The heart-stopping sting of the box jellyfish *p. 631*

Preserving aluminum-air battery shelf life *p. 658*

Facing the challenge of genetic privacy *p. 690*

Science

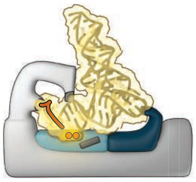
$15
9 NOVEMBER 2018
sciencemag.org

AAAS

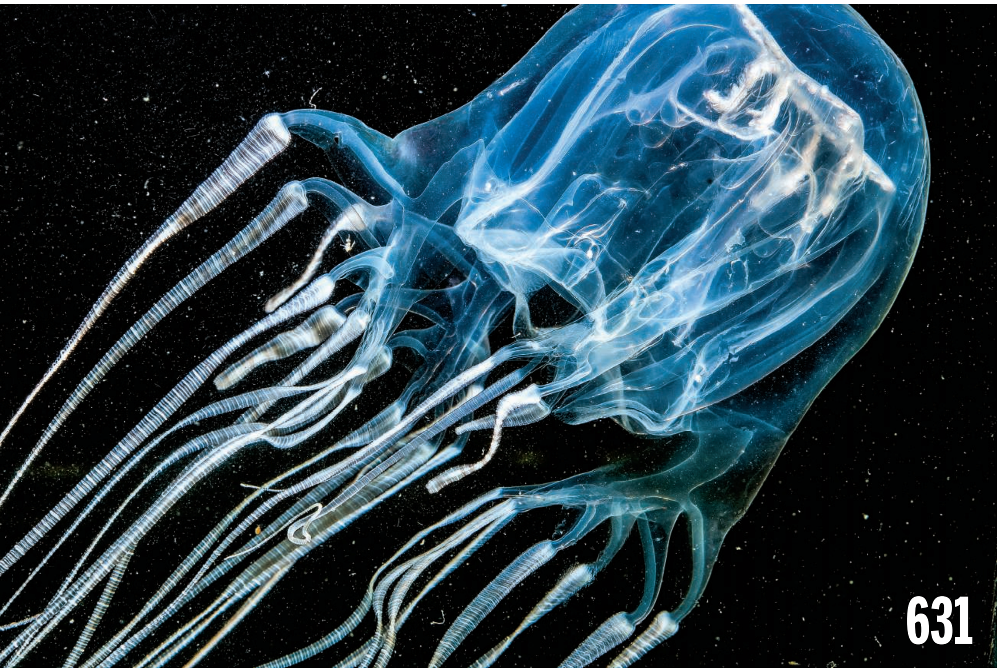
UNSAFE NESTS

Climate change disrupts predation patterns *p. 680*

CONTENTS



9 NOVEMBER 2018 • VOLUME 362 • ISSUE 6415



NEWS

IN BRIEF

IN DEPTH

FEATURE

INSIGHTS

POLICY FORUM

PERSPECTIVES

BOOKS

CREDITS: (TOP TO BOTTOM) V. ALTOUNIAN/*SCIENCE*; ANGEL YANAGIHARA; POSNOV/GETTY IMAGES

CONTENTS

640 & 656

Dissecting peanut allergy

9 NOVEMBER 2018 • VOLUME 362 • ISSUE 6415

LETTERS

RESEARCH

IN BRIEF

REVIEWS

RESEARCH ARTICLES

REPORTS

DEPARTMENTS

ON THE COVER

A red phalarope (*Phalaropus fulicarius*) returns to its nest in the Arctic tundra. Nest predation of shorebirds has increased worldwide over the past decades. Recent rises in predation rates are especially noticeable in the Arctic, in line with climate-induced alterations in predator-prey interactions. Declining shorebird populations are now further hindered by poor reproductive success, with likely detrimental consequences for their global numbers. See page 680. *Photo: JOEL SARTORE/National Geographic Creative*

PHOTO: BRUNO CRESCIA PHOTOGRAPHY INC/GETTY IMAGES

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. 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 on request. Canadian rates with GST available upon request, GST #125488122. 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. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Science
AAAS

1200 New York Avenue, NW, Washington, DC 20005
Clarendon House, Clarendon Road, Cambridge, UK CB2 8FH

Editor-in-Chief Jeremy Berg
Executive Editor Monica M. Bradford **News Editor** Tim Appenzeller
Editor, Insights Lisa D. Chong **Editors, Research** Valda Vinson, Jake S. Yeston

Research and Insights

DEPUTY EDITORS Julia Fahrenkamp-Uppenbrink(UK), Stella M. Hurtley(UK), Phillip D. Szuromi, Sacha Vignieri **SR. EDITORIAL FELLOW** Andrew M. Sugden(UK) **SR. EDITORS** Gemma Alderton(UK), Caroline Ash(UK), Pamela J. Hines, 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), Brad Wible, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Brent Grocholski, Priscilla N. Kelly, Tage S. Rai, 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 Kmec **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Amber Esplin, Chris Filiatreau, Cynthia Howe **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Shannon McMahon, Jerry Richardson, Alice Whaley(UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Ope Martins, Nida Masiulis, Dona Mathieu, Ronmel Navas, Hilary Stewart(UK), Alana Warnke, Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ASI DIRECTOR, OPERATIONS** Janet Clements(UK) **ASI SR. OFFICE ADMINISTRATOR** Jessica Waldock(UK)

News

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** Martin Enserink **DEPUTY NEWS EDITORS** Elizabeth Culotta, Lila Guterman, David Grimm, Eric Hand, David Malakoff, Leslie Roberts **SR. CORRESPONDENTS** Daniel Clery(UK), Jon Cohen, Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brainard, Catherine Matacic **NEWS WRITERS** Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Kelly Servick, Robert F. Service, Erik Stokstad(Cambridge, UK), Paul Voosen, Meredith Wadman **INTERN** Frankie Schembri **CONTRIBUTING CORRESPONDENTS** Warren Cornwall, Ann Gibbons, Mara Hvistendahl, Sam Kean, Eli Kintisch, Kai Kupferschmidt(Berlin), Andrew Lawler, Mitch Leslie, Eliot Marshall, Virginia Morell, Dennis Normile(Shanghai), Charles Piller, Tania Rabesandratana(London), Emily Underwood, Gretchen Vogel(Berlin), Lizzie Wade(Mexico City) **CAREERS** Donisha Adams, Rachel Bernstein(Editor), Katie Langin **COPY EDITORS** 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 **ASSOCIATE DIRECTOR, PRODUCT MANAGMENT** Kris Bishop **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

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

WEB TECHNOLOGIES TECHNOLOGY DIRECTOR David Levy **PROJECT MANAGER** Dean Robbins **DEVELOPER** Liana Birke

DIGITAL MEDIA DIRECTOR OF ANALYTICS Enrique Gonzales **DIGITAL REPORTING ANALYST** Timothy Frailey **MULTIMEDIA MANAGER** Sarah Crespi **MANAGING WEB PRODUCER** Kara Estelle-Powers **DIGITAL PRODUCER** Jessica Hubbard **VIDEO PRODUCERS** Chris Burns, Meagan Cantwell **SOCIAL MEDIA PRODUCER** Brice Russ

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

DESIGN DIRECTOR Beth Rakouskas **DESIGN MANAGING EDITOR** Marcy Atarod **SENIOR DESIGNER** Chrystal Smith **DESIGNER** Christina Aycock **GRAPHICS MANAGING EDITOR** Alberto Cuadra **GRAPHICS EDITOR** Nirja Desai **SENIOR SCIENTIFIC ILLUSTRATORS** Valerie Altounian, Chris Bickel **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

SENIOR EDITOR, CUSTOM PUBLISHING Sean Sanders: 202-326-6430 **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst: 202-326-6463 **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, FAX 617-507-8189 **WEST COAST/W. CANADA** Lynne Stickrod: 415-931-9782, FAX 415-520-6940 **MIDWEST** Jeffrey Dembksi: 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

ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING ASSOCIATE** Aimee Aponte **MARKETING MANAGER, JOURNALS** Shawana Arnold **MARKETING ASSOCIATES** Mike Romano, Tori Velasquez **SENIOR DESIGNER** Kim Huynh **TRADE SHOW COORDINATOR** Andrew Clamp

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

AAAS BOARD OF DIRECTORS, CHAIR Susan Hockfield **PRESIDENT** Margaret A. Hamburg **PRESIDENT-ELECT** Steven Chu **TREASURER** Carolyn N. Ainslie **CHIEF EXECUTIVE OFFICER** Rush D. Holt **BOARD** Cynthia M. Beall, May R. Berenbaum, Rosina M. Bierbaum, Kaye Husbands Fealing, Stephen P.A. Fodor, S. James Gates, Jr., Michael S. Gazzaniga, Laura H. Greene, Robert B. Millard, Mercedes Pascual, William D. Provine

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 **S**)

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. Andreae, *Max Planck Inst. Mainz*
Paola Arlotta, *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, *NIAID, NIH*
Philip Benfey, *Duke U.*
Gabriele Bergers, *VIB*
Bradley Bernstein, *Mass. General Hospital*
Peer Bork, *EMBL*
Chris Bowler, *École Normale Supérieure*
Ian Boyd, *U. of St. Andrews*
Emily Brodsky, *U. of California, Santa Cruz*
Ron Brookmeyer, *U. of California, Los Angeles* (**S**)
Christian Büchel, *UKE Hamburg*
Dennis Burton, *Scripps Research*
Carter Tribley Butts, *U. of California, Irvine*
Gyorgy Buzsaki, *New York U. School of Med.*
Blanche Capel, *Duke U.*
Nick Chater, *U. of Warwick*
Ib Chorkendorff, *Denmark TU*
James J. Collins, *MIT*
Robert Cook-Deegan, *Arizona State U.*
Lisa Coussens, *Oregon Health & Science U.*
Alan Cowman, *Walter & Eliza Hall Inst.*
Carolyn Coyne, *U. of Pittsburgh*
Roberta Croce, *VU Amsterdam*
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 Penn.*
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, *U. of Oxford*
Todd Ehlers, *U. of Tübingen*
Jennifer Elisseeff, *Johns Hopkins U.*
Tim Elston, *U. of North Carolina at Chapel Hill*
Nader Engheta, *U. of Penn.a*
Barry Everitt, *U. of Cambridge*
Vanessa Ezenwa, *U. of Georgia*
Ernst Fehr, *U. of Zürich*
Michael Feuer, *The George Washington U.*
Toren Finkel, *U. of Pittsburgh Med. Ctr.*
Kate Fitzgerald, *U. of Mass.*
Peter Fratzl, *Max Planck Inst. Potsdam*
Elaine Fuchs, *Rockefeller U.*
Eileen Furlong, *EMBL*
Jay Gallagher, *U. of Wisconsin*
Susan Gelman, *U. of Michigan*
Daniel Geschwind, *U. of California, Los Angeles*
Karl-Heinz Glassmeier, *TU Braunschweig*
Marta Gonzalez, *U. of California, Berkeley*
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, *Yale U.*
Wolf-Dietrich Hardt, *ETH Zürich*
Louise Harra, *U. College London*
Michael Hasselmo, *Boston U.*
Jian He, *Clemson U.*
Martin Heimann, *Max Planck Inst. Jena*
Carl-Philipp Heisenberg, *IST Austria*
Ykä Helariutta, *U. of Cambridge*
Janet G. Hering, *Eawag*
Kai-Uwe Hinrichs, *U. of Bremen*
David Hodell, *U. of Cambridge*
Lora Hooper, *UT Southwestern Med. Ctr.*
Fred Hughson, *Princeton U.*
Randall Hulet, *Rice U.*
Auke Ijspeert, *EPFL*
Akiko Iwasaki, *Yale U.*
Stephen Jackson, *USGS and U. of Arizona*
Kai Johnsson, *EPFL*
Peter Jonas, *IST Austria*
Matt Kaeberlein, *U. of Washington*
William Kaelin Jr., *Dana-Farber Cancer Inst.*
Daniel Kammen, *U. of California, Berkeley*
Abby Kavner, *U. of California, Los Angeles*
Masashi Kawasaki, *U. of Tokyo*
V. Narry Kim, *Seoul Nat. U.*
Robert Kingston, *Harvard Med. School*
Nancy Knowlton, *Smithsonian Institution*
Etienne Koechlin, *École Normale Supérieure*
Alexander Kolodkin, *Johns Hopkins U.*
Thomas Langer, *U. of Cologne*
Mitchell A. Lazar, *U. of Penn.*
David Lazer, *Harvard U.*
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 biomaGUNE*
Jonathan Losos, *Harvard U.*
Ke Lu, *Chinese Acad. of Sciences*
Christian Lüscher, *U. of Geneva*
Fabienne Mackay, *U. of Melbourne*
Anne Magurran, *U. of St. Andrews*
Oscar Marín, *King's College London*
Charles Marshall, *U. of California, Berkeley*
Christopher Marx, *U. of Idaho*
Geraldine Masson, *CNRS*
C. Robertson McClung, *Dartmouth College*
Rodrigo Medellín, *U. of Mexico*
Graham Medley, *London School of Hygiene & Tropical Med.*
Jane Memmott, *U. of Bristol*
Edward Miguel, *U. of California, Berkeley*
Tom Misteli, *NCI, NIH*
Yasushi Miyashita, *U. of Tokyo*
Richard Morris, *U. of Edinburgh*
Alison Motsinger-Reif, *NC State U.* (**S**)
Daniel Nettle, *Newcastle U.*
Daniel Neumark, *U. of California, Berkeley*
Kitty Nijmeijer, *TU Eindhoven*
Helga Nowotny, *Austrian Council*
Rachel O'Reilly, *U. of Warwick*
Harry Orr, *U. of Minnesota*
Pilar Ossorio, *U. of Wisconsin*
Andrew Oswald, *U. of Warwick*
Isabella Pagano, *Istituto Nazionale di Astrofisica*
Margaret Palmer, *U. of Maryland*
Elizabeth Levy Paluck, *Princeton U.*
Jane Parker, *Max Planck Inst. Cologne*
Giovanni Parmigiani, *Dana-Farber Cancer Inst.* (**S**)
Samuel Pfaff, *Salk Inst. for Biological Studies*
Julie Pfeiffer, *UT Southwestern Med. Ctr.*
Matthieu Piel, *Institut Curie*
Kathrin Plath, *U. of California, Los Angeles*
Martin Plenio, *Ulm U.*
Albert Polman, *FOM Institute for AMOLF*
Elvira Poloczanska, *Alfred-Wegener-Inst.*
Philippe Poulin, *CNRS*
Jonathan Pritchard, *Stanford U.*
David Randall, *Colorado State U.*
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 Nat. Lab*
Nitin Samarth, *Penn. State U.*
Jürgen Sandkühler, *Medical U. of Vienna*
Alexander Schier, *Harvard U.*
Wolfram Schlenker, *Columbia U.*
Susannah Scott, *U. of California, Santa Barbara*
Vladimir Shalaev, *Purdue U.*
Beth Shapiro, *U. of California, Santa Cruz*
Jay Shendure, *U. of Washington*
Brian Shoichet, *U. of California, San Francisco*
Robert Siliciano, *Johns Hopkins U. School of Med.*
Uri Simonsohn, *U. of Penn.*
Lucia Sivilotti, *U. College London*
Alison Smith, *John Innes Centre*
Richard Smith, *U. of North Carolina at Chapel Hill* (**S**)
Mark Smyth, *QIMR Berghofer*
Pam Soltis, *U. of Florida*
John Speakman, *U. of Aberdeen*
Tara Spires-Jones, *U. of Edinburgh*
Allan C. Spradling, *Carnegie Institution for Science*
Eric Steig, *U. of Washington*
Paula Stephan, *Georgia State U.*
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.*
Kathleen Vohs, *U. of Minnesota*
David Wallach, *Weizmann Inst. of Science*
Jane-Ling Wang, *U. of California, Davis* (**S**)
David Waxman, *Fudan U.*
Jonathan Weissman, *U. of California, San Francisco*
Chris Wikle, *U. of Missouri* (**S**)
Terrie Williams, *U. of California, Santa Cruz*
Ian A. Wilson, *Scripps Research* (**S**)
Yu Xie, *Princeton U.*
Jan Zaanen, *Leiden U.*
Kenneth Zaret, *U. of Penn. School of Med.*
Jonathan Zehr, *U. of California, Santa Cruz*
Maria Zuber, *MIT*

