resettlement locations in the United States offer better economic and social opportunities that can result in higher levels of refugee employment (Fig. 1A). In addition, refugees with certain characteristics, such as language and educational skills, are more likely to succeed economically regardless of the resettlement location to which they are sent (Fig. 1B). Finally, the expected employment returns associated with personal characteristics can vary across different resettlement locations (Fig. 1C). This indicates that there are synergies between places and people; certain

characteristics will make a refugee a better match for a particular location. In Switzerland, for example, we find that the ability to speak French (i.e., among French-speaking African refugees) results in a larger payoff for refugees assigned to French-speaking cantons than for those assigned to German-speaking cantons (fig. S2).

Host countries' current procedures for determining how to allocate refugees across domestic resettlement sites do not fully leverage synergies between refugees and geographic locations. For instance, in the United States, refugees without existing U.S. ties are primarily assigned to resettlement locations according to the capacity of local resettlement offices at the time of arrival, without a systematic assessment of the local employment rate for refugees of similar profiles. In Switzerland, where most refugees initially enter as asylum seekers, the federal government attempts to reduce fiscal and social strain on individual localities by making assignment random and proportional across regions.

Prior research has proposed different schemes for refugee assignment both across countries (10, 11) and within countries (12, 13). These proposals include two-sided matching markets in which an optimized assignment is determined on the basis of match efficiency and/or the preferences of refugees and host locations (14). Although these approaches are theoretically appealing, there are practical barriers to their implementation, including a lack of systematic data on refugee preferences and the need for extensive political coordination.

We have developed a data-driven approach that, in contrast, can be immediately implemented by using existing data to optimize integration outcomes. Our algorithm has three stages: modeling, mapping, and matching. The modeling stage involves a supervised machine learning process that predicts the expected success for any quantifiable metric—for example, early employment—of new

refugee arrivals across all possible resettlement locations. We designated historical resettlement data for model training, in which the unit of observation was a single refugee and which contained information on the refugees' background characteristics (e.g., country of origin, language skills, gender, age, etc.), time of arrival, assigned location, and measured employment success. These training data were then used to build a bundle of supervised learning models that predicted refugees' expected employment success as a function of their background characteristics. A separate model was fit for subgroups of refugees assigned to each location, thus yielding different models for each location and allowing for the discovery of refugee/location synergies. These fitted models were then applied to new, out-of-sample refugee arrival data to predict the expected employment success of each new arrival at each possible resettlement location.

The mapping stage involves transforming the refugee-level predictions from the modeling stage to a case-level metric. Mapping to a case-level metric is necessary because refugees are often not assigned to locations on an individual basis, but rather on a case-level basis, with cases most often being family units. Various mapping functions can be used. Our preferred case-level metric was the predicted probability that at least one refugee in the case would find employment at the location in question. This metric uses a simplifying assumption that the probabilities of employment for refugees within a case are independent, although we also tested alternative mapping functions—namely the mean, maximum, and minimum predicted probability of employment within each case—that do not require this assumption (15).

Finally, the matching stage involves assigning each case to a specific location to fulfill a chosen optimality criterion subject to constraints. Our algorithm is flexible and can accommodate multiple criteria and constraints. The optimality criterion we used in our applications was to maximize the average of the case-level metric (i.e., the global average of the probability that at least one refugee in each family gains employment). We also imposed constraints that represent real-world assignment restrictions, such as how many cases can be sent to different locations. To solve this constrained optimization problem, we used an optimal matching procedure with the RELAX-IV minimum cost flow solver (16, 17); see supplementary materials and figs. S3 to S5 for details of the algorithm, data, measures, and statistical analysis (including out-of-sample classification accuracy and probability calibration).

For the algorithm to obtain reliable predictions, it is important that the historical assignment process not be determined by unobserved refugee characteristics. This criterion is currently met in many countries that assign refugees either randomly (according to burden-sharing constraints) or according to premeasured refugee characteristics that would serve as feature inputs into the algorithm. We assessed the performance of the

¹Department of Political Science, Stanford University, Stanford, CA 94305, USA. ²Immigration Policy Lab, Stanford University, Stanford, CA 94305, USA, and ETH Zurich, 8092 Zurich, Switzerland. ³Department of Government, Dartmouth College, Hanover, NH 03755, USA. ⁴Graduate School of Business, Stanford University, Stanford, CA 94305, USA. ⁵Center for Comparative and International Studies, ETH Zurich, 8092 Zurich, Switzerland. ⁶Department of Government, London School of Economics and Political Science, London WC2A 2AE, UK.

*These authors contributed equally to this work.

†Corresponding author. Email: jhain@stanford.edu

algorithm through applications in two such countries: the United States, where refugees are assigned primarily on the basis of capacity constraints, and Switzerland, where refugees are assigned randomly according to a proportional distribution key (see supplementary materials and tables S1 and S2 for details).

In the United States, reception and placement services (e.g., arranging location assignments,

housing, etc.) for refugees are implemented by nine voluntary agencies in cooperation with the Department of State. After refugees are allocated to one of the agencies, placement officers centrally assign refugees to the agency's resettlement locations subject to local capacity constraints (18). Placement officers make assignment decisions prior to refugees' arrival and without interviewing the refugees. The premeasured characteristics of a

case available to the placement officers can be viewed in the data, and hence can be used as feature inputs into the algorithm.

Refugees are granted work authorization upon arrival and encouraged to find employment as soon as possible. To track refugee resettlement success, the agencies are required to report the refugees' employment status at the end of the reception and placement period, 90 days after

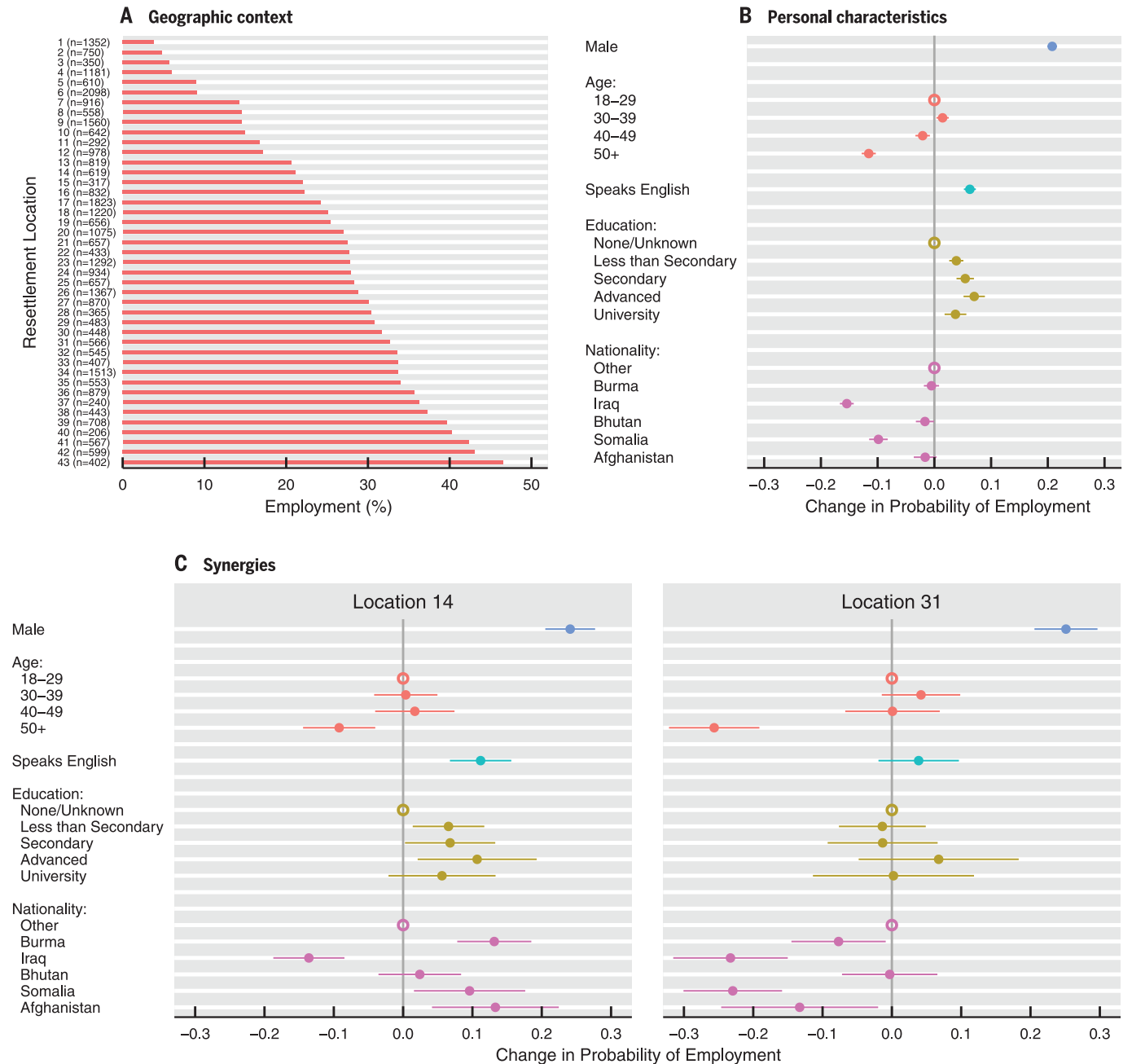


Fig. 1. Variation in refugee employment in the United States. (A to C) Refugee employment at 90 days after arrival varies as a function of refugees' assigned resettlement location (A), personal characteristics (pooled across refugees assigned to all locations) (B), and synergies between characteristics and locations (two example locations) (C). In (B) and (C), dots with horizontal lines indicate point estimates with robust

95% confidence intervals from ordinary least-squares regression. The open circles on the zero line denote reference categories. The data for all three panels include working-age refugees resettled by one of the largest U.S. resettlement agencies during the 2011–2016 period ($n = 33,782$). These results are replicated for only working-age refugees without U.S. ties (i.e., “free cases”) in fig. S1.

arrival. To assess whether an optimized assignment could improve refugee outcomes, we analyzed de-identified data from one of the largest resettlement agencies for working-age refugees (ages 18 to 64; $n = 33,782$) resettled during the 2011–2016 period. We split the data into training and test sets. For model training, we used data for the refugees who arrived from 2011 up to (but not including) the third quarter (Q3) of 2016, the most recent quarter with available data. We then applied the fitted models to predict the expected employment success at each location and determine the optimal assignment for the test set, refugees who arrived in 2016 Q3. For the test data, we focused on refugees who were free to be assigned to different resettlement locations ($n = 919$), in contrast to refugees who are assigned according to the location of family or other ties. We also imposed constraints on the assignment such that each location could only receive as many cases under the optimized assignment as were received in actuality.

Our algorithmic assignment considerably increased expected refugee employment over the status quo assignment (Fig. 2). The median refugee's predicted probability of employment in the United States more than doubled, increasing from approximately 25% to 50%. Our optimized assignment increased the probability of finding employment across the entire distribution of refugees, including those who were least likely and most likely to find work.

In addition, the algorithmic assignment yielded higher employment rates in almost every location, including locations that had higher and lower baseline employment rates. On average, the employment rate was 34% under the actual assignment and 48% under the optimized assignment, which means that the optimized assignment would increase the employment rate above the baseline by roughly 41%.

We conducted a second test in the context of Switzerland, whose asylum process is similar to that of other European countries belonging to the

Common European Asylum System. In Switzerland, asylum seekers who are not immediately rejected upon arrival are assigned to one of 26 cantons, where they wait for a decision on their asylum application. We focused on asylum seekers who received subsidiary protection status, which is Switzerland's largest refugee category (see supplementary materials). We drew upon data from the Swiss State Secretariat for Migration (SEM), which centrally manages the asylum process and assignment. In contrast to the U.S. case, the SEM uses a proportional random assignment of cases to locations, and tracks employment outcomes for several years after asylum seekers' arrival. This allowed us to benchmark our algorithm against a different status quo assignment mechanism and to optimize for a longer-term employment metric—specifically, refugees' employment at the end of their third calendar year in Switzerland. We focused on all working-age refugees who received subsidiary protection status and arrived from 1999 to 2013 ($n = 22,159$), with refugees arriving in 2013 who were free to be assigned to any canton as the test set ($n = 888$) and refugees arriving in all prior years as the training set. We also imposed the constraint that each canton gets assigned the same number of cases as in actuality, in which the number of cases, by law, is assigned in proportion to the population of the canton.

Our algorithmic assignment considerably increased expected refugee employment over the status quo assignment (Fig. 3). Similar to the U.S. context, our algorithm increased the predicted probability of finding employment across the entire distribution of refugees. On average, the third-year employment rate was 15% under the actual assignment and 26% under the optimized assignment. These results suggest that the data-driven assignment has the potential to increase third-year employment in the Swiss context by about 73%.

In the supplement, we present further results for both countries in which we applied alternative specifications for the algorithm. Specifically,

we replicated testing with different time periods (figs. S6 and S7), alternative mapping functions (fig. S8), shorter- and longer-term outcomes (fig. S9), and varying lengths of the training data period (fig. S10). The results from these tests all show considerable gains.

Our analysis demonstrated large potential improvements, but we did not test the algorithm prospectively. Ideally, it should be tested in a randomized controlled trial design. In addition, further research is needed to determine whether it is more effective to optimize for short-term or long-term outcomes. In Switzerland, for example, we find considerable gains regardless of whether we optimize for second-, third-, or fourth-year employment (see supplementary materials). In the United States, however, longer-term employment outcomes are currently not tracked. Still, early employment is often highly predictive of long-term employment (19), and the use of shorter-term outcomes in the algorithm allows for faster learning of emerging and declining synergies based on more recent data, possibly resulting in a more effective assignment.

In contrast to more expensive interventions (such as language or job training programs) that are sometimes implemented long after refugees' arrival, our approach is cost-efficient and implemented before refugees' arrival, giving them the strongest foundation possible from which to integrate into host societies. Furthermore, our approach modifies an existing policy process, facilitating its immediate implementation, and it is dynamic in that it adapts to synergies over time. Because of the algorithm's data-driven learning capacity, policy-makers do not need to invest in identifying the precise sources of those synergies—local economic conditions, social environments, resettlement office efficacy, etc.—to harness their benefits.

Our approach also preserves the ability of policy-makers to set their own parameters and priorities. Specifically, policy-makers can choose their

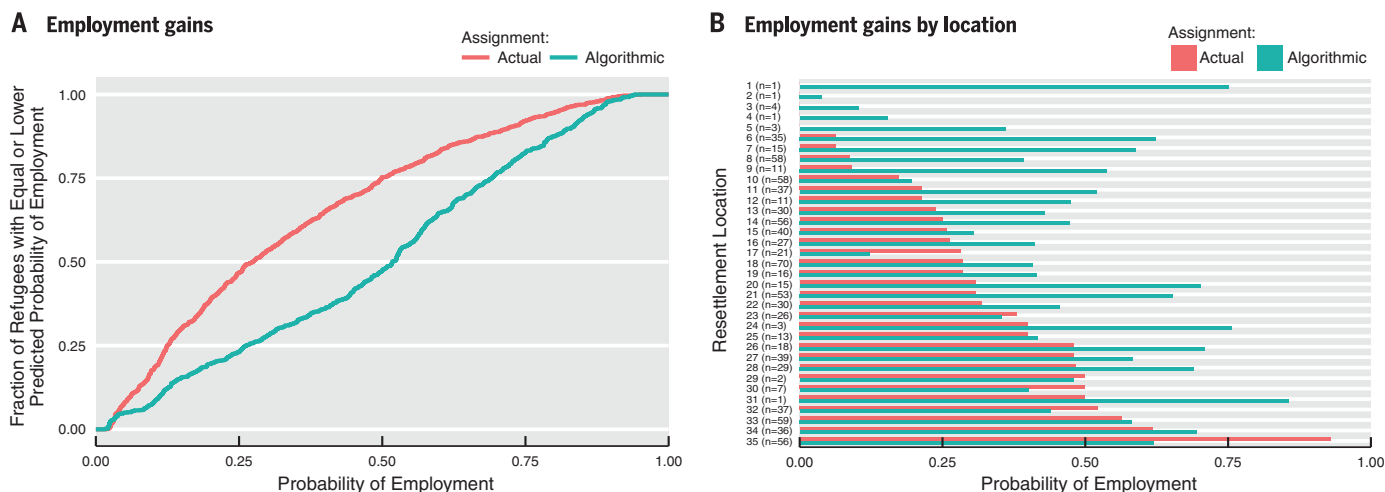


Fig. 2. Employment gains from data-driven refugee assignment in the United States. (A) Empirical cumulative distribution functions (ECDFs) of the refugees' predicted 90-day employment probabilities under their actual and algorithmic assignments. (B) Actual and algorithmic employment rates by resettlement location.

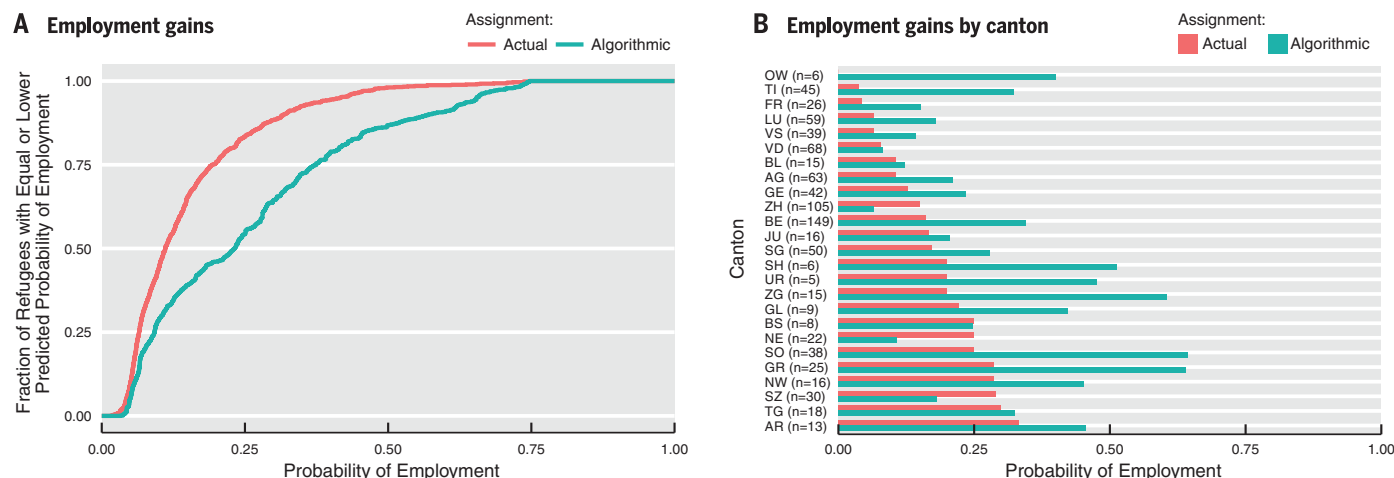


Fig. 3. Employment gains from data-driven refugee assignment in Switzerland. (A) ECDFs of the refugees' predicted third-year employment probabilities under their actual and algorithmic assignments. **(B)** Actual and algorithmic employment rates by canton. See table S3 for canton names.

own preferred integration success metrics, optimality criteria, and constraints for the assignment. For example, in our U.S. application, we find that the average employment gains achieved under our optimized assignment could be improved even further by slightly relaxing the constraints and allowing the algorithm to increase or reduce the number of cases assigned to different resettlement locations (fig. S11). In addition, if systematic data on refugee preferences regarding their geographic placement would become available, these data could also be incorporated into the algorithmic assignment by, for instance, optimizing for a weighted average of preferences and predicted integration success. Lastly, the algorithm could also be used to complement rather than replace placement officers' expertise. For instance, in a computer-assisted assignment process, the algorithm might provide several recommendations, and placement officers could use their own discretion to determine the final assignment or override any suggestions.

REFERENCES AND NOTES

- J. Smith, L. Daynes, *Lancet Glob. Health* **4**, e85–e86 (2016).
- Médecins Sans Frontières, *The Illness of Migration: Ten Years of Medical Humanitarian Assistance to Migrants in Europe and in Transit Countries* (2013); www.aerzte-ohne-grenzen.de/sites/gemany/files/attachments/msf-the-illness-of-migration-2013.pdf.
- P. Connor, *J. Refug. Stud.* **23**, 377–397 (2010).
- J. Hainmueller, D. Hangartner, D. Lawrence, *Sci. Adv.* **2**, e1600432 (2016).
- M. Marbach, J. Hainmueller, D. Hangartner, *The Long-Term Impact of Employment Bans on the Economic Integration of Refugees* (Stanford-Zurich Immigration Policy Lab Working Paper 17-03, 2017); <https://ssrn.com/abstract=3078172>.
- Although our focus is on refugee resettlement, the same issues also apply to the assignment of asylum seekers. Resettled refugees are persons who have officially been granted refugee status in advance of their arrival into the host country. In contrast, an asylum seeker is a person who has fled his or her home country and submitted a formal request for asylum in a host country. Authorities in that country then process the request and decide whether to grant official refugee status to the asylum seeker. This determination can take several years to conclude, and in the interim, the asylum seeker is typically placed within a specific location in the host country.
- P.-A. Edin, P. Fredriksson, O. Åslund, *Q. J. Econ.* **118**, 329–357 (2003).
- L. A. Beaman, *Rev. Econ. Stud.* **79**, 128–161 (2011).
- A. P. Damm, *J. Labor Econ.* **27**, 281–314 (2009).
- J. Fernández-Huertas Moraga, H. Rapoport, *J. Public Econ.* **115**, 94–108 (2014).
- J. Fernández-Huertas Moraga, H. Rapoport, *CESifo Econ. Stud.* **61**, 638–672 (2015).
- T. Andersson, L. Ehlers, *Assigning Refugees to Landlords in Sweden: Stable Maximum Matchings* (Lund University, 2016); http://project.nek.lu.se/publications/workpap/papers/wp16_18.pdf.
- D. Delacrétaz, S. D. Kominers, A. Teytelboym, *Refugee Resettlement* (University of Melbourne, 2016); www.t8el.com/jmp.pdf.
- Fernández-Huertas Moraga and Rapoport (10, 11) couple an auction mechanism for tradeable refugee quotas with a preference-matching algorithm that optimizes over refugees' preferences for resettlement countries and countries' preferences over refugee types. Andersson and Ehlers (12) focus on within-country matching and develop an algorithm to find a stable maximum matching of refugees to landlords given induced preferences for landlords and refugee families. Delacrétaz et al. (13) provide algorithms that optimize match efficiency subject to multidimensional capacity constraints and incorporate refugee preferences and location priorities.
- The formula we use is $y_{gi} = 1 - \prod_{j \in g} (1 - \alpha_{ij})$, where α_{ij} corresponds to the predicted probability of a positive employment outcome for refugee i at location j , and g denotes a particular case. Note that this formula uses a simplifying assumption that the probabilities of employment for refugees within a case are independent. See the supplement for additional details and alternative mapping functions—the mean, maximum, and minimum predicted probability of employment within each case—that do not require this assumption.
- B. B. Hansen, S. O. Klopfer, *J. Comput. Graph. Stat.* **15**, 609–627 (2006).
- D. P. Bertsekas, P. Tseng, *RELAX-IV: A Faster Version of the RELAX Code for Solving Minimum Cost Flow Problems* (MIT Laboratory for Information and Decision Systems, 1994).
- Each location can only accommodate a limited number of cases each year. Other less common restrictions include the inability of certain locations to accept cases with severe medical conditions or particular languages. The assignment algorithm is designed to incorporate such constraints.
- In the Swiss data, employment at the end of the first calendar year after arrival is associated with a 62–percentage point ($P < 0.0001$) increase in the probability of employment at the end of the second year and a 43–percentage point ($P < 0.0001$) increase at the end of the third calendar year; the P values are from a linear regression of actual second- or third-year employment on first-year employment, respectively.

ACKNOWLEDGMENTS

We acknowledge funding from the Ford Foundation for operational support of the Stanford Immigration Policy Lab, and from the Swiss National Science Foundation and the Carnegie Corporation of New York (J.H.). The funders had no role in the data collection, analysis, decision to publish, or preparation of the manuscript. The U.S. refugee registry data were provided to us under a collaboration research agreement with the Lutheran Immigration and Refugee Service (LIRS). This agreement requires that we do not transfer or disclose the data. Researchers can request access to the data from LIRS. The Swiss ZEMIS database is provided to us under a data use agreement with the Swiss State Secretariat for Migration (SEM), which requires that we do not disclose the individual-level data. Researchers can request access to the ZEMIS data from the SEM. We thank the SEM and LIRS for access to data and guidance, and G. Imbens, D. Laitin, R. Reich, L. Stanczyk, and S. Wager for helpful advice. For replication code, see Harvard Dataverse (doi:10.7910/DVN/MS8XES). The institutional review boards at Stanford University (IRB-40212) and Dartmouth College (00030198) approved this research. The U.S. resettlement agency that provided the data requested that we not disclose specific locations within the article.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6373/325/suppl/DC1
Supplementary Text
Figs. S1 to S11
Tables S1 to S3
References (20–37)

24 July 2017; accepted 13 December 2017
10.1126/science.aao4408

Cite as: N. K. Sinha *et al.*, *Science*
10.1126/science.aag0921 (2017).

Dicer uses distinct modules for recognizing dsRNA termini

Niladri K. Sinha, Janet Iwasa, Peter S. Shen,* Brenda L. Bass*

Department of Biochemistry, University of Utah, Salt Lake City, UT 84112, USA.

*Corresponding author. Email: bbass@biochem.utah.edu (B.L.B.); peter.shen@biochem.utah.edu (P.S.S.)

Invertebrates rely on Dicer to cleave viral double-stranded RNA (dsRNA), and *Drosophila* Dicer-2 distinguishes dsRNA substrates by their termini. Blunt termini promote processive cleavage, while 3' overhanging termini are cleaved distributively. To understand this discrimination, we used cryo-electron microscopy to solve structures of *Drosophila* Dicer-2 alone and in complex with blunt dsRNA. While the Platform-PAZ domains have been considered the only Dicer domains that bind dsRNA termini, unexpectedly, we found that the helicase domain is required for binding blunt, but not 3' overhanging, termini. We further showed that blunt dsRNA is locally unwound and threaded through the helicase domain in an ATP-dependent manner. Our studies reveal a previously unrecognized mechanism for optimizing antiviral defense and set the stage for discovery of helicase-dependent functions in other Dicers.

Dicer ribonucleases cleave double-stranded RNA (dsRNA) precursors to generate microRNAs (miRNAs) and small interfering RNAs (siRNAs) (1, 2). In concert with Argonautes, these small RNAs bind complementary mRNAs to down-regulate their expression. miRNAs are processed by Dicer from small hairpins, while siRNAs are typically processed from longer dsRNA, from endogenous sources (3), or exogenous sources such as viral replication intermediates (4–6). Some organisms, such as *Homo sapiens* and *Caenorhabditis elegans*, encode one Dicer that generates miRNAs and siRNAs, but other organisms have multiple Dicers with specialized functions.

Dicers exist throughout eukaryotes, and a subset have an N-terminal helicase domain of the RIG-I-like receptor (RLR) subgroup (7) (Fig. 1A and fig. S1A). RLRs often function in innate immunity (8), and Dicer helicase domains sometimes show differences in activity that correlate with roles in immunity. For example, *Drosophila melanogaster* expresses two Dicers, one specialized for processing miRNAs (dmDcr-1), and a second for siRNAs (dmDcr-2) (9). dmDcr-1 has a degenerate helicase domain and is an ATP-independent enzyme (10), while dmDcr-2, with dedicated antiviral roles (11–13), has a conserved helicase domain that hydrolyzes ATP (14–17). Under certain conditions *Homo sapiens* Dicer-1 (hsDcr-1) also generates viral siRNAs (18, 19). However, despite conservation of its helicase domain, hsDcr-1 does not hydrolyze ATP in vitro (20), and its helicase domain is not implicated in viral siRNA biogenesis in vivo (19). Differences in activities of the helicase domain of vertebrate and invertebrate Dicers may reflect distinct roles in antiviral defense.

dmDcr-2 activity depends on termini of its dsRNA substrates (15, 16). Blunt (BLT) termini promote a processive reaction whereby multiple siRNAs are produced before dmDcr-

2 dissociates, and this reaction requires a functional helicase domain and ATP (15, 17). In contrast, dsRNAs with 3' overhanging (3' ovr) termini promote an ATP-independent, distributive cleavage, whereby dmDcr-2 dissociates after each cleavage. hsDcr-1 does not require ATP for processing BLT or 3' ovr termini (fig. S1B, C and D, lanes 5, 7, 10, 12), suggesting that, at least in vitro, cleavage of BLT dsRNA is not dependent on its helicase domain.

To understand the mechanism of termini discrimination by dmDcr-2, we used cryo-electron microscopy (cryo-EM) to determine structures of dmDcr-2 alone and in complex with a BLT 52 base pair (bp) dsRNA (52 dsRNA) and ATP- γ S (Fig. 1, figs. S2 to S10, and table S1). We used full-length dmDcr-2 with a point mutation in each RNase III domain to preclude dsRNA cleavage (dmDcr-2^{RIII}) (Fig. 1A). ATP hydrolysis is required for processive cleavage of BLT dsRNA, and dmDcr-2 cannot hydrolyze ATP- γ S, which stabilizes a helicase-dependent conformation of dmDcr-2 (16).

The structure of apo-dmDcr-2^{RIII} (Fig. 1, B and C, and figs. S2 to S5) reiterated the “L shape” of lower resolution (~15–30 Å) EM reconstructions of hsDcr-1 (21–23). Our 7.1 Å EM density map (Fig. 1B and fig. S2C) enabled fitting (fig. S4, B to F) and homology modeling (Fig. 1B) of Platform-PAZ domains at the Cap and tandem RNase III domains in the Core. An additional round of 3D classification (fig. S3) revealed an 8.7 Å map (fig. S2C) allowing assignment of the helicase domain at the base (Fig. 1C and fig. S4A). Fitting of related apo-helicases into the EM density is consistent with the helicase domain adopting an open conformation (fig. S5, A to D).

The 2D class averages of the dmDcr-2 complex revealed protein with well-resolved secondary structure features bound to the BLT dsRNA terminus (fig. S6, B and C). Some protein density was missing, and since control experiments

indicated protein on the grid was intact (fig. S7A), this was likely due to inherent flexibility (fig. S7B). Measurements of the dsRNA, guided by major grooves, showed visible protein footprinted ~8-9 bps (fig. S6B). The crystal structure of RIG-I's helicase domain bound to dsRNA has a similar footprint (24, 25), suggesting dmDcr-2's helicase domain bound to BLT dsRNA termini. Indeed, our 6.8 Å reconstruction of the complex (figs. S6E and S8) resembled RIG-I in a closed conformation (Fig. 1D and fig. S9). The Hel1 and Hel2 subdomains of RIG-I's helicase, along with the pincer helices (fig. S1A) could be fit as a single rigid body (fig. S9A). The reconstruction also revealed a helical bundle characteristic of the Hel2i subdomain that could be fitted separately as a rigid body (fig. S9B). These fittings enabled a homology model of dmDcr-2's helicase domain bound to BLT dsRNA (Fig. 1D and fig. S9, D and E).

Our models of dmDcr-2's helicase in open (apo) and closed (substrate-bound) conformations implied clamping of the helicase on BLT dsRNA termini (Fig. 1E and movies S1 and S2). In the open conformation, Hel2 and Hel2i extend away from Hel1 creating a C-shaped opening for substrate engagement (Fig. 1C). In the BLT dsRNA-bound state, Hel2 and Hel2i swivel toward Hel1 to clamp on the terminus (Fig. 1, D and E, and movies S1 and S2).

Within the helicase domain, density was only observed for one RNA strand, indicative of local unwinding (Fig. 1F and fig. S9C). Unwinding would likely require ATP hydrolysis, and possibly was enabled by contaminating ATP in commercial preparations of ATP-γS (16). Using dsRNAs with a nick in sense or antisense strands (fig. S11A), we performed in vitro unwinding assays (fig. S11B). With ATP, dmDcr-2^{RIII}, but not the ATPase-defective Walker A mutant dmDcr-2^{RIII,K34A} (16), unwound BLT dsRNA termini (fig. S11B; compare lanes 3, 7, top and bottom panels).

The unwound single strand maintained an A-form conformation (Fig. 1F), likely to minimize entropic costs of reannealing before cleavage in RNase III sites. Whether the RIG-I helicase unwinds dsRNA is controversial (26, 27), but related helicases exhibit unwinding activity (28). Local unwinding may facilitate dmDcr-2's helicase domain in binding and translocating along dsRNA.

The Platform-PAZ domains have been considered the only Dicer domains that bind dsRNA termini (29–32), but our structures suggested the helicase domain also binds termini. To test this idea, we purified dmDcr-2^{ΔHel,RIII} (Fig. 2A), which lacked the helicase domain (Fig. 1A). Consistent with previous studies (16), in gel-shift assays full-length dmDcr-2^{RIII} bound both BLT and 3' ovr dsRNA (Fig. 2B, dsRNA design; and Fig. 2C, top); ATP increased affinity for BLT dsRNA and decreased affinity for 3' ovr dsRNA (Fig. 2, C and D and table S2). However, while dmDcr-2^{ΔHel,RIII} bound 3' ovr dsRNA with an affinity similar to dmDcr-2^{RIII}, its binding to BLT dsRNA

was not detected (Fig. 2C, bottom panel, Fig. 2D, and table S2). The inability of dmDcr-2^{ΔHel,RIII} to bind BLT dsRNA was not due to the absence of ATP hydrolysis because dmDcr-2^{RIII,K34A} bound BLT dsRNA efficiently (fig. S12, A and B, and table S2). Thus, the helicase domain is required for binding BLT, but not 3' ovr, dsRNA.

Single-turnover cleavage assays showed that neither dmDcr-2^{WT} nor dmDcr-2^{ΔHel} (fig. S12C) cleaved BLT dsRNA without ATP (Fig. 2, E and F, lanes 5 and 6). With ATP, cleavage of BLT dsRNA by dmDcr-2^{WT} gave heterogeneous cleavage products characteristic of Dicer enzymes with ATPase activity (16, 33), but strikingly, dmDcr-2^{ΔHel} was incapable of cleaving BLT dsRNA (Fig. 2, E and F; compare lanes 9 and 10). As expected (15, 16), cleavage of 3' ovr dsRNA was independent of ATP, and with both dmDcr-2^{WT} and dmDcr-2^{ΔHel} produced a single siRNA-sized 22 nt band (Fig. 2E; compare lanes 7, 8, 11, 12). dmDcr-2^{ΔHel} cleaved 3' ovr dsRNA more efficiently than dmDcr-2^{WT} (Fig. 2, E and F, compare lanes 7, 8, 11, 12), suggesting the helicase domain hinders cleavage of 3' ovr dsRNA. This observation is reminiscent of autoinhibition by hsDcr-1's helicase domain in processing 3' ovr dsRNAs (34).

Our biochemical and structural studies indicated that dmDcr-2 has two modes of substrate recognition and cleavage: one mediated by Platform-PAZ domains and resulting in precise cleavage of 3' ovr dsRNAs into 22mer siRNAs, and a second mediated by the helicase domain and resulting in heterogeneous cleavage of BLT dsRNAs. We searched for amino acids that might separately affect cleavage of a 3' ovr or BLT dsRNA. We created one variant of dmDcr-2 (Fig. 1A, PP) containing five point mutations in the Platform and PAZ domains (35). Multiple crystal structures show a phenylalanine in the C-terminal domain (CTD) of RIG-I recognizes BLT dsRNA by stacking on the terminal base pair (24). Dicer enzymes do not have a CTD, but for the second variant we searched for regions in dmDcr-2 with sequence similarity to the CTD. Within the region identified (fig. S13, A to C) we mutated a single phenylalanine to a glycine (dmDcr-2^{F225G}).

We compared activities of purified dmDcr-2^{PP} and dmDcr-2^{F225G} (fig. S13D) to dmDcr-2^{WT} using single-turnover cleavage assays (Fig. 3, A and B). As expected, cleavage of BLT dsRNA was not observed without ATP (Fig. 3, A and B, lanes 5 to 7). However, with ATP, cleavage of BLT dsRNA by dmDcr-2^{WT} or dmDcr-2^{PP} appeared nearly identical (Fig. 3, A and B, lanes 11 and 13), while cleavage was completely disrupted by the helicase point mutation in dmDcr-2^{F225G} (Fig. 3, A and B, lane 12). [At least part of this effect is due to weakened BLT dsRNA binding (fig. S13, E and F, and table S2)]. By contrast, cleavage of 3' ovr dsRNA was independent of ATP and minimally affected by the F225G helicase mutation (Fig. 3, A and B, lanes 8, 9, 14, 15). However, cleavage was eliminated by mutations in the Platform-PAZ domains (Fig. 3, A and B, lanes 10 and

16). These data reiterate that cleavage of 3' ovr dsRNA is mediated by Platform-PAZ domains, while the helicase domain coordinates recognition and cleavage of BLT dsRNA.

While dmDcr-2^{PP} cleaved BLT dsRNA to yield a pattern nearly identical to dmDcr-2^{WT}, levels of 22 nt siRNA decreased (Fig. 3, A and B, lanes 11 and 13). Since 22 nt siRNA was not observed with dmDcr-2^{F225G} (Fig. 3A, lane 12), we hypothesized this species derived from dsRNA that was threaded through the helicase domain until the BLT terminus encountered the Platform-PAZ domains. To confirm that smaller products (<22 nt) did not result from degradation of 22 nt siRNA, we monitored cleavage of chimeric dsRNAs containing deoxynucleotides at positions 21-23 from the 5' terminus (Fig. 3C). Cleavage of 3' ovr dsRNA was eliminated with chimeric molecules (Fig. 3, D and E; compare lanes 8, 9, 15, 16), as expected for Platform-PAZ-mediated cleavage. However, for BLT dsRNA, while 22 nt siRNA was absent, all other fragments were visible (Fig. 3, D and E; compare lanes 6 and 13), consistent with a helicase-mediated threading mechanism.

Studies of Dicer from other organisms indicate Platform-PAZ domains bind termini of 3' ovr dsRNA to allow measuring to RNase III active sites and production of an siRNA length (29, 30). Our cryo-EM structure of the dmDcr-2 complex and subsequent biochemical studies suggested that BLT dsRNA is cleaved differently, and in an ATP-dependent manner, threaded through the helicase domain to encounter the RNase III active sites. We tested the threading model by designing dsRNA with blocks at specific positions (Fig. 4A). Measurements using our EM density maps predicted that BLT dsRNAs are threaded through the helicase domain ~20 bp before encountering RNase III domains. To trap threading intermediates, we put biotin-dT analogs on both strands of 52 BLT dsRNA, at positions 28 or 37, counting from the 5' end of the sense strand (Fig. 4A); there was no significant difference in cleavage of these modified dsRNAs (figs. S14, A and B). However, we hypothesized that addition of streptavidin to biotin-dT-substituted dsRNAs (Block dsRNAs) would arrest threading of dsRNAs through the helicase. When dsRNA was incubated with streptavidin before initiating cleavage with dmDcr-2^{WT} and ATP, we trapped early (<11 nt, Block-28) and intermediate (11-20 nt, Block-37) threading products, without observing 22 nt siRNAs (Fig. 4, B and C; see fig. S14C for schematic). By contrast, cleavage by hsDcr-1 was unaffected by blocks, indicating that, at least under these conditions, hsDcr-1 cannot thread dsRNA through its helicase domain (Fig. 4D; compare lanes 12-14 to lanes 15-17).

We anticipated that short threading intermediates (<22 nts) might be unique to the initial cleavage event. However, threading intermediates were observed with dmDcr-2^{WT} under multiple-turnover conditions using internally ³²P-labeled dsRNAs, increasing proportionally with 22mers through the

reaction time course (fig. S14, D and E). Thus, at least in vitro, threading intermediates are recurring by-products of processive cleavage and not specific to the initial cleavage. dmDcr-2's highly efficient, helicase-dependent, processive cleavage is likely advantageous in antiviral defense. The generation of heterogeneous cleavage products during processive cleavage is predicted to dampen the phasing signal of viral siRNAs and is consistent with the overlapping and discontinuous viral siRNAs observed in invertebrate cells (6, 13, 36).

The dsRNA binding protein (dsRBP) Loquacious-PD (Loqs-PD) allows dmDcr-2 to cleave independent of termini (16, 37) and is required for processing endogenous siRNAs (38), but not for an antiviral response (13). This suggests dmDcr-2's intrinsic termini preferences function in viral defense, while Loqs-PD allows processing of endogenous dsRNA with diverse termini. By monitoring cleavage of dsRNAs with 5' ovr termini, or overhangs on both strands (fig. S15, B to E), we determined that dsRNA with an accessible 3' terminus is preferentially recognized by the Platform-PAZ domain, and without this feature, is processed by threading through the helicase domain.

RIG-I distinguishes capped termini of cellular transcripts from tri- and di-phosphorylated termini of viral transcripts, and this is inferred to allow self versus nonself discrimination (8). We found that, like RIG-I, dmDcr-2 cannot efficiently process dsRNAs capped at the 5'-terminus, although the phosphorylation state does not affect cleavage (fig. S16, A and B, and table S3). These results may reflect dmDcr-2's ability to process precursors of both endogenous and viral siRNAs.

We show that dmDcr-2 has two modes of cleavage (Fig. 4E and movies S3 and S4). dmDcr-2 is capable of using its Platform-PAZ domain to recognize 3' ovr dsRNAs in vitro, but it is unknown if dmDcr-2 processes such dsRNAs in vivo. dmDcr-2's cognate dsRBP, R2D2, may inhibit recognition and processing of substrates with 3' ovr termini (17). As such, the Platform-PAZ domains of dmDcr-2 may function solely on dsRNAs that are threaded through the helicase domain. At least in vitro, hsDcr-1 does not distinguish termini and does not exhibit helicase-dependent threading. Unlike dmDcr-2, hsDcr-1 may rely on the Platform-PAZ domain for generating viral siRNAs. Indeed, mutations to the Platform-PAZ domains of hsDcr-1 disrupt viral siRNA biogenesis (19). However, given the conservation of hsDcr-1's helicase domain, it is intriguing to consider that, under certain conditions, perhaps with additional factors, hsDcr-1 might mediate processive cleavage by threading of dsRNA through the helicase domain.

REFERENCES AND NOTES

1. M. Ha, V. N. Kim, Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509–524 (2014). [doi:10.1038/nrm3838](https://doi.org/10.1038/nrm3838) [Medline](#)
2. R. C. Wilson, J. A. Doudna, Molecular mechanisms of RNA interference. *Annu. Rev. Biophys.* **42**, 217–239 (2013). [doi:10.1146/annurev-biophys-083012-130404](https://doi.org/10.1146/annurev-biophys-083012-130404)

- [Medline](#)
3. K. Okamura, E. C. Lai, Endogenous small interfering RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **9**, 673–678 (2008). [doi:10.1038/nrm2479 Medline](#)
 4. R. Aliyari, Q. Wu, H.-W. Li, X.-H. Wang, F. Li, L. D. Green, C. S. Han, W.-X. Li, S.-W. Ding, Mechanism of induction and suppression of antiviral immunity directed by virus-derived small RNAs in *Drosophila*. *Cell Host Microbe* **4**, 387–397 (2008). [doi:10.1016/j.chom.2008.09.001 Medline](#)
 5. A. Flynt, N. Liu, R. Martin, E. C. Lai, Dicing of viral replication intermediates during silencing of latent *Drosophila* viruses. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5270–5275 (2009). [doi:10.1073/pnas.0813412106 Medline](#)
 6. L. R. Sabin, Q. Zheng, P. Thekkat, J. Yang, G. J. Hannon, B. D. Gregory, M. Tudor, S. Cherry, Dicer-2 processes diverse viral RNA species. *PLOS ONE* **8**, e55458 (2013). [doi:10.1371/journal.pone.0055458 Medline](#)
 7. M. E. Fairman-Williams, U.-P. Guenther, E. Jankowsky, SF1 and SF2 helicases: Family matters. *Curr. Opin. Struct. Biol.* **20**, 313–324 (2010). [doi:10.1016/j.sbi.2010.03.011 Medline](#)
 8. S. Ahmad, S. Hurr, Helicases in antiviral immunity: Dual properties as sensors and effectors. *Trends Biochem. Sci.* **40**, 576–585 (2015). [doi:10.1016/j.tibs.2015.08.001 Medline](#)
 9. Y. S. Lee, K. Nakahara, J. W. Pham, K. Kim, Z. He, E. J. Sontheimer, R. W. Carthew, Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004). [doi:10.1016/S0092-8674\(04\)00261-2 Medline](#)
 10. A. Tsutsumi, T. Kawamata, N. Izumi, H. Seitz, Y. Tomari, Recognition of the pre-miRNA structure by *Drosophila* Dicer-1. *Nat. Struct. Mol. Biol.* **18**, 1153–1158 (2011). [doi:10.1038/nsmb.2125 Medline](#)
 11. X.-H. Wang, R. Aliyari, W.-X. Li, H.-W. Li, K. Kim, R. Carthew, P. Atkinson, S.-W. Ding, RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science* **312**, 452–454 (2006). [doi:10.1126/science.1125694 Medline](#)
 12. D. Galiana-Arnoux, C. Dostert, A. Schneemann, J. A. Hoffmann, J.-L. Imler, Essential function in vivo for Dicer-2 in host defense against RNA viruses in *Drosophila*. *Nat. Immunol.* **7**, 590–597 (2006). [doi:10.1038/ni1335 Medline](#)
 13. J. T. Marques, J.-P. Wang, X. Wang, K. P. V. de Oliveira, C. Gao, E. R. G. R. Aguiar, N. Jafari, R. W. Carthew, Functional specialization of the small interfering RNA pathway in response to virus infection. *PLOS Pathog.* **9**, e1003579 (2013). [doi:10.1371/journal.ppat.1003579 Medline](#)
 14. Q. Liu, T. A. Rand, S. Kalidas, F. Du, H. E. Kim, D. P. Smith, X. Wang, R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* **301**, 1921–1925 (2003). [doi:10.1126/science.1088710 Medline](#)
 15. N. C. Welker, T. S. Maity, X. Ye, P. J. Aruscavage, A. A. Krauchuk, Q. Liu, B. L. Bass, Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. *Mol. Cell* **41**, 589–599 (2011). [doi:10.1016/j.molcel.2011.02.005 Medline](#)
 16. N. K. Sinha, K. D. Trettin, P. J. Aruscavage, B. L. Bass, *Drosophila* dicer-2 cleavage is mediated by helicase- and dsRNA termini-dependent states that are modulated by Loquacious-PD. *Mol. Cell* **58**, 406–417 (2015). [doi:10.1016/j.molcel.2015.03.012 Medline](#)
 17. E. S. Cenik, R. Fukunaga, G. Lu, R. Dutcher, Y. Wang, T. M. Tanaka Hall, P. D. Zamore, Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. *Mol. Cell* **42**, 172–184 (2011). [doi:10.1016/j.molcel.2011.03.002 Medline](#)
 18. Y. Qiu, Y. Xu, Y. Zhang, H. Zhou, Y.-Q. Deng, X.-F. Li, M. Miao, Q. Zhang, B. Zhong, Y. Hu, F.-C. Zhang, L. Wu, C.-F. Qin, X. Zhou, Human virus-derived small RNAs can confer antiviral immunity in mammals. *Immunity* **46**, 992–1004.e5 (2017). [doi:10.1016/j.immuni.2017.05.006 Medline](#)
 19. Y. Li, M. Basavappa, J. Lu, S. Dong, D. A. Cronkite, J. T. Prior, H.-C. Reinecker, P. Hertzog, Y. Han, W.-X. Li, S. Cheloufi, F. V. Karginov, S.-W. Ding, K. L. Jeffrey, Induction and suppression of antiviral RNA interference by influenza A virus in mammalian cells. *Nat. Microbiol.* **2**, 16250 (2016). [doi:10.1038/nmicrobiol.2016.250 Medline](#)
 20. H. Zhang, F. A. Kolb, V. Brondani, E. Billy, W. Filipowicz, Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* **21**, 5875–5885 (2002). [doi:10.1093/emboj/cdf582 Medline](#)
 21. P.-W. Lau, C. S. Potter, B. Carragher, I. J. MacRae, Structure of the human Dicer-TRBP complex by electron microscopy. *Structure* **17**, 1326–1332 (2009). [doi:10.1016/j.str.2009.08.013 Medline](#)
 22. P.-W. Lau, K. Z. Guiley, N. De, C. S. Potter, B. Carragher, I. J. MacRae, The molecular architecture of human Dicer. *Nat. Struct. Mol. Biol.* **19**, 436–440 (2012). [doi:10.1038/nsmb.2268 Medline](#)
 23. D. W. Taylor, E. Ma, H. Shigematsu, M. A. Cianfrocco, C. L. Noland, K. Nagayama, E. Nogales, J. A. Doudna, H.-W. Wang, Substrate-specific structural rearrangements of human Dicer. *Nat. Struct. Mol. Biol.* **20**, 662–670 (2013). [doi:10.1038/nsmb.2564 Medline](#)
 24. D. Luo, S. C. Ding, A. Vela, A. Kohlway, B. D. Lindenbach, A. M. Pyle, Structural insights into RNA recognition by RIG-I. *Cell* **147**, 409–422 (2011). [doi:10.1016/j.cell.2011.09.023 Medline](#)
 25. E. Kowalinski, T. Lunardi, A. A. McCarthy, J. Louber, J. Brunel, B. Grigorov, D. Gerlier, S. Cusack, Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* **147**, 423–435 (2011). [doi:10.1016/j.cell.2011.09.039 Medline](#)
 26. K. Takahashi, M. Yoneyama, T. Nishihori, R. Hirai, H. Kumeta, R. Narita, M. Gale Jr., F. Inagaki, T. Fujita, Nonself RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses. *Mol. Cell* **29**, 428–440 (2008). [doi:10.1016/j.molcel.2007.11.028 Medline](#)
 27. S. Myong, S. Cui, P. V. Cornish, A. Kirchhofer, M. U. Gack, J. U. Jung, K.-P. Hopfner, T. Ha, Cytosolic viral sensor RIG-I is a 5'-triphosphate-dependent translocase on double-stranded RNA. *Science* **323**, 1070–1074 (2009). [doi:10.1126/science.1168352 Medline](#)
 28. P. Linder, E. Jankowsky, From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* **12**, 505–516 (2011). [doi:10.1038/nrm3154 Medline](#)
 29. H. Zhang, F. A. Kolb, L. Jaskiewicz, E. Westhof, W. Filipowicz, Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**, 57–68 (2004). [doi:10.1016/j.cell.2004.06.017 Medline](#)
 30. I. J. Macrae, K. Zhou, F. Li, A. Repic, A. N. Brooks, W. Z. Cande, P. D. Adams, J. A. Doudna, Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195–198 (2006). [doi:10.1126/science.1121638 Medline](#)
 31. J.-E. Park, I. Heo, Y. Tian, D. K. Simanshu, H. Chang, D. Jee, D. J. Patel, V. N. Kim, Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* **475**, 201–205 (2011). [doi:10.1038/nature10198 Medline](#)
 32. Y. Tian, D. K. Simanshu, J.-B. Ma, J.-E. Park, I. Heo, V. N. Kim, D. J. Patel, A phosphate-binding pocket within the platform-PAZ-connector helix cassette of human Dicer. *Mol. Cell* **53**, 606–616 (2014). [doi:10.1016/j.molcel.2014.01.003 Medline](#)
 33. S. U. Colmenares, S. M. Buker, M. Buhler, M. Dlakić, D. Moazed, Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. *Mol. Cell* **27**, 449–461 (2007). [doi:10.1016/j.molcel.2007.07.007 Medline](#)
 34. E. Ma, I. J. MacRae, J. F. Kirsch, J. A. Doudna, Autoinhibition of human dicer by its internal helicase domain. *J. Mol. Biol.* **380**, 237–243 (2008). [doi:10.1016/j.jmb.2008.05.005 Medline](#)
 35. S. K. Kandasamy, R. Fukunaga, Phosphate-binding pocket in Dicer-2 PAZ domain for high-fidelity siRNA production. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14031–14036 (2016). [doi:10.1073/pnas.1612393113 Medline](#)
 36. Q. Wu, Y. Luo, R. Lu, N. Lau, E. C. Lai, W.-X. Li, S.-W. Ding, Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1606–1611 (2010). [doi:10.1073/pnas.0911353107 Medline](#)
 37. K. D. Trettin, N. K. Sinha, D. M. Eckert, S. E. Apple, B. L. Bass, Loquacious-PD facilitates *Drosophila* Dicer-2 cleavage through interactions with the helicase domain and dsRNA. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7939–E7948 (2017). [doi:10.1073/pnas.1707063114 Medline](#)
 38. R. Zhou, B. Czech, J. Brennecke, R. Sachidanandam, J. A. Wohlschlegel, N. Perrimon, G. J. Hannon, Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform. *RNA* **15**, 1886–1895 (2009). [doi:10.1261/rna.1611309 Medline](#)
 39. N. K. Sinha, B. L. Bass, Overexpression and purification of Dicer and accessory proteins for biochemical and structural studies. *Methods* **126**, 54–65 (2017). [doi:10.1016/j.ymeth.2017.07.012 Medline](#)
 40. S. D. Rose, D. H. Kim, M. Amarzguoui, J. D. Heidel, M. A. Collingwood, M. E. Davis, J. J. Rossi, M. A. Behlke, Functional polarity is introduced by Dicer processing of short substrate RNAs. *Nucleic Acids Res.* **33**, 4140–4156 (2005). [doi:10.1093/nar/gki732 Medline](#)
 41. R. Fukunaga, C. Colpan, B. W. Han, P. D. Zamore, Inorganic phosphate blocks

- binding of pre-miRNA to Dicer-2 via its PAZ domain. *EMBO J.* **33**, 371–384 (2014). doi:10.1002/embj.201387176 Medline
42. C. L. Noland, E. Ma, J. A. Doudna, siRNA repositioning for guide strand selection by human Dicer complexes. *Mol. Cell* **43**, 110–121 (2011). doi:10.1016/j.molcel.2011.05.028 Medline
 43. D. Goubau, M. Schlee, S. Deddouche, A. J. Pruijssers, T. Zillinger, M. Goldeck, C. Schuberth, A. G. Van der Veen, T. Fujimura, J. Rehwinkel, J. A. Iskarpatyoti, W. Barchet, J. Ludwig, T. S. Dermody, G. Hartmann, C. Reis E Sousa, Antiviral immunity via RIG-I-mediated recognition of RNA bearing 5'-diphosphates. *Nature* **514**, 372–375 (2014). doi:10.1038/nature13590 Medline
 44. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005). doi:10.1016/j.jsb.2005.07.007 Medline
 45. S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017). doi:10.1038/nmeth.4193 Medline
 46. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016). doi:10.1016/j.jsb.2015.11.003 Medline
 47. G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, S. J. Ludtke, EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007). doi:10.1016/j.jsb.2006.05.009 Medline
 48. S. H. W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012). doi:10.1016/j.jsb.2012.09.006 Medline
 49. X.-C. Bai, E. Rajendra, G. Yang, Y. Shi, S. H. W. Scheres, Sampling the conformational space of the catalytic subunit of human γ -secretase. *eLife* **4**, e11182 (2015). doi:10.7554/eLife.11182 Medline
 50. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015). doi:10.1038/nprot.2015.053 Medline
 51. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi:10.1002/jcc.20084 Medline
 52. T. D. Goddard, C. C. Huang, T. E. Ferrin, Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007). doi:10.1016/j.jsb.2006.06.010 Medline
 53. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). doi:10.1107/S0907444910007493 Medline
 54. P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, P. D. Adams, Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012). doi:10.1107/S0907444912001308 Medline
 55. S. Wu, Y. Zhang, MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins* **72**, 547–556 (2008). doi:10.1002/prot.21945 Medline
 56. B. Webb, A. Sali, Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* **86**, 2.9.1–2.9.37 (2016). doi:10.1002/cpps.20
 57. F. Jiang, A. Ramanathan, M. T. Miller, G.-Q. Tang, M. Gale Jr., S. S. Patel, J. Marcotrigiano, Structural basis of RNA recognition and activation by innate immune receptor RIG-I. *Nature* **479**, 423–427 (2011). doi:10.1038/nature10537 Medline
 58. J. Kawaoka, E. Jankowsky, A. M. Pyle, Backbone tracking by the SF2 helicase NPH-II. *Nat. Struct. Mol. Biol.* **11**, 526–530 (2004). doi:10.1038/nsmb771 Medline
 59. F. Armougom, S. Moretti, O. Poirot, S. Audic, P. Dumas, B. Schaeli, V. Keduas, C. Notredame, Expresso: Automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.* **34** (Web Server), W604–W608 (2006). doi:10.1093/nar/gkl092 Medline
 60. A. Marchler-Bauer, S. Lu, J. B. Anderson, F. Chitsaz, M. K. Derbyshire, C. DeWeese-Scott, J. H. Fong, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, J. D. Jackson, Z. Ke, C. J. Lanczycki, F. Lu, G. H. Marchler, M. Mullokandov, M. V. Omelchenko, C. L. Robertson, J. S. Song, N. Thanki, R. A. Yamashita, D. Zhang, N. Zhang, C. Zheng, S. H. Bryant, CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39** (Database), D225–D229 (2011). doi:10.1093/nar/gkq1189 Medline
 61. J. Gan, J. E. Tropea, B. P. Austin, D. L. Court, D. S. Waugh, X. Ji, Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell* **124**, 355–366 (2006). doi:10.1016/j.cell.2005.11.034 Medline
 62. Z. Du, J. K. Lee, R. Tjhen, R. M. Stroud, T. L. James, Structural and biochemical insights into the dicing mechanism of mouse Dicer: A conserved lysine is critical for dsRNA cleavage. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2391–2396 (2008). doi:10.1073/pnas.0711506105 Medline
 63. S. C. Kwon, T. A. Nguyen, Y.-G. Choi, M. H. Jo, S. Hohng, V. N. Kim, J.-S. Woo, Structure of human DROSHA. *Cell* **164**, 81–90 (2016). doi:10.1016/j.cell.2015.12.019 Medline
 64. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014). doi:10.1093/nar/gku316 Medline
 65. T. Nishino, K. Komori, D. Tsuchiya, Y. Ishino, K. Morikawa, Crystal structure and functional implications of *Pyrococcus furiosus* Hef helicase domain involved in branched DNA processing. *Structure* **13**, 143–153 (2005). doi:10.1016/j.str.2004.11.008 Medline
 66. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017). doi:10.1038/nmeth.4169 Medline

ACKNOWLEDGMENTS

We thank D. Cazalla and E. Cao for critique of the manuscript and P. J. Aruscavage for technical assistance. EM was performed at University of Utah EM Core Laboratory, with computational support from Utah Center for High Performance Computing. RNA was synthesized by the DNA/Peptide facility (Health Sciences Center Cores at University of Utah). This work was supported by funding from the National Institute of General Medical Sciences (R01GM121706) and the H. A. and Edna Benning Presidential Endowed Chair (to B.L.B.). The authors declare no competing financial interests. The models and cryo-EM maps are available via the following accession numbers: PDB 6BUA, EMD-7291, EMD-7292 (apo-dmDcr-2^{R113}); PDB 6BU9, EMD-7290 (dmDcr-2^{R113}•52 BLT dsRNA•ATP- γ S complex).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aag0921/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S16
Tables S1 to S3
References (39–66)

29 September 2017; accepted 11 December 2017
Published online 21 December 2017
10.1126/science.aag0921

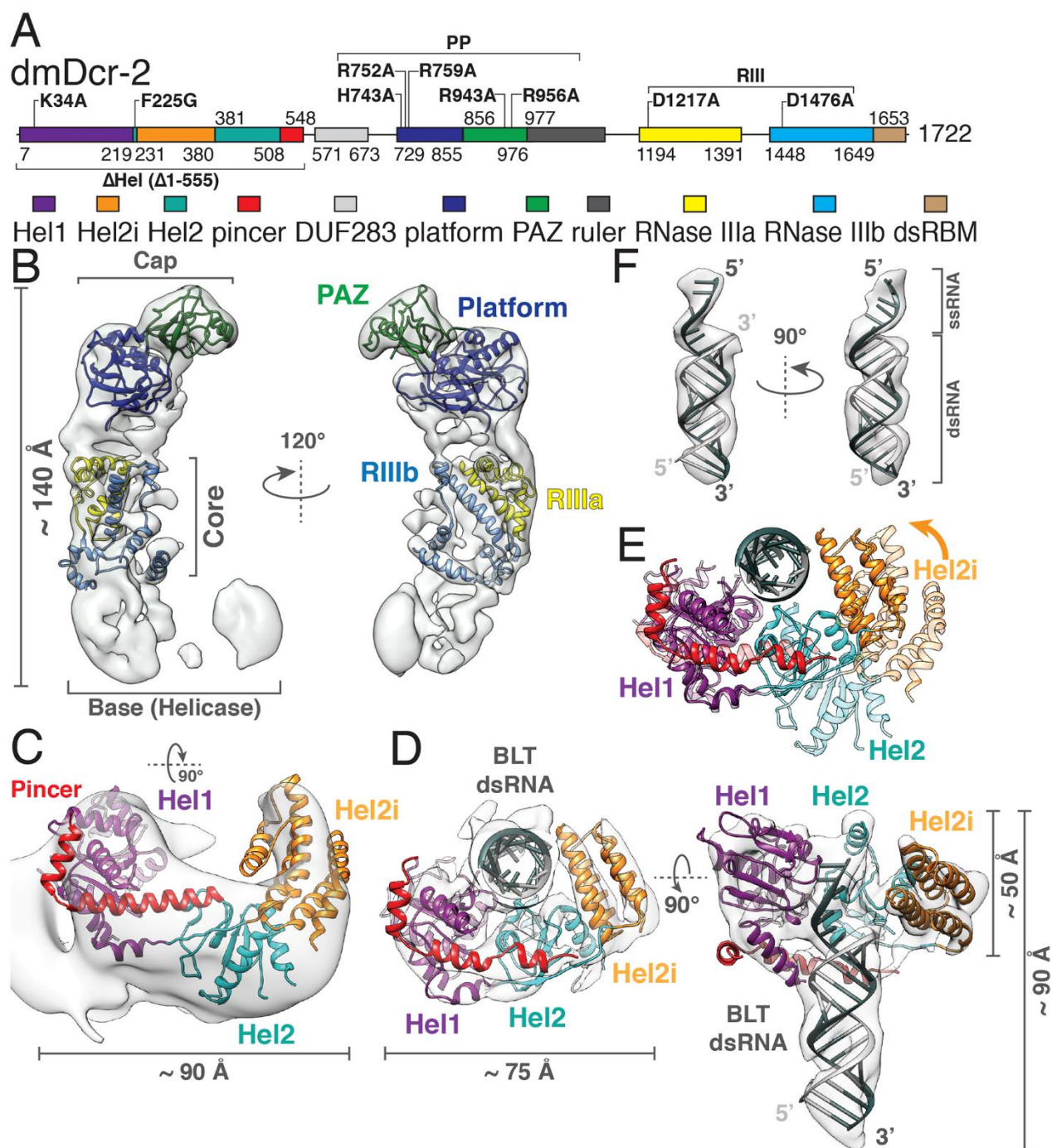


Fig. 1. Cryo-EM reconstructions of apo-dmDcr-2 and dmDcr-2•BLT dsRNA•ATP-γS. (A) dmDcr-2 domains numbered at boundaries. Mutations/deletions in bold, and designated in text as superscripts. All mutations were in the context of the full-length protein, unless specified by Δ . Two full-length variants had multiple mutations: PP (H743A, R752A, R759A, R943A, R956A); RIII (D1217A, D1476A). (B) Cryo-EM density map of apo-dmDcr-2 (7.1 Å) fitted with homology models of subdomains. RIIIa, RNase IIIa; RIIIb, RNase IIIb. (C) Homology model of apo-dmDcr-2 in open conformation, based on apo-RIG-I, and fitted as rigid body into 8.7 Å cryo-EM density map (also figs. S4A and S5, B to D). (D) Cryo-EM reconstruction of dmDcr-2^{RIII}•BLT dsRNA•ATP-γS showing helicase in closed, ligand-bound conformation. (E) Superimposition of open (light) and closed (dark) helicase conformations showing clamping of Hel2 and Hel2i on BLT dsRNA. Arrow, direction of clamping. (F) EM density and modeling of BLT single and double-stranded RNA.