Published by AAAS

EDITORIAL

Facing hatred

I came to the United States in 2016—from Paris to Pittsburgh—in the aftermath of a horrific series of terrorist attacks in France that targeted everyone. As a clinician-scientist, I wanted to continue focusing my efforts on therapies to restore vision rather than constantly check the news. As a Jew, I wanted to live according to ethical principles nurtured by my identity in an open society that does not consider deviation from secularism as questionable. As a father, I wanted my children and grandchildren to grow within a humanistic perspective and at the same time to honor their dual heritage. In 2012, the very mild reaction to the killing of soldiers and Jews, especially of children in a Jewish school, in Toulouse, was saddening, and there was great uncertainty about how a society infected by anger, fear, and intolerance would evolve. Two weeks ago, the mass killing at a Pittsburgh synagogue proved what I knew but wanted to forget—that no place on Earth is "safe" from hatred. But regardless of where any one of us lives and works, we are faced with the same, immense challenge: The quest for facts, enlightenment, and care versus ignorance and hatred matters more than ever.

In the early 1930s, Albert Einstein asked Sigmund Freud to contribute to an initiative launched by the League of Nations that sought prominent individuals to promote peace and its values. Their exchange, written under the title *Why War*?, fell short of finding ways to counteract violence. It was published after Hitler had already been appointed chancellor. The rest is history. Should we simply believe that today there is a new cycle of history taking place? Moreover, in contrast to the League of Nations, nobody is asking the scientific community for help today. Science, with its fundamental quest for truth, or at least facts, and its open discourse, seems to have lost its iconic status, despite its contributions to societal well-being. The scientific community should not passively watch the disastrous rise of hatred worldwide.


Shalekhet (Fallen Leaves), by M. Kadishman, located in the Memory Void in the Jewish Museum Berlin, Germany, is an art installation dedicated to all victims of violence and war.

We are tasked with building a society of knowledge and care, where truth, integrity, and respect for all prevail. The heartening responses of health care providers to incidents in the United States and France epitomize this ideal. In 2015, after the mass killing at the Bataclan theater in Paris, nurses and physicians spontaneously converged on hospitals to help the victims, limiting the massive toll of the attacks. Likewise, in Pittsburgh, nurses and physicians treated the wounded, including the presumed killer, with efficiency and humanity. Helplines offering information and support were set up for a whole city in mourning. Religious and political leaders of all faiths and backgrounds, and community members from across the cultural spectrum, were united against hatred. In both cities, the responses were deeply rooted in society's best, most inspiring traits. I specifically recall Dr. Jerry Rabinowitz, a colleague and friend to many, who was a victim in the Pittsburgh tragedy. He was a pioneer in providing treatment during the early years of the AIDS epidemic—one who saved many lives and gave selflessly to all. He was shot as he volunteered to help the wounded.

Caring means that each life matters, and that we all can and should be supported to grow and give back to society. The French Jewish philosopher Emmanuel Levinas asserted that looking into the face of one's fellow man invokes the imperative: "Thou shalt not kill." This sounds naïve and far too simplistic in the face of guns and strongly held prejudices. Yet, is there anything else in the world more meaningful than looking into human faces and listening?

If we are truly an enlightened and caring society, then our response to violence must be to reject resignation and to include actions by those who seek truth and fact. This is now, as ever, our inheritance.

– Jose-Alain Sahel

Jose-Alain Sahel *is chair of the Department of Ophthalmology at the University of Pittsburgh School of Medicine, Pittsburgh, PA, USA, and professor of ophthalmology in the Faculty of Medicine at Sorbonne Université, Paris, France. sahelja@upmc.edu*

10.1126/science.aav9391

CREDITS: (INSET) ADAM EASTLAND/ALAMY STOCK PHOTO; (RIGHT) UPMC

NEWS

IN BRIEF

Edited by **Jeffrey Brainard**

WILDLIFE CONSERVATION

Court rules for red wolf recovery

The U.S. Fish and Wildlife Service (FWS) must do a better job of protecting the endangered red wolf (*Canis rufus*), a federal judge ruled this week. The decision overturns several controversial management decisions that wolf advocates say would have doomed the 40 or so remaining red wolves. Only one population exists in the wild, in eastern North Carolina, where FWS in 1987 started to reintroduce red wolves from zoos. Some local residents object to the wolves roaming onto their land, so FWS—which has more flexibility in managing this "experimental" population than other endangered species—in 2014 began to allow red wolves to be shot even if they weren't threatening people or causing problems. After North Carolina officials requested an end to the program, FWS also stopped releasing red wolves and halted its long-standing effort to prevent interbreeding with coyotes. U.S. District Court Judge Terrence Boyle found that the changes violated the Endangered Species Act.

PHOTO: ABESELOM ZERIT/ALAMY STOCK PHOTO

Foundations back open access

PUBLISHING | Two major research funders this week announced they will change their policies to generally match Plan S, the open-access initiative unveiled in September by 11 European funders to require grantees to publish in journals that make all their content freely available. The Wellcome Trust in London and the Bill & Melinda Gates Foundation in Seattle, Washington, said as of January 2020 they will no longer pay for grantees to publish in so-called hybrid open-access journals, which have both paywalled content and free content. If a grantee wants to publish in a paywalled journal, they must simultaneously add their accepted manuscript to the open repositories PubMed Central or Europe PubMed Central. (Many high-impact subscription journals, including *Science* and *Nature*, do not allow the placement of papers in these repositories until 6 months after publication.) The two funders add substantial funding muscle to Plan S; the existing 11 supporters collectively spend about $8.7 billion annually on research, whereas Wellcome provides $1.3 billion annually and the Gates Foundation spends more than $1.2 billion.

NSF to pilot LGBT questions

DIVERSITY | The U.S. National Science Foundation (NSF) is moving toward asking questions about sexual orientation and gender identity in its national science, technology, engineering, and math (STEM) workforce surveys, starting with the National Survey of College Graduates as early as 2021. At a meeting last week, NSF said it first plans to pilot the questions with test audiences. The move was catalyzed by a letter submitted to NSF in October by 243 scientists and engineers and 17 scientific organizations, including AAAS (publisher of *Science*). Past versions of the surveys did not ask any questions on those topics but did pose other demographic questions, such as age and race. The letter writers argued that comprehensive, nationwide data on lesbian, gay, bisexual, and transgender scientists and engineers are needed because the group experiences "disadvantages and disparities in STEM fields similar to [those experienced by] other under-represented groups, such as racial and ethnic minorities and women."

Diarrhea vaccine in short supply

PUBLIC HEALTH | A half-million children in four West African countries risk contracting a deadly diarrheal disease because Merck & Co. is running short of a rotavirus vaccine used to prevent it. Gavi, the Vaccine Alliance, a public-private partnership based in Geneva, Switzerland, that supplies the Merck rotavirus vaccine to Burkina Faso, Ivory Coast, Mali, and São Tomé and Príncipe, said last week that the company could supply Gavi with only two-thirds of the doses it had committed for those countries for 2018 and 2019. Seth Berkley, Gavi's head, says Burkina Faso has already run out of its stock of the vaccine. Merck, based in Kenilworth, New Jersey, cited manufacturing issues. Media reports have suggested Merck ran short because it started to sell the vaccine to China for a much higher price, but the company told *Science* the supply is not being shifted from West Africa because of that new market. Other companies make the vaccine, but Berkley says it will take them time to scale up and fill the gap.

Earliest figurative painting found


A painting at least 40,000 years old shows what may be a wild cow.

ARCHAEOLOGY | Daubed in ochre at least 40,000 years ago, a meter-long image of what may be a wild cow in a cave on the Indonesian island of Borneo is now the world's oldest known figurative painting. It adds to evidence that by this time, artistic traditions had emerged at opposite ends of Eurasia. Images of animals in France's Chauvet Cave that are about 35,000 years old were once thought to be the earliest figurative artworks. Then in 2014, archaeologists led by Maxime Aubert of Griffith University in Gold Coast, Australia, dated calcite crusts on top of cave paintings of animals on the Indonesian island of Sulawesi to 35,400 years ago. Now, Aubert's Indonesian-Australian team reports in the 7 November issue of *Nature* that they used the same uranium-thorium radiometric dating method on crusts covering a painting of what may be a banteng, a Southeast Asian wild cattle, in Borneo's Lubang Jeriji Saléh cave—and found it to be even older.

Tighter test dooms some results

BIOMEDICINE | Nearly one-third of clinical trial results in three leading medical journals that met the existing gold standard for statistical significance would fail a proposed tighter one, a study reports. The authors examined findings of phase III studies published in 2017 in *The Journal of the American Medical Association* (*JAMA*), *The Lancet*, or *The New England Journal of Medicine*. They identified 174 measured outcomes that met a probability value (p-value) of at least 0.05; of these outcomes, about 29% failed to meet a p-value of at least 0.005. These results could be deemed only "suggestive," the authors suggest in the 6 November issue of *JAMA*. Some scientists have argued in recent years for adopting this tighter statistical standard, saying it would greatly reduce the reporting of false-positive results—studies that claim to find an effect when there is none.

Tracking authors' work, bit by bit

PUBLISHING | A proposal to develop a new online platform that publishes chunks of work smaller than a typical article won a competition sponsored by the Royal Society in London last month. Alexandra Freeman of the University of Cambridge in the United Kingdom, who came up with the idea, says she hoped to alleviate pressures associated with traditional publishing. Researchers could use the searchable platform, called Octopus, to publish eight types of contributions, including hypotheses, protocols, data,

BIODIVERSITY

Seed banking faces limits

Seed banks, where seeds are dried and then frozen indefinitely, play by far the biggest role in a global effort to conserve 75% of the world's threatened plants outside their natural habitat. But reaching that goal may be impossible, plant researchers have concluded. Banks have found about 10% of plants have seeds that don't survive this procedure. A new study offers an even more troubling estimate: Using a computer model for predicting the behavior of stored seeds, John Dickie, a seed biologist at the Royal Botanic Gardens, Kew, in Ardingly, U.K., and colleagues concluded in the 2 November issue of *Nature Plants* that up to a third of critically endangered plants, for example, won't tolerate the drying and freezing process. "We will have to look at other methods," Dickie says. They include cryopreservation, which involves storing material at very low temperatures, or redoubling efforts to protect endangered species where they normally live.