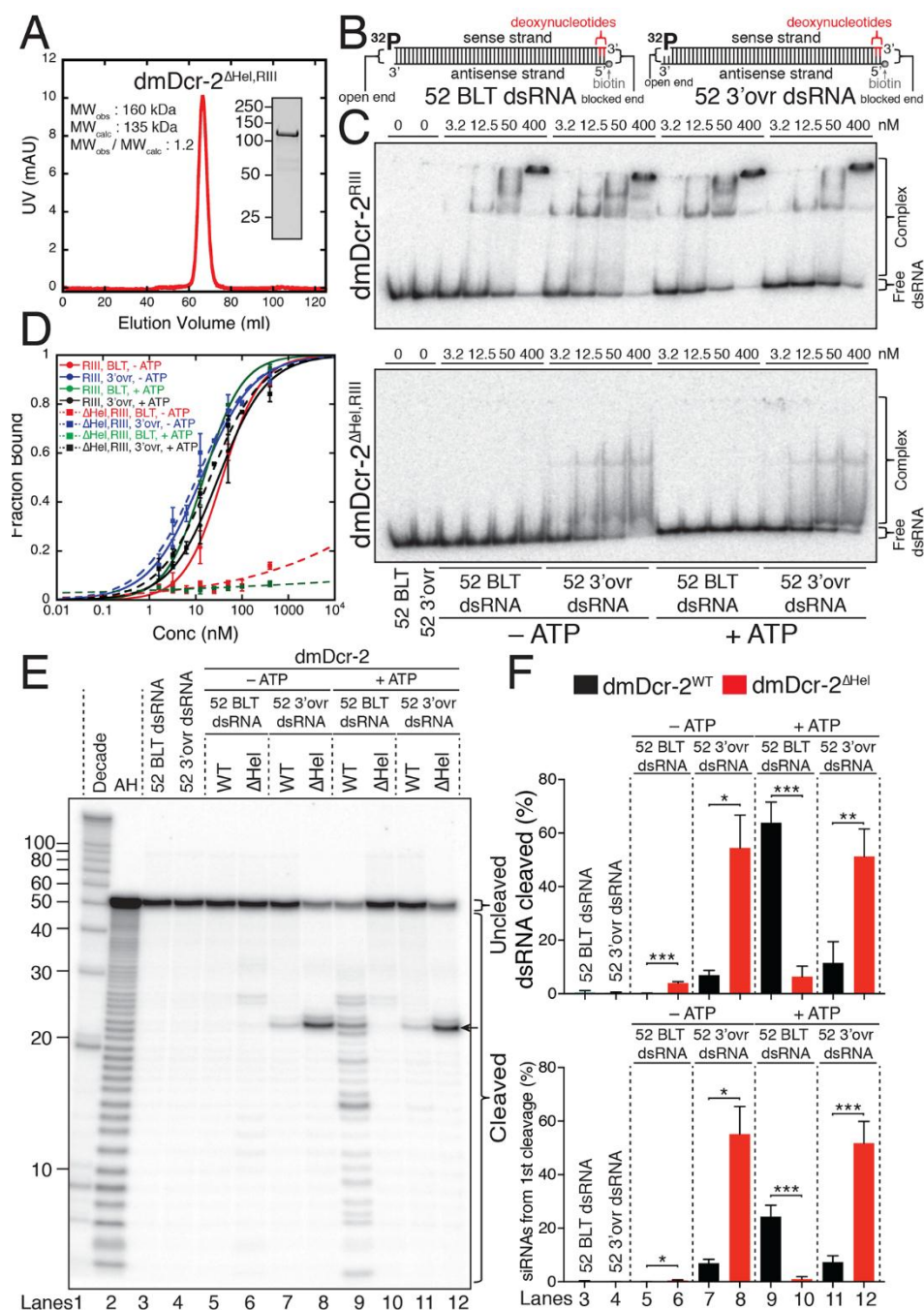


Fig. 2. dmDcr-2's helicase domain is required to recognize and cleave BLT dsRNA. (A) Gel-filtration and SDS-PAGE analyses of dmDcr-2^{ΔHel,RIII}. (B) Cartoon of dsRNA used in (C-F), showing modifications that block binding at one end. (C) Gel mobility shift assays of dmDcr-2^{RIII} (top) and dmDcr-2^{ΔHel,RIII} (bottom) with 52 BLT or 3'ovr dsRNA, +/- 5 mM ATP ($n \geq 3$). (D) Binding curves using data as in (C). Data points, mean \pm SD ($n = 3$). (E) Single-turnover cleavage assays of 52 BLT or 3'ovr dsRNA (1 nM) with dmDcr-2^{WT} or dmDcr-2^{ΔHel} (30 nM), +/- 5 mM ATP ($n = 3$). Only initial cleavage is monitored, since this removes 5' ³²P. Arrow, 22 nucleotide (nt) siRNA product. AH, alkaline hydrolysis. Left, nt lengths. (F) Quantification of cleavage assays as in (E). Percent dsRNA cleaved (all dsRNA except uncleaved) and percent siRNAs (21-23 nt products) resulting from 1st cleavage were quantified. Data points, mean \pm SD ($n = 3$). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

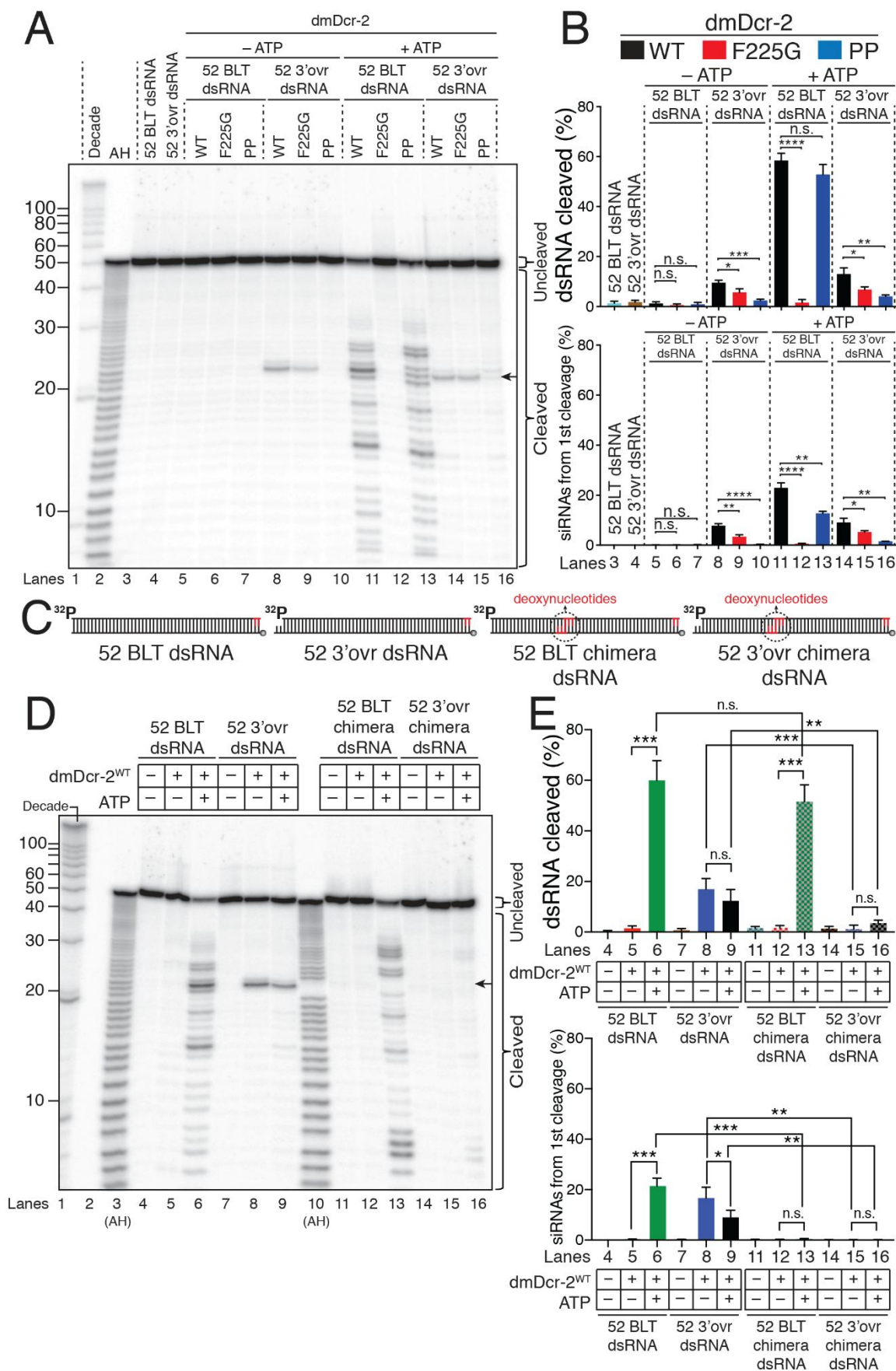


Fig. 3. Helicase and Platform-PAZ domains differentially contribute to cleavage of BLT and 3'ovr dsRNA. (A) Single-turnover cleavage assays of 52 BLT or 3'ovr dsRNA (1 nM) with dmDcr-2^{WT}, dmDcr-2^{F225G} and dmDcr-2^{PP} (30 nM), -/+ 5 mM ATP ($n = 3$). Substrates were as described in Fig. 2, B and E. AH, arrow, as in Fig. 2E. (B) Quantification of cleavage assays as in (A). Data points, mean \pm SD ($n = 3$). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$; $P > 0.05$, n.s. (nonsignificant). (C) Substrates used in (D) and (E), with additional details in Fig. 2, B and E. In chimeras, nts 21-23 from 5' end of sense strand, and nts 32-34 from 5' end of antisense strand were deoxyribonucleotides (dashed circles). (D) Single-turnover cleavage assays of regular or chimeric 52 BLT or 3'ovr dsRNA (1 nM) with dmDcr-2^{WT} (30 nM), -/+ 5 mM ATP ($n = 4$). (E) Quantification of cleavage assays as in (D). Data points, mean \pm SD ($n = 4$).

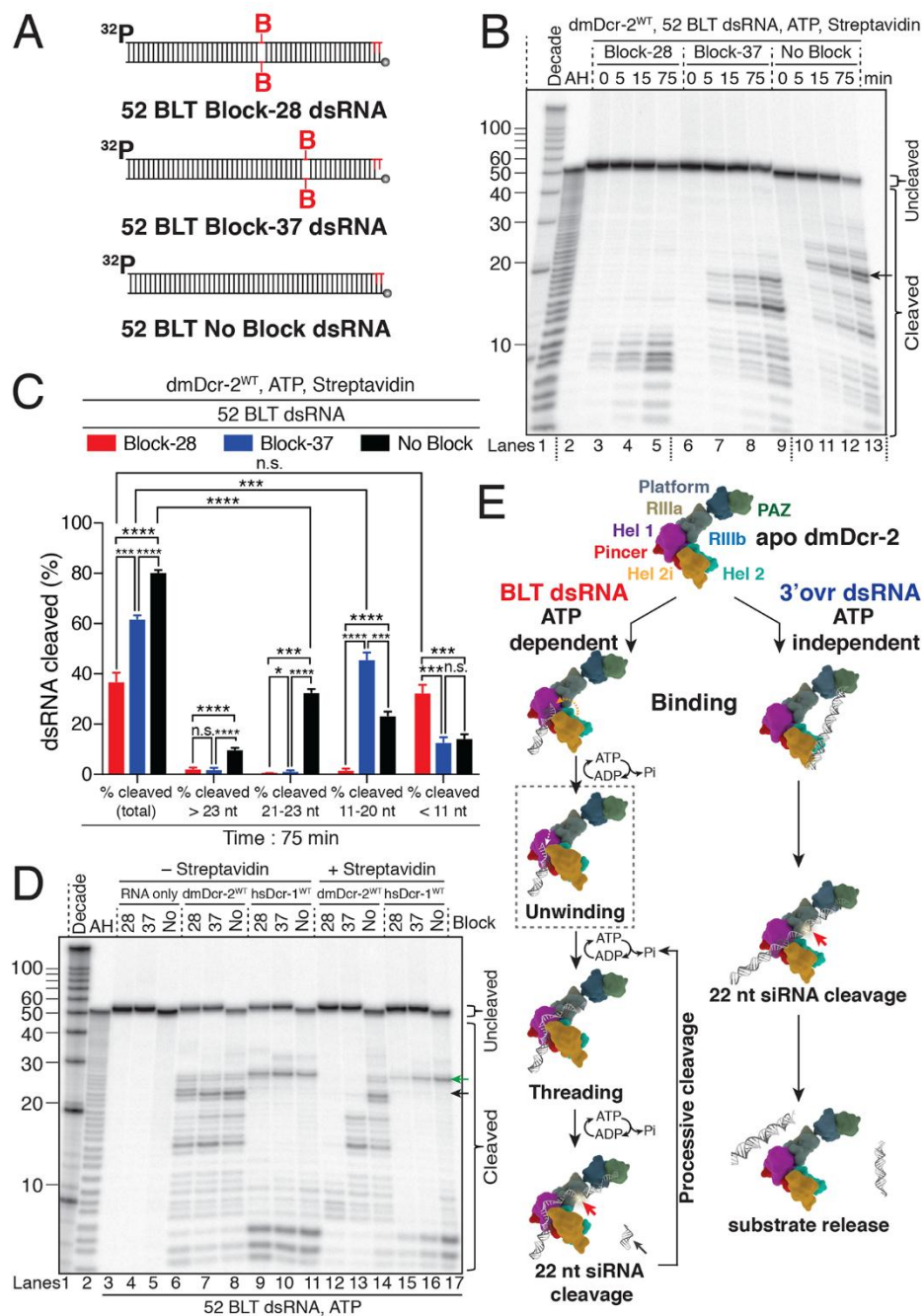


Fig. 4. BLT dsRNA threads through helicase domain. (A) Substrates for (B to D); some features described in Fig. 2, B and E. Blocked dsRNAs contained biotin-dT (red B) on both strands, with 28 and 37 indicating position from 5' end of sense strand. (B) Single-turnover cleavage assays of blocked or unblocked 52 BLT dsRNA (1 nM) with dmDcr-2^{WT} (30 nM), 5 mM ATP and 80 nM streptavidin ($n = 3$). dsRNA was pre-incubated with streptavidin before adding dmDcr-2^{WT}. Arrow, AH, as in Fig. 2E. (C) Quantification of cleavage at 75 min as in (B). dsRNA cleaved (%) is plotted based on all products (total), those >23 nt, siRNAs (21-23 nt products), those 11-20 nts, and those <11 nts. Data points, mean \pm SD ($n = 3$). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$; $P > 0.05$, n.s. (nonsignificant). (D) Single-turnover cleavage assays as in (B) with dmDcr-2^{WT} or hsDcr-1^{WT}, $-/+$ 80 nM streptavidin ($n = 3$). Black arrow, siRNA product (22 nt) with dmDcr-2^{WT}; green arrow, siRNA product (26 nt) with hsDcr-1^{WT}. (E) Model for recognition and cleavage of BLT and 3'ovr dsRNA by dmDcr-2. Dotted yellow arrow, clamping of helicase on BLT dsRNA; dotted white arrow, unwinding; dotted gray box, dmDcr-2^{WT}•BLT dsRNA•ATP- γ S complex shown in Fig. 1D (also see fig. S10); red arrow, cleavage; gray arrow, threading intermediates (see text and movie S3 for details). Model for 3'ovr recognition from data reported here and elsewhere (also see movie S4) (23, 34).

CHEMICAL BIOLOGY

Phosphoethanolamine cellulose: A naturally produced chemically modified cellulose

Wiriya Thongsomboon,¹ Diego O. Serra,² Alexandra Possling,² Chris Hadjineophytou,^{2,3*} Regine Hengge,^{2†} Lynette Cegelski^{1†}

Cellulose is a major contributor to the chemical and mechanical properties of plants and assumes structural roles in bacterial communities termed biofilms. We find that *Escherichia coli* produces chemically modified cellulose that is required for extracellular matrix assembly and biofilm architecture. Solid-state nuclear magnetic resonance spectroscopy of the intact and insoluble material elucidates the zwitterionic phosphoethanolamine modification that had evaded detection by conventional methods. Installation of the phosphoethanolamine group requires BcsG, a proposed phosphoethanolamine transferase, with biofilm-promoting cyclic diguanylate monophosphate input through a BcsE-BcsF-BcsG transmembrane signaling pathway. The *bcsEFG* operon is present in many bacteria, including *Salmonella* species, that also produce the modified cellulose. The discovery of phosphoethanolamine cellulose and the genetic and molecular basis for its production offers opportunities to modulate its production in bacteria and inspires efforts to biosynthetically engineer alternatively modified cellulosic materials.

Cellulose is the most abundant biopolymer on Earth. Plants rely on the tensile strength and mechanical properties of cellulose to stand upright (1). Chemically, cellulose is a linear polysaccharide composed of β -1,4-linked glucosyl residues. Individual strands participate in strong hydrogen-bonding networks with neighboring strands and contribute to the physical and chemical integrity of plant cell walls and cellulosic materials (2). Microorganisms are also major producers of cellulose (3). The essential genetic and protein machinery for cellulose production in bacteria include the cellulose synthase genes, termed *bcsA* and *bcsB*, which encode cellulose synthase subunits BcsA and BcsB (4). BcsA is an integral membrane protein containing the catalytic active site. BcsB interacts with BcsA at the periplasmic face of the inner membrane in Gram-negative bacteria, with the two subunits forming a channel for cosynthetic secretion of cellulose. Cellulose biosynthesis requires activation by the ubiquitous bacterial second messenger cyclic diguanylate monophosphate (c-di-GMP) (5), which directly binds to BcsA (6). Intense curiosity has emerged in understanding the diversity of additional genes in cellulose biosynthesis operons that are present in many microorganisms (3). Here we report on the determination of the structure of a modified cellulose, phosphoethanolamine (pEtN) cellulose, produced naturally by *Escherichia*

coli and other Gram-negative bacteria. We provide the genetic basis for its production and the functional implications of gene-directed pEtN cellulose synthesis.

E. coli and *Salmonella* are among the best-studied microorganisms reported to produce cellulose. These include human pathogens such as uropathogenic and enterohemorrhagic *E. coli*. Functionally, the exopolysaccharide cellulose is a major component of the self-produced extracellular matrix in biofilms, which represent physiologically heterogeneous and spatially structured bacterial communities (7, 8). Biofilm formation is of high medical relevance, as it confers enhanced resistance to antibiotics and host defenses during infection (9). Within the biofilm matrix, cellulose forms a nanocomposite with amyloid curli fibers that encapsulates individual cells in supramolecular basketlike structures, enmeshes the bacterial community, and confers cohesion and elasticity that allow biofilms to fold and buckle up in a tissuelike manner (10–12). Biochemical and solid-state nuclear magnetic resonance (NMR) measurements with the clinically important uropathogenic *E. coli* strain UTI89 established that the matrix was composed of curli fibers and cellulosic material in a 6:1 ratio by mass. During this bottom-up analysis involving ¹³C and ¹⁵N NMR analysis of the purified components, we also discovered that the cellulose portion appears to be modified in some way with an aminoethyl functionality (11).

Solid-state NMR analysis of the intact cellulosic material, complemented by solution-state NMR and mass spectrometry analysis of acid-digested material, has now enabled the determination of the chemical structure of the modified cellulose as a polymer containing glucose and glucose-6-phosphoethanolamine (Fig. 1A). The ¹³C cross-polarization/magic angle spinning (CPMAS)

NMR spectrum of isolated cellulosic material contains carbon contributions from the glucose backbone plus two additional carbons associated with the modification (Fig. 1B). A comparison of this spectrum with that of the cellulosic material isolated with the aid of the dye Congo red (CR) is provided in Fig. 1C, as the extraction and yield of cellulosic material is enhanced in the presence of CR and is sometimes used for the preparation of larger samples in subsequent analysis (13). CR is also commonly used as a supplement in nutrient agar plates for evaluation of *E. coli* and *Salmonella* community phenotypes because both curli and cellulosic polymers bind the dye (14). The use of CR does not result in any changes to the cellulosic carbon composition (Fig. 1C). C{N} rotational-echo double-resonance (REDOR) NMR was used to experimentally select directly bonded carbon-nitrogen pairs and identified the 41-parts per million (ppm) carbon peak as the C-8 carbon, the only carbon directly bonded to nitrogen (fig. S1). ³¹P CPMAS revealed the presence of ³¹P as a phosphate in the polymer (fig. S2), and we hypothesized that the attachment site was at the C-6 position because of the downfield shift of a carbon in the C-6 region (as in glucose-6-phosphate). C{P} REDOR NMR revealed that phosphorous was indeed positioned closest to the C-6' and C-7 carbons and next closest to the C-5 and C-8 carbons (Fig. 1D), suggesting the full structural assignment as pEtN cellulose. A ¹³C CP-array NMR experiment enabled the quantitative accounting of carbon contributions to the pEtN ¹³C CPMAS spectrum and determined that approximately one-half of the cellulose glucose units in the intact polymer are modified (fig. S3).

Additional solution-based analyses were performed to complement the solid-state NMR analysis, although this required acid digestion of the pEtN cellulose to release soluble components into solution. Solution-state NMR analysis of acid-digested pEtN cellulose supported the assignments from solid-state NMR and confirmed that the modification occurs at the C-6 position. However, a standard 48-hour HCl hydrolysis leads to degradation of the modification, and only soluble glucose, glucose-6-phosphate, and ethanolamine were detected (figs. S4 to 7), observations that explain the difficulty of identifying the modified cellulose using digestions and solution-based methods. To attempt to capture an intact modified glucose unit, shorter hydrolysis times were used and liquid chromatography-mass spectrometry was used for analysis. The intact pEtN glucose was detected, as well as released glucose, glucose-6-phosphate, and ethanolamine (fig. S8). The detection of pEtN glucose likely escaped detection in previous research because of the instability of the modification upon acid digestion and isolation protocols designed to specifically detect cellulose. Typical approaches do not use the careful purification protocol we developed and, instead, rely on cell lysis and polysaccharide enrichment followed by harsh hydrolysis and either colorimetric detection upon reaction with a sulfuric acid solution of anthrone (15) or chemical derivatization and mass spectrometry detection (16) to support the presence

¹Department of Chemistry, Stanford University, Stanford, CA 94305, USA. ²Institute of Biology, Microbiology, Humboldt-Universität zu Berlin, 10115 Berlin, Germany. ³Department of Chemistry and Molecular Biology, University of Gothenburg, 41296 Gothenburg, Sweden.

*Present address: Department of Biosciences, University of Oslo, 0371 Oslo, Norway.

†Corresponding author. Email: cegelski@stanford.edu (LC.); regine.hengge@hu-berlin.de (RH.)

of significant glucose content as a reporter for cellulose production. The presence of glucose-6-phosphate, ethanolamine, and other compounds could escape detection or be ascribed to residual cellular components.

A biosynthetically modified cellulose has wide-ranging implications and potential applications. Among these, the specifically modified cellulose could be essential for the formation and function of bacterial biofilms containing the polymer, could exhibit attractive properties for new cellulosic materials, and could potentially be introduced into other organisms, if gene directed. Thus, we sought to identify the genes involved in the installation of the cellulose modification. The *bcsEFG* operon, which is part of the cellulose gene cluster in *E. coli*, had not been ascribed a definitive role in cellulose synthesis. The ^{13}C CPMAS NMR comparison of the isolated cellulose from a *bcsG* deletion mutant ($\Delta bcsG$) revealed that the *bcsG* gene was indispensable for the cellulose modification (Fig. 1E). The spectrum lacks the contributions from the 41-ppm C-8 carbon and the 63-ppm C-7 carbon. As expected, the sugar C-6 carbon contribution appears only at the upfield ^{13}C position of 63 ppm, corresponding to unmodified glucose. Complementation of the $\Delta bcsG$ mutant with *bcsG* on a plasmid restored production of the modified cellulose (fig. S9). The prevalence of the modification was reduced in the in-frame nonpolar $\Delta bcsF$ mutant and more strongly reduced or abolished in the $\Delta bcsE$ mutant, indicating that BcsE and BcsF may play accessory, and possibly regulatory, roles in the installation of pEtN by BcsG (fig. S10).

Similar macrocolony phenotypes of nonpolar $\Delta bcsE$, $\Delta bcsF$, and $\Delta bcsG$ mutants (fig. S11), as well as their coexpression from a single operon in the *bcs* gene cluster, also suggested functional cooperation of these three proteins. Notably, BcsG and BcsF are membrane inserted, whereas BcsE is a soluble c-di-GMP-binding protein (17). To further elucidate the molecular basis of BcsG function and the roles of BcsE and BcsF, we tested for potential direct interactions between these proteins as well as with the cellulose synthase subunits BcsA and BcsB. Using a bacterial two-hybrid assay that is based on the reconstitution of adenylate cyclase from two separate domains fused to proteins that potentially interact (18), we observed strong interactions in vivo of BcsG with BcsF as well as with BcsA (Fig. 2A). Thus, BcsG operates in close proximity to the cellulose synthesizing BcsA-BcsB complex. Moreover, BcsF also showed strong interaction with BcsE (Fig. 2A), suggesting a BcsE-BcsF-BcsG pathway that controls cellulose modification through direct protein-protein interactions. Qualitative evaluation of cellulose production by CR binding in *E. coli* strain AR3110 $\Delta csgBA$, which produces no curli fibers but cellulose only, and its $\Delta bcsE$, $\Delta bcsF$, and $\Delta bcsG$ mutant derivatives additionally revealed comparable CR binding for AR3110 $\Delta csgBA$ and AR3110 $\Delta csgBA\Delta bcsG$, but reduced CR binding for AR3110 $\Delta csgBA\Delta bcsE$ and AR3110 $\Delta csgBA\Delta bcsF$ (fig. S12). The decrease in cellulose production by the $\Delta bcsE$ and $\Delta bcsF$ mutants was corroborated

by a low yield in isolatable material for NMR analysis and could indicate that BcsE and BcsF contribute to stability of the BcsA-BcsB machinery in AR3110, influencing efficiency of cellulose synthesis.

BcsG is composed of 559 amino acids and has been predicted to be an integral membrane protein. A hydropathy plot analysis of BcsG supported the presence of several putative transmembrane-spanning regions in the N-terminal 160 amino acids followed by a large hydrophilic C-terminal domain. However, charge distribution flanking the hydrophobic amino acid stretches did not allow us to unequivocally predict the number and orientation of the transmembrane regions (19) and thus the localization of the C-terminal domain. Therefore, we generated a set of translational reporter fusions to β -galactosidase (LacZ) and alkaline phosphatase (PhoA) to identify cytoplasmic and periplasmic regions of BcsG. This approach is based on the observation that the signal sequence-driven attempt to export the normally cytoplasmic LacZ results in the jamming of secretion machinery and interferes with LacZ ac-

tivity, whereas PhoA becomes active only upon its secretion to the periplasm, where DsbAB-mediated disulfide bond formation occurs (20). Accordingly, fusions inserted after codon 1 of *bcsG* showed high LacZ activity, but low PhoA activity (Fig. 2B). By contrast, when the entire N-terminal BcsG domain of approximately 160 amino acids—including all hydrophobic stretches—was present in the hybrid proteins, low LacZ and high PhoA activities were observed, consistent with the large C-terminal domain of BcsG residing on the periplasmic side of the inner membrane (Fig. 2, B and C). A similar analysis with BcsF, a 63-amino acid peptide, showed that its N-terminal single-transmembrane region, which is preceded by negative charges and followed by positive charges, inserts into the membrane such that the small hydrophilic C-terminal domain remains on the cytoplasmic side of the membrane (Fig. 2, B and C).

We addressed the question of the substrate for BcsG-mediated pEtN modification of cellulose emerging from the BcsA-BcsB complex. We noticed that, with respect to overall size (559 and 563

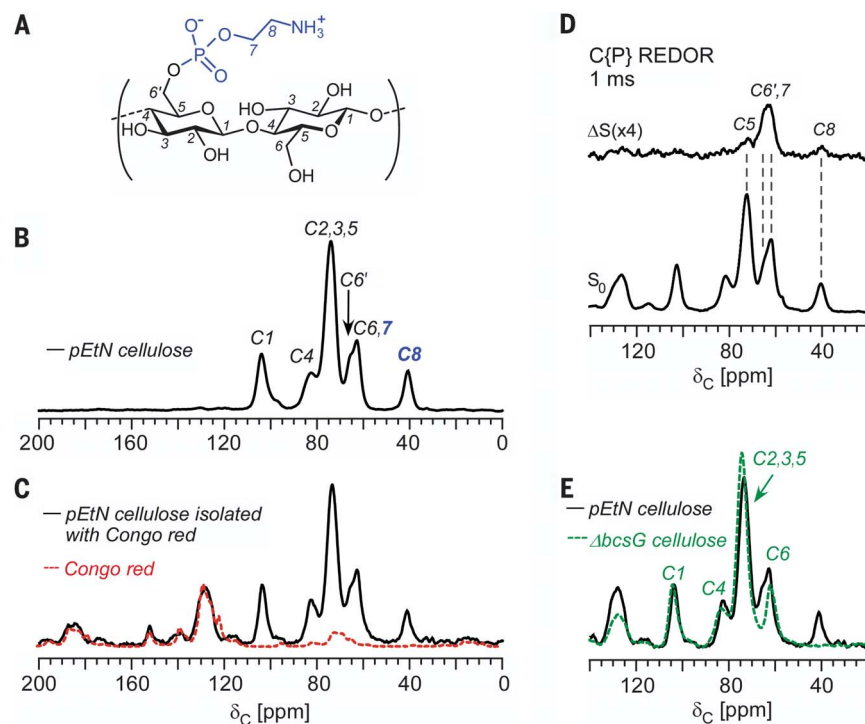


Fig. 1. *E. coli* produces phosphoethanolamine cellulose. (A) Representation of the chemical structure of glucose and pEtN glucose units in pEtN cellulose. (B) ^{13}C CPMAS solid-state NMR spectra of the pure modified cellulose with two additional carbon contributions, C-7 (63 ppm) and C-8 (41 ppm). The C-6 carbon appears at 62 ppm for the unmodified glucose units and at 66 ppm for the modified glucose units (figs. S1 and S4). (C) ^{13}C CPMAS spectra of the modified cellulose compared with that of the modified cellulose isolated from cells grown in the presence of CR. The pure CR spectrum is provided as an overlay (dashed red line). The comparison demonstrates that purification with CR does not influence the polysaccharide composition. δ_{C} , carbon chemical shift. (D) The C-6' and C-7 carbon chemical-shift region exhibited the strongest dephasing in the 1-ms C{P} REDOR NMR measurement, followed by that of the C-5' and C-8 carbons, suggesting the full structural assignment as pEtN cellulose, further confirmed by solution-state NMR and mass spectrometry (figs. S4 to S8). S_0 , REDOR full-echo spectrum; ΔS , REDOR difference spectrum. (E) The ^{13}C CPMAS spectrum of the cellulosic material isolated from the *bcsG* derivative lacked modification carbons and contained only the ^{13}C chemical shifts expected for standard amorphous cellulose.

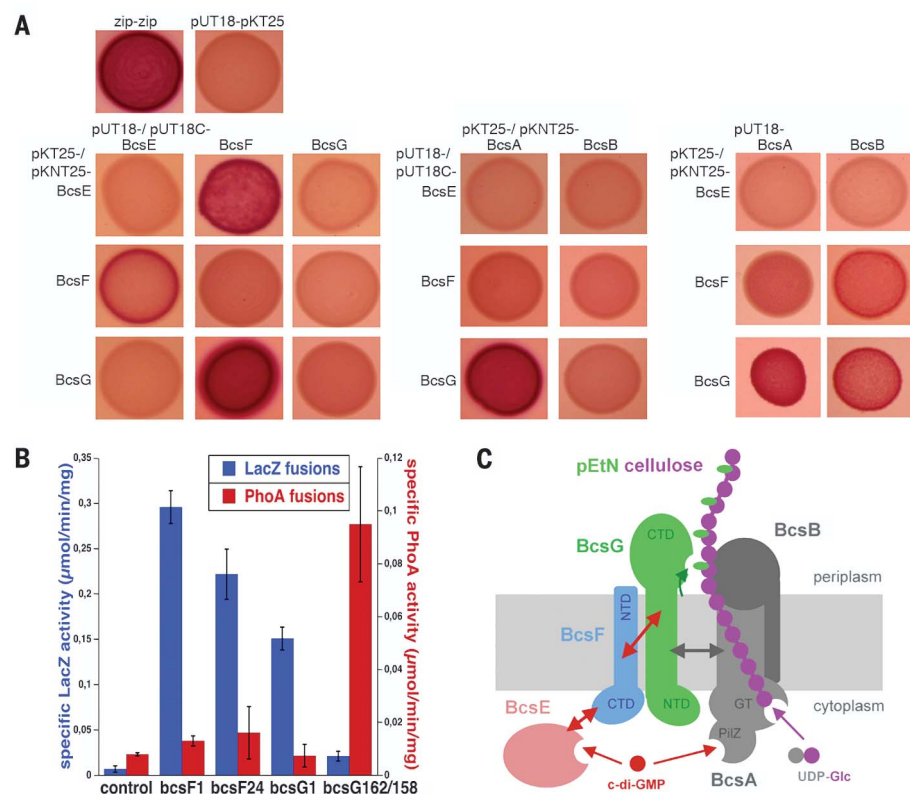


Fig. 2. BcsG directly interacts with cellulose synthase and communicates with the c-di-GMP-binding BcsE via the transmembrane peptide BcsF. (A) Interactions of the indicated proteins were tested

using a bacterial two-hybrid (2H) system based on the reconstitution of adenylate cyclase (AC) (18), which allows the utilization of maltose by W3110Δcya, resulting in red color on MacConkey agar plate. The 2H vector plasmids allow the attachment of the respective AC domain tags (18, 25), either at the N terminus (pKT25, pUT18c) or the C terminus (pKNT25, pUT18) of a protein. For BcsA, BcsB, BcsE, and BcsF, the tags were located at the C terminus; for BcsG, the tags were located at the N terminus. Zip-zip, leucine zipper domain of the yeast GCN4 protein, used as a positive control. **(B)** Transmembrane orientation of BcsF and BcsG was determined by assaying enzymatic activities of hybrid proteins between N-terminal parts from BcsF and BcsG fused to LacZ and PhoA expressed from low-copy number plasmids in strains W3110Δlac (I-A) and W3110ΔphoA. Fusion joints were after codon 1 (all combinations), codon 24 (of bcsF fused to both reporter genes), codon 162 (bcsG::lacZ), or codon 158 (bcsG::phoA). **(C)** A schematic model of the directly interacting modules for cellulose synthesis (BcsAB) and modification (BcsEFG) summarizes the protein-protein interactions (double-headed arrows) detected in (A), the transmembrane orientation of BcsG and BcsF as tested in (B), and dual control by the second messenger c-di-GMP, which binds to both BcsA (6) and BcsE (17). NTD, N-terminal domain; CTD, C-terminal domain; UDP-Glc, uridine diphosphate-glucose.

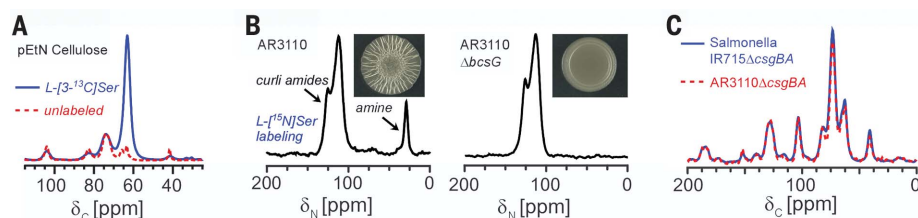


Fig. 3. Phosphoethanolamine cellulose production is detected in curli-integrated *E. coli* biofilm matrices with isotopic serine labeling and is also produced by *Salmonella enterica*. (A) Isotopic

labeling with L-[3-¹³C]Ser-supplemented YESCA nutrient medium resulted in enrichment of the pEtN cellulose C-7 carbon in an isolated pEtN sample, consistent with routing through a possible substrate such as phosphatidylethanolamine. **(B)** Isotopic labeling with L-[¹⁵N]Ser was evaluated by ¹⁵N CPMAS NMR on extracellular matrix samples containing both curli and cellulosic material. The ¹⁵N-amide signals correspond to curli amides. The loss of the ¹⁵N-amine signal in the bcsG derivative (right) confirmed the amine nitrogen assignment as that from pEtN cellulose. Loss of the modification was accompanied by loss of the wrinkled macrocolony morphology (inset photographs). **(C)** The ¹³C CPMAS spectrum of the cellulosic material isolated from *Salmonella enterica* serovar Typhimurium strain IR715ΔcsgBA matched that of pEtN cellulose from AR3110ΔcsgBA.

residues, respectively) and length and transmembrane orientation of domains, BcsG resembles EptB, a pEtN transferase that uses the phospholipid phosphatidylethanolamine (PE) to modify bacterial lipopolysaccharide (21). Two similar pEtN transferases, EptC from *Campylobacter jejuni* and LptA from *Neisseria meningitidis*, also have their active site domains on the periplasmic side of the cytoplasmic membrane; they equally use PE as a substrate, and their crystal structures show two or three histidine residues coordinating zinc, which is essential for activity (22). We therefore hypothesized that histidine residues in BcsG may play a similar role and that pEtN modification by BcsG may also originate from PE. We isolated a mutant version of BcsG (BcsG^{mut}) in which His³⁹⁶, His⁴⁰⁰, and His⁴⁴³ in the periplasmic domain were replaced by alanine residues. BcsG^{mut} did not complement the ΔbcsG mutation with respect to macrocolony formation (fig. S13) and was inactive with respect to cellulose modification (fig. S14), suggesting that BcsG enzymatic activity resides in the periplasm and is related to these known pEtN transferases, despite the absence of clear primary-sequence similarity. If BcsG also uses PE as a substrate, the modified cellulose should have atoms derived from serine, which serves as a direct precursor for the ethanolamine moiety of PE. Thus, pEtN cellulose was prepared from cells grown on agar medium supplemented with 25 mg/liter of L-[3-¹³C]Ser to detect whether pEtN cellulose would be enriched through incorporation of the serine label. The expected C-7 carbon in the pEtN cellulose spectrum was indeed enhanced as a result of label incorporation from serine (Fig. 3A). The routing of serine into the modification is consistent with PE serving as a substrate for BcsG. Labeling with L-[¹⁵N]Ser was also successful and is additionally valuable in identifying the pEtN cellulose modification in the context of complex extracellular matrix samples that contain both pEtN cellulose and curli. Labeling with L-[¹⁵N]Ser yielded the anticipated ¹⁵N-amine contribution from pEtN cellulose and the shift-resolved curli ¹⁵N-amide contributions from serine as well as glycine residues that result from serine conversion in glycine biosynthesis (Fig. 3B). In this way, the potential presence of pEtN cellulose can be determined in intact extracellular matrix preparations from different *E. coli* strains and different organisms.

Overall, we propose a model in which BcsG acts as a pEtN transferase with its catalytic domain in the periplasm, modifying cellulose after its emergence from the BcsA-BcsB machinery (Fig. 2C). Given the association of BcsG with the c-di-GMP-binding cytoplasmic protein BcsE and the transmembrane peptide BcsF, we also propose that cellulose modification is controlled by c-di-GMP in a transmembrane signaling pathway that involves BcsE and BcsF, with BcsF serving as a direct link between BcsE and BcsG (Fig. 2C). Thus, the biofilm-promoting second messenger c-di-GMP plays a dual role by activating both the synthesis and the modification of cellulose via the PilZ domain of BcsA and the BcsE-BcsF-BcsG transmembrane signaling pathway, respectively.

Notably, the K_d values for c-di-GMP binding to BcsA and BcsE are 8 (23) and 2.4 μ M (17), respectively, which ensures efficient modification whenever c-di-GMP levels are high enough to support cellulose synthesis.

Production of the biofilm matrix by wild-type *E. coli* involves the coproduction and tight association of amyloid curli fibers and what has been considered to be cellulose (8, 12). Yet, we have now determined that *E. coli* strains such as strain AR3110, which is a direct derivative of the widely studied K-12 laboratory strain W3110 (12), as well as the classical uropathogenic UTI89 (24) produce pEtN cellulose. Thus, we sought to test whether this cellulose modification is functionally important for biofilm-matrix architecture and function. Macrocolony morphotypes, the matrix fine structure as analyzed by in situ fluorescence and electron microscopy, and the multicellular cohesion of macrocolonies when challenged by shear stress were evaluated. BcsG was required for the buckling into radial ridges and wrinkles typically observed for the otherwise very flat AR3110 macrocolonies (Fig. 4A, compare panels i and iii). The ring-shaped curli-only-driven macrocolony architecture of AR3110 Δ bcsG (Fig. 4A, panel iii) resembles that of cellulose-free strains such as AR3110 Δ bcsA (Fig. 4A, panel ii) or W3110 (12). Furthermore, fluorescence microscopy of vertical cryosections through macrocolonies grown in the presence of thioflavin S, acting as a matrix dye that binds to cellulose and curli, revealed that BcsG was required for cellulose to assemble into long, thick, and straight filaments. These were most clearly observed in the absence of curli, i.e., in AR3110 Δ csgB (Fig. 4A, panel iv). Compared to these long filaments, only short, thin, and curled filaments were detected in AR3110 Δ csgB Δ bcsG macrocolonies (Fig. 4A, panel v, and fig. S15). Scanning electron microscopy further supported a role of modified cellulose in matrix architecture. The matrix at the surface of a AR3110 Δ bcsG macrocolony resembles that of the curli-only strain W3110 (Fig. 4B). Thus, cellulose modification is required to form the extended composite structure with curli fibers that nearly fully covers the surface of a macrocolony biofilm (as visible with strain AR3110 in Fig. 4B). Finally, an extracellular matrix consisting of either the cellulose-curli nanocomposite or of cellulose alone is known to provide cohesion to the cellular community. The resulting tissue-like behavior includes the ability not only to fold and buckle up but also to resist shear stress as a cohesive community (12). The latter can be tested by submerging a macrocolony in liquid and exposing it to gentle shaking. In this assay, the curli-free AR3110 Δ csgBA strain detached from the agar phase as an entire macrocolony, whereas colonies of the corresponding cellulose modification-deficient strain just dissolved into flares of loose cells (fig. S16). Thus, pEtN modification of cellulose is required for community behavior based on the formation of the connective matrix consisting of either long-range cellulose fibers or the curli-cellulose nanocomposite network that envelops and connects bacterial cells during biofilm formation.

The production of pEtN cellulose may also have consequences for virulence of pathogenic strains. Amyloid curli fibers are proinflammatory (25, 26) and a virulence factor for various types of pathogenic *E. coli* (27, 28), yet the tight association of curli with “cellulose” was reported to counteract these properties (27, 29). Therefore, mutations in the *bcsEFG* operon resulting in nonmodified cellulose that cannot form the nano-

composite with curli fibers may enhance the contribution of curli to virulence of pathogenic *E. coli*. Consistent with this notion, the 2011 European outbreak O104:H4 strain, an enteroaggregative and enterohemorrhagic *E. coli* which not only produced Shiga toxin but also high amounts of curli at 37°C while being a *bcsE* mutant (30), was of unprecedented virulence for this pathotype of *E. coli* (31).

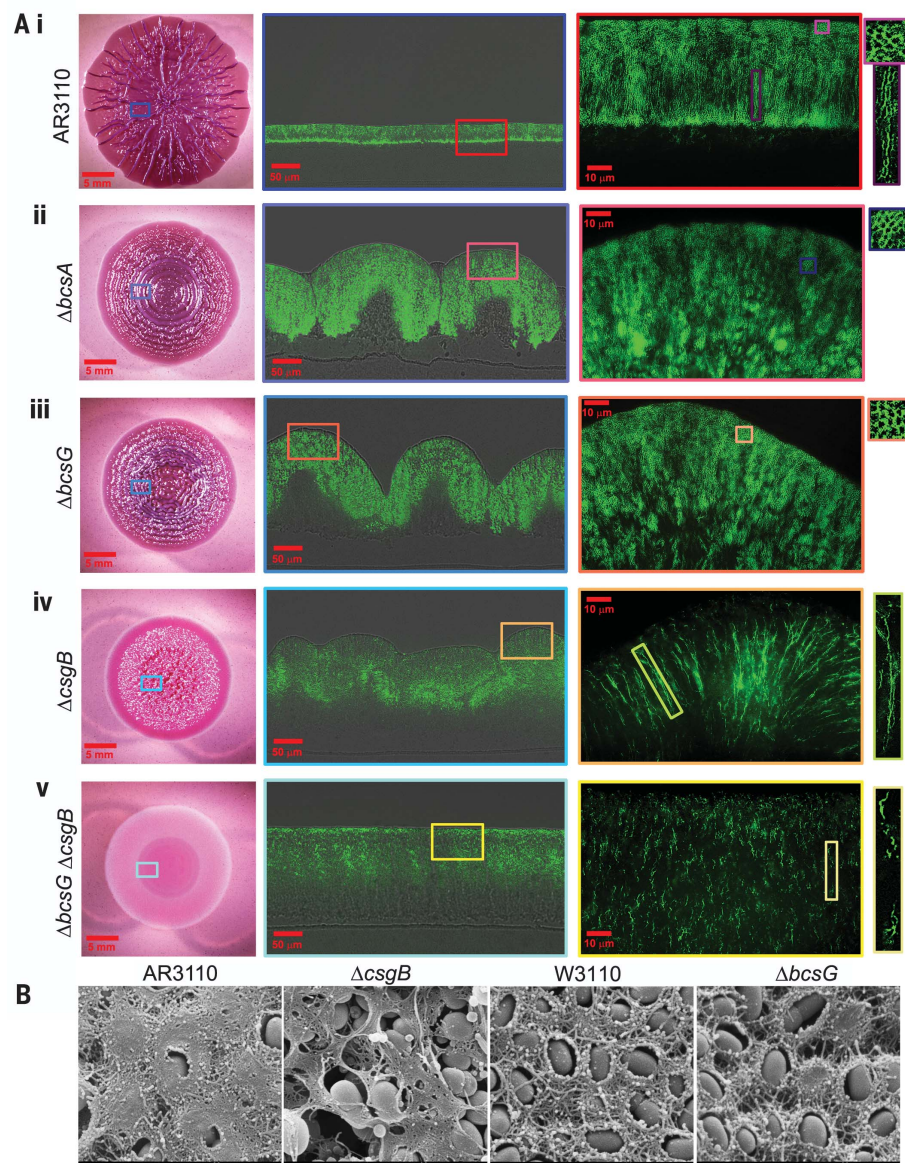


Fig. 4. Eliminating BcsG changes macroscopic morphology and microscopic matrix architecture of *E. coli* macrocolony biofilms. (A) Macrocolonies of strain AR3110, which produces both cellulose and curli fibers, and the indicated mutant derivatives (i to v) were grown for three days on salt-free LB agar plates containing either Congo red or the green-fluorescent thioflavin S, which stain cellulose and curli without affecting the overall matrix architecture and colony morphotype. The microscopic architecture of thioflavin S-stained matrix was visualized in thin cross sections of macrocolonies, with color-coded boxed areas being further enlarged adjacently. (B) The surface of macrocolonies was visualized at high resolution by scanning electron microscopy. The classical *E. coli* K-12 lab strain W3110 is isogenic to AR3110, except for a *bcsQ*^{stop} mutation, which eliminates the ability to produce cellulose (12). Because of polarity, the Δ csgB mutation also eliminates the expression of CsgA (from the *csgBA* operon), i.e., both curli subunits encoded by *csgBA* are not produced.

Finally, together with core cellulose genes, *bcsEFG* genes have been shown to occur in many γ - and β -proteobacteria (3). We isolated the cellulose material from *Salmonella enterica* serovar Typhimurium strain IR715 Δ csgBA (32), a curli mutant, and discovered that it also produces pEtN cellulose. The ^{13}C CPMAS NMR spectrum of isolated cellulose from *Salmonella* matches that of pEtN cellulose from *E. coli* (Fig. 3C). Thus, the pEtN modification of cellulose is likely to be common in the γ and β branches of proteobacteria. Notably, however, *Komagataeibacter xylinus* (formerly known as *Acetobacter xylinum*), which produces excessive amounts of cellulose such that pEtN modification would predictably lead to a depletion of the headgroups of the phospholipid membrane, does not possess the *bcsEFG* genes (3). Other bacterial species that do not possess *bcsEFG* genes, but feature accessory *bcs* genes of unknown function, could possibly use alternative modes of cellulose modification.

Modified cellulose is produced by strains that have been assumed in the literature to be producing standard amorphous cellulose on the basis of simple calcofluor-white staining procedures and conventional isolation methods designed for the detection of glucose from hydrolyzed cellulose. However, these methods involve harsh hydrolysis protocols and crude purification or enrichment methods, followed by chromatography and mass spectrometry, and, thus, a complete accounting of the intact material has not been attempted. Solid-state NMR analysis of the relevant intact polysaccharide was able to identify this biologically important pEtN modification that evaded detection by conventional approaches. pEtN cellulose is a newly identified zwitterionic polymer, and, to our knowledge, our study provides the first definitive evidence so far of a naturally postsynthetically modified cellulose. In the extracellular matrix of bacterial biofilms, pEtN modification of cellulose seems to be multifunctional: It is required for the formation of long cellulose fibrils and a tight nanocomposite with amyloid curli fibers that

generate tissue-like cohesive and elastic behavior of biofilms, it may confer resistance against attacks by cellulase-producing microorganisms (e.g., fungi) in the environment, and it can prevent amyloid curli fibers from hyperstimulating immune responses, which, in the long run, may contribute to pathogen fitness in the host. Moreover, cellulose biosynthesis and postsynthetic modification are coregulated by the ubiquitous biofilm-promoting second messenger c-di-GMP via a transmembrane c-di-GMP signaling pathway. Inhibition of BcsG could offer new opportunities to control biofilm formation, in particular by Gram-negative pathogens associated with chronic infections. Furthermore, the identification of the gene-directed biosynthetic machinery also inspires the generation of engineered systems to produce alternately modified cellulosic materials.

REFERENCES AND NOTES

- Klemm, B.; Heublein, H. P.; Fink, A. Bohn, *Angew. Chem. Int. Ed.* **44**, 3358–3393 (2005).
- Nishiyama, J.; Sugiyama, H.; Chanzy, P.; Langan, J. *Am. Chem. Soc.* **125**, 14300–14306 (2003).
- Römling, M.; Y. Galperin, *Trends Microbiol.* **23**, 545–557 (2015).
- O. Ormadjela et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17856–17861 (2013).
- U. Jenal, A. Reinders, C. Lori, *Nat. Rev. Microbiol.* **15**, 271–284 (2017).
- J. L. Morgan, J. T. McNamara, J. Zimmer, *Nat. Struct. Mol. Biol.* **21**, 489–496 (2014).
- P. S. Stewart, M. J. Franklin, *Nat. Rev. Microbiol.* **6**, 199–210 (2008).
- D. O. Serra, A. M. Richter, G. Klauk, F. Mika, R. Hengge, *mBio* **4**, e00103-13 (2013).
- G. G. Anderson, G. A. O'Toole, *Curr. Top. Microbiol. Immunol.* **322**, 85–105 (2008).
- C. Hung et al., *mBio* **4**, e00645-13 (2013).
- O. A. McCrate, X. Zhou, C. Reichhardt, L. Cegelski, *J. Mol. Biol.* **425**, 4286–4294 (2013).
- D. O. Serra, A. M. Richter, R. Hengge, *J. Bacteriol.* **195**, 5540–5554 (2013).
- C. Reichhardt et al., *Anal. Bioanal. Chem.* **408**, 7709–7717 (2016).
- U. Römling, *Cell. Mol. Life Sci.* **62**, 1234–1246 (2005).
- D. M. Updegraff, *Anal. Biochem.* **32**, 420–424 (1969).
- X. Zogaj, M. Nimtz, M. Rohde, W. Bokranz, U. Römling, *Mol. Microbiol.* **39**, 1452–1463 (2001).
- X. Fang et al., *Mol. Microbiol.* **93**, 439–452 (2014).
- G. Karimova, J. Pidoux, A. Ullmann, D. Ladant, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5752–5756 (1998).
- J. Nilsson, B. Persson, G. von Heijne, *Proteins* **60**, 606–616 (2005).
- C. Manoil, J. Beckwith, *Science* **233**, 1403–1408 (1986).
- C. M. Reynolds, S. R. Kalb, R. J. Cotter, C. R. H. Raetz, *J. Biol. Chem.* **280**, 21202–21211 (2005).
- C. D. Fage, D. B. Brown, J. M. Boll, A. T. Keatinge-Clay, M. S. Trent, *Acta Crystallogr. D Biol. Crystallogr.* **70**, 2730–2739 (2014).
- I. S. Pultz et al., *Mol. Microbiol.* **86**, 1424–1440 (2012).
- S. L. Chen et al., *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5977–5982 (2006).
- Z. Bian, A. Brauner, Y. Li, S. Normark, *J. Infect. Dis.* **181**, 602–612 (2000).
- C. Tükel et al., *Cell. Microbiol.* **12**, 1495–1505 (2010).
- Y. Kai-Larsen et al., *PLoS Pathog.* **6**, e1001010 (2010).
- L. Cegelski et al., *Nat. Chem. Biol.* **5**, 913–919 (2009).
- X. Wang et al., *Cell. Mol. Life Sci.* **63**, 2352–2363 (2006).
- A. M. Richter, T. L. Povolotsky, L. H. Wieler, R. Hengge, *EMBO Mol. Med.* **6**, 1622–1637 (2014).
- H. Karch et al., *EMBO Mol. Med.* **4**, 841–848 (2012).
- C. Tükel et al., *Mol. Microbiol.* **58**, 289–304 (2005).

ACKNOWLEDGMENTS

We thank J. Schaefer for assistance with ^{31}P NMR measurements; S. Lynch and A. Banerjee for assistance with solution-state NMR; and the Vincent Coates Foundation Mass Spectrometry Laboratory, Stanford University Mass Spectrometry (T. McLaughlin) for mass spectrometry assistance. We thank F. Yildiz and D. Zamorano-Sanchez for providing the pMMB956 plasmid; J. Rolff (Freie Universität Berlin) for the ability to use the scanning electron microscope in his laboratories; and C. Tükel for the *Salmonella enterica* serovar Typhimurium strain IR715 Δ csgBA. Financial support was provided by the NSF CAREER award 1453247 (to L.C.), the Precourt Institute for Energy Seed Grant (to L.C.), and the Deutsche Forschungsgemeinschaft grants He1556/17-1 and He1556/21-1 (to R.H.). C.H. was supported by an ERASMUS fellowship from the European Union. W.T. and L.C. have filed a U.S. patent application relating to the work in this manuscript, no. PCT/US2017/047511 filed 18 August 2017 (production and applications of pEtN cellulose).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6373/334/suppl/DC1
Materials and Methods
Figs. S1 to S16
Table S1
References (33–38)

17 July 2017; accepted 13 November 2017
10.1126/science.aao4096

Cite as: S. Chittori *et al.*, *Science*
10.1126/science.aar2781 (2017).

Structural mechanisms of centromeric nucleosome recognition by the kinetochore protein CENP-N

Sagar Chittori,^{1*} Jingjun Hong,^{2*} Hayden Saunders,² Hanqiao Feng,² Rodolfo Ghirlando,³ Alexander E. Kelly,^{2†} Yawen Bai,^{2†} Sriram Subramaniam^{1†}

¹Laboratory of Cell Biology, CCR, NCI, NIH, Bethesda, MD 20892, USA. ²Laboratory of Biochemistry and Molecular Biology, CCR, NCI, NIH, Bethesda, MD 20892, USA.

³Laboratory of Molecular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, NIH, Bethesda, MD 20892, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: baiyaw@mail.nih.gov (Y.B.); alexander.kelly@nih.gov (A.E.K.); ss1@nih.gov (S.S.)

Accurate chromosome segregation requires the proper assembly of kinetochore proteins. A key step in this process is the recognition of the histone H3 variant CENP-A in the centromeric nucleosome by the kinetochore protein CENP-N. We report cryo-EM, biophysical, biochemical, and cell biological studies of the interaction between the CENP-A nucleosome and CENP-N. We show that human CENP-N confers binding specificity through interactions with the L1 loop of CENP-A, stabilized by electrostatic interactions with the nucleosomal DNA. Mutational analyses demonstrate analogous interactions in *Xenopus*, which is further supported by residue-swapping experiments involving the L1 loop of CENP-A. Our results are consistent with co-evolution of CENP-N and CENP-A, and establish the structural basis for recognition of the CENP-A nucleosome to enable kinetochore assembly and centromeric chromatin organization.

The correct transfer of genetic material from mother to daughter cells requires the accurate segregation of chromosomes during mitosis. Mis-segregation of chromosomes can lead to aneuploidy, a hallmark of cancer (1). Chromosome segregation requires the assembly of kinetochore proteins at the centromere, which is epigenetically marked by nucleosomes containing the histone H3 variant CENP-A (2). The centromeric nucleosome associates with a group of 16 inner kinetochore proteins, collectively known as the constitutive centromere associated network (CCAN) (3). Among the CCAN proteins, CENP-C and CENP-N can directly recognize the CENP-A nucleosome core region and are essential for kinetochore assembly and chromosome segregation (4). Two regions of CENP-C, the central region and the CENP-C motif, bind the CENP-A nucleosome by interacting with the C-terminal tail of CENP-A and the acidic patch on histones H2A and H2B in the centromeric nucleosome (5, 6). The CENP-N N-terminal region (residues 1-286 in human) is responsible for recognition of the CENP-A nucleosome (7, 8) while its C-terminal region (residues 287-339) interacts with the CCAN through CENP-L (9, 10). The recruitment of CENP-N to the centromere requires recognition of the L1 loop of CENP-A (11).

To determine structural aspects of the binding between CENP-A and CENP-N, we carried out cryo-EM analyses of the complexes formed between residues 1-286 of human CENP-N

(tagged at its N terminus with His₆ and maltose binding protein (MBP); hereafter referred to as hCENP-N₁₋₂₈₆) and the CENP-A nucleosome containing “601” DNA (figs. S1 and S2). CENP-N binding to the CENP-A nucleosome is independent of DNA sequence (8), and we observed that CENP-A nucleosomes containing either “601” DNA or human centromeric α -satellite DNA bound hCENP-N₁₋₂₈₆ with similar affinity and stoichiometry (fig. S3). When excess hCENP-N₁₋₂₈₆ was used, CENP-N/CENP-A nucleosome complexes with 1:1 or 2:1 stoichiometry are observed in the cryo-EM 2D class averages (fig. S4 and table S1, dataset 1). The complex with 2:1 stoichiometry is a smaller fraction of the overall population, and the density observed for the second site is weaker than at the first site. In the absence of excess hCENP-N₁₋₂₈₆, most of the 2D class averages displayed only one molecule of hCENP-N₁₋₂₈₆ bound to the CENP-A nucleosome (fig. S5C). We refined this population to obtain a 3D reconstruction at an overall resolution of 3.9 Å for the complex formed between hCENP-N₁₋₂₈₆ and the CENP-A nucleosome (Fig. 1; figs. S5 and S6; and table S1, dataset 2).

The density map shows an identifiable nucleosome and a bi-lobed assembly corresponding to hCENP-N₁₋₂₈₆ (Fig. 1, A to C). The resolution of this map is highest in the core of the CENP-A nucleosome and lower in the periphery and in the bound CENP-N (Fig. 1D and fig. S6). The map of the CENP-A nucleosome core is at a resolution (~3.5 Å) adequate to interpret directly in terms of an atomic model since side-chain densities are clearly visualized (Fig. 1D and fig. S7). The final

structure of the CENP-A nucleosome core is similar to the crystal structure of the free CENP-A nucleosome (PDB 3AN2) (12). While the DNA near the entry and exit sites can be visualized (fig. S8) and is displaced further away from the core histones in comparison with the canonical nucleosome structure (PDB 3LZ0) (13), the significance of these minor differences with previously reported structures is currently unclear.

The map of the CENP-N region is at a lower resolution (~4.5 Å; fig. S6, C to F). At this resolution, densities for several α -helical regions as well as β -strands are well-resolved, but since side chain densities are not delineated, it is not possible to thread the sequence unambiguously into the density map. In order to identify CENP-N residues that may form the interface with the CENP-A nucleosome, we carried out an extensive, iterative process to predict and fit various stretches of secondary structure elements and subdomains into the density map (see Methods). These efforts resulted in a predicted molecular model with a sequence assignment based on the modeled subdomains (figs. S9 and S10 and table S1). From this model, we identified residues likely to form the interface with the CENP-A nucleosome.

In our model, residues 1-185 of hCENP-N are organized into distinct N-terminal and central domains (Fig. 1 and figs. S9 and S10). The MBP-moiety, which is attached to the N terminus of hCENP-N₁₋₂₈₆ through a di-peptide (Ala-Ala) linker, does not interfere with the hCENP-N/CENP-A nucleosome interaction based on binding studies (fig. S3, C to G). The apparent K_D values for MBP-tagged and untagged hCENP-N₁₋₂₈₆ are $0.042 \pm 0.007 \mu\text{M}$ and $0.046 \pm 0.008 \mu\text{M}$, respectively. This is consistent with our assignment of the location of MBP to the weak density to the side of hCENP-N₁₋₂₈₆ that is distal from the core (Fig. 1C and figs. S5E and S6A). There is unstructured density consistent with regions beyond residue 185 of hCENP-N, but they could not be assigned presumably due to conformational flexibility (fig. S9, A and B). The N-terminal hCENP-N domain, comprising residues 1-81, is composed of five antiparallel α -helices that show structural conservation to proteins of the death domain superfamily, while the central domain comprising residues 101-185 contains both α helices and β strands with similarity to regions of the appendage domain of $\beta 2$ adaptin (fig. S9, F to H) (14, 15).

The structure of the complex shows that hCENP-N₁₋₂₈₆ has direct interactions with the DNA of the nucleosome, as well as with the L1 loop of CENP-A (Fig. 2A and fig. S11, A and B). We identified several positively charged residues of hCENP-N proximal to nucleosomal DNA that likely form stabilizing interactions with the backbone phosphates of DNA (Fig. 2B). Consistent with our structural model, mutations of hCENP-N residues R11, K15 and K45 have already been shown to reduce its binding to the CENP-A nucleosome, and further, mutation of R11 disrupts centromeric localization and the recruitment

of other CCAN proteins in human cells (8). To further verify the role of these and other residues identified at the interface, we constructed hCENP-N₁₋₂₈₆ mutants in the N-terminal domain (R11A, R44A, K45A, K81A) and in the central domain (K148A, R169A, and R170A) and analyzed the binding by electrophoretic mobility shift analysis (EMSA); these mutations impair hCENP-N₁₋₂₈₆ binding to the CENP-A nucleosome (Fig. 2C). Mutation of residues H77, H79, or K90, which are located on the linker connecting the N-terminal and central domains, also resulted in decreased binding affinity to various extents (fig. S12A), likely due to perturbations of CENP-N tertiary structure. As controls, we also tested the effects of mutating several positively charged residues predicted to be distal from the nucleosomal DNA (R60A, K109A/K110A, and K143A) or in residues C-terminal to the central domain (K207A, R236A, and R250A). All of these mutations showed minimal effects on hCENP-N₁₋₂₈₆ binding affinity (Fig. 2D and fig. S12A) (7).

To examine the physiological significance of CENP-N interactions with the nucleosomal DNA, we mutated conserved, and analogous residues within *Xenopus laevis* CENP-N (xCENP-N) (fig. S13A), and tested their effects on the centromeric localization of exogenously expressed MBP-xCENP-N/CENP-L complex in interphase *Xenopus* egg extracts (fig. S14). Mutation of xCENP-N residues R29, K99 and K165 (which correspond to residues R11, K81 and K148 in hCENP-N) led to a significant reduction in the localization of the xCENP-N/L complex to centromeres (Fig. 2, E and F). Mutation of K108 (K90 in hCENP-N) in the linker region caused minor defects in localization (fig. S14, E and F). Furthermore, mutation of K223 (K207 in hCENP-N), which is C-terminal to the central domain, did not diminish centromeric localization (fig. S14, E and F). These results are in agreement with our findings on hCENP-N (Fig. 2 and fig. S14I), and suggest that the interactions with nucleosomal DNA identified in the structure of the hCENP-N₁₋₂₈₆/CENP-A nucleosome complex are conserved across *Xenopus* and human centromeres.

In our structural model, the N-terminal helix of hCENP-N₁₋₂₈₆ is in contact with the L1 loop of CENP-A, in which residues R80 and G81 constitute an insertion compared to canonical H3.1 (fig. S11B) (12). Previous studies have shown that hCENP-N fails to localize to the centromere upon a double mutation of R80A and G81A in CENP-A (11), and substitution of the centromere targeting domain (CATD) of CENP-A including the L1 loop and $\alpha 2$ helix to the corresponding region in H3 (H3^{CATD}) is sufficient for recruitment of CENP-N (8, 16). We further corroborated these findings by in vitro mutational analysis (fig. S12B), and by cryo-EM structure determination of the hCENP-N₁₋₂₈₆/H3.1^{CATD} chimeric nucleosome complex (fig. S15 and table S1), which revealed that hCENP-N₁₋₂₈₆ binding to the chimeric nucleosome is structurally similar to its binding to CENP-A containing nucleosome (fig. S15E). In our structural model of the hCENP-N₁₋₂₈₆/CENP-A nucleosome

complex, conserved residues E3 and E7 from helix 1 of hCENP-N are proximal to R80 of CENP-A, and residue T4 of hCENP-N is close to V82 of CENP-A (Fig. 3, A and B, and fig. S11). The structure thus suggests that both electrostatic and hydrophobic interactions are involved at the interface. Gel shift assays show that the E3A/E7A and T4A mutants of hCENP-N₁₋₂₈₆ show reduced binding to the CENP-A nucleosome, while the E3K, E7K, E3K/E7K and E3A/T4A/E7A mutants show aberrant binding (Fig. 3C and fig. S12C). Further, the E3K/E7K and E3A/T4A/E7A mutants exhibit loss of specificity for CENP-A over H3.1 nucleosomes (fig. S12, D and E). Together, these results are consistent with a model in which a small number of residues of hCENP-N that interact with the L1 loop of CENP-A provide interaction specificity, while broad electrostatic interactions with the backbone phosphates of DNA enhance binding affinity.

To further evaluate the physiological significance of CENP-N interactions with the L1 loop of CENP-A, we examined the effects of the xCENP-N double mutants E21A/E25A and E21K/E25K (which correspond to mutations E3A/E7A and E3K/E7K in hCENP-N) on centromeric localization in *Xenopus* egg extracts (Fig. 3, D and E, and fig. S13A). Complete loss of centromeric localization was observed in the E21K/E25K mutant, but not in the E21A/E25A mutant (Fig. 3, D and E, and fig. S14C). Furthermore, the W22A mutation in xCENP-N (corresponding to mutation of T4 in hCENP-N) severely disrupted centromeric localization (Fig. 3, F to H, and fig. S14D). Thus, xCENP-N localization appears to depend more on hydrophobic rather than electrostatic interactions as compared to hCENP-N, where the E3A/E7A mutation leads to a total loss of centromeric localization in human cells (fig. S14, G to I). Our results support a model in which residues within helix 1 of hCENP-N (E3, T4, and E7) are critical for the recognition of the L1 loop within the CENP-A nucleosome.