PHOTOS: (TOP TO BOTTOM) LUC-HENRI FAGE; JAMES KING-HOLMES/SCIENCE SOURCE

including hypotheses, protocols, data, analysis, and interpretation. It would function something like a preprint server for sections of papers, publishing them separately as authors completed them and allowing the contributions to be recognized and critiqued independently. Freeman says she will use her £1000 prize, awarded for the best idea to improve the culture of research, to start to develop the platform.

U.S. gender proposal decried

POLICY | More than 1600 scientists have criticized the "grave harm" they believe would be caused by a plan under consideration by President Donald Trump's administration to legally define gender as male or female based on genitalia at birth or genetic tests later. The proposal, drafted last spring and reported by *The New York Times* in September, calls for "a biological basis that is clear [and] grounded in science." If enacted, the definition could undo education and health policies created under former President Barack Obama that recognize gender, and being transgender, largely as an individual's choice. In the open letter posted online at not-binary.org, the scientists argue the new proposal is "fundamentally inconsistent" with scientific findings. Some people's genitalia differs from what clinicians would predict from their sex chromosomes, it notes; imaging studies indicate the brains of transgender individuals differ from those of control men and women.

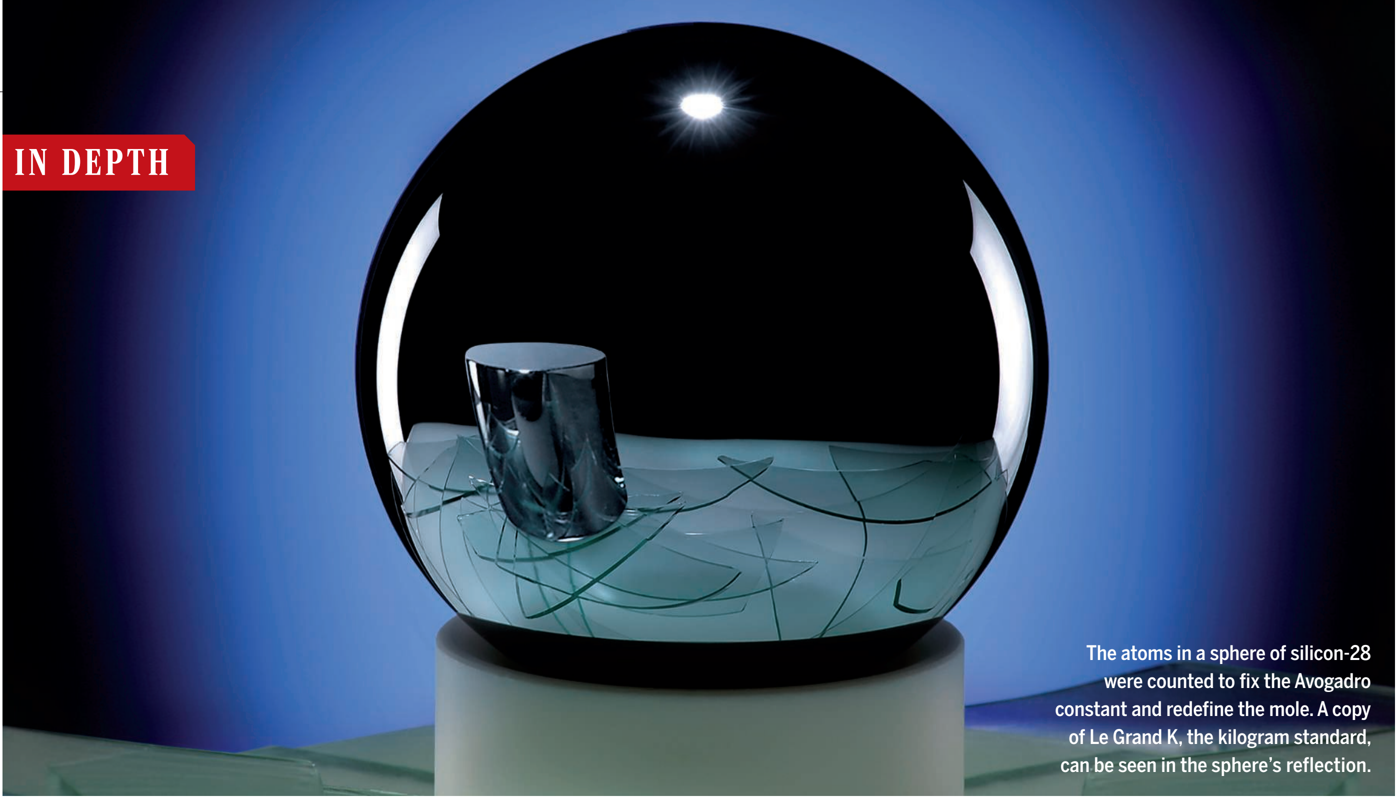
Whales identified from space

WILDLIFE CONSERVATION | New analyses of high-resolution satellite images have provided unprecedented details of individual whales, a step toward making comprehensive, automated population counts of whales worldwide, researchers at the British Antarctic Survey (BAS) in Cambridge report. Satellites can survey more of the ocean than boats and planes can. But in the past, low resolution meant whales appeared more like blobs. The new images' high resolution—a square 31 centimeters on a side—allows specialists to distinguish body sizes, shapes, and features, such as flukes and fins, which are useful for identifying species. BAS graduate student Hannah Cubaynes and colleagues noticed the details while examining images from the new, privately owned WorldView-3 satellite, they reported online 27 October in *Marine Mammal Science*. In images from four ocean patches where fin, gray, humpback, or southern right whales congregate, Cubaynes's team counted 200 whales seen in enough detail to assign them to one of those species. Next, the team will use the data to develop a computer program able to flag images that might show whales.

U.K. currency to feature scientist

HISTORY | Citizens of the United Kingdom will soon be taking science to the bank. The Bank of England announced last week that the new £50 note will feature a prominent British scientist. Over the next 6 weeks, members of the public can nominate any scientist, mathematician, engineer, or technologist through a form on the bank's website—as long as they are deceased and not fictional. (Sorry, Dr. Who.)

SCIENCEMAG.ORG/NEWS
Read more news from *Science* online.



The atoms in a sphere of silicon-28 were counted to fix the Avogadro constant and redefine the mole. A copy of Le Grand K, the kilogram standard, can be seen in the sphere's reflection.

METROLOGY

World poised to adopt new metric units

Kilogram and other units will be derived from fixed constants of nature

By **Adrian Cho**

Like an aging monarch, Le Grand K is about to bow to modernity. For 130 years, this gleaming cylinder of platinum-iridium alloy has served as the world's standard for mass. Kept in a bell jar and locked away at the International Bureau of Weights and Measures (BIPM) in Sèvres, France, the weight has been taken out every 40 years or so to calibrate similar weights around the world. Now, in a revolution far less bloody than the one that cost King Louis XVI his head, it will cede its throne as the one, true kilogram.

When the 26th General Conference on Weights and Measures (CGPM) convenes next week in Versailles, France, representatives of the 60 member nations are expected to vote to redefine the International System of Units (SI) so that four of its base units—the kilogram, ampere, kelvin, and mole—are defined indirectly, in terms of physical constants that will be fixed by fiat. They'll join the other three base units—the second, meter, and candela (a measure of a light's perceived brightness)—that are already defined that way. The rewrite eliminates the last physical artifact used to define a unit, Le Grand K.

The shift aims to make the units more stable and allow investigators to develop ever more precise and flexible techniques for converting the constants into measurement units. "That's the beauty of the redefinition," says Estefanía de Mirandés, a physicist at BIPM. "You are not limited to one technology." But even proponents of the arcane changes acknowledge they may bewilder nonexperts. "Cooler heads have said, 'What are we going to do about teaching people to use this?'" says Jon Pratt, a physicist at the U.S. National Institute of Standards and Technology (NIST) in Gaithersburg, Maryland.

Metric makeover

An impending vote is expected to redefine metric base units in terms of fixed physical constants.

METRIC UNIT	QUANTITY	DEFINING CONSTANT
Kilogram	Mass	Planck constant
Meter	Distance	Speed of light
Second	Time	Cesium radiation frequency
Ampere	Current	Electron's charge
Kelvin	Temperature	Boltzmann constant
Mole	Amount of substance	Avogadro constant
Candela	Luminous intensity	Efficacy of light of a specific frequency

The new SI generalizes the trade-off already exploited to define the meter more precisely in terms of the speed of light. Until 1983, light's speed was something to be measured in terms of independently defined meters and seconds. However, that year, the 17th CGPM defined the speed of light as exactly 299,792,458 meters per second. The meter then became the measurable thing: the distance light travels in 1/299,792,458 seconds. (The second was pegged to the oscillations of microwave radiation from cesium atoms in 1967.)

The new SI plays the same game with the other units. For example, it defines the kilogram in terms of the Planck constant, which pops up all over quantum mechanics. The constant is now fixed as exactly 6.62607015x10^{-34} kilogram meters squared per second. Because the kilogram appears in that definition, any experiment that previously measured the constant becomes a way to measure out a kilogram instead.

Such experiments are much harder than clocking light speed, a staple of undergraduate physics. One technique employs a device called a Kibble balance, which is a bit like the mythical scales of justice. A mass on one side is counterbalanced by the electric force produced by an electrical coil on the other side, hanging in a magnetic field.

CREDITS: (PHOTO) PTB; (DATA) INTERNATIONAL BUREAU OF WEIGHTS AND MEASURES

Corrected 7 November 2018. See full text.

Published by AAAS

To balance the weight, a current must run through the coil. Researchers can equate the mass to that current times an independent voltage generated when they remove the mass and move the coil up and down in the magnetic field.

The real trickiness enters in sizing up the current and voltage, with quantum mechanical devices that do it in terms of the charge of the electron and the Planck constant. Now that the new SI has fixed those constants, the balance can be used to mete out a slug with a mass of exactly 1 kilogram. The redefinition also effectively makes the quantum techniques the SI standards for measuring voltages and currents, says James Olthoff, a NIST physicist. Until now, the SI has defined the ampere impractically, in terms of the force between infinitely long current carrying wires separated by a meter.

But applying the complex new definitions will baffle anybody without an advanced degree in physics, argues Gary Price, a metrologist in Sydney, Australia, who used to advise Australia's National Standards Commission. In fact, he argues, the new SI fails to meet one of the basic requirements of a units system, which is to specify the amount of mass with which to measure masses, the amount of length with which to measure lengths, and so on. "The new SI is not weights and measures at all," Price says.

Metrologists considered more intuitive redefinitions, Olthoff says. For example, you could define the kilogram as the mass of some big number of a particular atom. But such a standard would be impractical, Olthoff says. Somewhat ironically, researchers have already counted the atoms in exquisitely round, 1-kilogram spheres of silicon-28 to fix an exact value for the mole, formerly defined as the measurable number of carbon-12 atoms in 12 grams of the stuff.

If approved, the new SI goes into effect in May 2019. In the short term, little will change, Pratt says. NIST will continue to propagate weight standards by calibrating its kilogram weights—although now it will do so with its Kibble balance. Eventually, Pratt says, researchers could develop tabletop balances that companies could use to calibrate their own microgram weights.

Next up is a rethink of the second. Metrologists are developing more precise atomic clocks that use optical radiation with higher frequencies than the current cesium standard (*Science*, 2 March, p. 968). They should form the basis for a finer definition of the second, De Mirandés says, perhaps in 2030.

As for Le Grand K, BIPM will keep it and will periodically calibrate it as a secondary mass standard, De Mirandés says. That's a fairly dignified end for a deposed French king. ■

PHOTO: TAYLOR WEIDMAN/LIGHTROCKET/GETTY IMAGES

HUMAN EVOLUTION

Early Mongolians ate dairy, but lacked the gene to digest it

Ancient DNA and protein help solve evolutionary riddle

By **Andrew Curry**

More than 3000 years ago, herds of horses, sheep, and cows or yaks dotted the steppes of Mongolia. Their human caretakers ate the livestock and honored them by burying the animal bones with their own. Now, analysis of deposits on ancient teeth shows that early Mongolians milked their animals as well. That may not seem surprising. But the DNA of the same ancient individuals shows that as adults they lacked the ability to digest lactose, a key sugar in milk.

The findings deepen an emerging puzzle, challenging an oft-told tale of how people evolve lactase persistence, the ability to produce a milk-digesting enzyme as adults. From other studies, "We know now dairying was practiced 4000 years before we see lactase persistence," says Christina Warinner of the Max Planck Institute for the Science of Human History in Jena, Germany. With its long history and culture of dairying, "Mongolia shows us how."

As University of Copenhagen paleoproteomicist Matthew Collins, who was not on the team, puts it, "We thought we understood everything, but then we got more data and see how naïve we were."

Most people in the world lose the ability to digest lactose after childhood. But in pastoralist populations, the story went, culture and DNA changed hand in hand. Mutations that allowed people to digest milk as adults would have given their carriers an advantage, enabling them to access a rich, year-round source of fat and protein. Dairying spread along with the adaptation, explaining why it is common in herding populations in Europe, Africa, and the Middle East.