CENP-A and centromeric DNA sequences show evidence for rapid evolution (17, 18), and alignments of the CENP-N N-terminal region and the CENP-A L1 loop across eukaryotes show a high degree of variability (fig. S13, B to D). To test if CENP-N may have co-evolved with CENP-A to maintain kinetochore assembly, we created a chimeric CENP-A nucleosome in which R80 and V82 of hCENP-A were respectively replaced by Cys and Met to mimic the *Xenopus* CENP-A L1 loop sequence (Fig. 3F). Wild-type hCENP-N does not interact with the chimeric nucleosome (Fig. 3I), in agreement with decreased centromeric localization of a W22T xCENP-N mutant (fig. S14, E and F). However, chimeric nucleosome binding could be enabled by a complementary hCENP-N T4W mutant, which also retained its affinity for wild-type hCENP-A nucleosomes (Fig. 3I). Together, these results suggest that CENP-N has likely co-evolved with CENP-A to enable kinetochore assembly utilizing hydrophobic and electrostatic interactions to recognize centromeric chromatin across different

species.

Centromere identity and function are specified by a two-step mechanism wherein the deposition of CENP-A at the centromere by the CENP-A chaperone HJURP allows for the recognition of CENP-A by the kinetochore proteins CENP-C and CENP-N (19). Interestingly, structures of HJURP (20–22), CENP-C fragments (5) and CENP-N (this study) in complex with different forms of CENP-A show that in all three cases, only a small number of residues are involved in the specific recognition of CENP-A while additional interactions are used for further stabilization (Fig. 4A, and fig. S11, C and D). These residues on CENP-A have both conserved and necessary functions, despite the variability of their amino acid sequence across eukaryotes. Our data suggest the CENP-A-interacting residues of CENP-N co-evolve with CENP-A, which could reflect a common strategy for maintaining CENP-A specificity (17). Previous studies (7, 23) and our results (Fig. 4B and fig. S2, A and D) show that CENP-N and fragments of CENP-C that include one of its two CENP-A nucleosome binding domains (fig. S2E; central domain (7, 23); or CENP-C motif (this work)) can simultaneously bind to the same face of the CENP-A nucleosome. Furthermore, the kinetochore protein CENP-L has been shown to mediate interactions between CENP-C and CENP-N (9, 23–25). Thus, the valency of CENP-C-CENP-N-CENP-L interactions could facilitate clustering of sparse and non-adjacent CENP-A nucleosomes (Fig. 4C) (9, 26), which might help establish the folding of centromeric chromatin (27, 28), and/or the integrity of the kinetochore (29, 30).

REFERENCES AND NOTES

1. G. J. Kops, B. A. Weaver, D. W. Cleveland, On the road to cancer: Aneuploidy and the mitotic checkpoint. *Nat. Rev. Cancer* **5**, 773–785 (2005). [doi:10.1038/nrc1714](https://doi.org/10.1038/nrc1714) [Medline](#)
2. K. L. McKinley, I. M. Cheeseman, The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol.* **17**, 16–29 (2016). [doi:10.1038/nrm.2015.5](https://doi.org/10.1038/nrm.2015.5) [Medline](#)
3. J. S. Verdaasdonk, K. Bloom, Centromeres: Unique chromatin structures that drive chromosome segregation. *Nat. Rev. Mol. Cell Biol.* **12**, 320–332 (2011). [doi:10.1038/nrm3107](https://doi.org/10.1038/nrm3107) [Medline](#)
4. C. W. Carroll, K. J. Milks, A. F. Straight, Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J. Cell Biol.* **189**, 1143–1155 (2010). [doi:10.1083/jcb.201001013](https://doi.org/10.1083/jcb.201001013) [Medline](#)
5. H. Kato, J. Jiang, B.-R. Zhou, M. Rozendaal, H. Feng, R. Ghirlando, T. S. Xiao, A. F. Straight, Y. Bai, A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. *Science* **340**, 1110–1113 (2013). [doi:10.1126/science.1235532](https://doi.org/10.1126/science.1235532) [Medline](#)
6. A. Guse, C. W. Carroll, B. Moree, C. J. Fuller, A. F. Straight, In vitro centromere and kinetochore assembly on defined chromatin templates. *Nature* **477**, 354–358 (2011). [doi:10.1038/nature10379](https://doi.org/10.1038/nature10379) [Medline](#)
7. L. Y. Guo, P. K. Allu, L. Zandarashvili, K. L. McKinley, N. Sekulic, J. M. Dawicki-McKenna, D. Fachinetti, G. A. Logsdon, R. M. Jamiolkowski, D. W. Cleveland, I. M. Cheeseman, B. E. Black, Centromeres are maintained by fastening CENP-A to DNA and directing an arginine anchor-dependent nucleosome transition. *Nat. Commun.* **8**, 15775 (2017). [doi:10.1038/ncomms15775](https://doi.org/10.1038/ncomms15775) [Medline](#)
8. C. W. Carroll, M. C. Silva, K. M. Godek, L. E. Jansen, A. F. Straight, Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. *Nat. Cell Biol.* **11**, 896–902 (2009). [doi:10.1038/ncb1899](https://doi.org/10.1038/ncb1899) [Medline](#)

9. K. L. McKinley, N. Sekulic, L. Y. Guo, T. Tsinman, B. E. Black, I. M. Cheeseman, The CENP-L-N complex forms a critical node in an integrated meshwork of interactions at the centromere-kinetochore interface. *Mol. Cell* **60**, 886–898 (2015). [doi:10.1016/j.molcel.2015.10.027](https://doi.org/10.1016/j.molcel.2015.10.027) [Medline](#)
10. S. M. Hinshaw, S. C. Harrison, An Imi3-Chl4 heterodimer links the core centromere to factors required for accurate chromosome segregation. *Cell Reports* **5**, 29–36 (2013). [doi:10.1016/j.celrep.2013.08.036](https://doi.org/10.1016/j.celrep.2013.08.036) [Medline](#)
11. J. Fang, Y. Liu, Y. Wei, W. Deng, Z. Yu, L. Huang, Y. Teng, T. Yao, Q. You, H. Ruan, P. Chen, R.-M. Xu, G. Li, Structural transitions of centromeric chromatin regulate the cell cycle-dependent recruitment of CENP-N. *Genes Dev.* **29**, 1058–1073 (2015). [doi:10.1101/gad.259432.115](https://doi.org/10.1101/gad.259432.115) [Medline](#)
12. H. Tachiwana, W. Kagawa, T. Shiga, A. Osakabe, Y. Miya, K. Saito, Y. Hayashi-Takanaka, T. Oda, M. Sato, S.-Y. Park, H. Kimura, H. Kurumizaka, Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* **476**, 232–235 (2011). [doi:10.1038/nature10258](https://doi.org/10.1038/nature10258) [Medline](#)
13. D. Vasudevan, E. Y. Chua, C. A. Davey, Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J. Mol. Biol.* **403**, 1–10 (2010). [doi:10.1016/j.jmb.2010.08.039](https://doi.org/10.1016/j.jmb.2010.08.039) [Medline](#)
14. S. F. Sukits, L.-L. Lin, S. Hsu, K. Malakian, R. Powers, G.-Y. Xu, Solution structure of the tumor necrosis factor receptor-1 death domain. *J. Mol. Biol.* **310**, 895–906 (2001). [doi:10.1006/jmbi.2001.4790](https://doi.org/10.1006/jmbi.2001.4790) [Medline](#)
15. Y. Wu, T. R. Albrecht, D. Baillat, E. J. Wagner, L. Tong, Molecular basis for the interaction between Integrator subunits IntS9 and IntS11 and its functional importance. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4394–4399 (2017). [doi:10.1073/pnas.1616605114](https://doi.org/10.1073/pnas.1616605114) [Medline](#)
16. B. E. Black, D. R. Foltz, S. Chakravarthy, K. Luger, V. L. Woods Jr., D. W. Cleveland, Structural determinants for generating centromeric chromatin. *Nature* **430**, 578–582 (2004). [doi:10.1038/nature02766](https://doi.org/10.1038/nature02766) [Medline](#)
17. H. S. Malik, S. Henikoff, Adaptive evolution of Cid, a centromere-specific histone in *Drosophila*. *Genetics* **157**, 1293–1298 (2001). [Medline](#)
18. S. Henikoff, Chromosomes on the move. *Trends Genet.* **17**, 689–690 (2001). [doi:10.1016/S0168-9525\(01\)02527-6](https://doi.org/10.1016/S0168-9525(01)02527-6) [Medline](#)
19. D. Fachinetti, H. D. Folco, Y. Nechemia-Arbely, L. P. Valente, K. Nguyen, A. J. Wong, Q. Zhu, A. J. Holland, A. Desai, L. E. T. Jansen, D. W. Cleveland, A two-step mechanism for epigenetic specification of centromere identity and function. *Nat. Cell Biol.* **15**, 1056–1066 (2013). [doi:10.1038/ncb2805](https://doi.org/10.1038/ncb2805) [Medline](#)
20. Z. Zhou, H. Feng, B.-R. Zhou, R. Ghirlando, K. Hu, A. Zwolak, L. M. Miller Jenkins, H. Xiao, N. Tjandra, C. Wu, Y. Bai, Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3. *Nature* **472**, 234–237 (2011). [doi:10.1038/nature09854](https://doi.org/10.1038/nature09854) [Medline](#)
21. H. Hu, Y. Liu, M. Wang, J. Fang, H. Huang, N. Yang, Y. Li, J. Wang, X. Yao, Y. Shi, G. Li, R.-M. Xu, Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. *Genes Dev.* **25**, 901–906 (2011). [doi:10.1101/gad.204511](https://doi.org/10.1101/gad.204511) [Medline](#)
22. U. S. Cho, S. C. Harrison, Recognition of the centromere-specific histone Cse4 by the chaperone Scm3. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9367–9371 (2011). [doi:10.1073/pnas.1106389108](https://doi.org/10.1073/pnas.1106389108) [Medline](#)
23. J. R. Weir, A. C. Faesen, K. Klare, A. Petrovic, F. Basilico, J. Fischböck, S. Pentakota, J. Keller, M. E. Pesenti, D. Pan, D. Vogt, S. Wohlgemuth, F. Herzog, A. Musacchio, Insights from biochemical reconstitution into the architecture of human kinetochores. *Nature* **537**, 249–253 (2016). [doi:10.1038/nature19333](https://doi.org/10.1038/nature19333) [Medline](#)
24. H. Nagpal, T. Hori, A. Furukawa, K. Sugase, A. Osakabe, H. Kurumizaka, T. Fukagawa, Dynamic changes in CCAN organization through CENP-C during cell-cycle progression. *Mol. Biol. Cell* **26**, 3768–3776 (2015). [doi:10.1091/mbc.F15-07-0531](https://doi.org/10.1091/mbc.F15-07-0531) [Medline](#)
25. K. Klare, J. R. Weir, F. Basilico, T. Zimniak, L. Massimiliano, N. Ludwigs, F. Herzog, A. Musacchio, CENP-C is a blueprint for constitutive centromere-associated network assembly within human kinetochores. *J. Cell Biol.* **210**, 11–22 (2015). [doi:10.1083/jcb.201412028](https://doi.org/10.1083/jcb.201412028) [Medline](#)
26. D. L. Bodor, J. F. Mata, M. Sergeev, A. F. David, K. J. Salimian, T. Panchenko, D. W. Cleveland, B. E. Black, J. V. Shah, L. E. T. Jansen, The quantitative architecture of centromeric chromatin. *eLife* **3**, e02137 (2014). [doi:10.7554/eLife.02137](https://doi.org/10.7554/eLife.02137) [Medline](#)
27. M. D. Blower, B. A. Sullivan, G. H. Karpen, Conserved organization of centromeric chromatin in flies and humans. *Dev. Cell* **2**, 319–330 (2002). [doi:10.1016/S1534-5807\(02\)00135-1](https://doi.org/10.1016/S1534-5807(02)00135-1) [Medline](#)
28. S. A. Ribeiro, P. Vagnarelli, Y. Dong, T. Hori, B. F. McEwen, T. Fukagawa, C. Flors, W. C. Earnshaw, A super-resolution map of the vertebrate kinetochore. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10484–10489 (2010). [doi:10.1073/pnas.1002325107](https://doi.org/10.1073/pnas.1002325107) [Medline](#)
29. A. Suzuki, B. L. Badger, X. Wan, J. G. DeLuca, E. D. Salmon, The architecture of CCAN proteins creates a structural integrity to resist spindle forces and achieve proper intrakinetochore stretch. *Dev. Cell* **30**, 717–730 (2014). [doi:10.1016/j.devcel.2014.08.003](https://doi.org/10.1016/j.devcel.2014.08.003) [Medline](#)
30. G. Vargiu, A. A. Makarov, J. Allan, T. Fukagawa, D. G. Booth, W. C. Earnshaw, Stepwise unfolding supports a subunit model for vertebrate kinetochores. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 3133–3138 (2017). [doi:10.1073/pnas.1614145114](https://doi.org/10.1073/pnas.1614145114) [Medline](#)
31. Y. Tanaka, M. Tawaramoto-Sasanuma, S. Kawaguchi, T. Ohta, K. Yoda, H. Kurumizaka, S. Yokoyama, Expression and purification of recombinant human histones. *Methods* **33**, 3–11 (2004). [doi:10.1016/j.ymeth.2003.10.024](https://doi.org/10.1016/j.ymeth.2003.10.024) [Medline](#)
32. S. J. Falk, L. Y. Guo, N. Sekulic, E. M. Smoak, T. Mani, G. A. Logsdon, K. Gupta, L. E. T. Jansen, G. D. Van Duyne, S. A. Vinogradov, M. A. Lampson, B. E. Black, CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. *Science* **348**, 699–703 (2015). [doi:10.1126/science.1259308](https://doi.org/10.1126/science.1259308) [Medline](#)
33. P. N. Dyer, R. S. Edayathumangalam, C. L. White, Y. Bao, S. Chakravarthy, U. M. Muthurajan, K. Luger, Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* **375**, 23–44 (2003). [doi:10.1016/S0076-6879\(03\)75002-2](https://doi.org/10.1016/S0076-6879(03)75002-2) [Medline](#)
34. A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, M. Mann, In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2006). [doi:10.1038/nprot.2006.468](https://doi.org/10.1038/nprot.2006.468) [Medline](#)
35. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997). [doi:10.1038/38444](https://doi.org/10.1038/38444) [Medline](#)
36. P. Schuck, On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* **320**, 104–124 (2003). [doi:10.1016/S0003-2697\(03\)00289-6](https://doi.org/10.1016/S0003-2697(03)00289-6) [Medline](#)
37. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995). [doi:10.1007/BF00197809](https://doi.org/10.1007/BF00197809) [Medline](#)
38. M. Wühr, R. M. Freeman Jr., M. Presler, M. E. Horb, L. Peshkin, S. Gygi, M. W. Kirschner, Deep proteomics of the *Xenopus laevis* egg using an mRNA-derived reference database. *Curr. Biol.* **24**, 1467–1475 (2014). [doi:10.1016/j.cub.2014.05.044](https://doi.org/10.1016/j.cub.2014.05.044) [Medline](#)
39. A. W. Murray, Cell cycle extracts. *Methods Cell Biol.* **36**, 581–605 (1991). [doi:10.1016/S0091-679X\(08\)60298-8](https://doi.org/10.1016/S0091-679X(08)60298-8) [Medline](#)
40. I. Krizaic, S. J. Williams, P. Sánchez, M. Rodríguez-Corsino, P. T. Stukenberg, A. Losada, The distinct functions of CENP-C and CENP-T/W in centromere propagation and function in *Xenopus* egg extracts. *Nucleus* **6**, 133–143 (2015). [doi:10.1080/19491034.2014.1003509](https://doi.org/10.1080/19491034.2014.1003509) [Medline](#)
41. A. Tighe, V. L. Johnson, S. S. Taylor, Truncating APC mutations have dominant effects on proliferation, spindle checkpoint control, survival and chromosome stability. *J. Cell Sci.* **117**, 6339–6353 (2004). [doi:10.1242/jcs.01556](https://doi.org/10.1242/jcs.01556) [Medline](#)
42. A. Bartesaghi, D. Matthies, S. Banerjee, A. Merk, S. Subramaniam, Structure of β -galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11709–11714 (2014). [doi:10.1073/pnas.1402809111](https://doi.org/10.1073/pnas.1402809111) [Medline](#)
43. S. H. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012). [doi:10.1016/j.jsb.2012.09.006](https://doi.org/10.1016/j.jsb.2012.09.006) [Medline](#)
44. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010). [doi:10.1107/S0907444910007493](https://doi.org/10.1107/S0907444910007493) [Medline](#)
45. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010). [doi:10.1107/S0907444909052925](https://doi.org/10.1107/S0907444909052925) [Medline](#)
46. D. T. Jones, Protein secondary structure prediction based on position-specific

- scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999). [doi:10.1006/jmbi.1999.3091](https://doi.org/10.1006/jmbi.1999.3091) [Medline](#)
47. Y. Wang, J. Wang, R. Li, Q. Shi, Z. Xue, Y. Zhang, ThreaDomEx: A unified platform for predicting continuous and discontinuous protein domains by multiple-threading and segment assembly. *Nucleic Acids Res.* **45**, W400–W407 (2017). [Medline](#)
 48. D. Xu, Y. Zhang, Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins* **80**, 1715–1735 (2012). [Medline](#)
 49. V. B. Chen, W. B. Arendall III, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *MolProbity*: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010). [doi:10.1107/S0907444909042073](https://doi.org/10.1107/S0907444909042073) [Medline](#)
 50. B. A. Barad, N. Echols, R. Y.-R. Wang, Y. Cheng, F. DiMaio, P. D. Adams, J. S. Fraser, EMRinger: Side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015). [doi:10.1038/nmeth.3541](https://doi.org/10.1038/nmeth.3541) [Medline](#)
 51. J. B. Heymann, Bsoft: Image and molecular processing in electron microscopy. *J. Struct. Biol.* **133**, 156–169 (2001). [doi:10.1006/jsbi.2001.4339](https://doi.org/10.1006/jsbi.2001.4339) [Medline](#)
 52. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). [doi:10.1002/jcc.20084](https://doi.org/10.1002/jcc.20084) [Medline](#)
 53. The PyMOL Molecular Graphics System, Version 1.8, Schrödinger, LLC. (DeLano Scientific, 2002).
 54. T. Ishida, K. Kinoshita, PrDOS: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **35** (suppl. 2), W460–W464 (2007). [doi:10.1093/nar/gkm363](https://doi.org/10.1093/nar/gkm363) [Medline](#)
 55. L. Holm, P. Rosenström, Dali server: Conservation mapping in 3D. *Nucleic Acids Res.* **38** (suppl. 2), W545–W549 (2010). [doi:10.1093/nar/gkq366](https://doi.org/10.1093/nar/gkq366) [Medline](#)

ACKNOWLEDGMENTS

This research was supported by the intramural programs of the NCI, and NIDDK, NIH. We thank Tara Fox and Ulrich Baxa at the Center for Molecular Microscopy for help with cryo-EM data collection, Alberto Bartesaghi for advice with cryo-EM analysis, Lisa Jenkins for mass spectrometry, Shipeng Li for DNA plasmid purifications, and Veronica Falconieri for assistance with figure preparation. This study utilized the computational resources of the High-Performance Computing (HPC) Biowulf cluster at NIH, Bethesda, MD (<http://hpc.nih.gov>). The cryo-EM density map and atomic coordinates for the hCENP-N₁₋₂₈₆/CENP-A nucleosome complex have been deposited in the EMDB and PDB under accession codes EMD-7293 and 6BUZ, respectively. S.C., J.H., A.E.K., Y.B., and S.S. were involved in all stages of experiment design and interpretation of results; J.H. prepared nucleosome complex samples and performed biochemical analysis; H.F. and J.H. performed NMR experiments; R.G. and J.H. performed AUC experiments; H.S. and A.E.K. performed in vivo experiments; S.C. carried out cryo-EM grid preparation, data collection, and image processing; S.C. built the structural model with inputs from J.H.; S.C., Y.B., A.E.K., and J.H. carried out structural analysis; S.C., J.H., A.E.K., Y.B., and S.S. integrated all of the data, and wrote the manuscript, with help from all authors. The authors declare no competing financial interests.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aar2781/DC1

Materials and Methods

Figs. S1 to S15

Table S1

References (31–55)

20 October 2017; accepted 12 December 2017

Published online 21 December 2017

10.1126/science.aar2781

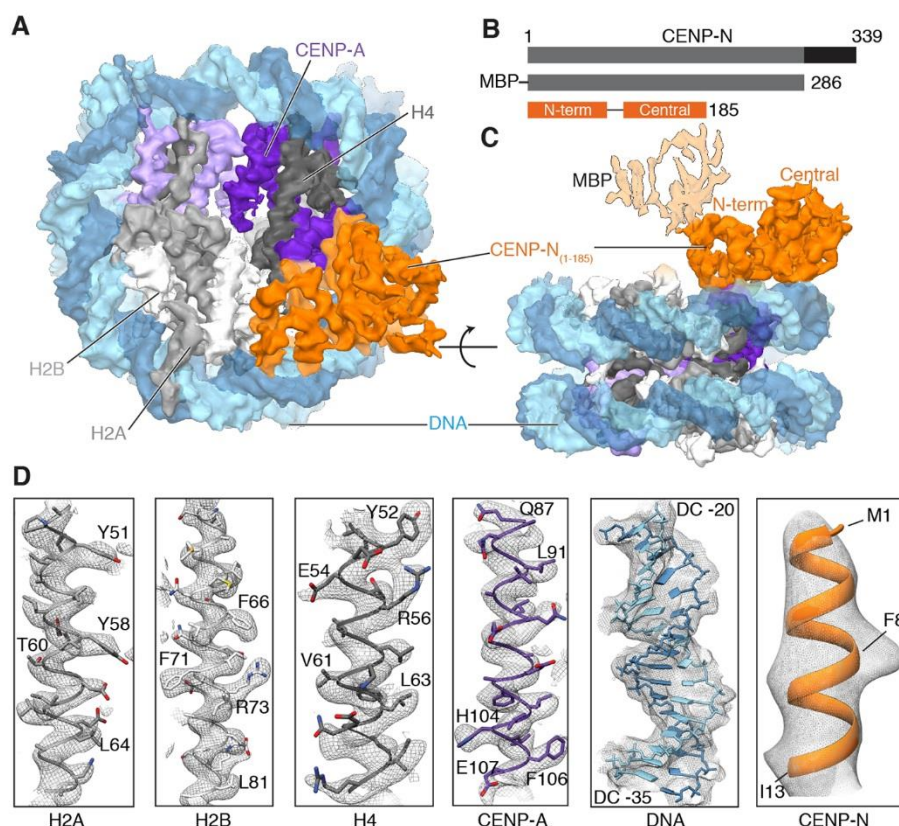


Fig. 1. Structure of the human CENP-N/CENP-A nucleosome complex. (A) Cryo-EM density map of the hCENP-N₁₋₂₈₆/CENP-A nucleosome complex viewed down the axis of the DNA supercoil. (B) Schematic of the functional domains of CENP-N known to bind the CENP-A nucleosome (gray) and CENP-L (black) (top panel). The CENP-N construct used for the present structural analysis (hCENP-N₁₋₂₈₆), and the regions of the sequence whose structure we report here (N-terminal domain: 1-81, and central domain: 101-185; hCENP-N₁₋₁₈₅) are shown in the middle and bottom panels, respectively. (C) Cryo-EM density map of the hCENP-N₁₋₂₈₆/CENP-A nucleosome complex as viewed from the side, at an orientation 90° to the view shown in (A). This view also depicts the extra density connected to the N-terminal domain that we assign to MBP, shown with lighter shading. (D) Representative regions of the cryo-EM density map to illustrate map quality (from left to right) for canonical histones H2A, H2B, and H4, centromere specific H3 variant CENP-A, nucleosomal DNA, and CENP-N.

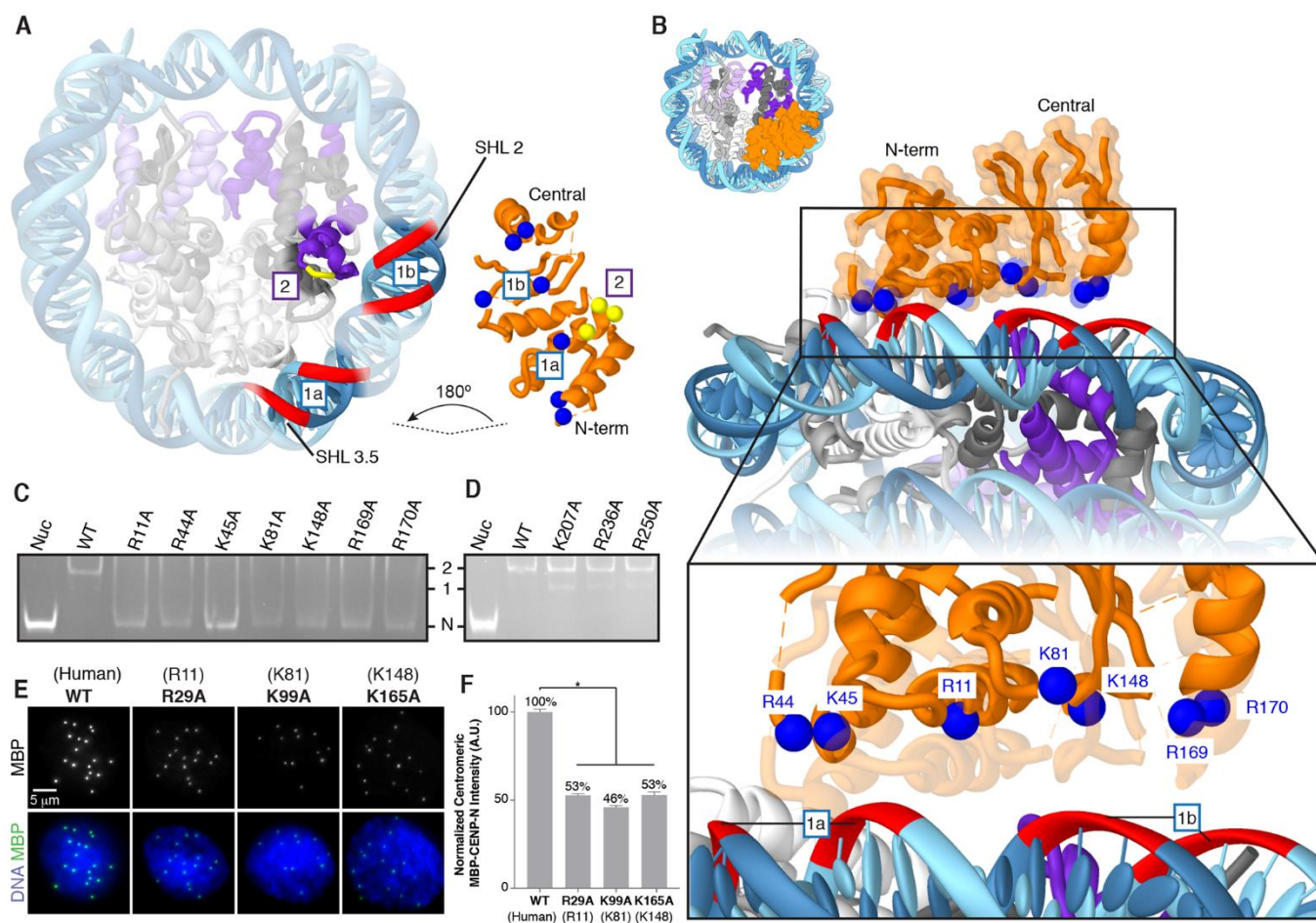


Fig. 2. Interaction of CENP-N with nucleosomal DNA. (A) Cut-away view of the hCENP-N₁₋₂₈₆/CENP-A nucleosome model to highlight interfaces involved in complex formation (please also see fig. S11, A and B). For the CENP-N/DNA interface (labeled “1a” and “1b”), nucleosomal DNA is shown as a red ribbon, while positively charged residues of CENP-N that are proposed to interact with it are shown as blue spheres. For the CENP-N/CENP-A interface (labeled “2”), CENP-A residues (R80-G81-V82) are marked by the short yellow ribbon while interacting CENP-N residues (E3, T4 and E7) are shown as yellow spheres. (B) View of the CENP-N/DNA interface at different magnifications to highlight details of interactions between the nucleosomal DNA and positively charged residues of CENP-N. (C) Gel mobility shift experiment to examine the effects of CENP-N mutations (indicated at top of the gel) on binding to the CENP-A nucleosome. Impaired binding is reflected by increased intensity of the free nucleosome band, concomitant with the disappearance of defined 1:1 and 2:1 bands. Labels: N, indicates migration position of the free CENP-A nucleosome; “1” and “2”, indicate migration positions of CENP-A nucleosomes bound with either one or two molecules of CENP-N, respectively. (D) Similar analysis to that in (C), carried out with a set of CENP-N mutations involving residues distal from the binding interface. (E) Images of interphase nuclei in *Xenopus* egg extracts with exogenous MBP-xCENP-N and xCENP-L protein containing the indicated mutations (with analogous human mutations in parentheses), stained with an antibody for MBP (green) and Hoechst (blue). (F) Centromeric MBP fluorescence intensity normalized as a percentage of that observed for wildtype MBP-xCENP-N. Error bars represent SEM ($n > 200$ centromeres).

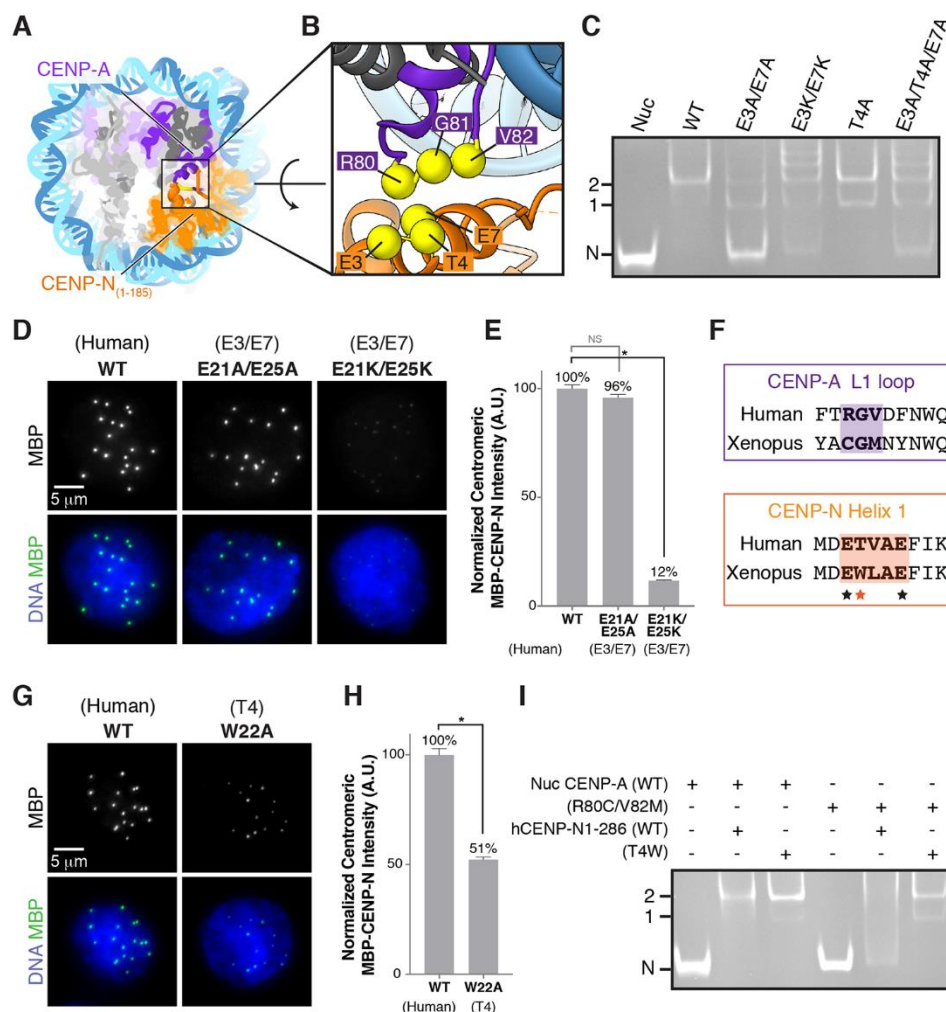


Fig. 3. Interaction between the L1 loop of CENP-A and helix 1 of CENP-N. (A and B) Overall (A) and close-up (B) view of the hCENP-N₁₋₂₈₆/CENP-A interface formed between R80-G81-V82 on the L1 loop of CENP-A, and E3, T4, and E7 on helix 1 of CENP-N. (C) Gel mobility shift experiment to examine effects of CENP-N mutations (indicated at top of the gel) on binding to the CENP-A nucleosome. (D) Images of interphase nuclei in *Xenopus* egg extracts with exogenous MBP-xCENP-L and xCENP-L protein containing the indicated mutations of xCENP-N residues E21 and E25 (corresponding to residues E3 and E7 in hCENP-N), stained with an antibody for MBP (green) and Hoechst (blue). (E) Centromeric MBP fluorescence intensity normalized as a percentage of that observed for wildtype MBP-xCENP-N. Error bars represent SEM ($n > 200$ centromeres). (F) Alignment of human and *Xenopus laevis* sequences corresponding to the L1 loop of CENP-A and helix 1 of CENP-N. Closely interacting segments of the L1 loop of CENP-A and helix 1 of CENP-N are highlighted by shading. The asterisks indicate conserved Glu residues (black asterisks) and variability in the hydrophobic residue corresponding to position T4 (red asterisk) of human CENP-N. (G) Images of interphase nuclei in *Xenopus* egg extracts with exogenous MBP-xCENP-N and xCENP-L protein containing the indicated mutations of xCENP-N, as in (D). (H) Centromeric MBP fluorescence intensity, determined as in (E). (I) Gel mobility shift experiment to examine the effects of correlated amino acid substitutions between the L1 loop of CENP-A and helix 1 of CENP-N.