But a closer look at cultural practices around the world has challenged that picture. In modern Mongolia, for example, traditional herders get more than a third of their calories from dairy products. They milk seven kinds of mammals, yielding diverse cheeses, yogurts, and other fermented milk products, including alcohol made from mare's milk. "If you can milk it, they do in Mongolia," Warinner says. And yet 95% of those people are lactose intolerant.

Warinner wondered whether dairying arose recently in Mongolia or whether early Mongolians had lactase persistence and then lost it in a population turnover. Ancient people there might have picked up such mutations from the famed Yamnaya herders—about a third of whom were lactase persistent—who swept east and west from central Eurasia 5000 years ago.

To find answers, her team analyzed human remains from six sites of the Deer Stone-Khirigsuur Complex, a Mongolian culture that between 1300 B.C.E. and 900 B.C.E. built burial mounds marked with standing stones. Because those nomads rarely built permanent structures, and constant winds strip

Corrected 7 November 2018. See full text.

Published by AAAS

For millennia, people on the Mongolian steppes have milked their animals despite being lactose intolerant.

away the soil along with pot fragments and trash pits, archaeological evidence for diet is scarce. So Warinner's Max Planck colleague Shevan Wilkin took dental calculus—the hard plaque that builds up on teeth—from nine skeletons and tested it for key proteins.

The calculus yielded milk proteins from sheep, goats, and bovines such as yak or cow. Yet DNA from teeth and leg bones showed the herders were lactose intolerant. And they carried only a trace of DNA from the Yamnaya, the team reports this week in the *Proceedings of the National Academy of Sciences*. "They're exploiting these animals for dairying even though they're not lactase persistent," Collins says.

That disconnect between dairy and DNA isn't limited to Mongolia. Researchers recently found milk proteins on pots at Çatalhöyük in Turkey, which at 9000 years old dates to the beginnings of domestication, 4 millennia before lactase persistence appears. "There seem to be milk proteins popping up all over the place, and the wonderful evolution we expected to see isn't happening," Collins says.

Modern Mongolians digest dairy by using bacteria to digest lactose for them, turning milk into yogurt and cheese, along with a rich suite of dairy products unknown in the Western diet. Ancient pastoralists may have adopted similar strategies. "Control and manipulation of microbes is core to this whole transformation … that enables them to have a dairying culture," Warinner says.

Geneticists are going back to the drawing board to understand why lactase persistence is common—and apparently selected for—in some dairying populations but absent in others. "Why is there a signal of natural selection if there was already a cultural solution?" asks Joachim Burger, a geneticist at Johannes Gutenberg University in Mainz, Germany.

How dairying reached Mongolia is also a puzzle. The Yamnaya's widespread genetic signature shows they replaced many European and Asian Bronze Age populations. But they seem to have stopped at the Altai Mountains west of Mongolia. "Culturally, it's a really dynamic period, but the people themselves don't seem to be changing," Warinner says. She thinks even though the Yamnaya didn't contribute their genes to East Asia, they did spread their culture, including dairying. "It's a local population that has adopted the steppe way of life."

Given these surprising results, Warinner has a new goal: To figure out just which microbes helped Mongolians digest milk. ■

Andrew Curry is a journalist in Berlin.

HUMAN EVOLUTION

Ancient DNA tracks migrations around Americas

Trove of new samples reveals expansion of Clovis hunters and mysterious 9000-year-old population turnover

By **Lizzie Wade**

For decades, scientists could describe the peopling of the Americas only in broad strokes, leaving plenty of mysteries about when and how people spread across the continents. Now, state of the art ancient DNA methods, applied to scores of new samples from around the Americas, are filling in the picture. Two independent studies, published in *Cell* and online in *Science*, find that ancient populations expanded rapidly across the Americas about 13,000 years ago. They also emphasize that the story continued in the thousands of years since, revealing previously undocumented, large-scale movements between North and South America.

The data include 64 newly sequenced ancient DNA samples from Alaska to Patagonia, spanning more than 10,000 years of genetic history. "The numbers [of samples] are just extraordinary," says Ben Potter, an archaeologist at the University of Alaska in Fairbanks. Prior to these studies, only six genomes older than 6000 years from the Americas had been sequenced. As a result, says Jennifer Raff, an anthropological geneticist at the University of Kansas in Lawrence, "The [genetic] models that we've been using to explain the peopling of the Americas have always been oversimplified."

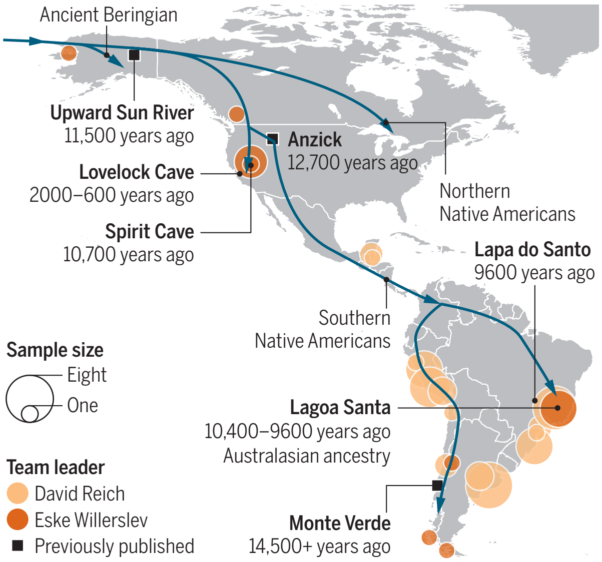
Eske Willerslev, an evolutionary geneticist at the University of Copenhagen who led the *Science* team, worked closely with the Fallon Paiute-Shoshone Tribe in Nevada to gain access to some of the new samples. The tribe had been fighting to repatriate 10,700-year-old remains found in Nevada's Spirit Cave and had resisted destructive genetic testing. But when Willerslev visited the tribe in person and vowed to do the work only with their permission, the tribe agreed, hoping the result would bolster their case for repatriation.

It did. Willerslev found that the remains from Spirit Cave are most closely related to living Native Americans. That strengthened the Fallon Paiute-Shoshone Tribe's claim to the bones, which were returned to them in 2016 and reburied. Willerslev's study validates that "this is our homeland, these are our ancestors," says Rochanne Downs, the tribe's cultural coordinator.

Willerslev added the Spirit Cave data to 14 other new whole genomes from sites scattered from Alaska to Chile and ranging from 10,700 to 500 years old. His data join an even bigger trove published in *Cell* by a team led by population geneticist David Reich of Harvard Medical School in Boston. They analyzed DNA from 49 new samples from Central and South America dating from 10,900 to 700 years old, at more than 1.2 million positions across the genome. All told, the

A trail of DNA

Two new papers add DNA from 64 ancient individuals to the sparse genetic record of the Americas. They show that people related to the Anzick child, part of the Clovis culture, quickly spread across both North and South America about 13,000 years ago.



CREDITS: (MAP) C. POSTH ET AL., CELL, 175 (2018) ADAPTED BY J. YOU/SCIENCE; (DATA) J. MORENO-MAYAR ET AL., SCIENCE 10.1126/SCIENCE.AAV2621

data decisively dispel suggestions, based on the distinctive skull shape of a few ancient remains, that early populations had a different ancestry from today's Native Americans. "Native Americans truly did originate in the Americas, as a genetically and culturally distinctive group. They are absolutely indigenous to this continent," Raff says.

The two studies also provide an unprecedented view of how ancient Americans moved across the continent beginning about 13,000 years ago. Previous genetic work had suggested the ancestors of Native Americans split from Siberians and East Asians about 25,000 years ago, perhaps when they entered the now mostly drowned landmass of Beringia, which bridged the Russian Far East and North America. Some populations stayed isolated in Beringia, and Willerslev sequenced one new example of such an "Ancient Beringian," 9000-year-old remains from Alaska's Seward peninsula. Meanwhile, other groups headed south. At some point, those that journeyed south of the ice sheets split into two groups—"Southern Native Americans" and "Northern Native Americans" (also sometimes called Ancestral A and B lineages), who went on to populate the continents (see map, p. 627).

> ***"This is our homeland, these are our ancestors."***
>
> **Rochanne Downs**, Fallon Paiute-Shoshone Tribe

By looking for genetic similarities between far-flung samples, both papers add detail—some of it puzzling—to this pattern. The 12,700-year-old Anzick child from Montana, who is associated with the mammoth-hunting Clovis culture, known for their distinctive spear points, provided a key reference point. Willerslev detected Anzick-related ancestry in both the Spirit Cave individual—who is associated with western stemmed tools, a tradition likely older than Clovis—and 10,000-year-old remains from Lagoa Santa in Brazil. Reich's team found an even closer relationship between Anzick and 9300- to 10,900-year-old samples from Chile, Brazil, and Belize.

Those close genetic affinities at similar times but across vast distances suggest people must have moved rapidly across the Americas, with little time to evolve into distinct genetic groups. Reich's team argues that Clovis technology might have spurred this rapid expansion. But anthropological geneticist Deborah Bolnick of the University of Connecticut in Storrs notes the Anzick-related ancestry group may have been broader than the Clovis people, and doubts that the culture was a driver.

Willerslev also finds traces of this Anzick-related ancestry in later samples from South America and Lovelock Cave in Nevada. But in Reich's data it fades starting about 9000 years ago in much of South America, suggesting "a major population replacement," he says.

After that population turnover in South America, both teams see striking genetic continuity in many regions. But that doesn't mean no one moved around. Reich's group sees a new genetic signal entering the central Andes about 4200 years ago, carried by people who are most closely related to ancient inhabitants of the Channel Islands, off Southern California. Meanwhile, Willerslev's team detects ancestry related to the present-day Mixe, an Indigenous group from Oaxaca in Mexico, spreading to South America about 6000 years ago and North America about 1000 years ago. Neither of these migrations replaced local communities, but rather mixed with them. Both teams say they could be seeing the same signal, but "without comparing the data, it's really hard to tell," says archaeogeneticist Cosimo Posth of the Max Planck Institute for the Science of Human History in Jena, Germany, the first author of the *Cell* paper.

Just as mysterious is the trace of Australasian ancestry in some ancient South Americans. Reich and others had previously seen hints of it in living people in the Brazilian Amazon. Now, Willerslev has provided more evidence: telltale DNA in one person from Lagoa Santa in Brazil, who lived 10,400 years ago. "How did it get there? We have no idea," says geneticist José Víctor Moreno-Mayar of the University of Copenhagen, first author of the Willerslev paper.

The signal doesn't appear in any other of the team's samples, "somehow leaping over all of North America in a single bound," says co-author and archaeologist David Meltzer of Southern Methodist University in Dallas, Texas. He wonders whether that Australasian ancestry was confined to a small population of Siberian migrants who remained isolated from other Native American ancestors throughout the journey through Beringia and the Americas. That suggests individual groups may have moved into the continents without mixing.

Delighted as they are with the data in the new studies, scientists want more. Meltzer points out that none of the new samples can illuminate what's happening at pre-Clovis sites such as Chile's Monte Verde, which was occupied 14,500 years ago. And Potter notes that, "We have a huge, gaping hole in the central and eastern North American [sampling] record. … These papers aren't the final words." ■

INFECTIOUS DISEASES

Indonesian fatwa causes immunization rates to drop

Clerics declare measles and rubella vaccine made with pork components impure

By **Dyna Rochmyaningsih**, *in Sungai Karang, Indonesia*

As the bell rang on a recent morning at an elementary school here and pupils filled the classrooms, anxious adults crowded the corridors outside. It was vaccination day, but many parents in this North Sumatra village did not want their children immunized with a new measles-rubella (MR) vaccine. Some told the teacher their children were at home, not feeling well. Others were there to make sure their kids didn't get the jab. They whispered the reason with disgust: The vaccine "contains elements of pork." By the time the vaccination team left, only six out of 38 students had been immunized.

Millions of parents around Indonesia have eschewed the vaccine in recent months, after Islamic clerics declared the MR vaccine "*haram*," or forbidden under Islamic law because pig components are used in its manufacturing. Vaccine coverage has plummeted as a result, alarming public health experts who worry that the world's largest Muslim-majority country could see new waves of measles and more miscarriages and birth defects resulting from rubella infections during pregnancy.

Indonesia has long used a locally produced measles vaccine as part of its childhood vaccination scheme, but coverage has been patchy, and until recently, the country had one of the highest measles burdens in the world, according to the World Health Organization (WHO). Last year, as part of a WHO-led plan to eliminate measles and rubella globally by 2020, Indonesia switched to a combined MR vaccine, produced by the Serum Institute of India in Mumbai. The Ministry of Health launched an ambitious catchup campaign targeting 67 million children aged 9 months to 15 years. The first phase, in 2017 on the island of Java, was a success; all six provinces

reached the 95% coverage target, and measles and rubella cases dropped by more than 90%.

But the rollout to the rest of the country, originally scheduled for August and September of this year, ran into trouble. Just before it began, the Indonesian Ulama Council (MUI) of the Riau Islands, a provincial Islamic body, raised concerns that the new MR vaccine had not been certified as "halal," or lawful, by the central MUI in Jakarta, the highest authority in such matters. The letter asked for vaccinations to be postponed. The news quickly spread throughout the country, stoking distrust among parents.