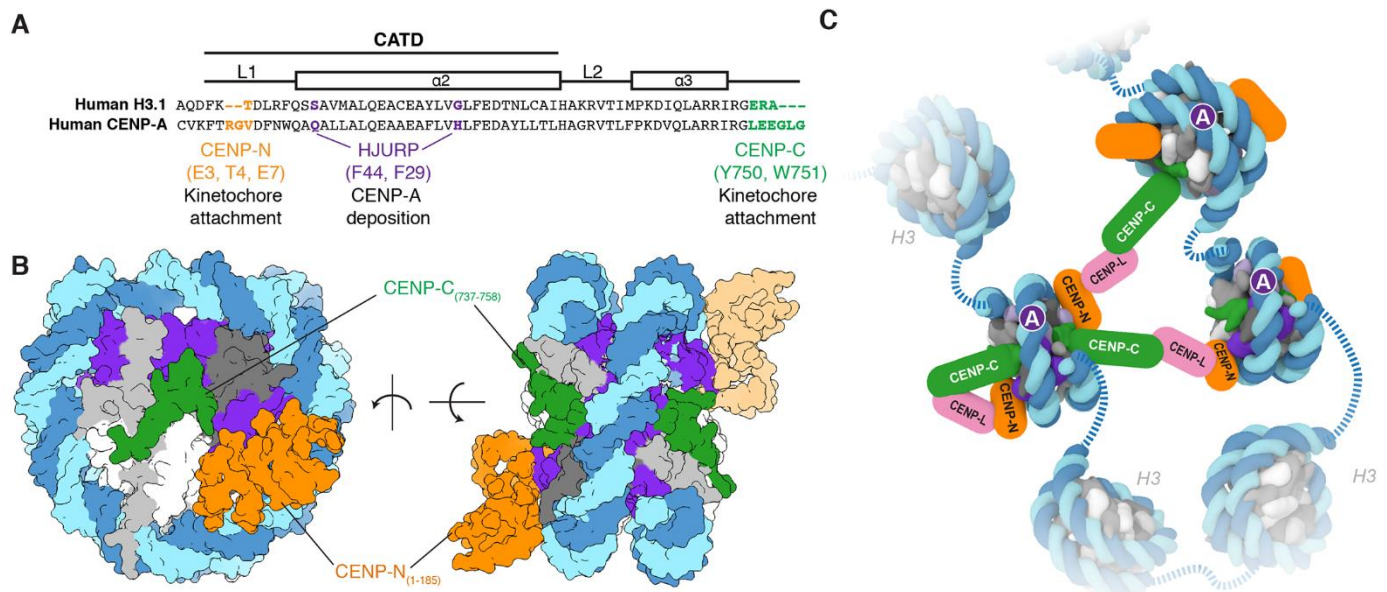


Fig. 4. Structural determinants of kinetochore assembly on the CENP-A nucleosome. (A) Sequence alignment between human H3.1 and CENP-A to highlight unique CENP-A motifs involved in deposition and recognition of CENP-A at centromeric chromatin (also see fig. S11, C and D). (B) Two different views of the CENP-A nucleosome bound to hCENP-N and a modeled CENP-C motif peptide (5) to highlight potential dual binding of full-length CENP-C and CENP-N proteins on the CENP-A nucleosome. The second CENP-N (shown with lighter shading) is modeled based on the cryo-EM density map obtained in the presence of excess hCENP-N₁₋₂₈₆ (fig. S4), while the CENP-C motif peptides (human numbering shown for clarity) on each face of the nucleosome are positioned based on the crystal structure of the nucleosome in complex with the rat CENP-C motif (5). (C) Schematic view to highlight recognition and possible enrichment of CENP-A nucleosomes by the CCAN proteins, CENP-C, CENP-N, and CENP-L. Other kinetochore proteins, and the dimerization of CENP-C, have been omitted for clarity.

SYNTHETIC BIOLOGY

Multiplexed gene synthesis in emulsions for exploring protein functional landscapes

Calin Plesa,^{1*} Angus M. Sidore,^{2*} Nathan B. Lubock,¹ Di Zhang,³ Sriram Kosuri^{1,4†}

Improving our ability to construct and functionally characterize DNA sequences would broadly accelerate progress in biology. Here, we introduce DropSynth, a scalable, low-cost method to build thousands of defined gene-length constructs in a pooled (multiplexed) manner. DropSynth uses a library of barcoded beads that pull down the oligonucleotides necessary for a gene's assembly, which are then processed and assembled in water-in-oil emulsions. We used DropSynth to successfully build more than 7000 synthetic genes that encode phylogenetically diverse homologs of two essential genes in *Escherichia coli*. We tested the ability of phosphopantetheine adenyltransferase homologs to complement a knockout *E. coli* strain in multiplex, revealing core functional motifs and reasons underlying homolog incompatibility. DropSynth coupled with multiplexed functional assays allows us to rationally explore sequence-function relationships at an unprecedented scale.

The scale at which we can build and functionally characterize DNA sequences sets the pace at which we explore and engineer biology. The recent development of multiplexed functional assays allows for the facile testing of thousands to millions of sequences for a wide array of biological functions (1, 2). Currently, such assays are limited by their ability to build or access DNA sequences to test. Natural or mutagenized DNA sequences (3, 4) allow for large libraries but are not easily programmed and thus limit hypotheses, applications, and engineered designs. Alternatively, researchers can use low-cost microarray-based oligo pools that allow for large libraries of designed ~200-nucleotide (nt) sequences (5), but their short lengths limit many other applications. Gene synthesis is capable of creating long-length sequences, but high costs currently prohibit building large libraries of designed sequences (6–9).

We developed a gene synthesis method we term DropSynth: a multiplexed approach capable of building large pooled libraries of designed gene-length sequences. DropSynth uses microarray-derived oligo libraries to assemble gene libraries at vastly reduced costs. We and others have developed robust parallel processes to build genes from oligo arrays, but because each gene must be assembled individually, costs are prohibitive for large gene libraries (6, 10). In these efforts,

the ability to isolate and concentrate DNA from the background pool complexity was paramount for robust assemblies (11). Previous efforts to multiplex such assemblies have not isolated reactions from one another and thus suffered from short assembly lengths, highly biased libraries, the inability to scale, and constraints on sequence homology (12–15).

DropSynth works by pulling down only those oligos required for a particular gene's assembly onto barcoded microbeads from a complex oligo pool. By emulsifying this mixture into picoliter droplets, we isolate and concentrate the oligos before gene assembly, overcoming the critical roadblocks for proper assembly and scalability (Fig. 1A and movie S1). The microbead barcodes are distinct 12-nt sequences that all oligos for a particular assembly share, and pair with complementary strands displayed on the microbead. Within each droplet, sequences are released from the bead by using Type IIs restriction enzyme sites and assembled through polymerase cycling assembly (PCA) into full-length genes. Last, the emulsion is broken, and the gene library is recovered. To test and optimize the protocol, we built model assemblies that were different but shared common overlap sequences. As a result, any contaminating oligo would still participate in the assembly reaction, allowing us to monitor assembly specificity and library coverage. We optimized each aspect of the protocol by trying to assemble 24-, 96-, and 288-member libraries composed of 3, 4, 5, and 6 oligos at once, based on how often we saw intended targets versus their expected frequency given random (bulk) assembly (Fig. 1B). Over many iterations, we achieved high enrichment rates (~10⁸) by modifying the amount of beads, presence of size selection after assembly, ligase used for capture, and bead attachment chemistry. We ultimately found that using streptavidin bead chemistry, Taq ligase for bead capture, and size-selection after assembly

yielded the highest enrichment rates. Using these protocols, we were able to build libraries of up to six oligos that produced correctly sized bands (Fig. 1C), and the resulting assembly distributions were not overly skewed (Fig. 1D and fig. S1).

To test the scalability of DropSynth, we attempted assembly of 12,672 genes ranging in size from 381 to 669 base pairs (bp) that encode homologs of two bacterial proteins from across the tree of life (Fig. 2A and fig. S2). A total of 33 libraries of 384 genes each encoded 5775 homologs of dihydrofolate reductase (DHFR) with two different codon usages (11,520 DHFR genes), as well as 1152 homologs of the enzyme phosphopantetheine adenyltransferase (PPAT) (fig. S3, A and B). DHFR genes were assembled from either four or five 230-nt oligos, whereas PPAT genes were assembled from five 200-nt oligos. We obtained correctly sized bands for 31 of 33 assemblies, with one failing because of oligo amplification issues and the other because of low yield on the oligo processing steps, in contrast to attempts using bulk assembly that produced shorter failed by-products (fig. S3C). Three of the libraries (5x 230-nt oligomers) were too long to verify by using our barcoding approach, but the resulting synthesis showed correct band formation (fig. S4).

We cloned the libraries into an expression plasmid containing a random 20-bp barcode (assembly barcode) and sequenced the remaining 28 libraries consisting of 10,752 designs (figs. S3D, S4, and S5). For the PPAT 5x 200-nt oligo assemblies, sequencing revealed that a total of 872 genes (75%) had assemblies corresponding to a perfect amino acid sequence represented by at least one assembly barcode, with a median of two reads per assembly barcode and 56 assembly barcodes per homolog (Fig. 2B and fig. S6, A and B). This coverage increased when including sequences with deviations from the designed sequences, with 1002 genes (87%) represented within five amino acids from the designed sequences (all homologs have some alignments regardless of distance) (fig. S6D). For the DHFR 4x 230-nt oligo assemblies, we observed perfect sequences for 65% (6271) of the designed homologs, and 75% have at least one assembly within a two-amino-acid difference from design. Because there are two codon usages per homolog, when combined over homologs we observed that 3950 (79%) have at least one perfect, and 88% have at least one assembly in a distance of two amino acids (Fig. 2C). We see a strong correlation [Pearson correlation coefficient (ρ) = 0.73, P value = 3.4×10^{-5}] between the amount of DNA used to load the DropSynth beads and the resulting library coverage (fig. S7A). We also found 15 microbead barcodes that have more dropouts than would be expected by chance (fig. S7B). For constructs with at least 100 assembly barcodes, we observed a median of 1.9% (σ = 2.9%) and 3.9% (σ = 3.8%) perfect protein assemblies (Fig. 2A and figs. S6C and S8) for PPAT and DHFR libraries, respectively. The nearly double the rate of perfects for DHFR libraries compared with PPAT can be

¹Department of Chemistry and Biochemistry, University of California, Los Angeles (UCLA), Los Angeles, CA, USA.

²Department of Chemical and Biomolecular Engineering, UCLA, Los Angeles, CA, USA. ³Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁴UCLA–U.S. Department of Energy Institute for Genomics and Proteomics, Molecular Biology Institute, Quantitative and Computational Biology Institute, Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, Jonsson Comprehensive Cancer Center, UCLA, Los Angeles, CA, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: sri@ucla.edu

attributed to using longer oligos (230 versus 200 nt) that only require four oligos instead of five to assemble the gene (fig. S9A). Increasing the oligo length provides a way to assemble longer genes without large decreases in the resulting yields (fig. S9B). Furthermore, the distribution of perfect assemblies in the PPAT libraries is not overly skewed (fig. S6D), and most library members have assemblies with high identity to their

respective designed homologs (fig. S6F). The resultant error profiles were consistent with Taq-derived mismatch and assembly errors that we have observed previously (fig. S10) (16).

We sought to show how DropSynth-assembled libraries could be easily coupled as inputs into multiplex functional assays by probing how well the PPAT homologs of various evolutionary distance to *Escherichia coli* could rescue a knockout

phenotype. PPAT is an essential enzyme, encoded by the gene *coaD*, which catalyzes the second-to-last step in the biosynthesis of coenzyme A (CoA) (fig. S11) (17) and is an attractive target for the development of novel antibiotics (18). Assembled PPAT variants on the barcoded expression plasmid were transformed into *E. coli* $\Delta coaD$ cells and screened for complementation by growing the library in batch culture through three serial 1000-fold

Fig. 1. DropSynth assembly and optimization.

(A) We amplified array-derived oligos and exposed a single-stranded region that acts as a gene-specific microbead barcode. Barcoded beads display complementary single-stranded regions that selectively pull down the oligos necessary to assemble each gene. The beads are then emulsified, and the oligos are assembled by means of PCA. The emulsion is then broken, and the resultant assembled genes are barcoded and cloned. (B) We used a model gene library that allowed us to monitor the level of specificity and coverage of the assembly process. We then optimized various aspects of the protocol—including purification steps, DNA ligase, and bead couplings—in order to improve the specificity of the assembly reaction. Enrichment is defined as the number of specific assemblies observed relative to what would be observed by random chance in a full combinatorial assembly. (C) We attempted 96-plex gene assemblies with three, four, five, or six oligos, and the resultant libraries displayed the correct-sized band on an agarose gel. (D) The distribution of read counts for all 96 assemblies (four-oligo assembly) as determined with NGS.

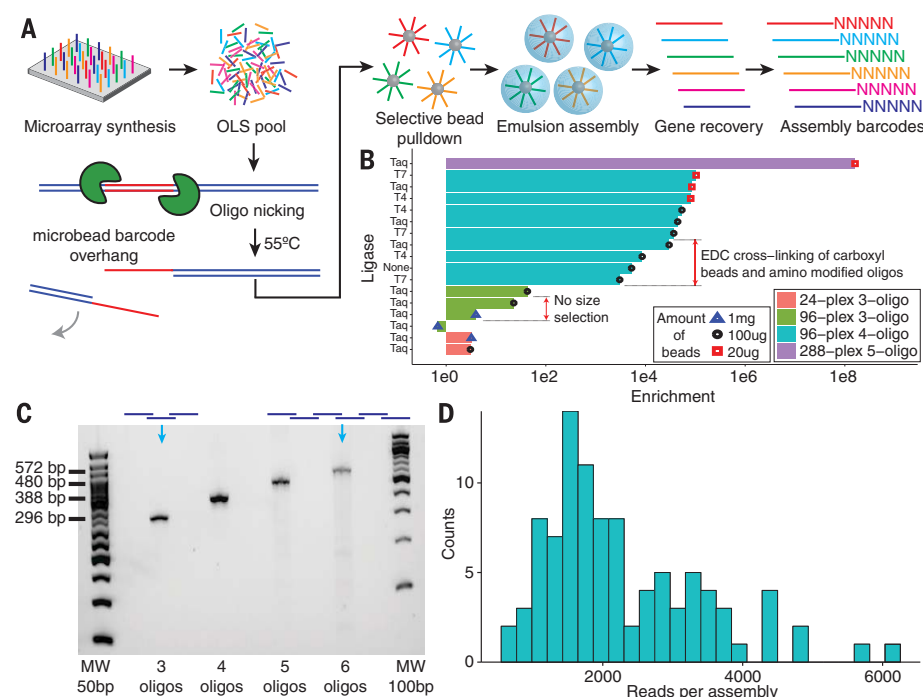
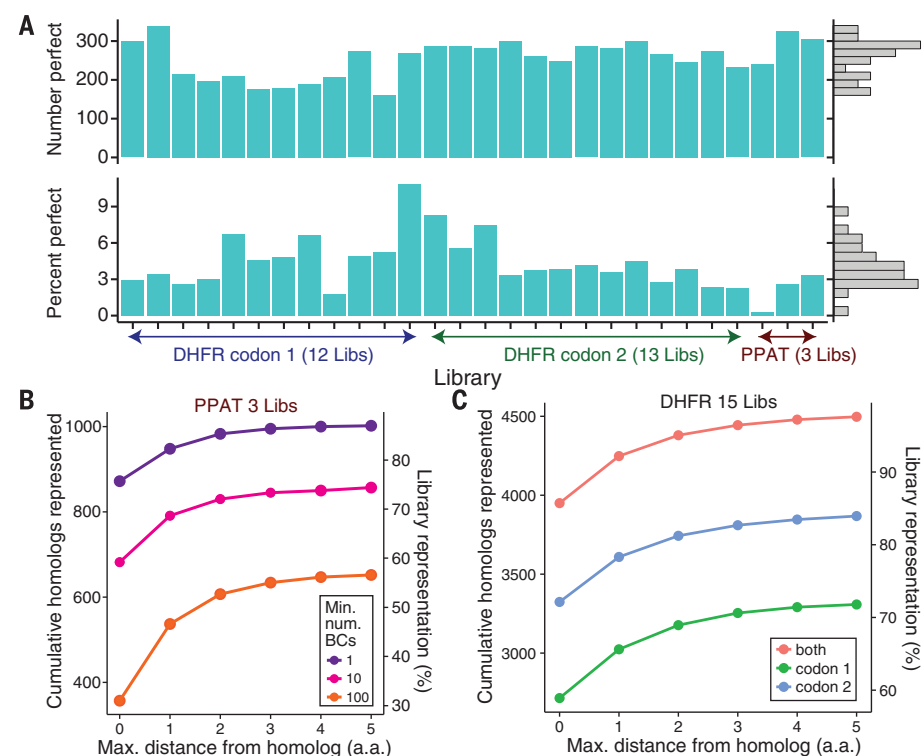


Fig. 2. DropSynth assembly of 10,752 genes.

(A) We used DropSynth to assemble 28 libraries of 10,752 genes representing 1152 homologs of PPAT and 4992 homologs of DHFR. The number of library members with at least one perfect assembly and the median percent perfects determined by using constructs with at least 100 barcodes is shown for each library. (B) We observed that 872 PPAT homologs (75%) had at least one perfect assembly, and 1002 homologs (87%) had at least one assembly within a distance of five amino acids from design. (C) We assembled two codon variants for each designed DHFR homolog, allowing us to achieve higher coverage.



dilutions (Fig. 3A and table S1), while a rescue plasmid was simultaneously heat-cured (fig. S12). Assembly barcode sequencing of the resulting populations provided a reproducible estimate for the fitness of all homologs successfully assembled without error (biological replicates $\rho = 0.94$; Pearson, $P < 2.2 \times 10^{-16}$) (figs. S13A and S14A). Individual barcodes can display considerable noise, so having many assembly barcodes per construct improved confidence (fig. S14, B and C). Negative controls and sequences containing indels show strong depletion (figs. S13A, S15A, and S16), and fitness is reduced with increasing numbers of mutations ($\rho = -0.38$; Spearman, $P < 2.2 \times 10^{-16}$) (fig. S15, B and C). Pooled fitness scores also correlated well with measured growth rates of individually tested controls [Spearman correlation coefficient (r_s) = 0.86, $P = 5.9 \times 10^{-12}$] (fig. S17). Approximately 14% percent of the homologs show strong depletion (fitness below -2.5), whereas 70% have a positive fitness value in the pooled assay. Low-fitness homologs are evenly distributed throughout the phylogenetic tree, with only minor clustering of clades (Fig. 3B and figs. S13B, S18, and S19A). There are several reasons homologs could have low fitness, including environmental mismatches, improper folding, mismatched metabolic flux, interactions with other cytosolic components, or gene dosage toxicity effects resulting from improperly high expression (supplementary text) (19).

Errors during the oligo synthesis or DropSynth assembly give us mutational data across all the

homologs, which we can further analyze to better understand function. We selected all 497 homologs that showed some degree of complementation (fitness greater than -1) as well as their 71,061 mapped mutants within a distance of five amino acids and carried out a multiple sequence alignment in order to find equivalent residue positions. For each amino acid and position, we found the median fitness among all of these homologs and mutants. The resulting data was projected onto the *E. coli* PPAT sequence (Fig. 4, A and B), providing data similar to deep mutational scanning approaches (20, 21). We term this approach broad mutational scanning (BMS). The average BMS fitness for each position shows strong constraints in the catalytic site, at highly conserved sites ($\rho = -0.64$; Pearson, $P < 2.2 \times 10^{-16}$), and at buried residues compared with solvent-accessible ones ($\rho = 0.42$; Pearson, $P = 3.9 \times 10^{-8}$) (fig. S20, A and B, and supplementary text). Surprisingly, some residues that are known to interact with either adenosine 5'-triphosphate (ATP) or 4'-phosphopantetheine turn out to be relatively promiscuous when averaged over a large number of homologs. Furthermore, when mapped onto the *E. coli* structure (Fig. 4B), positions known to be involved with allosteric regulation by CoA or dimer formation show relatively little constraint, highlighting the diversity of distinct approaches used among different homologs while maintaining the same core function. We implemented a simple binary classifier to predict the sign of the BMS fitness

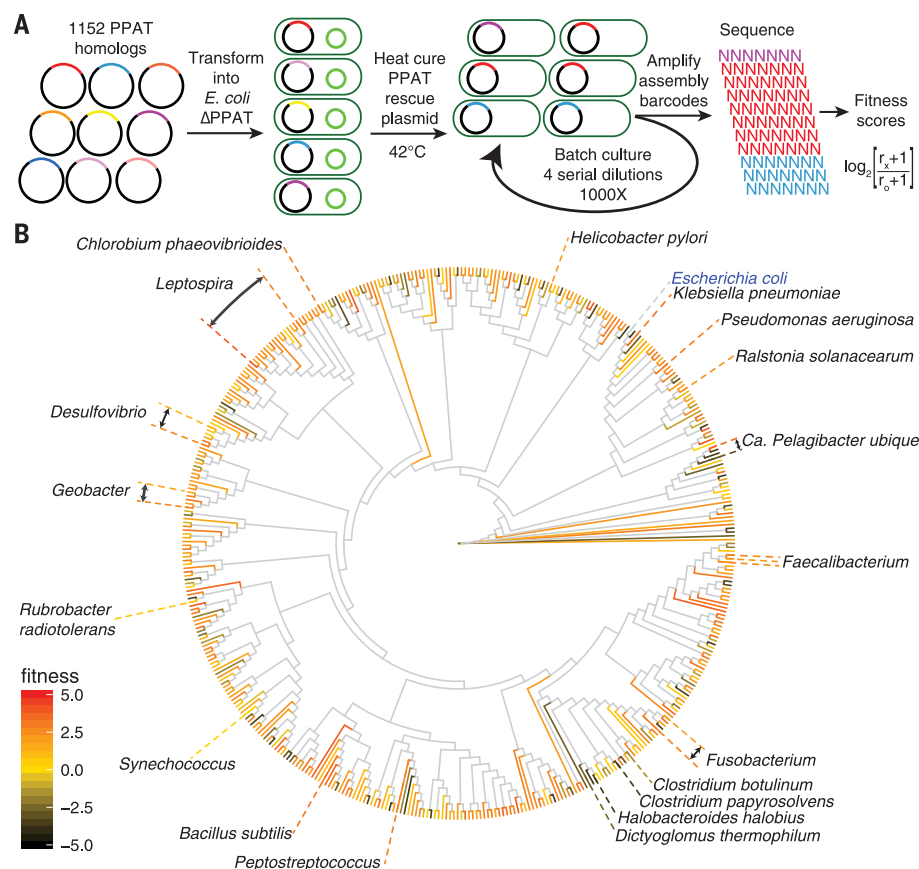
value on the basis of a number of features, achieving an accuracy of 0.825 (fig. S21).

Additionally, we can search for gain-of-function (GOF) mutations among those homologs that did not complement. A total of 385 GOF mutants out of 4658 were found for 55 homologs out of 129 low-fitness homologs (fitness < -2.5). By aligning these mutations to the *E. coli* sequence, the eight statistically significant residues (34, 35, 64, 68, 69, 103, 134, and 135) shown in Fig. 4C localize to four small regions in the protein structure (fig. S22 and supplementary text). We retrieved six GOF mutants of six different homologs from the library, each with fitness determined from only a single assembly barcode, and individually tested their growth rates. Five of the six mutants showed strong growth, and one failed to complement (fig. S17B). We also tested two of the corresponding low-fitness homologs, finding increases in the growth rate of 10 and 42% for their GOF mutants (table S2).

Broad mutational scanning enabled by DropSynth is a useful tool with which to explore protein functional landscapes. By analyzing many highly divergent homologs, individual steric clashes, which might be important to a particular sequence, become averaged across the homologs. More broadly, DropSynth allows for building large designed libraries of gene-length sequences, with no specialized equipment and estimated total costs below \$2 per gene (tables S3 and S4). We also show that DropSynth can be combined with dial-out polymerase chain reaction (15), which

Fig. 3. PPAT complementation assay.

(A) We used DropSynth to assemble a library of 1152 homologs of PPAT, an essential enzyme catalyzing the second-to-last step in CoA biosynthesis, and functionally characterized them using a pooled complementation assay. The barcoded library was transformed into *E. coli* $\Delta coaD$ cells containing a curable rescue plasmid expressing *E. coli* *coaD*. The rescue plasmid was removed, allowing the homologs and their mutants to compete with each other in batch culture. We tracked assembly barcode frequencies over four serial 1000-fold dilutions and used the frequency changes to assign a fitness score. (B) This phylogenetic tree shows 451 homologs each with at least five assembly barcodes, a subset of the full data set, in which leaves are colored by fitness. Despite having a median 50% sequence identity, we found that the majority of PPAT homologs are able to complement the function of the native *E. coli* PPAT, with 70% having positive fitness values, whereas low-fitness homologs are dispersed throughout the tree, without much clustering of clades.



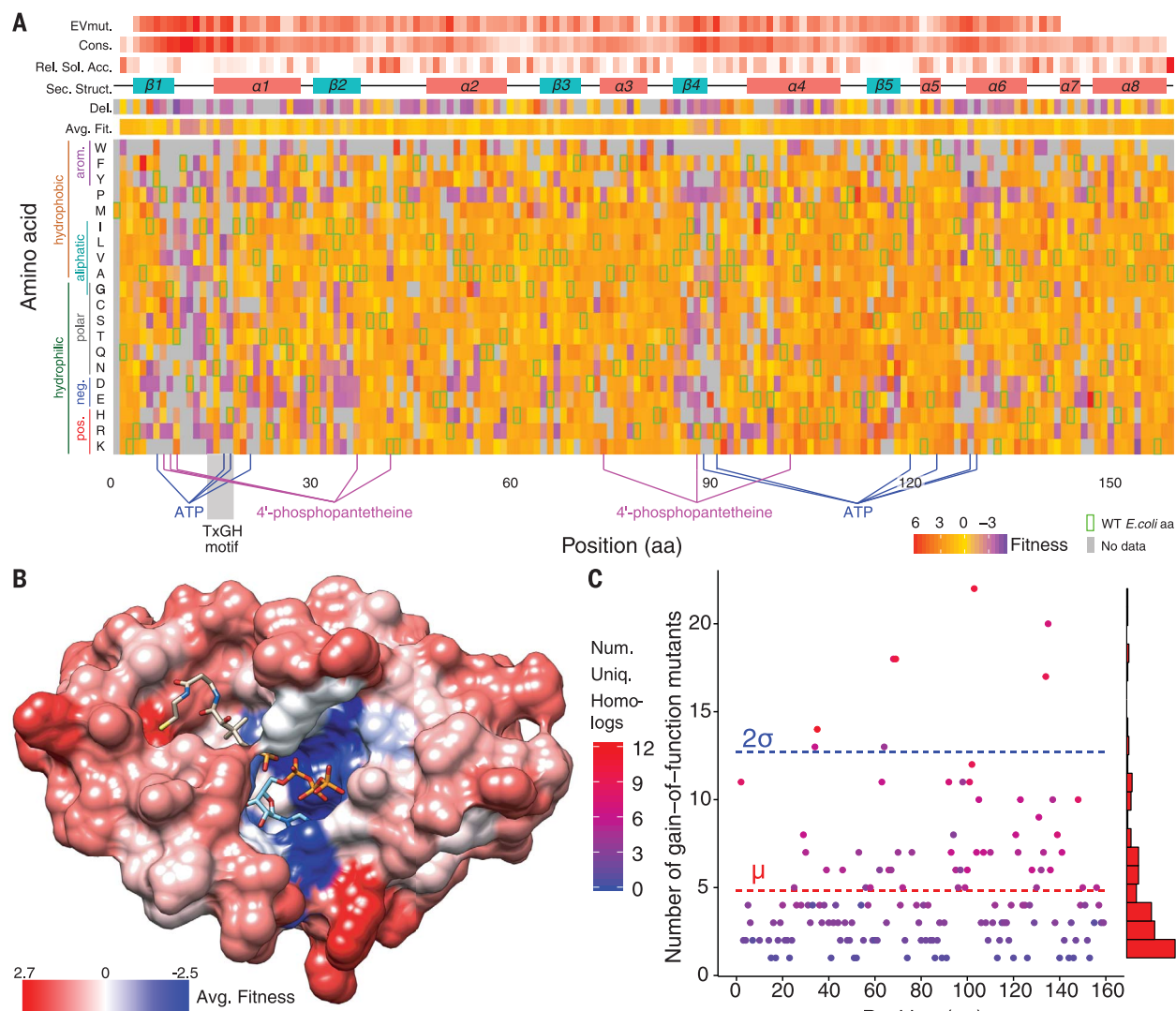


Fig. 4. Broad mutational scanning analysis. (A) The fitness landscape of 497 complementing PPAAT homologs and their 71,061 mutants (within a distance of five amino acids) is projected onto the *E. coli* PPAAT sequence, with each point in the heatmap showing the average fitness over all sequences containing that amino acid at each aligned position. Mutations are highly constrained at a core group of residues involved in catalytic function. Other positions show relatively little loss of function, when averaged over many homologs, despite known interactions with the substrates. The *E. coli* wild-type (WT) sequence is indicated by green squares, and the average position fitness, fitness of a residue deletion, mean EVmutation evolutionary statistical energy (22), site conservation, relative solvent accessibility, and secondary structure information is shown above. (B) The average fitness at each position, with blue

and red representing low and high fitness, respectively, overlaid on the *E. coli* PPAAT [Protein Data Bank 1QJC and 1GN8 (23)] structure complexed with 4'-phosphopantetheine and ATP. We observed loss of function for mutations occurring at the active site, whereas other residues involved with allosteric regulation by CoA or dimer interfaces show large promiscuity, highlighting different strategies used among homologs. (C) In addition to complementing homologs, we can also analyze mutants of the 129 low-fitness (<2.5) homologs, finding 385 GOF mutants across 55 homologs. We project this data onto the *E. coli* PPAAT sequence and plot the number of GOF mutants at each position, shaded by the number of different homologs represented. We found a total of eight statistically significant positions (residues 34, 35, 64, 68, 69, 103, 134, and 135) corresponding to four regions in the PPAAT structure.

could be expanded for gene synthesis applications for which perfect sequences are paramount. The scale, quality, and cost of DropSynth libraries can likely be improved further with investment in algorithm design, better polymerases, and larger barcoded bead libraries.

REFERENCES AND NOTES

1. F. Inoue, N. Ahituv, *Genomics* **106**, 159–164 (2015).
2. M. Gasperini, L. Starita, J. Shendure, *Nat. Protoc.* **11**, 1782–1787 (2016).
3. K. S. Sarkisyan *et al.*, *Nature* **533**, 397–401 (2016).
4. D. M. Fowler, S. Fields, *Nat. Methods* **11**, 801–807 (2014).
5. G. J. Rocklin *et al.*, *Science* **357**, 168–175 (2017).
6. S. Kosuri, G. M. Church, *Nat. Methods* **11**, 499–507 (2014).
7. S. Ma, N. Tang, J. Tian, *Curr. Opin. Chem. Biol.* **16**, 260–267 (2012).
8. J. Quan *et al.*, *Nat. Biotechnol.* **29**, 449–452 (2011).
9. R. A. Hughes, A. D. Ellington, *Cold Spring Harb. Perspect. Biol.* **9**, a023812 (2017).
10. S. Kosuri *et al.*, *Nat. Biotechnol.* **28**, 1295–1299 (2010).
11. A. Y. Borovkov *et al.*, *Nucleic Acids Res.* **38**, e180 (2010).
12. J. C. Klein *et al.*, *Nucleic Acids Res.* **44**, e43 (2016).
13. H. Kim *et al.*, *Nucleic Acids Res.* **40**, e140 (2012).
14. T. H.-C. Hsiao *et al.*, *PLOS ONE* **10**, e0119927 (2015).
15. J. J. Schwartz, C. Lee, J. Shendure, *Nat. Methods* **9**, 913–915 (2012).
16. N. B. Lubock, D. Zhang, A. M. Sidore, G. M. Church, S. Kosuri, *Nucleic Acids Res.* **45**, 9206–9217 (2017).
17. T. Izard, A. Geerlof, *EMBO J.* **18**, 2021–2030 (1999).
18. B. L. M. de Jonge *et al.*, *Antimicrob. Agents Chemother.* **57**, 6005–6015 (2013).
19. S. Bhattacharyya *et al.*, *Eng. Life Sci.* **5**, e20309 (2016).