To salvage the campaign, the health ministry in August lobbied the central MUI to issue a fatwa—a ruling under Islamic law—declaring the vaccine halal. Instead, the council declared the MR vaccine *haram*, based on its ingredients and manufacturing process. Like many vaccines, it is made using several porcine components. Trypsin, an enzyme, helps separate the cells in which the vaccine viruses are grown from their glass container. Gelatin derived from pigs' skin serves as a stabilizer, protecting vaccine viruses as they are freeze-dried.

MUI took pains not to block the vaccination campaign. It ruled that parents could still have their children vaccinated, given the need to protect public health. "Trusted experts have explained the dangers posed by not being immunized," MUI said, a message it reiterated at a public consultation with Health Minister Nila Moeloek on 18 September.

But local clerics and confused parents have drawn their own conclusions. In contrast to the success on Java, coverage of children on other islands has reached only 68% so far, according to the health ministry, which did not respond to interview requests. In some regions it is far worse—just 8% in Aceh, for example, a province ruled by sharia law.

A spokesperson for WHO's country office in Jakarta notes that Indonesia is hardly the only country where trust in vaccination has eroded and says WHO remains optimistic about the campaign. Although the fatwa "has caused some confusion at local levels, it is in fact clear in its directive and ultimately supportive" of vaccination, the spokesperson wrote in an email. WHO is working with the Indonesian government, which has extended the catch-up campaign until December, to expand the coverage.

Failure could be a major setback for public health. Measles can cause deafness, blindness, seizures, permanent brain damage, and even death; vaccination coverage needs to be at 95% to reach herd immunity, in which even nonvaccinated people are protected. That threshold is about 80% for rubella. At lower levels, a paradoxical effect can occur: Some women who would otherwise have an innocuous infection early in life now catch the virus while pregnant, raising their risk of miscarriage or giving birth to babies with congenital rubella syndrome—whose symptoms include blindness, deafness, heart defects, and mental disabilities. "We can't play" with the MR vaccine, says Elizabeth Jane Soepardi, an independent public health expert who until January was director of disease surveillance and quarantine at the health ministry. Low vaccination rates "could mean a boomerang for us," she says.

There is no ready alternative; no MR vaccines have been certified as halal anywhere. (Indonesia's previous measles vaccine didn't have a halal certificate either, which has not hampered its use.) Arifianto Apin, a Muslim pediatrician in Jakarta who advocates for vaccination within the Indonesian Pediatric Society, says education may help. Clerics in many Muslim countries have concluded that gelatin in vaccines is halal because it has undergone hydrolysis, a chemical transformation that purifies it under an Islamic legal concept called *istihalah*. And in 2013, the Islamic Religious Council of Singapore declared a rotavirus vaccine halal despite the use of trypsin; it ruled that the enzyme had been made pure by dilution and the addition of other pure compounds, which is known as *istihlak*. If Muslim parents learn about the diverse legal views within Islam, Apin says, "they won't hesitate to vaccinate their children."

If that doesn't happen, the only solution is to develop a halal vaccine as soon as possible, says Art Reingold, an epidemiologist at the University of California, Berkeley. Neni Nurainy, the lead scientist at Indonesia's state-owned vaccine company, Bio Farma, in Bandung notes that nonporcine vaccine stabilizers exist, for instance; the company plans to start to investigate bovine gelatin as a replacement. But development and clinical trials could take 6 to 10 years, she says. "In the meantime, many will be made ill and some may die avoidable deaths," Reingold says.

WHO, however, is steering clear of the religious debate and won't recommend the development of a halal vaccine. "WHO works with regulatory authorities and manufacturers to ensure vaccines have the highest standards of safety and efficacy," the spokesperson says. "We don't assess vaccines on other criteria." ■

Children are immunized against measles and rubella at a school in Aceh in Indonesia, where coverage so far is only 8%.

PHOTO: FACHRUL REZA/NURPHOTO/GETTY IMAGES

Dyna Rochmyaningsih is a science journalist in North Sumatra in Indonesia.

ANIMAL WELFARE

U.S. labs using a record number of monkeys

Nonhuman primates grow in research importance, but trend alarms animal activists

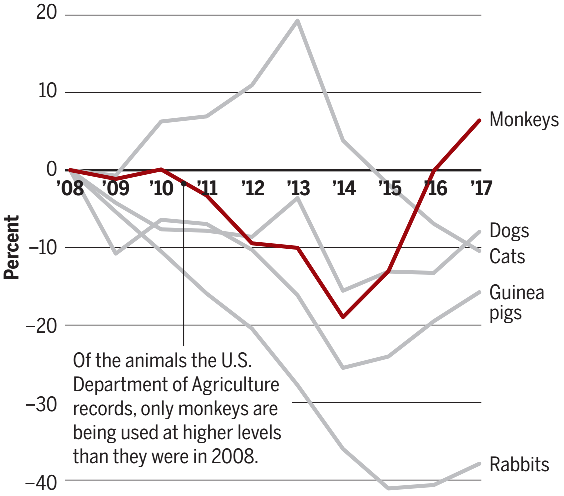
By **David Grimm**

The number of monkeys used in U.S. biomedical research reached an all-time high last year, according to data released in late September by the U.S. Department of Agriculture (USDA).

The uptick—to nearly 76,000 nonhuman primates in 2017—appears to reflect growing demand from scientists who believe nonhuman primates are more useful than other animals, such as mice or dogs, for testing drugs and studying diseases that also strike humans.

Monkeys on the rise

Trends in the use of various research animals as a percentage change over their 2008 numbers.



Of the animals the U.S. Department of Agriculture records, only monkeys are being used at higher levels than they were in 2008.

"I think the numbers are trending up because these animals give us better data. … We need them more than ever," says Jay Rappaport, director of the Tulane National Primate Research Center in Covington, Louisiana, which houses about 5000 monkeys. The increase also comes amidst a surge in funding from the National Institutes of Health (NIH), which supports much of the nonhuman primate research in the United States.

The figures have surprised and disappointed groups seeking to reduce the use of lab animals. The biomedical community has said it is committed to reducing the use of research animals by finding replacements and using these animals more selectively, says Thomas Hartung, director of Johns Hopkins University's Center for Alternatives to Animal Testing in Baltimore, Maryland. But the new numbers suggest "people are just blindly running toward the monkey model without critically evaluating how valuable it really is."

Nonhuman primate research faces intensifying scrutiny. Harvard University closed its national primate research center—one of eight in the country—in 2015, after a federal investigation into the deaths of four of its animals. That same year, NIH ended its support of all invasive chimpanzee studies, citing a report that found these animals were no longer essential to biomedical research. And in 2016, Congress directed NIH to hold a workshop on the utility and ethics of monkey research.

Public opposition to animal research has been rising—with a recent Pew Research Center poll finding that a record 52% of Americans oppose such studies. And importing monkeys to the United States has become increasingly difficult as almost all commercial air carriers now refuse to fly the animals (*Science*, 5 October, p. 15).

Yet according to the new USDA figures, scientists used 75,825 nonhuman primates for research last year, up 22% since 2015 and 6% since 2008 (see graph, below). In contrast, researchers are using cats, dogs, rabbits, and other animals recorded by USDA at lower numbers than they were a decade ago. (Nonhuman primates constitute just 0.5% of all animals used in U.S. biomedical research; about 95% are rats and mice, which are not reported by USDA.) The total number of monkeys in labs—which also includes those bred in colonies and those not currently being used in research—has remained fairly steady for the past decade, with about 110,000 recorded last year.

The uptick in monkey research "represents both the state of the science and the importance of nonhuman primates," NIH said in a statement. Nearly two-thirds of the nonhuman primates the agency supports are rhesus macaques, with cynomolgus macaques (15%), baboons (6%), and a dozen other monkey species making up the remainder. The demand for rhesus macaques appears to be driven by researcher on HIV/AIDS, the brain, Alzheimer's disease, and addiction, according to an NIH report released in September.

The rise might also reflect the agency's expanding investment in these studies. NIH gave 249 grants in 2017 that supported nonhuman primate research, up from 171 in 2013. The agency expects the number of nonhuman primates it supports to grow in coming years.

That forecast frustrates Hartung, who says NIH should launch a review of the need for monkeys, similar to the one that led it to end its support for chimpanzee research. He challenges the idea, for instance, that nonhuman primates are more useful for drug testing than rats or mice. Nonhuman primates are more genetically variable than rodents, he argues, and researchers typically use relatively few monkeys for studies of drug efficacy and safety. As a result, those experiments could yield skewed data on how the drugs will act in humans. Scientists embracing monkey experiments, he says, are at risk of "repeating the mistakes of the past."

Animal advocates hope the new statistics will move members of Congress to put pressure on U.S. agencies to reduce nonhuman primate use. "I think when Congress sees these numbers, things are going to come to a head," says Mike Ryan, director of policy and government affairs at the New England Anti-Vivisection Society in Boston. Last week, Representative Brendan Boyle (D–PA) reacted to an investigation into the Food and Drug Administration (FDA) by the Washington, D.C.–based animal activist White Coat Waste Project by sending a bipartisan letter asking FDA to review all studies involving the more than 300 nonhuman primates it oversees. "Painful primate testing … has no place in the 21st century," Boyle tells *Science*. "Federal agencies are still not doing enough to curb this appalling practice."

In the meantime, Rappaport says nonhuman primate facilities like his are simply struggling to meet the demand. Some scientists report that they have delayed studies by at least 6 months because they can't obtain animals, the NIH report notes. The growing demand could sharpen the tensions surrounding animal research. "The public wants more cures, but fewer animals," says Cindy Buckmaster, board chair of the Washington, D.C.–based Americans for Medical Progress, which supports animal studies. "They can't have it both ways." ■

CREDITS: (GRAPHIC) J. YOU/*SCIENCE*; (DATA) USDA

FEATURES

One of the deadliest box jellyfish, *Chironex yamaguchii*, collected by Angel Yanagihara off the Philippines.

STOPPING THE STING

Jellyfish almost killed Angel Yanagihara. Now, she is on a mission to save others from their fatal venom

By **Yao-Hua Law**, *in Talao-talao, the Philippines*

PHOTO: ANGEL YANAGIHARA

On 17 June, several families were celebrating Father's Day here at Dalahican Beach, a popular bathing spot near Lucena, a city on Luzon island. A steady breeze blew across sand that looked like fine brown sugar. Children splashed in the dark green water. Suddenly, people started to scream as a toddler was lifted unconscious from the water, his lips pale. A witness recalled that dark lashes crawled across the toddler's thighs—the telltale marks of a jellyfish sting. The boy's family simply held him and cried. Shortly after, Prince Gabriel Mabborang, 18 months old, was dead—one of at least three children killed in the Philippines this summer by the stings of box jellyfish.

On a midmorning 3 weeks later, Angel Yanagihara, who studies jellyfish venom at the University of Hawaii (UH) in Honolulu, arrived at Dalahican Beach. After slipping into a full-body wetsuit, she slung a box over her shoulder, put on gloves, and walked into the sea. No reminders of the recent tragedy were present; children were playing in the shallows, clapping their hands to

Filipino songs. “Hello! What’s your name?” they giggled as Yanagihara, 58, walked by. Yanagihara spent almost 3 hours wading in waist-deep waters, hoping to catch box jellyfish for her studies of their venom. One of the nearly transparent animals swam to the surface, almost within reach, but then escaped as she approached. She emerged empty-handed, but villagers had brought her two specimens earlier that day.

Among the world’s public health problems, jellyfish stings may seem trivial, affecting millions of people each year but known to kill only a few dozen. But many deaths may go unrecorded, and in some places, jellyfish stings take a real toll. Prince Gabriel was the second child killed on the same beach in the past year, and many people in the area bear the scars of nonfatal attacks. After news of the boy’s death spread rapidly on social media, Lucena health officials invited Yanagihara to talk about jellyfish venom and how to save sting victims, a service she provided for free. She spoke at a basketball court by the beach, and as she flipped to her slide on first aid, cellphones rose in a wave, snapping photos.

Her message was clear—and controversial. Yanagihara has staked out one corner in a debate over how the venom of box jellyfish kills, stopping the heart in as little as 5 minutes. What she calls her unified field theory holds that the venom contains proteins that puncture red blood cells and release potassium, disrupting the electrical rhythms that keep the heart beating. Her conclusions, and the treatments she has based on them, emerged from 20 years of science that colleagues praise as thorough and imaginative. Yanagihara “has done a great favor to the field in doing systematic comparisons” of methods to collect and study the venom, says Kenneth Winkel, a former director of The University of Melbourne’s Australian Venom Research Unit who is now at the university’s Melbourne School of Population and Global Health.

Angel Yanagihara preparing for a dive off the shore of Honolulu. She decided to study jellyfish venom after being stung during a morning swim in 1997.

PHOTO: LAURA AGUON

But nobody has independently replicated Yanagihara’s methods and findings or tested her treatments. Some jellyfish researchers say other compounds in the venom are the real killers and that different remedies—or none at all—are more likely to work. “Jellyfish venom is a graveyard for simplistic causation and therapy,” Winkel says.

Research that would resolve the debates is scarce. Worldwide, only about five research groups study jellyfish venom. Funders prefer to focus on bigger public health problems—although Yanagihara thinks the stings exact a much higher death toll than most people assume. So she and her few colleagues and competitors struggle on with small budgets to study the threat, develop remedies, and educate communities at risk.