20. D. S. Marks, T. A. Hopf, C. Sander, *Nat. Biotechnol.* **30**, 1072–1080 (2012).
21. N. Halabi, O. Rivoire, S. Leibler, R. Ranganathan, *Cell* **138**, 774–786 (2009).
22. T. A. Hopf et al., *Nat. Biotechnol.* **35**, 128–135 (2017).
23. T. Izard, *J. Mol. Biol.* **315**, 487–495 (2002).

ACKNOWLEDGMENTS

This work was supported by the funds from the Human Frontier Science Program (LT000068/2016 to C.P.), Netherlands Organisation for Scientific Research Rubicon fellowship (to C.P.), National Science Foundation Graduate Research Fellowship under grant 2016211460 (to

A.M.S.), a Ruth L. Kirschstein National Research Service Award (GM007185 to N.L.), National Institutes of Health New Innovator Award (DP2GM114829 to S.K.), Searle Scholars Program (to S.K.), U.S. Department of Energy (DE-FC02-02ER63421 to S.K.), UCLA, and L. Wudl and F. Wudl. We thank J. Sampson and P. Anderson at Agilent Technologies for oligo pools and critical advice. We thank G. Church and R. Terry for guidance during the early developments and S. Feng, the UCLA Broad Stem Cell Research Center Sequencing Core, and the Technology Center for Genomics and Bioinformatics for providing next-generation sequencing (NGS) services. S.K. and D.Z. are named inventors on a patent application on the DropSynth method (US14460496). The scripts required to generate DropSynth oligos are available at <https://github.com/kosurilab/DropSynth>. Sequencing

data are available from the sequencing read archive (SRA) with the accession no. SRP126669.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6373/343/suppl/DC1
Materials and Methods
Figs. S1 to S25
Tables S1 to S14
References (24–48)
Movie S1

28 July 2017; accepted 18 December 2017
10.1126/science.aao5167



Check out our exciting research Prize!



\$25, 000 Grand Prize!
Get published in *Science*!

The Prize is a new highly competitive international prize which honors scientists for their excellent contributions to neuromodulation research. For purposes of the Prize, neuromodulation is any form of alteration of nerve activity through the delivery of physical (electrical, magnetic, optical) stimulation to targeted sites of the nervous system.”

For full details, judging criteria and eligibility requirements, visit:

www.sciencemag.org/prizes/pins

Submission Deadline: March 15, 2018

Science
AAAS



PINS 品驰



2018

AAAS MARTIN AND
ROSE WACHTEL
CANCER RESEARCH

AWARD

Recognize the work of an early career scientist who has performed outstanding work in the field of cancer research. Award nominees must have received their Ph.D. or M.D. within the last 10 years. The winner will deliver a public lecture on his or her research, receive a cash award of **\$25,000**, and publish a Focus article in *Science Translational Medicine*.

For more information visit
www.aaas.org/aboutaaas/awards/wachtel
or e-mail wachtelprize@aaas.org.
Deadline for submission: **February 1, 2018.**

Science Translational Medicine |  AAAS

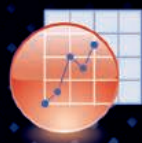
Submit Your Research for
Publication in *Science Robotics*

ScienceRobotics.org

ScienceRobotics



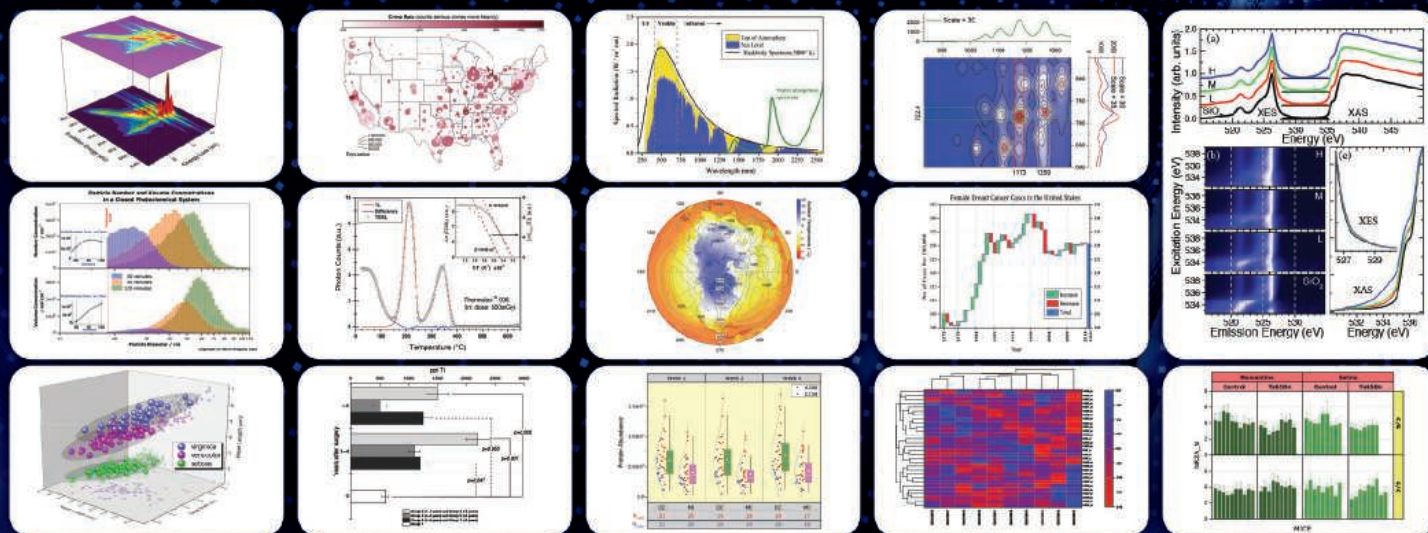
Send pre-submission inquiries
and expressions of interest to
sciroboteditors@aaas.org.



ORIGIN[®] 2018

Graphing & Analysis

New Version!



Over 75 New Features &
Apps in Origin 2018!

Over 500,000 registered users worldwide in:

- 6,000+ Companies including 20+ Fortune Global 500
- 6,500+ Colleges & Universities
- 3,000+ Government Agencies & Research Labs

For a FREE 60-day
evaluation, go to
OriginLab.com/demo
and enter code: 7564

OriginLab[®]

25+ years serving the scientific & engineering community



Confocal Laser Scanning Microscopes

The Olympus FLUOVIEW FV3000 and FV3000RS confocal laser scanning microscopes combine high-performance imaging capabilities with ease of use, so researchers can collect publication-quality imaging

data quickly and efficiently. Available as a hybrid system, the FV3000RS is equipped with both a galvanometer scanner and an extremely accurate resonant scanner that can capture dynamic physiological events at up to 438 frames per second. TruSpectral technology is featured on every detector, providing the flexibility of spectral imaging while maintaining efficient light transmission for excellent sensitivity. Built for fast, stable, and accurate measurements of biological reactions within living cells or tissues, these microscopes allow even novice users to generate superior data and images.

Olympus

For info: 704-877-8801

www.olympus-lifescience.com/en/laser-scanning/fv3000

Cell Analyzer

With the Muse Cell Analyzer, you can now achieve highly quantitative results at a fraction of the price, effort, and time. It is a compact [footprint of only 8 in. x 10 in. (20 cm x 25 cm)], easy-to-use benchtop device, making flow cytometry accessible to anyone, anytime. A user-friendly, integrated touchscreen interface, intuitive software for data acquisition and analysis, and optimized Muse assays help to simplify your research. The Muse's microcapillary flow cell is engineered for acquisition of both suspension and adherent cells of 2 μm –60 μm in diameter. It uses fluorescent reagents and detection to measure three parameters for every cell, with little or no sample preparation required. Muse assays are available for precision cell counts as well as single-cell measurement of critical cell parameters, including viability, apoptosis, autophagy, oxidative stress, and cell signaling.

EMD Millipore

For info: 800-645-5476

www.emdmillipore.com

Focused Ion Beam Scanning Electron Microscope

The ZEISS Crossbeam 550 is a focused ion beam scanning electron microscope (FIB-SEM) that features an increase in resolution for imaging and material characterization and a speed gain in sample preparation. Nanostructures such as composites, metals, biomaterials, or semiconductors can be investigated with analytical and imaging methods. The Crossbeam 550 allows simultaneous modification and monitoring of samples, resulting in fast sample preparation and high throughput (e.g., for cross-sectioning, transmission electron microscope lamella preparation, or nanopatterning). It provides quality 2D and 3D images. The Tandem decel mode enables enhanced resolution along with maximization of image contrast at low landing energies. Gemini II electron optics deliver optimum resolution at low voltage and high probe current. The FIB column

combines the highest available FIB current of 100 nA with the FastMill mode, allowing for precise, efficient material processing and imaging. Automated emission recovery increases convenience and optimizes the FIB column for reproducible results during long-term experiments.

ZEISS

For info: 858-790-7700

www.zeiss.com/crossbeam

Proteomics Reagents

Get higher throughput, greater robustness, and more convenience with an Agilent Jet Stream proteomics solution. Based on Agilent's revolutionary Jet Stream source, Agilent Jet Stream proteomics solutions achieve near-nanoflow analytical sensitivity using 2.1-mm columns and standard-flow chromatography. And when sample size is limited and the highest sensitivity is required, the new Nanodaptor solution converts conventional high-performance liquid chromatography to a nanoflow system. Agilent's comprehensive proteomics portfolio of hardware, software, automated sample preparation, and columns is designed to meet your application needs.

Agilent Technologies

For info: 877-424-4536

www.agilent.com

AAV Biosensors

AMS Biotechnology's adeno-associated virus (AAV) biosensor products come as ready-to-use AAV viruses. The viruses encode your chosen biosensor, either calcium or glutamate, and are ready for in vivo injection. Biosensors are genetically engineered fluorescent proteins attached to an additional protein sequence that makes them sensitive to small biomolecules (e.g., Ca^{2+}) or other intracellular processes. These biosensors are introduced to cells, tissues, or organisms to detect changes through fluorescence microscopy. Many biosensors permit long-term imaging and can be engineered to specifically target cellular compartments or organelles. Additionally, biosensors permit signaling pathway exploration or allow the measurement of a biomolecule—they do all this while preserving both spatial and temporal cellular processes.

AMS Biotechnology

For info: +44-(0)-1235-828200

www.amsbio.com/cellular-metabolism.aspx

Automated Microscope

The Lionheart FX Automated Microscope is a compact, inclusive microscopy system for a broad range of imaging workflows. It offers up to 100X air- and oil-immersion magnification, with fluorescence, brightfield, color-brightfield, and phase-contrast channels for maximum application reach. An optional environmental-control cover provides incubation to 40°C and effective containment for CO_2/O_2 control; a humidity chamber optimizes conditions for long-term, live-cell imaging applications; and an available dual reagent injector facilitates rapid kinetic assays. Automated image preprocessing optimizes images for downstream analysis, from cell counting to characterization of subcellular details. The filter/LED cubes and objectives are accessed from the front panel, and tool-free connections for injectors, gas control, and the environmental cover add to the easy installation, which takes only 15 minutes. Its small size means that minimal benchtop space is used.

BioTek

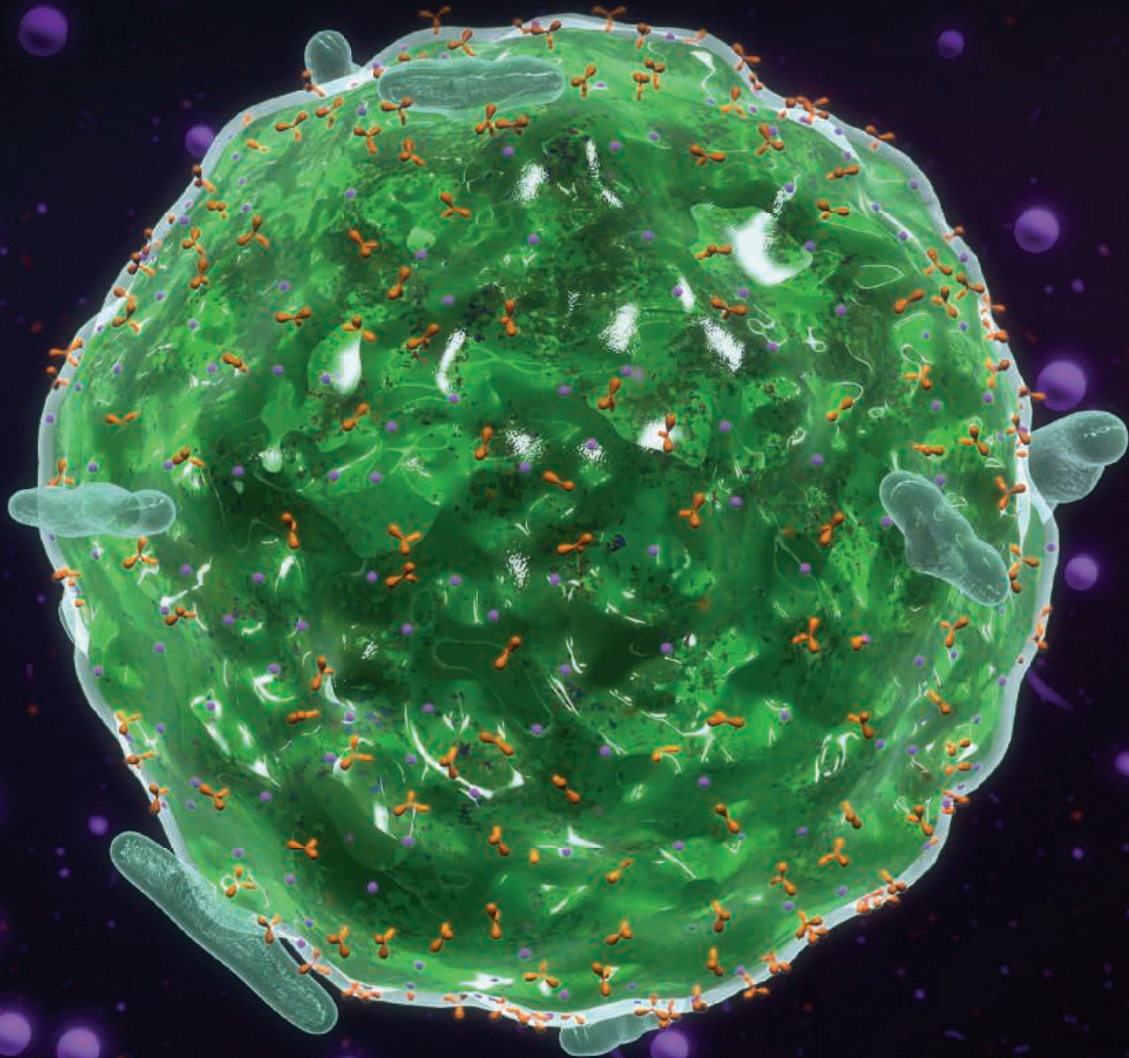
For info: 888-451-5171

www.biotek.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

Submit your high-impact research
to ***Science Immunology***



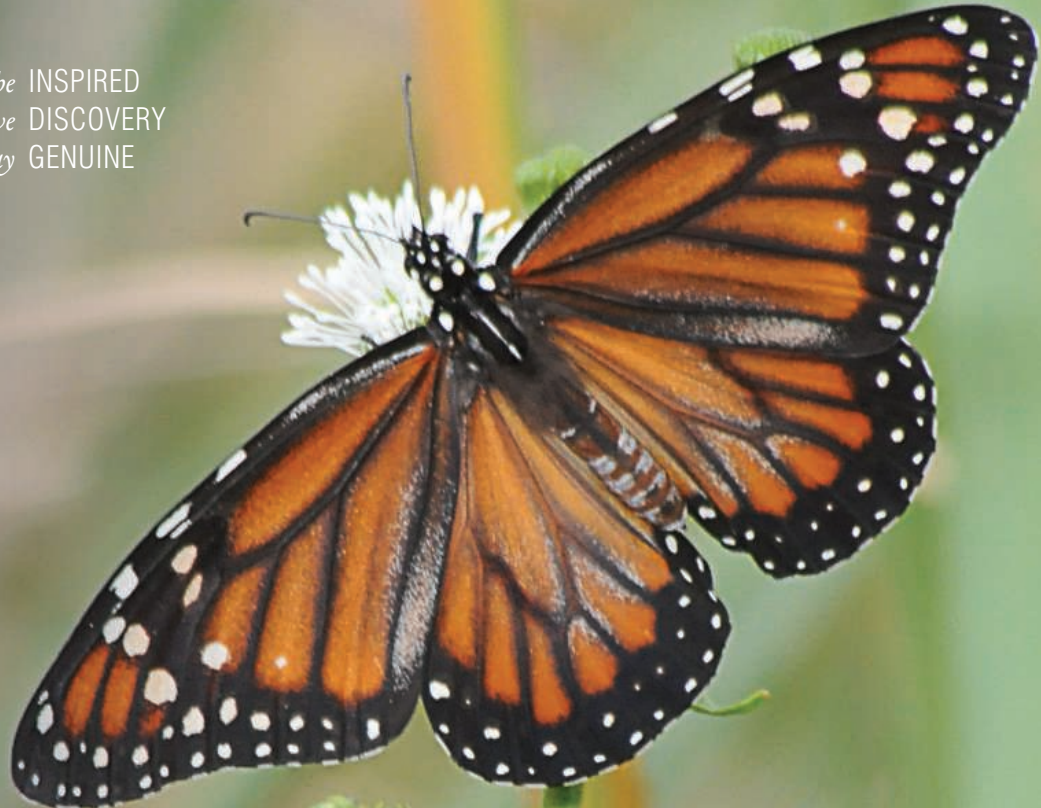
Science Immunology publishes original, peer-reviewed, science-based research articles that report critical advances in all areas of immunological research, including important new tools and techniques. Share your research with *Science Immunology* global readership and submit your manuscript today!

What will your discovery be?

Submit your manuscript today at
ScienceImmunology.org

ScienceImmunology

 AAAS



We've spread our wings.

Monarch[®] Nucleic Acid Purification Kits Now available for DNA & RNA

Designed with sustainability in mind, Monarch[®] Nucleic Acid Purification Kits are the perfect complement to many molecular biology workflows. Available for DNA & RNA purification, with buffers and columns available separately, Monarch kits are optimized for excellent performance, convenience and value. Quickly and easily recover highly pure, intact DNA and RNA in minutes. Available kits include:

- Monarch Plasmid Miniprep Kit
- Monarch DNA Gel Extraction Kit
- Monarch PCR & DNA Cleanup Kit (5 µg)
- **NEW** MONARCH TOTAL RNA MINIPREP KIT – optimized for use with a variety of sample types, including cells, tissues, blood, and more

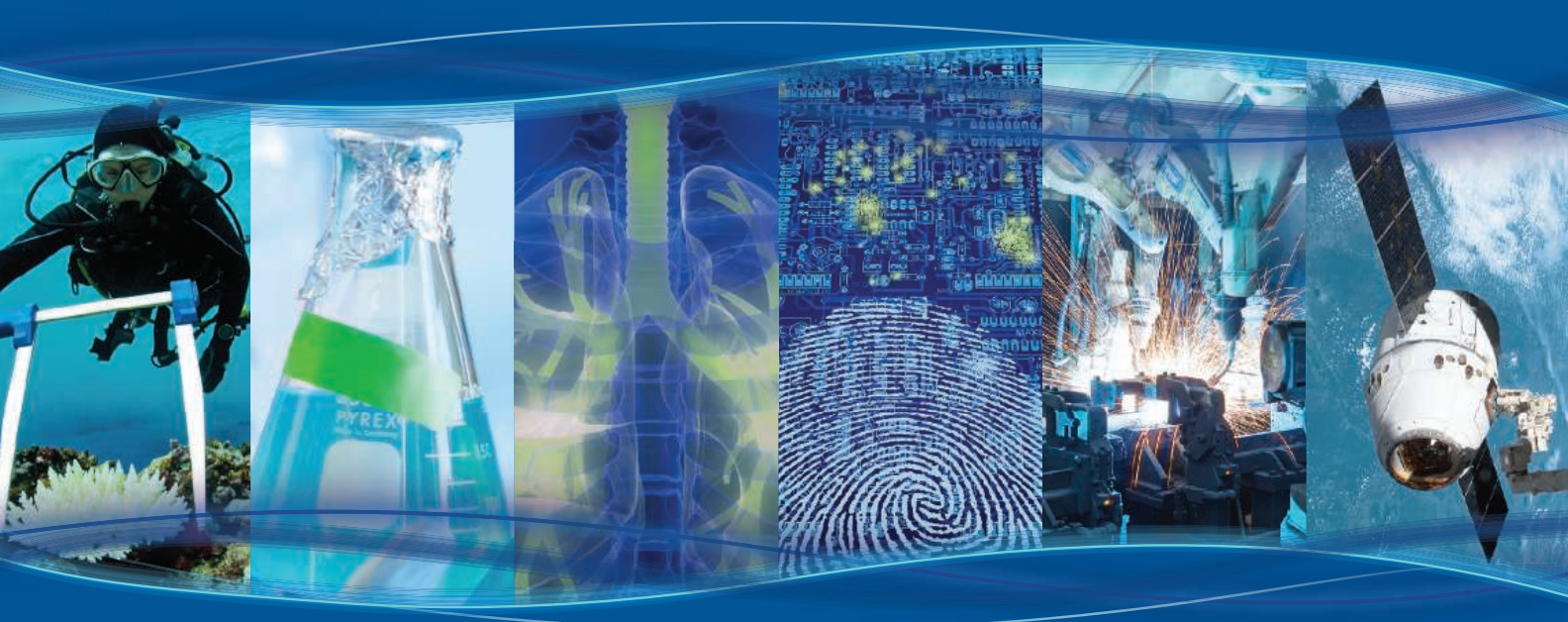
Make the change and migrate to Monarch today.

Interested in trying a sample
of our new Monarch Total
RNA Miniprep Kit?



Learn more at

NEBMonarch.com



ANNUAL MEETING
Feb 15–19 2018 ★ Austin, TX

ADVANCING SCIENCE | DISCOVERY TO APPLICATION

Advance registration available until January 24

The 2018 meeting theme highlights the critical roles of academia, government, and industry in moving ideas into innovations.

aaas.org/meetings

See Inside for Details:

President's Address / Registration Rates

Plenary Lectures / Topical Lectures

Seminars / Session Tracks / Flash Talk Program

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Join us in Austin

Learn about the critical roles of academia, government, and industry in moving ideas into innovations.

- Seminars on technology and innovation; the future of artificial intelligence; diversity and inclusion; and communicating science
- 120+ scientific sessions in 14 disciplinary tracks covering the latest research advances
- 5 flash talk sessions: dynamic presentations and discussion
- Network with colleagues and attend career development workshops

Connect with us

 @AAASmeetings #AAASmtg

 facebook.com/AAAS.Science

aaas.org/meetings

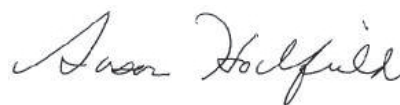
Reporters: The AAAS Annual Meeting Newsroom will be hosted on EurekAlert! at eurekalert.org/aaasnewsroom

Dear Colleague:

In these changing times, it is critical that academia, government, and industry continue to work together to move ideas into innovative advancements. This is why the theme of the 2018 AAAS Annual Meeting is *Advancing Science: Discovery to Application*.

On behalf of the AAAS Board of Directors, I urge you to join us in Austin February 15-19, where this theme will be explored through interdisciplinary scientific sessions, renowned speakers, and one-on-one discussions. The AAAS Annual Meeting is the most widely reported global science gathering and the premier event at which you can network with future collaborators across disciplines.

We look forward to seeing you in Austin. Registration and housing details are now available online.



Susan Hockfield

AAAS President

President Emerita and Professor of Neuroscience,
Massachusetts Institute of Technology

PRESIDENT'S ADDRESS



Susan Hockfield
Thursday, February 15

Dr. Susan Hockfield served as president of the Massachusetts Institute of Technology from 2004 to 2012. As the first woman and the first biologist in that role, she highlighted the importance of building diversity all along the talent pipeline. She fostered cross-disciplinary, cross-institutional, and cross-national initiatives, among them the Koch Institute for Integrative Cancer Research, the MIT Energy Initiative, and the Massachusetts Green High-Performance Computing Center, and she co-chaired the White House's Advanced Manufacturing Partnership. By expanding MIT's international education and research activities, including the launch of edX, she amplified MIT's global engagement. Hockfield avidly advocates increasing interactions across the academy, industry, and government.

Hockfield earned her Ph.D. in anatomy and neuroscience from the Georgetown University School of Medicine. She

was a National Institutes of Health postdoctoral fellow at the University of California, San Francisco, and a member of the scientific staff at the Cold Spring Harbor Laboratory in New York before joining the faculty at Yale University in 1985. At Yale, Hockfield was named the William Edward Gilbert Professor of Neurobiology and served as dean of the Yale University Graduate School of Arts and Sciences and then as provost of the university.

Hockfield was among the first scientists to apply molecular biology to neuroscience, using monoclonal antibodies to study brain structure and development. She demonstrated that early experience leads to lasting changes in the molecular structure of the brain and discovered a gene involved in the spread of brain cancer cells into healthy brain tissue.

Dr. Hockfield became a member of AAAS in 1975, was elected as a Fellow in 2005, and currently serves as president of AAAS.



PLENARY LECTURES

**Ellen Ochoa**

Director,
Johnson Space Center,
National Aeronautics
and Space
Administration,
Houston, TX

***The International
Space Station:
A Laboratory in Space***

Friday, February 16

**Cori Bargmann**

President,
Chan Zuckerberg
Science, Palo Alto, CA

***The Chan Zuckerberg
Initiative: Accelerating
Science***

Saturday, February 17

**Katherine Hayhoe**

Professor,
Texas Tech University,
Lubbock

***When Facts Are
Not Enough***

Sunday, February 18

TOPICAL LECTURES

Friday, February 16

Jason De León

University of Michigan, Ann Arbor

Nina Kraus

Northwestern University, Evanston, IL

Nora Volkow

National Institute on Drug Abuse, National Institutes of
Health, Bethesda, MD

Saturday, February 17

Thomas Maina Kariuki

African Academy of Sciences, Nairobi, Kenya

Jed S. Rakoff

U.S. District Court for the Southern District of New York,
New York City

SARTON MEMORIAL LECTURE IN THE
HISTORY AND PHILOSOPHY OF SCIENCE

Bruce J. Hunt

University of Texas, Austin

Sunday, February 18

James P. Allison

University of Texas M.D. Anderson Cancer Center, Houston

Meg Urry

Yale University, New Haven

JOHN P. MCGOVERN AWARD LECTURE
IN THE BEHAVIORAL SCIENCES

Robert A. Bjork

University of California, Los Angeles

Updated program information will
be posted on aaas.org/meetings
as it becomes available.

SEMINARS

Communicating Science

Science and technology are integral to modern life, and many critical decisions facing society require finding common ground between scientists and members of the public. This annual seminar focuses on different aspects and approaches to communicating science, always emphasizing both theory and practice. The sessions provide a forum for scientists, science communication and public engagement professionals, and social scientists whose research can inform best practices to share their expertise and learn from one another. Participants gain actionable knowledge and join a growing community focused on public engagement with science.

Organized by Emily Cloyd and Elana Kimbrell, AAAS Center for Public Engagement with Science, Washington, DC

Reaching Beyond the Science-Interested Public

Developing a Narrative About Your Data

Advocating for Public Engagement With Science

Ideas to Innovation

This seminar brings together a diverse group of sessions around the 2018 AAAS Annual Meeting theme, “Advancing Science: Discovery to Application.” Each of these panels highlights different ways scientific research translates into applications that serve society, and how the various sectors within the scientific enterprise collaborate to achieve this goal. One session focuses on materials research and manufacturing, sharing examples of how companies choose what to invest in, and how partnerships with academia and government have contributed to new technologies. Another session explores the stages of technology development and transfer across several areas of chemistry, describing impacts in the fields of energy, transportation, and public health. A third panel traces the path from laboratory research to treatments for neurodegenerative diseases, and a fourth discusses collaboration in national security and defense between government, academia, and private industry, which leads not only to technological advancement but also to improvements in how social science is applied to public policymaking.

Materials Research for Manufacturing: Lessons From Industry on New Product Development

Organized by Lynnette D. Madsen, Svedberg Science Inc., Falls Church, VA

Technological Applications of Chemistry: Stages of Development and Societal Impacts

Organized by Jonathan Sessler, University of Texas, Austin

From New Discoveries to New Treatments for Neurodegenerative Diseases

Organized by Benjamin Wolozin, Boston University School of Medicine, MA

The Role of National Security in Strengthening the Science and Technology Pipeline

Organized by Taeyjuana Lyons, U.S. Department of Defense, Alexandria, VA

The Future of Artificial Intelligence

Artificial intelligence has potential applications and implications across nearly all areas of life and science. Convening three leaders in the field, one session addresses landscape-level questions of how artificial intelligence research and development has grown to where it is today and how it will meet future challenges to progress. One session demonstrates how artificial intelligence can improve water resource management decisions, and another addresses how to develop artificial intelligence technologies in socially responsible ways. A fourth session focuses on areas where humans will work with machines rather than be replaced by them, exploring how the unique strengths of each can be leveraged to maximize effectiveness.

Advancing Artificial Intelligence: From the Lab to the Street

Organized by Henry Kautz, University of Rochester, NY

Finding Water Management Solutions With Artificial Intelligence

Organized by Suzanne A. Pierce, Texas Advanced Computing Center, Austin; Yolanda Gil, University of Southern California, Marina del Rey

The Fourth Industrial Revolution: Supporting Societal Needs With Artificial Intelligence

Organized by Claire Craig, The Royal Society, London, United Kingdom

Artificial Intelligence Augmenting Not Replacing People

Organized by Ann Drobni, Computing Community Consortium, Washington, DC; Gregory Hager, Johns Hopkins University, Baltimore, MD

2018 FLASH TALK PROGRAM

Diversity and Inclusion

Broadening access to science and scientific careers, and supporting equity in treatment as well as opportunity are critical to individual civil rights, although they also benefit science and society at large. These efforts include both outreach and recruitment of historically marginalized groups (such as people with disabilities, underrepresented minorities, women, and those identifying as LGBTQ+), as well as changes to the culture and structure of scientific institutions and endeavors to ensure these individuals are accepted, able to thrive, and are reflected in how science is viewed and conducted. This seminar convenes sessions discussing various aspects of STEM diversity and inclusion to share projects and data and gather additional perspectives on the range of ways to achieve these goals through funding, programs, and everyday action.

A More Inclusive Science: Examples, Tools, and Strategies

Organized by Jarita Holbrook, National Science Foundation, Arlington, VA

Changing Expectations: The Future of Careers in STEM

Organized by Claire Craig, The Royal Society, London, United Kingdom

Communication Challenges and Opportunities for Women in STEM

Organized by Christine O'Connell, Alan Alda Center for Communicating Science, Stony Brook, NY; Amy Landis, Colorado School of Mines, Golden

LGBTQ+ Identities in STEM Fields: Research and Implications

Organized by Rochelle Diamond, National Association of Gay and Lesbian Scientists and Technical Professionals, Pasadena, CA; Allison Mattheis, California State University, Los Angeles

MANAGING INNOVATION

Speakers in this flash talk session discuss the range of ways research and technology development can be fostered and regulated. One talk focuses on a public-private partnership, another on a national government effort, and a third on the web of existing and potential options for managing one important area of science. After these short talks, all attendees participate in a group discussion about strategies and considerations for efficiently and effectively encouraging innovation.

TECHNOLOGY TO SERVE THE WORLD

This session brings together flash talks oriented toward the interface between technological advances and societal needs. Speakers share a diversity of topics, from space technology to medicine-quality screening to applying the Internet of Things in agriculture. Each considers how these technologies serve the public good in just and equitable ways. After these short talks, a group discussion provides an opportunity to learn more about the specific technologies and projects, as well as consider broader questions about technology development and its benefits.

ADVOCATING FOR SCIENCE

These flash talks share a variety of motivations and strategies for engaging in science policy and science advocacy. Topics include a program promoting the value of science to society as a method of garnering political support, and research conducted during the March for Science to help place the event in a larger context and examine its outcomes. Another speaker will provide action-oriented tips for researchers interested in policy engagement. After these short talks, all participants join for a 30-minute group discussion.

COMMUNICATION AND PERCEPTION

This session considers human communication from anthropological, linguistic, and cognitive perspectives. It touches on continuing studies and theories of how language evolved, how our speech affects the way we perceive and receive others, and the science communication implications of research on how humans reason, process information, and make decisions. After three short talks, a 30-minute discussion provides an opportunity for all participants to engage with the speakers and one another.

DEVELOPING ROBOTICS TO ASSIST HUMANS

In this flash talk session, experts in robotics share several new applications of robot technology and explore the associated technological and societal questions and challenges. One speaker focuses on how autonomous intelligent robots are being programmed to learn and interact, while another talk centers on the use of robots in disasters and the reasons these haven't been deployed very successfully in the United States. A third flash talk shares recent research on how machines can infer the mental state of humans and thereby enhance human performance on military, healthcare, and manufacturing tasks. After three short talks, a 30-minute discussion provides an opportunity for all participants to engage with the speakers and one another.