MOST OF THE 4000 species of jellyfish cause only pain and discomfort when they sting humans. Only Cubozoans, or box jellyfish, of which some 50 species inhabit tropical and temperate seas around the globe, are fatal. They take their name from their cubic body, which has between four and 15 tentacles up to 3 meters long growing from each of the four corners. The tentacles are carpeted with hundreds of thousands of specialized cells, each harboring a capsule called a nematocyst that can fire a microscopic harpoon at speeds of more than 60 kilometers per hour. The harpoon carries a spiny hollow tube that injects venom after it strikes a victim (see graphic, p. 633).

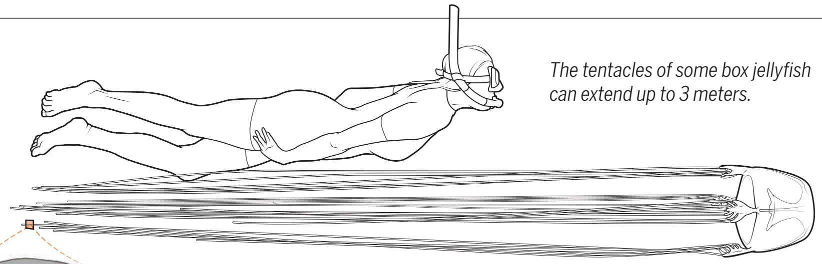
Yanagihara, born in Alaska, hadn’t planned to study jellyfish. But in 1997, the year she obtained her Ph.D. at UH for research on cellular ion channels, the jellyfish found her. One day that year, Yanagihara swam out to sea before dawn—“My father taught me to swim before I walked,” she says—when she encountered a swarm of box jellyfish some 500 meters offshore. She felt needles burning into her neck and arms and her lungs collapsing; her arms began to fail. She switched to a breathing technique she had learned for childbirth and clawed back to shore in agony, “like an automaton.” The pain kept her in bed for 3 days. After she recovered, she wanted to know what almost killed her.

In some cases, box jellyfish venom causes Irukandji syndrome, in which an overload of stress hormones and inflammation proteins produces pain and nausea for days, as well as high blood pressure that can lead to brain hemorrhage and death. Most sting casualties, however, die within minutes from cardiac arrest. The prevailing hypothesis 20 years ago was that the culprits are ion channel blockers, molecules that disrupt movement of ions in and out of cells. The blockage shuts down nerve and muscle cells, including those that keep the heart pumping.

To test the idea, Yanagihara followed a standard procedure for studying jellyfish venom: She dissolved the tentacles in water to release the nematocysts and broke them with a mortar and pestle or glass beads to release the venom. Then she exposed immature frog egg cells—a common model in cell physiology—to the venom and measured ion movement using electrophysiological techniques. But the experiments kept failing. After scrutinizing every part of her experimental setup, she began to wonder whether her venom preparation was

KILLING MECHANISM

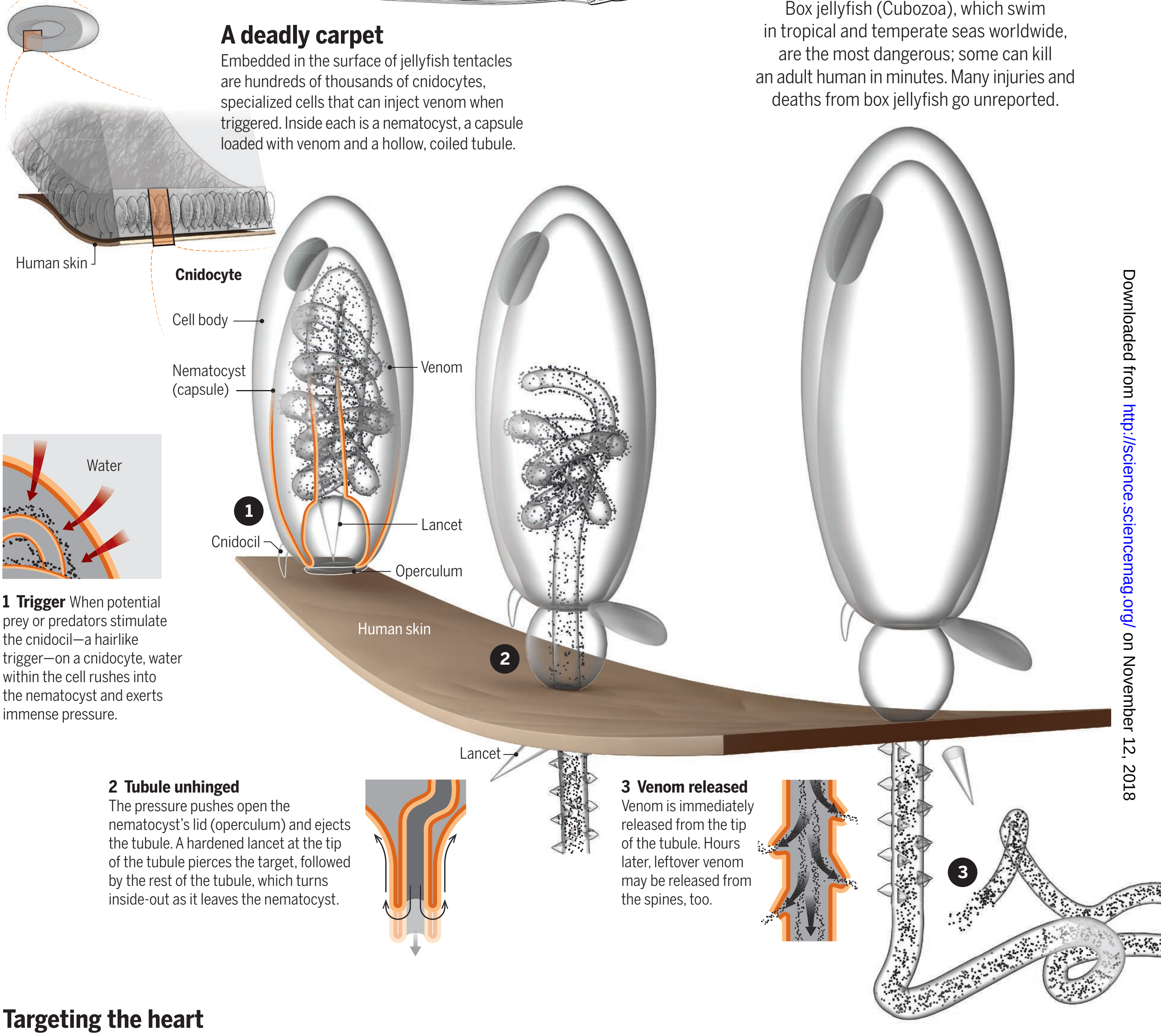
Jellyfish use venom to capture prey and to defend themselves from predators. Box jellyfish (Cubozoa), which swim in tropical and temperate seas worldwide, are the most dangerous; some can kill an adult human in minutes. Many injuries and deaths from box jellyfish go unreported.



The tentacles of some box jellyfish can extend up to 3 meters.

A deadly carpet

Embedded in the surface of jellyfish tentacles are hundreds of thousands of cnidocytes, specialized cells that can inject venom when triggered. Inside each is a nematocyst, a capsule loaded with venom and a hollow, coiled tubule.



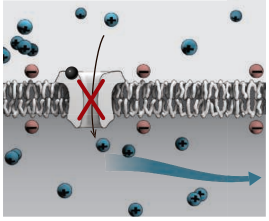
1 Trigger When potential prey or predators stimulate the cnidocil—a hairlike trigger—on a cnidocyte, water within the cell rushes into the nematocyst and exerts immense pressure.

2 Tubule unhinged The pressure pushes open the nematocyst's lid (operculum) and ejects the tubule. A hardened lancet at the tip of the tubule pierces the target, followed by the rest of the tubule, which turns inside-out as it leaves the nematocyst.

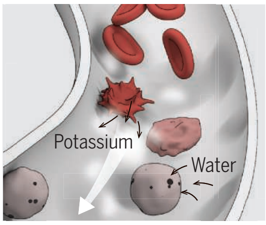
3 Venom released Venom is immediately released from the tip of the tubule. Hours later, leftover venom may be released from the spines, too.

Targeting the heart

Scientists have three different theories to explain how jellyfish venom, which attacks nerves, blood, and muscle, can cause cardiac arrest, the main cause of death after a sting.



Ion channel blockers may disrupt the flow of ions across the membranes of nerve and muscle cells, including those that keep the heart beating.



Pore-forming proteins called porins poke holes in red blood cells, releasing a flood of potassium into the blood that may cause cardiac arrest. (Hemoglobin leaves the cells as well, causing them to lose their color.)

Specific proteins in the venom may cause cardiac arrest by directly attacking muscle cells in the heart.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

too impure to reveal its secrets. She realized that crushing the nematocysts produced a crude mix of venom and cellular debris—akin to putting "a rattlesnake in a blender" to get its venom, she says.

Taking a cue from a 1970s study, she developed a new method that uses citrate, an acidic compound, to dislodge the nematocysts without breaking them. She then puts them in a French press, in which a piston forcibly ruptures all the nematocysts at once. A minuscule harvest of venom squeezes out through a tiny outlet that filters larger cellular components.

The yield is excruciatingly low: some 10 milliliters of venom from 1000 box jellyfish. (Yanagihara collects a species named *Alatina alata*, often called the sea wasp, en masse in Hawaii.) But the result, she says, is a much purer venom. In it she found not only ion channel blockers, but also many porins, proteins that puncture cells, allowing their contents to leak out. She suspected hemolysis—the destruction of red blood cells by porins—might be the fatal mechanism.

Studies supported that hunch. In a 2012 paper in *PLOS ONE*, Yanagihara and a colleague reported that venom of *Chironex fleckeri*, one of the deadliest jellyfish species, rapidly punctures red blood cells, causing them to leak a huge amount of potassium ions. A high level of potassium in the blood, or hyperkalemia, causes cardiac arrest, and when Yanagihara injected mice with high doses of venom, their hearts quickly stopped. The same happened when she injected only the porins from the venom.

In human jellyfish sting victims, however, autopsies show no signs of hemolysis, says Jamie Seymour, a prominent toxinologist at James Cook University in Cairns, Australia. He is skeptical that porins are the killers. In venom from *C. fleckeri*, his team instead found two distinct protein groups that specifically attack and kill human heart cells; those proteins are "the bit that will kill you," he says.

Seymour says he has unpublished evidence that Yanagihara's technique for collecting venom deactivates the heart toxins along with other components. Winkel, too, is skeptical. He doesn't contest that porins puncture red blood cells, but agrees with Seymour that hemolysis is not usually seen in sting victims. Porins should be tested on heart cells and tissues, he says, to find out whether they directly affect the heart.

Chironex fleckeri, one of the deadliest box jellyfish species, has left its mark on a patient's leg in North Queensland in Australia.

PHOTO: AUSCAPE/UIG/GETTY IMAGES

Yanagihara acknowledges that jellyfish venom contains other toxins, including molecules that break down lipids and proteins, but her studies convinced her that porins are the main and fastest killer. Recently, she and U.S. military researchers began to study how the venom affects piglets, which are physiologically much closer to humans than mice are. At a 2017 meeting in Florida, the group presented results showing they could reproduce both rapid death and Irukandji syndrome, depending on the dose of venom injected; the as-yet-unpublished findings also supported Yanagihara's porin hypothesis.

That hypothesis pointed to a remedy. In the *PLOS ONE* paper, Yanagihara showed that zinc gluconate inhibits porins and prolongs survival when injected into mice that had received a lethal dose of porins. Later, she found that copper gluconate works even better.

On the basis of those findings—and heeding instructions from the U.S. Department of Defense, which had funded her work—Yanagihara developed two patented products under the brand name Sting No More to counter jellyfish envenoming. A spray helps remove tentacles clinging to the skin; it contains urea, which is thought to make tentacles less sticky, and vinegar, which older studies and Yanagihara's own work had shown can deactivate unfired nematocysts. A cream containing copper gluconate is then applied to inhibit the injected venom. The products are used by U.S. military divers and sold on her website; dive shops in Hawaii carry them as well. She says she has yet to recoup her startup costs, in part because she gives the products away in developing countries.

Yanagihara has also developed simpler ways to test how well her products and other interventions inhibit porins, including a bioassay consisting of human blood suspended in agar (a gelatin derived from seaweed) overlaid with a membrane from pig intestine. A live tentacle placed on the membrane immediately pierces it and injects venom into the agar; blood cells destroyed by porins show up as white patches against the vibrant red. Winkel calls the test "the closest we have to human skin and blood, short of getting an experiment on human volunteers," and Yanagihara says it confirms her treatment's effectiveness.

"I was really impressed by the scientific rigor" in Yanagihara's methods, says jellyfish ecologist Thomas Doyle at University College Cork in Ireland. In 2016, he worked with Yanagihara to test treatments for several species in Irish waters, including the lion's mane (*Cyanea capillata*) and the Portuguese man-of-war (*Physalia physalis*), which resembles a jellyfish but belongs to a different class. Doyle and Yanagihara showed that treating stings with seawater and ice, as recommended in Irish guidelines that Doyle helped draft in 2008, actually worsens sting injury. He is now pushing to revise those guidelines.

HERE IN TALAO-TALAO, the day before her talk, Yanagihara's hotel room smelled of vinegar. Neat rows of empty spray bottles stood beside a big plastic box on the floor. Her Filipino collaborator poured 23 liters of vinegar into the box, followed by a base solution—made separately by mixing water with a blue powder—and voilà, the Sting No More spray was ready. They pumped the solution into the bottles with a long siphon, ready to be handed out.

Her talk offered an unexpected chance for a real-world test. As she started to speak, a young man who had heard about her quest for box jellyfish walked in with a live one the size of a baseball cap. Wearing only boxers—he had just come from the beach—he held the relatively harmless cubic top in his hand, at arm's length, the tentacles dangling to his knees. The audience froze in tension, while Yanagihara grabbed her spray. The man thrust the jellyfish into a bag and then jumped back when a tentacle grazed his hand. It hurt so badly that he wanted to scratch his hand off, he said. Yanagihara quickly applied her spray and cream. Three minutes later, the man said the pain had eased. He sat through the 90-minute talk.

So far, Yanagihara has only such anecdotal evidence—along with hundreds of testimonials, she says—that her products work. Together with a clinician and two nurses in Hawaii, she has started a clinical trial in which 48 volunteers will be stung on both arms with centimeter-long pieces of tentacle from *A. alata*—small enough to cause only minor damage at the sting site. One arm will then be treated with vinegar and a hot pack, the other with either Yanagihara's products or a combination of vinegar and a cold pack. (Yanagihara says she will take no part in the data collection and analysis.)

Seymour questions whether Yanagihara's antiporin cream can save lives, and he argues that her vinegar-based spray may even harm sting victims. In a 2014 paper in the journal *Diving and Hyperbaric Medicine*, he and his colleagues reported that vinegar causes nematocysts that have already fired to release more venom. He now recommends no treatment at all for sting victims suffering cardiac arrest, except cardiopulmonary resuscitation, which can help keep blood pumping to the brain until the heart starts to beat again. "I don't care if they are screaming in pain 20 minutes later," Seymour says, "as long as they are alive."

In a letter in the same journal, Yanagihara, along with a statistician, criticized Seymour's vinegar study for flaws in the design and statistics; a group of Australian physicians published a critical letter as well. Yanagihara has also blasted "wildly extrapolative" reports of the study by Australian media, which claimed vinegar might kill.

SETTLING THE DEBATE will require more research on venom pathology and treatment—plus funding, which might be easier to win if researchers could point to hard numbers on the toll of stings. Studies and media reports often cite an estimate of 150 million stings each year worldwide and 20 to 40 deaths in the Philippines annually. Those figures surfaced in a 2008 report from the U.S. National Science Foundation, but what they are based on is unclear. In a 1998 review, clinicians estimated that jellyfish kill up to 50 people in the Philippines every year, "based on personal experience," without further explanation. More recent studies tallied at least two dozen fatal and severe jellyfish stings in Malaysia and Thailand combined since 2000, almost all in tourists from abroad.

Most researchers believe the real number is much higher. The Philippines has a long, populated coastline dotted with estuaries where box jellyfish like to breed. In almost every coastal community Yanagihara has visited, locals lifted their shirts, sleeves, or pants to show scars from stings and recalled the deaths of friends and family from jellyfish. Many such cases don't make it into official statistics. Seymour says he had the same experience in the Southeast Asian nation of Timor-Leste 20 years ago: Villagers "said they get stung all the time but didn't bring the victims to the hospital," he recalls. "They pointed to a tree and said they just buried them there."

Yanagihara and her collaborators are examining health surveillance records and surveying villagers and health workers in the Philippines. "We can triangulate these results to get a better idea of the burden," says Catherine Pirkle, a UH public health epidemiologist on the project. Getting the study underway wasn't easy. The National Institutes of Health twice rejected a grant application, Yanagihara says, and local institutes and health units initially were lukewarm as well. Part of the problem may be that many communities accept the danger as part of life. "Although our fishermen and children are often stung by box jellyfish, we don't think it's a serious problem," says Reil Briones, Talaotalao's village chief, who was stung by a jellyfish at age 11 and carries a scar on his arm.

Yanagihara says the sentiment is now changing. On her latest trip, she spoke to full rooms of policymakers, health workers, and researchers, and many asked to collaborate with her team. Photos of Prince Gabriel circulating on social media may have played a role. "It's a big issue if people are dying from jellyfish," says Janet Gendrano, who leads the Disaster Risk Reduction and Management Office in Lucena. She says the tragedy was a wake-up call and wants to join the survey project; once the data are in, her office might propose an ordinance requiring beach resort operators to take first-aid training for stings and to put up warning signs.

Yanagihara hopes the study will get jellyfish the attention they deserve. "If you are a pony on this racetrack of human suffering," as many believe jellyfish stings to be, "you want to stand down," she says. "But I have nothing but evidence to the contrary." ■

Yao-Hua Law is a science journalist in Kuala Lumpur.

Angel Yanagihara gives a presentation about box jellyfish biology and stings at a rural health unit in Tagalag, a town on Samar island in the Philippines. Education and first aid can help reduce injuries and deaths, she says.

PHOTO: NOEL SAGUIL

PHOTO: ENVIRONMENTAL PROTECTION AGENCY

SCIENTIFIC INTEGRITY

Whose science? A new era in regulatory "science wars"

Proposed reforms show a clear break with historical norms

By **Wendy Wagner,[1] Elizabeth Fisher,[2] Pasky Pascual[3]**

Good laws need good science; however, good science is never guaranteed. Debate over the use of science in law is nearly as old as the laws themselves. With "science wars" waging in health and environmental regulation for at least three-quarters of a century, it is tempting to conclude that recent proposals for reforming regulatory science are similar to what has occurred in the past. They are not. They mark a sharp departure with the past because they legally constrain how agency scientists conduct the initial literature review and synthesis informing policy. Because the reforms generally take the form of legislation or regulation, they do not simply suggest best practices for conducting scientific analyses but establish legal lines that cannot be crossed. Moreover, even though they create legal ground rules for scientific deliberations, the reforms have not been developed by the scientific community, but by members of Congress and political officials. In providing a birds'-eye view of the legal developments in regulatory science over the past 50 years, we identify just how idiosyncratic these current reforms are and why the scientific community needs to be aware of their implications.

Although the agency's underlying scientific analysis is often subject to scrutiny by stakeholders and political officials and review by the courts, these new proposals cut deeper and dictate in part how the formative scientific assessments themselves must be done (*1*, *2*). For example, these proposals require the exclusion of potentially relevant research during agencies' initial review of the literature, dictate the types of computational models that must be considered in analyzing that information, and exclude respected scientists from peer reviewing the analysis (*1*, *2*). If the agency does not respect these legal lines, the agency's review of the scientific literature is legally invalid and technically illegal. This contrasts with present practice where norms governing scientific analyses are rebuttable and subject to modification in light of specific contexts and scientific progress. The proposals thus reach down to control and limit the scientific record.

The scientific community has been vocal in pointing out how the rules diverge from normal scientific practices, even while the legal requirements—some of which are still proposed and others which are final—purport to advance common goals, like data transparency and reproducibility (*3*). Editors of several journals (including *Science*), for example, recently observed how one of these proposals conflicts with current scientific norms and practices: "It does not strengthen policies based on scientific evidence to limit the scientific evidence that can inform them; rather, it is paramount that the full suite of relevant science vetted through peer review, which includes ever more rigorous features, inform the landscape of decision making"

---

[1] *Richard Dale Endowed Chair, University of Texas School of Law, Austin, TX, USA.* [2] *Professor of Environmental Law, Faculty of Law and Corpus Christi College, University of Oxford, Oxford, UK.* [3] *Data Science Consultant, Silver Spring, MD, USA. Email: wwagner@law.utexas.edu*

The Environmental Protection Agency helped to bring more science into regulation. Polluted air was a major challenge, as seen in Birmingham, AL, USA, 1972.

(*3*). Scientific analyses subject to these legally prescribed rules are thus at risk of being cordoned off from advancements within the global scientific community.

SCIENCE IN REGULATION

Science in regulation is a distinct scientific practice (*4*), in which scientists discursively interact and collaborate with lawyers, politicians, and regulatory agencies as part of a process to inform policy decisions. As public decision-makers, these parties also must be accountable. But the scientific aspects of this are scientific practices all the same.

Regulatory science generally follows a two-step process, although the steps are not always explicit (*5*). First, scientific staff review the available literature and provide a description of what it brings to the policy question at hand. Second, policy-makers can accept, ignore, rerun some of the analysis, or reinterpret the results. This bifurcation of the decision process, which produces a rigorous synthesis and analysis of the literature by scientists before the policy-makers take over, has been endorsed by the U.S. National Academies as the ideal way to make regulatory policy (*6*, p. 148).

Both steps involve judgments. Even at the first step, substantial discretion is embedded in the scientific review of the available literature. Agency experts have been called on to provide not just point estimates from their syntheses of the literature, but also explanations and descriptions about uncertainties, assumptions, and sources of judgment embedded in the analysis (*6*). The resulting regulatory science is not perfect, but when done well it signals to policy-makers where there is convergence in the available literature and the nature of the remaining uncertainties (see sidebar). It is also important that agencies' analytical methods keep up with scientific advancements, such as in computational power, or methods such as Bayesian modeling and expert elicitation.

HISTORY

The historic arc of regulatory science reveals important innovations and steady progress in providing means of holding agencies accountable for their scientific analyses (*7*). But this history also spotlights how the most recent proposals mark a departure from the reforms of the past. Rather than allowing a scientific record to be developed and then subjecting that analysis and research to scrutiny and adversarial debate, these current reforms seek to alter the agency's initial analysis of the scientific literature.

Early 1970s: Creation of science bureaucracies

The U.S. government has deployed science in the public interest since its earliest times, for example, with the creation of the patent office in 1802. Federal, science-based environmental laws were passed later, starting in the mid–20th century. Yet they were not well designed from a scientific standpoint. For example, only minimal data collection and analyses were conducted in support of federal water quality requirements. The early 1970s saw the creation of several agencies, like the Environmental Protection Agency (EPA) and other "science bureaucracies," to address serious environmental and public health problems (*4*).

Bipartisan legislation passed in the 1970s empowered these expert agencies to carry out research and set standards to advance public health and environmental protection. In these laws, Congress required agencies to base their decisions on science but provided little detail on what that meant. Scientific experts were thus entrusted with both diagnosing and solving society's environmental problems. At the same time, it quickly became clear that agencies needed to be held accountable. Agency scientists could not operate in secrecy.

Almost immediately after agencies began promulgating rules, stakeholders challenged them in court, arguing that certain rules were "arbitrary and capricious" on scientific grounds. Courts sometimes agreed and sent these defective rules back to the agency to justify its decision with evidence from its administrative record. Although a court would not rule on the science, it would insist on "the disclosure of the basis of the agency's action" so that it could determine the legal validity of the decision (*8*).

Mid-1970s to early 1990s: Emerging importance of accountability

These court decisions highlighted that agency decisions needed to be underpinned by visible scientific explanations. Regulatory agencies devoted more attention to working with the broader scientific community to ensure the scientific integrity of their work. This included the creation of the EPA's Science Advisory Board in 1974, composed of outside experts who review the agency's technical analyses (*4*). Agencies also relied on external peer review to ensure their decisions were consistent with the scientific evidence and were not unduly influenced by affected parties (*4*). The changes occurred against a background in which decisions of the EPA were becoming increasingly politically controversial.

Court challenges continued against EPA and other agencies' rules. These cases reinforced the imperative that the agencies not only provide a scientific basis for their regulatory action, but also explain how that evidentiary basis was used to make decisions. For example, if the agency's analyses revealed a range of exposures that might produce health hazards from a pollutant, the agency was expected to explain, with scientific support, why it chose an exposure standard outside the range (*7*).

Risk assessment was developed during this time as part of a way of structuring the overall regulatory process (*6*). The important point about the developments of this period is that they were not directly defining science per se. They were regulating the administrative processes to make them more accountable in legal and administrative terms. Judicial pressure impelled the EPA to develop more rigorous analytical processes in response to consistent legal challenges (*7*).

Early 1990s to mid-2010s: Increased opportunities to challenge the agency's scientific record and scientific analyses

Although court challenges allowed stakeholders to hold agencies accountable for their use of science, they did not allow stakeholders to challenge the scientific record itself as it was being developed. Reforms during this time period opened up opportunities for stakeholders to challenge the rigor of this underlying science, although generally not in ways that were enforced by courts. The Information Quality Act, for example, was passed in 2001 to afford aggrieved parties the right to challenge the reliability of information used by an agency at any point, regardless of whether that information was being used to inform a rule or policy. In another piece of legislation, Alabama Senator Richard Shelby inserted a single sentence into a 4000-page budget bill requiring federally funded researchers to provide their data to anyone who requested them under the Freedom of Information Act, explicitly targeting data underlying a controversial but well-regarded (and reviewed) epidemiological study of health impacts of fine particulate air pollutants (the Harvard Six Cities study) (*9*).

Political meddling with the agency's scientific record had been a concern in the previous time periods, but during this period, there were more frequent news reports of partisan intervention with agency scientists' own underlying analyses. For example, Julie A. MacDonald, former deputy assistant secretary at the U.S. Department of the Interior, was investigated for unduly influencing field biologists' assessments of the research informing the listing of endangered species and ultimately resigned (*10*). The White House, under several presidents, was caught altering or censoring staff technical reports in ends-oriented ways (*11*).