SESSION TRACKS

Organizers are listed under session titles.

BIOLOGY AND CHEMISTRY

Advancing Health and Environmental Science Through Standardized Laboratory Microbial Ecosystems

Organized by Trent Northen, Lawrence Berkeley National Laboratory, Berkeley, CA; Karsten Zengler, University of California, San Diego

Applying Conservation Genetics and Genomics to Wildlife and Fisheries Management

Organized by Abraham J. Miller-Rushing, Acadia National Park, Bar Harbor, ME; Kelly LaRue, Jackson Laboratory, Bar Harbor, ME

Applying Insights From Animal Behavior to Address Global Challenges

Organized by Vanessa Ezenwa, University of Georgia, Athens; John Swaddle, College of William and Mary, Williamsburg, VA

Applying Mass Spectrometry to Understanding Complex Cellular Processes

Organized by Livia Eberlin and Jennifer Brodbelt, University of Texas, Austin

Assessing Risk From Chemical Exposures: Advances in Technology and Gaps in Application

Organized by Ellen Mantus, U.S. National Academies of Sciences, Engineering, and Medicine, Washington, DC; David C. Dorman, North Carolina State University, Raleigh

Future Products of Biotechnology and Needs for Risk Analysis Science

Organized by Kara Laney, U.S. National Academies of Sciences, Engineering, and Medicine, Washington, DC

Opioid Addiction: Biology, Psychology, and Social Policy

Organized by Yasmin Hurd, Icahn School of Medicine at Mount Sinai, New York City

Synthetic Biology: From Technology Development to Risk Governance

Organized by Katherine Bowman, U.S. National Academies of Sciences, Engineering, and Medicine, Washington, DC; Igor Linkov, U.S. Army Engineer Research and Development Center, Concord, MA

The Science of Art Conservation: Preserving Cultural Heritage Objects

Organized by Eric Breitung, Metropolitan Museum of Art, New York City

CLASSROOM TO CAREER

Empirical Findings on Science Fairs: Experiencing the Nature of Science

Organized by Frederick Grinnell, University of Texas Southwestern Medical Center, Dallas

Evidence for More Versatile Graduate Education and Academic Culture

Organized by Linda Hyman, Boston University, MA; Muriel Poston, Pitzer College, Claremont, CA

Overcoming Barriers to Change: Applying Evidence-Based STEM Teaching Strategies

Organized by Emily Miller, Association of American Universities, Washington, DC

Understanding Your Roots: STEM Diversity and an Evidence-Based Curriculum

Organized by Elizabeth Wright and Nina Jablonski, Pennsylvania State University, University Park

Women in STEM at Historically Black Institutions: South Africa and the United States

Organized by Lindiwe Gama and Cecil Masoka, South African Department of Science and Technology, Pretoria

CLIMATE AND THE ENVIRONMENT

Applying Earth Science Models and Satellite Observation to Benefit and Engage Society

Organized by Margaret Hurwitz and Danielle Wood, NASA Goddard Space Flight Center, Greenbelt, MD

Informing Mitigation and Adaptation Options With the Climate Science Special Report

Organized by Donald J. Wuebbles, University of Illinois, Urbana

Involving Stakeholders to Improve Outcomes: Lessons From the Climate Science Centers

Organized by Renee McPherson, University of Oklahoma, Norman; Katharine Hayhoe, Texas Tech University, Lubbock

Longer-Term Models for Arctic Sea Ice and Global Earth System Predictions

Organized by Sim James, U.S. National Earth System Prediction Capability Interagency Program, Silver Spring, MD; Jessie Carman, U.S. National Oceanic and Atmospheric Administration, Silver Spring, MD

Mathematics of Planet Earth: Superbugs, Storm Surges, and Ecosystem Change

Organized by Hans Engler and Hans Kaper, Georgetown University, Washington, DC

Mitigating Methane From Super-Emitters: California Pilot Projects

Organized by Riley Duren, NASA Jet Propulsion Laboratory, Pasadena, CA

Understanding Causality to Inform Decision-Making Under Uncertainty

Organized by Adam Douglas Henry, University of Arizona, Tucson; Thomas Dietz, Michigan State University, East Lansing

COMMUNICATION, LANGUAGE, AND CULTURE

Cultural and Linguistic Insights From the Study of Immigrant Languages

Organized by Joseph Salmons, University of Wisconsin, Madison

Evaluation and Best Practices for Training in Science Communication

Organized by Anthony Dudo, University of Texas, Austin

Exploring Public Fears and Myths: Vaccine Hesitancy, Food Safety in Fukushima, and Bacteria

Organized by Miyoko O. Watanabe, Japan Science and Technology Agency, Tokyo; Mark Ferguson, Science Foundation Ireland, Dublin

Gender in Translation: How Speech Communicates Sex, Gender Identity, and Sexuality

Organized by Nan Bernstein Ratner, University of Maryland, College Park

Natural and Cultural Resource Stewardship: New Scientific Insights and Audiences

Organized by Suzanne M. Thurston, AAAS Office of Education and Human Resources Programs, Washington, DC

Strategies for Communities and Scientists to Collaborate Effectively

Organized by Cathryn A. Manduca, Carleton College, Northfield, MN

The Impact of Sputnik on Science, Technology, and the Public in the United States

Organized by Jon D. Miller, University of Michigan, Ann Arbor

Understanding and Responding to Climate Change Denial

Organized by Heather Akin, University of Missouri, Columbia; Matthew H. Slater, Bucknell University, Lewisburg, PA

Visual, Attentional, and Gestural Foundations of Signed Languages

Organized by Richard P. Meier, University of Texas, Austin

DATA AND COMPUTING

Estimating the Prevalence of Human Trafficking in the United States

Organized by Theresa L. Harris, AAAS Scientific Responsibility, Human Rights and Law Program, Washington, DC; Davina Durgana, Walk Free Foundation, Great Falls, VA

Experiencing the Future Internet: User-Centered Social Television and Multimedia

Organized by David Wizel and Agata Stasiak, European Commission, Brussels, Belgium

Exploring Universal and Industrial Quantum Computing

Organized by Charles W. Clark, Joint Quantum Institute, Gaithersburg, MD; Daniel Rogers, Terbium Labs, Baltimore, MD

How Honey Bees Set Web Servers Abuzz

Organized by Erin Heath, AAAS Office of Government Relations, Washington, DC; Josh Shiode, Semiconductor Industry Association, Washington, DC

Prospects for Long-Term Information Security in the Face of Quantum Computing

Organized by Scott Aaronson, University of Texas, Austin; Charles W. Clark, Joint Quantum Institute, Gaithersburg, MD

Rethinking Approaches to Disaster Management and Public Safety With Intelligent Infrastructure

Organized by Ann Drobni, Computing Community Consortium, Washington, DC; Daniel Lopresti, Lehigh University, Bethlehem, PA

Transforming Cities, Transportation, and Agriculture With Intelligent Infrastructure

Organized by Ann Drobni, Computing Community Consortium, Washington, DC; Elizabeth Mynatt, Georgia Institute of Technology, Atlanta

Using Wearable Device Data to Analyze and Improve Physical Activity and Health

Organized by Raymond Carroll, Texas A&M University, College Station

ENERGY

A Sustainable Energy Future With Next-Generation, Low-Cost Solar Cells

Organized by Juan-Pablo Correa-Baena, Massachusetts Institute of Technology, Cambridge; Michael Saliba, Federal Institute of Technology, Lausanne, Switzerland

Addressing Systemic and Societal Factors in Germany's Energy Transition

Organized by Stefan Stückrad, Institute for Advanced Sustainability Studies, Potsdam, Germany

Bio-Based Industries Joint Undertaking: A Model for Catalyzing Sustainable Bio-Based Economic Growth

Organized by Sarah Black, Biobased Industries Joint Undertaking, Brussels, Belgium; Eleni Zika, Biobased Industries Joint Undertaking, Brussels, Belgium

Breakthroughs in Cellulosic Biomass and Transportation Fuels

Organized by Elizabeth Hood, Arkansas State University, Jonesboro

Energy-Enabling Materials and the Smart Cities of Tomorrow

Organized by Olga Rio and Luca Polizzi, European Commission, Brussels, Belgium



2018 ANNUAL MEETING

ADVANCING
SCIENCE | DISCOVERY TO
APPLICATION

ADVANCE REGISTRATION RATES

AAAS Member Rates for members in good standing

New Member Includes a year of AAAS membership

Non-Member Rates for all other attendees

General Attendee	\$310	\$380	\$440
Postdoc	\$135	\$200	\$360
K-12 Teacher	\$135	\$200	\$360
Retired Professional	\$250	\$320	\$360
Student	\$65	\$70	\$95
One-Day	\$175	N/A	\$220

aaas.org/meetings

Advance registration rates are available until **January 24**

Solving Materials Science Challenges to Shape Our Energy Future

Organized by Peter Genzer and James Misewich, Brookhaven National Laboratory, Upton, NY

ENGINEERING AND TECHNOLOGY

Advancing Biopharmaceutical Manufacturing With Public-Private Partnerships

Organized by Mike Molnar, U.S. National Institute of Standards and Technology, Gaithersburg, MD

Analyzing Picasso: Scientific Innovation, Instrumentation, and Education

Organized by Marc Walton, Northwestern University, Evanston, IL; Francesca Casadio, Art Institute of Chicago, IL

Biomedical Sensors: Advances in Health Monitoring and Disease Treatment

Organized by Qiaoqiang Gan, State University of New York, Buffalo; Zakya Kafafi, Lehigh University, Bethlehem, PA

Closing the Innovation Gap Between Research and Industry

Organized by Kaoru Natori, Okinawa Institute of Science and Technology Graduate University, Japan

Future-Generation Cars and Drivers: Research, Moving to Market, and Policymaking

Organized by Ingrid Skogsmo, European Commission, Brussels, Belgium

Gateway to Discovery: Research Teams Exploring the Edge of Technical Feasibility

Organized by Ben Verschueren, General Electric Global Research, Niskayuna, NY

Harnessing the Transformative Potential of Photonics Research and Innovation

Organized by Alex van Nieuwland, EuroTech Universities Alliance, Brussels, Belgium; Sandra M.J. Buys, Eindhoven University of Technology, Netherlands

Improvements in Earthquake Science and Risk Reduction

Organized by John Anderson, University of Nevada, Reno; William Savage, Seismological Society of America, Las Vegas, NV

Open Innovation Ecosystem for Advancing Science Through Advanced Manufacturing Research and Development

Organized by James Garrett, Carnegie Mellon University, Pittsburgh, PA; Sudarsan Rachuri, U.S. Department of Energy, Washington, DC

Oil and Water Do Mix: The Fate of Dispersed Oil Droplets in the Sea

Organized by Edward Buskey, University of Texas Marine Science Institute, Port Aransas

Successful Innovation and Commercialization in Industrial Science and Technology

Organized by William Provine, DowDuPont, Wilmington, DE

Sustained Academic, Industry, and Clinician Collaboration in Pre-Competitive Medical Research

Organized by Zoran Zvonar, Analog Devices Inc, Wilmington, MA; Charles Sodini, Massachusetts Institute of Technology, Cambridge, MA

Using Real-Time GPS for a Global Tsunami Early Warning System

Organized by Michael Angove, U.S. National Oceanic and Atmospheric Administration, Silver Spring, MD; Gerald Bawden, U.S. National Aeronautics and Space Administration, Washington, DC

GLOBAL COLLABORATION

Building Research Capacity as a Critical Component of International Development

Organized by Matt Goode, U.K. Research and Innovation, Swindon, United Kingdom

Cuban Biomedical Science: The Role of Science Diplomacy in Translating Cures

Organized by Mark Rasenick, University of Illinois College of Medicine, Chicago

Embracing International Partnerships to Achieve Science-Based Development in Kuwait

Organized by Ameena Farhan and Layla al-Musawi, Kuwait Foundation for the Advancement of Sciences, Sharq

Instruments of Science and Diplomacy: The Importance of International Research Organizations

Organized by Jan Marco Müller, International Institute for Applied Systems Analysis, Laxenburg, Austria

Issues and Impacts of U.S. Global Food Security Policy and Research

Organized by Nora Lapitan, and Jerry Glover, U.S. Agency for International Development, Washington, DC

Migration: A Case for Science Diplomacy

Organized by Jan Marco Müller, International Institute for Applied Systems Analysis, Laxenburg, Austria

Science for Sustainable Development Goals: Key Lessons and Gaps

Organized by E. William Colglazier, AAAS Center for Science Diplomacy, Washington, DC

MEDICINE AND HEALTH

Additive Manufacturing and 3-D Printing: Medical Technology Applications and Impacts

Organized by Luca Polizzi, European Commission, Brussels, Belgium

Advanced Technology for Oral Health Care: Diagnosis, Prevention, and Treatment

Organized by Janet Moradian-Oldak, University of Southern California, Los Angeles

Applying the Science of Genomics in Precision Medicine and Cancer Treatment

Organized by William Beck, University of Illinois, Chicago

Effective, Safe, and Underutilized: The HPV Vaccine From Development to Implementation

Organized by Joseph Margolick, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; Melinda Wharton, U.S. Department of Health and Human Services, Washington, DC

Emerging Cancer Immunotherapies: Challenges of Developing Modality-Specific Drugs

Organized by Cris Kamperschroer, Pfizer Inc., Groton, CT; Gautham Rao, Genentech Inc., South San Francisco, CA

Emerging Epidemic Pathogens: Basic, Translational, and Social Science

Organized by Gerald Keusch, National Emerging Infectious Diseases Laboratory, Boston, MA

Emulating Human Biology: Organ Chips for Drug Development and Personalized Medicine

Organized by Geraldine Hamilton, Emulate Inc., Boston, MA

Evolutionary Arms Races: From Bacteria to Cancer

Organized by Susan M. Rosenberg, Baylor College of Medicine, Houston, TX

Faster Responses to Epidemics and Bioterrorism With Standardized Product Development

Organized by Jeffrey Fortman, U.S. Department of Defense, Washington, DC

Gene Editing for Xenogeneic Organ Production: Regenerating a Patient's Transplantation Organ

Organized by Alison Van Eenennaam, University of California, Davis

Harnessing the Human Microbiome as a Tool for Prevention and Treatment of Disease

Organized by Wendy Cozen, University of Southern California, Los Angeles; Rob Knight, University of California, San Diego

Statistical and Computational Challenges in Genomics and Precision Medicine

Organized by Michael Boehnke, University of Michigan, Ann Arbor; Michael Epstein, Emory University School of Medicine, Atlanta, GA

Strategies to Accelerate the Widespread Adoption of Pharmacogenomics in Healthcare

Organized by Elizabeth Woo, Thermo Fisher Scientific, Waltham, MA; Tricia Kenny, Thermo Fisher Scientific, Carlsbad, CA

NEUROSCIENCE

Advanced Data Analysis Techniques for Understanding Brain Function

Organized by Robert Kass, Carnegie Mellon University, Pittsburgh, PA

Brain Plasticity Revisited: How Special Is the Young Brain?

Organized by Barbara Landau, Johns Hopkins University, Baltimore, MD

Connecting Behavior and the Brain to Understand Mental Health

Organized by Nora Newcombe, Temple University, Philadelphia, PA

New Technologies Emerging From the BRAIN Initiative

Organized by Walter J. Koroshetz, U.S. National Institutes of Health, Bethesda, MD

Optimal Aging and Mechanisms of Neurocognitive Enrichment

Organized by Denise Park, University of Texas, Dallas

“To Sleep, Perchance to Dream”: Rewiring the Brain During Sleep

Organized by Ted Abel, University of Iowa, Iowa City

Uncovering Novel Epigenetic Modifications in Neuropsychiatric Diseases

Organized by Tracy Bale, University of Pennsylvania School of Veterinary Medicine, Philadelphia

PHYSICS AND ASTRONOMY

A Universe of Discoveries: Progress in Astronomy, Statistics, and Machine Learning

Organized by Chad Schafer, Carnegie Mellon University, Pittsburgh, PA

Asteroids for Research, Discovery, and Commerce

Organized by Martin Elvis, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA

Capturing Dark Matter With Quantum Devices

Organized by Maria Spiropulu, California Institute of Technology, Pasadena

Exoplanets Everywhere: Discovering and Characterizing Worlds Beyond Our Solar System

Organized by Jennifer Wiseman, NASA Goddard Space Flight Center, Greenbelt, MD

Innovation in 2-D: Discovering Novel Materials and Applications

Organized by Eva Andrei, Rutgers, The State University of New Jersey, Piscataway

Investigating the Mysteries of Antimatter

Organized by Saeko Okada, High Energy Accelerator Research Organization, Tsukuba, Japan; Arnaud Marsollier, European Organization for Nuclear Research, Geneva, Switzerland

Is There a Future for Humanity in Space?

Organized by Amanda Arnold, Arizona State University, Washington, DC

Revolutionizing Ultrasound Applications for Treating Disease

Organized by Mark Hamilton, University of Texas, Austin; Joel Mobley, University of Mississippi, University

Scientific and Engagement Outcomes of the 2017 Total Solar Eclipse

Organized by Angela Speck, University of Missouri, Columbia; Jay Pasachoff, Williams College, Williamstown, MA

The Chemistry and Physics of Nascent Planetary Systems

Organized by Mark T. Adams, National Radio Astronomy Observatory, Charlottesville, VA

The Technological Applications of Glass: From Smartphones to Eagle-Eye Vision

Organized by Barbara Jones, IBM, San Jose, CA

POLICY

Balancing Facts and Values in Public Policymaking

Organized by Milena Raykovska, European Commission Joint Research Center, Brussels, Belgium

Building Public Trust and Fostering Innovation With Transparency

Organized by Matt Goode, U.K. Research and Innovation, Swindon, United Kingdom

Exploring Perspectives on Open Science and Impacts on Scientific Discovery

Organized by Ruth Francis and Shane Canning, F1000, London, United Kingdom

Gene Editing and Human Identity: Promising Advances and Ethical Challenges

Organized by Se Y. Kim and Robert O'Malley, AAAS Dialogue on Science, Ethics, and Religion, Washington, DC

Industry and Research Infrastructures as Co-Creators of Innovation

Organized by Jana Pavlic, European Molecular Biology Laboratory, Heidelberg, Germany; Antonio Di Giulio, European Commission Directorate-General for Research and Innovation, Brussels, Belgium

Recommendations of the U.S. Commission on Evidence-Based Policymaking

Organized by Nick Hart, Bipartisan Policy Center, Washington, DC

Reimagining the Innovation Ecosystem: Experiments to Maximize the Impact of Hard Tech

Organized by Brenna Krieger, U.S. Department of Energy, Washington, DC

Research and Policy on Voter ID Laws and Voter Participation

Organized by David Marker, Westat, Rockville, MD

Science Activism: Advancing Science in a New Political Landscape

Organized by Carol L. Rogers, University of Maryland, Washington, DC

Science and the Fair Administration of Justice

Organized by Alicia Carriquiry, Iowa State University, Ames

Translating Engineering and Operations Analyses into Effective Homeland Security Policy

Organized by Sheldon Jacobson, University of Illinois, Urbana

When Regulation Drives Innovation

Organized by Gerald Epstein, Independent, Bethesda, MD

SOCIAL SCIENCES

Advancing Interdisciplinary Collaboration With Lessons From the Field

Organized by Nancy Nersessian, Harvard University, Cambridge, MA; Hanne Andersen, University of Copenhagen, Denmark

Behavioral Challenges and Solutions to Science-Based Action

Organized by Kateryna Wowk, Texas A&M University, Corpus Christi, TX

Economic Analysis of Links Between Patents, Publications, and Policy

Organized by Pierre Azoulay, Massachusetts Institute of Technology Sloan School of Management, Cambridge; Bhaven Sampat, Columbia University, New York City

Implications of Evidence About Drug Use Hot Spots, Gerrymandering, and Gang Violence

Organized by William Alex Pridemore, State University of New York, Albany

Pre-Election Polling Uncertainty: An Interdisciplinary Analysis

Organized by Andrew Gelman, Columbia University, New York City

Social Inequality and Obesity: Effects of Poverty and Uncertainty on Health

Organized by Gregory Pavela, University of Alabama, Birmingham

Survey Data Collection: Theoretical and Practical Perspectives

Organized by Florian Keusch, University of Mannheim, Germany; Frauke Kreuter, University of Maryland, College Park

The Role of Conspiracy Theories in Perceptions of Fake News About Science

Organized by Jason Reifler, University of Exeter, Devon, United Kingdom; Asheley Landrum, Texas Tech University, Lubbock

Using Social Networks to Improve Response to Natural Disasters

Organized by Branda Nowell, North Carolina State University, Raleigh; Toddi Steelman, University of Saskatchewan, Saskatoon, Canada

What Citizens Think About Science: Survey Data and Implications for Communicators

Organized by Cary Funk, Pew Research Center, Washington, DC; John C. Besley, Michigan State University, East Lansing

SUSTAINABILITY AND RESOURCES

Airborne Laser Mapping Research in Tropical Forests and Archaeological Sites

Organized by Timothy Beach and Sheryl Luzzadder-Beach, University of Texas, Austin

Dealing With Deadly Pests Through the Sterile Insect Technique

Organized by Erin Heath, AAAS Office of Government Relations, Washington, DC; Josh Shlode, Semiconductor Industry Association, Washington, DC

From Lab to Farm to Table, and Back

Organized by Elizabeth (Toby) Kellogg, Donald Danforth Plant Science Center, St. Louis, MO

From the Bioreactor to the Plate: Addressing the Global Food Challenge With Algae

Organized by Rahel Byland, ETH Zurich, Switzerland

Implications of Water Cycle Science and Technology for Resource Management

Organized by Marty Ralph, Scripps Institution of Oceanography, La Jolla, CA; Xubin Zeng, University of Arizona, Tucson

Overcoming Challenges to Using Environmental Big Data for Conservation

Organized by Richard Feldman, Center for Scientific Research, Yucatán, Mexico; Abraham J. Miller-Rushing, Acadia National Park, Bar Harbor, ME

Phytobiome Research to Improve Agricultural Productivity

Organized by Jan E. Leach, Colorado State University, Fort Collins; Kellye Eversole, Eversole Associates, Bethesda, MD

Preparing for Floods and Droughts With NASA's "Scale in the Sky"

Organized by Jay Famiglietti, California Institute of Technology, Pasadena

Science Across Borders: Bilateral Collaboration and the U.K.-China Agri-Tech Project

Organized by Terry O'Connor, U.K. Science and Technology Facilities Council, Swindon, United Kingdom

The Importance of Tradeoffs in Managing Food Security

Organized by Margaret C. Nelson, Arizona State University, Tempe

The View From Above: Exploring the Future of Drones

Organized by Robert Hickey, Central Washington University, Ellensburg

2018 Annual Meeting app is now available:

aaas.org/app



2018 ANNUAL MEETING

**ADVANCING
SCIENCE** | DISCOVERY TO
APPLICATION

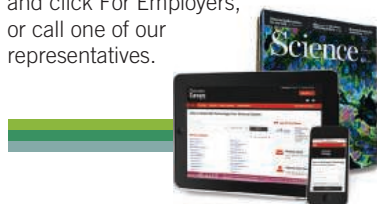
AAAS, publisher of *Science*, thanks the sponsors and supporters of the 2018 Annual Meeting



As of December 2017

SCIENCE CAREERS ADVERTISING

For full advertising details,
go to ScienceCareers.org
and click For Employers,
or call one of our
representatives.



AMERICAS

+1 202 326-6577
+1 202 326-6578
advertise@sciencecareers.org

EUROPE, INDIA, AUSTRALIA, NEW ZEALAND, REST OF WORLD

+44 (0) 1223 326527
advertise@sciencecareers.org

CHINA, KOREA, SINGAPORE, TAIWAN, THAILAND

+86 131 4114 0012
advertise@sciencecareers.org

JAPAN

+81 3-6459-4174
advertise@sciencecareers.org

CUSTOMER SERVICE

AMERICAS

+1 202 326-6577
REST OF WORLD
+44 (0) 1223 326528

advertise@sciencecareers.org

All ads submitted for publication must comply with applicable U.S. and non-U.S. laws. Science reserves the right to refuse any advertisement at its sole discretion for any reason, including without limitation for offensive language or inappropriate content, and all advertising is subject to publisher approval. Science encourages our readers to alert us to any ads that they feel may be discriminatory or offensive.

ScienceCareers

FROM THE JOURNAL SCIENCE 

ScienceCareers.org

Chair, Dept. of Cell & Regenerative Biology

The University of Wisconsin School of Medicine and Public Health (SMPH) invites applications and nominations for the position of Chair of the Department of Cell and Regenerative Biology (CRB).

The Department of Cell and Regenerative Biology is committed to understanding the fundamental mechanisms by which living systems operate at cellular and molecular levels of organization. By embracing a wide range of contemporary and emerging approaches and experimental systems, the department seeks to define signaling and regulatory pathways that provide the basis for understanding, diagnosis and treatment of human disease. Basic research is the centerpiece of the Department and serves as the driving force behind teaching and training efforts. The overarching research interests of the Department are highly interdisciplinary, emphasizing molecular, cellular and systems approaches to describe biological processes in molecular terms. To maintain its excellence and stature, the department is currently focusing on existing strengths in four research areas: Cell and Molecular Biology, Developmental Biology, Stem Cell and Regenerative Biology, and Cardiovascular Biology.

We seek a recognized leader with an outstanding academic background, strong research credentials, demonstrated commitment to education, experience in mentoring junior faculty, and proven leadership and management skills. The chair will provide professional and administrative leadership of the highest quality to this distinguished department in its teaching and research.

The successful candidate will have a compelling vision for the future of CRB in a leading academic medical center. Candidates must have a Ph.D., M.D., MD/Ph.D. or equivalent degree(s). They should possess substantial background and experience in administrative leadership, research, and teaching, and a strong academic background that would qualify for appointment as a tenured professor at the University of Wisconsin-Madison.

Please send nominations to: **Terri Young, MD, MBA, and David Gamm, MD, PhD, Co-Chairs of the CRB Chair Search Committee, c/o Staci Andersen, 4150Q HSLC, 750 Highland Avenue, Madison, WI, 53705- 2111, or use: slandersen2@wisc.edu**

To apply for this position, please use the University of Wisconsin applicant tracking system which is found at the link below. Click the "Apply Now" button. Applicants will be asked to upload a current cover letter, CV, and list of three references.

Application Link: jobs.hr.wisc.edu/crbchair

Applications from minorities and women are strongly encouraged. To receive full consideration, applications should arrive by **January 31, 2018**

Unless confidentiality is requested in writing, information regarding applicants must be released upon request. Finalists cannot be guaranteed confidentiality. Wisconsin Caregiver Law applies. The University of Wisconsin is an equal opportunity, affirmative action employer. For more information: med.wisc.edu, crb.wisc.edu.



Advance your
career with expert
advice from
Science Careers.



Download Free Career Advice Booklets!
[ScienceCareers.org/booklets](https://www.sciencecareers.org/booklets)

Featured Topics:

- Networking
- Industry or Academia
- Job Searching
- Non-Bench Careers
- And More



ScienceCareers

FROM THE JOURNAL SCIENCE  AAAS



ASSISTANT PROFESSOR Faculty Positions in Biochemistry and Molecular Biology

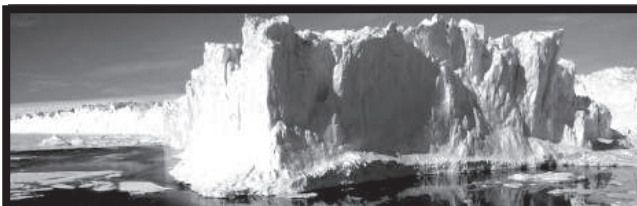
Two tenure-track, Assistant Professor positions are available in the Department of Biochemistry and Molecular Biology (<http://www.uams.edu/biochem/>) at the University of Arkansas for Medical Sciences. The positions include competitive salary, benefits, start-up packages, research space, and access to state-of-the-art core facilities, including a National Resource for Proteomics. The positions will be associated with the Winthrop P. Rockefeller Cancer Institute (<http://www.cancer.uams.edu>). Little Rock has an area population of 600,000, an affordable cost of living, many cultural amenities, and beautiful natural surroundings.

Candidates must possess a PhD and/or MD degree and postdoctoral experience. We seek highly qualified biochemists/molecular biologists who will establish an internationally leading research program on mechanisms of biomedically important processes and who will contribute to teaching medical and graduate students. Particular areas of interest include cancer biology, developmental biology, pediatric diseases, signal transduction, membrane traffic, DNA damage response, epigenetic mechanisms and systems biology. Cancer-related applications will be particularly responsive.

Applicants should submit curriculum vitae, a brief statement of proposed research, and arrange for three reference letters to be sent to (electronic submission preferred): **Biochemistry Search Committee, Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, 4301 W. Markham St., Little Rock, AR 72205.** E-mail: Biochemsearch@uams.edu.

*University of Arkansas for Medical Sciences is an Equal Opportunity/
Affirmative Action Employer.*

PRIZES



The 2018 Tinker-Muse Prize for Science and Policy in Antarctica

The "Tinker-Muse Prize for Science and Policy in Antarctica" is a USD \$100,000 unrestricted award presented to an individual in the fields of Antarctic science and/or policy who has demonstrated potential for sustained and significant contributions that will enhance the understanding and/or preservation of Antarctica. The Prize is inspired by Martha T. Muse's passion for Antarctica and is a legacy of the International Polar Year 2007-2008.

The prize-winner can be from any country and work in any field of Antarctic science and/or policy. The goal is to provide recognition of the important work being done by the individual and to call attention to the significance of understanding Antarctica in a time of change. A website with further details, including the process of nomination, closing date and selection of the Prize recipients, is available at www.museprize.org.

The Prize is awarded by the Tinker Foundation and administered by the Scientific Committee on Antarctic Research (SCAR).



**Nominations
open until
14 March 2018**



myIDP:
A career plan customized
for you, by you.



For your career in science, there's only one **Science**



**Recommended by
leading professional
societies and the NIH**

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests
- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track
- Articles and resources to guide you through the process
- Options to save materials online and print them for further review and discussion
- Ability to select which portion of your IDP you wish to share with advisors, mentors, or others
- A certificate of completion for users that finish myIDP.

Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers In partnership with:



From parade ground to PI

At the end of my postdoc, I excelled at moving tiny amounts of colorless fluid from one tube to another at strange times of day, and I had a vague sense of where I wanted to go scientifically. But my experience in the lab wasn't sufficient to equip me to run my own group. That preparation came from my double life as a British Army Reserve officer. Admittedly, I learned several things in the army that have not been so useful in a research lab, including how to march in time with others, shout very loudly, and swear with an unrivaled range of colorful expletives. Critically, though, my time in the army taught me how to be a leader—for good and ill.

One piece of military management advice that has stuck with me is that “it's not a popularity contest, sir.” In the army, I sometimes had to make unpopular decisions, such as who stayed behind to guard the barracks while everyone else went on rest and recreation. Getting the job done was more important than being universally liked. I now draw on this mindset in running my lab, whether I need to assign people to empty the garbage or decide who should get first authorship on a paper.

My military experience also helps me prioritize my time in the best interests of the lab. This often comes down to knowing when to lend a helping hand with a big experiment—the popular choice—and when the lab will be better served if I stay in the office to finish that paper. More broadly, the army taught me that sometimes there is no best decision, only a least worst decision. This helps me make difficult choices in science, such as when to ditch a study.

I also learned the importance of the team. Science can feel like a solitary activity. You may find yourself working in the vicinity of other people—the one who leaves the lab a mess, the one who books the key piece of equipment for 3 days solid and then doesn't use it, the one who talks too loudly about inappropriate things in the communal office—but this isn't the same as working *with* other people. In the army, on the other hand, working in stressful environments with people who were tired, cold, and wet fostered camaraderie and understanding that helped us be more effective individually and as a group.

So, as a group leader, I have prioritized team-building. I have found that getting to know my team and building rapport—breaking bread, going to the pub, doing things



“My time in the army taught me how to be a leader.”

together outside the lab—makes it possible for me to stress them from time to time, for example by making demands for data and paper revisions and setting deadlines for experiments. But, having stressed the team, I have also found that it is then necessary to rebuild. To do this, I apply the ABCs: acknowledgment (telling people they've done well), booze (a bottle of Prosecco goes a long way), and cake.

Of course, the lab is not the army. Among other differences, academia is less directive and more discursive. Discussion leads to better science, but I sometimes find the lack of hierarchy a challenge when it comes to the day-to-day work of running the lab.

People don't simply do things I tell them to just because of my position, and shouting at them in my best parade ground voice (plus expletives) is equally ineffective. Telling one of my students to “JFDI” (look it up) when we were arguing about the best way to do an assay did not have the desired effect—quite the opposite, and it took a bit of ABC to fix this. So, I have learned to temper my inner sergeant major, though he does tend to rear up when I am overworked and stressed.

Ninety percent of being a good lab head is being able to direct your team. Although I've needed to adjust my leadership style for different surroundings, having the opportunity to develop a leadership style in the first place has been a real advantage. The army is not for everyone. But if you want to learn to lead, I would recommend getting out of the lab. ■

John Tregoning is a senior lecturer at Imperial College London. Do you have an interesting career story? Send it to SciCareerEditor@aaas.org.