Published by AAAS

Evolution of regulatory science behind National Ambient Air Quality Standards (NAAQS)

For 50 years, EPA, working with the scientific community, has developed increasingly rigorous methods for synthesizing the literature to inform the agency's mandate to set ambient air quality standards to protect public health. Part of this evolution is credited to judicial oversight of the administrative process. When EPA failed to explain its scientific analyses or conclusions, courts overturned its standards. We summarize below the evolution of regulatory science of the NAAQS process for particulate matter (PM) using the four stages described in this paper.

Early 1970s

Creation of science bureaucracies. An ad hoc committee of experts from universities, industry, and government synthesized laboratory and epidemiological studies (primarily from New York and London data) and recommended ambient air quality standards for the nation.

Mid-1970s to early 1990s

Emerging importance of accountability. EPA staff took on the role of conducting scientific analyses, which political appointees then used to make decisions about appropriate standards. To ensure the rigor of the staff analysis, EPA and ultimately Congress created a formal, long-term science advisory board to peer review the staff's work [the Clean Air Scientific Advisory Committee (CASAC)]. EPA, working iteratively with the CASAC, collated literature on air pollution and summarized the evidence.

Early 1990s to mid-2010s

Increased opportunities to challenge the agency's scientific record and scientific analyses. Mounting litigation prompted the EPA to streamline and enhance the rigor and transparency of its scientific assessment process. Pursuant to a restructuring of the process in 2006, the EPA now holds a planning workshop for interested parties, conducts a scientific literature search, and prepares risk and exposure assessments with multiple scenarios. The EPA evaluates the strength of evidence of various studies and summarizes the distribution of health effects. The staff also summarizes the results of these technical assessments for nonscientists. All of this work is captured in separate reports, each of which undergoes external scientific and public review. Each of these staff assessments is also firewalled from political control and communication.

Mid-2010s to the present

Efforts to control how scientific analysis and peer review are conducted. If the proposed legal reforms discussed here are finalized, EPA's analyses would be legally altered at the literature review and synthesis stage by rules that proscribe exclusionary tests and other mandated practices that affect how scientists identify and synthesize the available literature. EPA is already making important changes to the composition and other features of external peer review by the CASAC.

The ensuing bad publicity underscored that even though it sometimes occurred, political manipulation of the agency's scientific record was considered off-limits (*11*).

Overall, this time period signaled a movement toward greater adversarial pressures on the scientific analysis used for regulation. Additional legal tools made the agency's scientific record more vulnerable to challenge, but it is important to note that these tools did not alter the initial scientific record itself. Scientists would still use their professional standards and methods to determine how to conduct literature searches and analyze the available literature.

During this period, some regulatory programs even required that the agency's initial scientific review and synthesis be firewalled from policy staff and political officials to maintain a strict separation, in keeping with the National Academies' recommendation (*5*), (*6*, p. 148). These developments reinforced the importance of scientific integrity. To ensure that no inappropriate pressure was brought to bear on a decision, there was a need to define the scientific basis of decision-making in clear terms. For example, this greater focus on scientific integrity is exemplified by the program to establish National Ambient Air Quality Standards (see sidebar). Additionally, one of President Obama's first actions when he took office in 2009 was to establish a scientific integrity initiative that sought to create stronger protections for the independence of agency scientists.

Mid-2010s to the present: Efforts to control how scientific analysis and peer review are conducted

Presently, the proposed reforms of regulatory science aim to change the nature of the scientific deliberations and underlying record itself. They target the agency's initial scientific analysis and synthesis and prescribe substantial constraints on how this literature review and synthesis must be done. They also alter the composition of the science advisory boards that review these staff analyses. Yet, despite reaching deep down into how scientists assess the available literature, these proposed reforms do not emerge from the scientific community. They are proposed by congressmen and political appointees in the agencies and crafted largely without input or advice of science advisers and mainstream scientific organizations (*3*).

For example, proposed reforms in Congress and the agencies prohibit agency scientists from including studies in their synthesis of the literature unless the "dose response data and models…are publicly available in a manner sufficient for independent validation" (*1*). Under the EPA's proposed version of this directive, any exceptions to this transparency requirement must be made by the EPA administrator. Exceptions are narrow and expressly limited to when the disclosure of data is infeasible because of "privacy, confidentiality, confidential business information, and… national and homeland security" (*1*).

H.R. 1430, legislation that passed the House without amendment in 2017, would establish even more far-reaching restrictions on the nature of information that the agency can consider. Under the terms of the bill, at least the following items underlying "all scientific and technical information" must be "publicly available online in a manner that is sufficient for independent analysis and substantial reproduction of research results" before that information can be used in the agency's scientific analysis supporting a decision: "(i) materials, data, and associated protocols necessary to understand, assess, and extend conclusions; (ii) computer codes and models involved in the creation and analysis of such information; (iii) recorded factual materials; and (iv) detailed descriptions of how to access and use such information."

These proposed legal standards would apply to all research used to inform a regulatory decision. They even extend to research conducted before the standards were put in place, apparently irrespective of whether compliance with the standards is feasible or even technically possible. The standards if passed as laws would inevitably be enforced by the courts. Lawyers and judges would use these legislated standards to determine what information is needed to allow a study to be capable of validation or replication. As laws, moreover, these legal pronouncements governing scientific deliberations would remain binding and enforceable until they are amended by Congress or formally revised by the appropriate agency.

Changes are also being made to how agency analyses are peer reviewed. Historically, when agencies establish peer-review panels, the composition of reviewers remained flexible and endeavored to enlist the nation's top experts. At the same time, agencies required reviewers to disclose potential conflicts of interest to advance the goals of transparency and balance (*4*).

By contrast, a 2017 EPA directive by Administrator Pruitt (*2*) and H.R. 1431, a second bill passed by the House without amendment in 2017, both impose a flat prohibition on the nature of the experts who can serve as peer reviewers of the EPA's analyses. EPA's directive, for example, decrees that "no member of a federal advisory committee [may] currently receive EPA grants, either as principal investigator or co-investigator" (*2*). In this rule, industry experts with a stake in the proceeding are not excluded from serving on matters in which they or their employers have a financial interest; only researchers holding EPA grants (regardless of the size of the grant) are excluded (*2*). To our knowledge, there is no precedent for such a unilateral exclusion of federal grantees as peer reviewers in either existing regulatory practice or in the practices of scientific publishers or federal granting agencies. Since EPA issued the directive, at least a few respected scientists have been removed from EPA science advisory boards because they were not willing to abandon their EPA-funded research.

> ***"...science should be allowed to bring its best work to the table."***

The impacts of the proposed reforms on scientific analysis and the newly enacted rules on peer review are not trivial. For example, if in the wake of EPA's proposed transparency rule, EPA considers a study in its analysis for reasons the scientific community might generally view as meritorious, but the data are not available to the satisfaction of the law, the scientific analysis cannot be used unless the administrator explicitly exempts it (*1*). And if the top researcher in the country is tapped to help review a staff analysis, but that researcher has an EPA grant, he or she is legally prohibited from doing so.

These initiatives would, in a legally enforceable manner, constrain agencies in determining the best science to fulfill their statutory mandates. They would also limit the ability of scientific staff to use scientific judgment in individual cases; to adopt science innovations that conflict with these proposed legal rules; or to work more generally with the global community of scientists. As such, the proposed legal rules would substantially alter the terms of the open-ended scientific deliberations running through the history of regulatory science. It is important to remember that throughout this history, the agencies were required to explain and disclose the scientific basis of their action as a matter of law.

Moreover, the new reforms introduce legalized requirements for scientific terms, like transparency and reproducibility, that currently are at best in flux and at worst still largely undefined within the scientific community (*3*). Even proposals within science that sound similar to the proposed reforms are actually very different. In 2014 for example, a committee of academicians, scientific publishers, and funding agencies developed principles to encourage sharing of data, code, research methods and materials, and replication of studies (*12*). But these principles apply only to encourage better practices for research—they do not suggest that synthesis of the literature, including past studies, should be limited to research that meets these criteria (*13*). Moreover, even in the research context, this committee recognized that these principles were not universally applicable to all science investigations and described them as aspirational, providing flexibility on how they would be implemented.

COMMUNAL PRODUCT

As of late September, EPA's proposed transparency rule garnered almost 600,000 comments on the agency's docket, as well as numerous commentaries in leading journals. This substantial feedback may explain in part why the agency recently announced that it will need more than a year to finalize the rule. Some commentators celebrate EPA's proposal, as well as the other reforms discussed here, as a move to replace EPA's open-ended scientific deliberations with legally mandated reproducible science standards. Others warn the proposal will "prohibit the agency from using a wide swath of high-quality, past and present scientific research" (*14*).

While debates rage about whether the proposed statutes and regulations for science reform will ultimately improve or diminish the quality of scientific evidence, what is clear is that in setting legal standards for what the scientific basis for decision-making is, these proposals mark a clear departure from the past. Scientists not only need to take notice of that fact, but they need to be part of the debate.

Regulatory science cannot and should not be isolated from policy, but science should be allowed to bring its best work to the table. We can all agree with former EPA Administrator Pruitt's statement that "[w]hatever science comes out of EPA, shouldn't be political science" (*15*). The issue before us is whether this goal will be compromised if legal reform reaches into the earliest stages of the agency's scientific synthesis to narrow the ground rules for these deliberations. Although defining good science has been assigned largely to agency experts, they did not do their work in isolation. It was not only "their" science. The agencies' work benefited from, and was peer reviewed by, the global science community. The agencies' analysis was also expected to keep pace with scientific progress. Finally, agency experts were held accountable to the courts for the choices they made. In the end, the agency's formative scientific analysis was essentially a communal product of science that attempted to summarize what the available scientific information suggests for pressing policy questions of the day. To ignore attempts by politically elected and appointed individuals to dictate how science should be conducted is to betray the very essence of science. ■

REFERENCES AND NOTES

1. *Fed. Reg.* **83**, 18768, 18773 (2018).
2. E. S. Pruitt, "Strengthening and Improving Membership on EPA Federal Advisory Committees" (EPA, 21 October 2017), p. 3.
3. J. Berg *et al.*, *Science* **360,** eaau0116 (2018).
4. S. Jasanoff, *The Fifth Branch: Science Advisers as Policymakers* (Harvard Univ. Press, 1998).
5. W. Wagner, "Science in Regulation: A Study of Agency Decisionmaking Approaches," U.S. ACUS Report (2013); www.acus.gov/report/science-regulation-final-report.
6. National Research Council, Risk Assessment in the Federal Government: Managing the Process (National Academies Press, 1983).
7. E. Fisher *et al.*, *Texas Law Review* **93**, 1681 (2015).
8. *Kennecott Copper Corp.* v. *Environmental Protection Agency* 462 F.2d 846, 849 (DC Cir. 1972).
9. B. Johnson, Republicans Wage Anti-'Secret Science' Campaign Against The EPA, *Huffington Post*, 25 June 2014.
10. U.S. Department of the Interior, Office of Inspector General, "Investigative Report:The Endangered Species Act and the Conflict between Science and Policy" (2008).
11. H. Doremus, *Texas Law Rev.* **86**, 1601 (2008).
12. B.A. Nosek *et al.*, *Science* **348**, 1422 (2015).
13. National Research Council, Access to Research Data in the 21st century: An Ongoing Dialogue Among Interested Parties: Report of a Workshop (National Academies Press, 2002).
14. A. Q. Hoy, Scientific, Medical, Academic Groups Urge EPA to Drop 'Transparency' Rule (16 July 2018); www.aaas.org/news/scientific-medical-academic-groups-urge-epa-drop-transparency-rule.
15. U.S. EPA, Press Release (21 October 2017); www.epa.gov/newsreleases/administrator-pruitt-issues-directive-ensure-independence-geographic-diversity.

ACKNOWLEDGMENTS

We thank anonymous reviewers and participants at Columbia University's Interdisciplinary Conference on Evidence held in April 2017 for valuable feedback and comments. P. P. was a former employee at the U.S. Environmental Protection Agency. W.W. and E. F. declare no competing interests.

10.1126/science.aau3205

Food allergens, such as peanuts, can cause anaphylaxis.

PERSPECTIVES

IMMUNOLOGY

Dangerous liaisons in anaphylaxis

Dendritic cells deliver allergens to mast cells in the skin via microvesicles

By **Francesca Levi-Schaffer**[1]
and **Jörg Scheffel**[2]

In 1902, Richet and Portier tried to immunize a dog with a toxic sea anemone (*Actinia*) extract. However, to their surprise, the dog died 25 minutes after the second injection. They named the harmful effect anaphylaxis (from Greek, lack of protection) (*1*). Awarded with the Nobel Prize in 1913, Richet explained that the two necessary and sufficient features for anaphylaxis to develop are an incubation time in between injections and two injections of the same “poison.” Thereafter, the often-life-threatening phenomenon of systemic anaphylaxis (or anaphylactic shock) was found to be caused rapidly after exposure to several environmental antigens (allergens) that enter the circulation and are usually derived from certain foods, drugs, insect venoms, latex, and immunotherapy injections. Similar to all allergic reactions, anaphylaxis is induced by rapid, immunoglobulin E (IgE) antibody–mediated release of immune mediators, such as histamine and platelet-activating factor (PAF) from tissue-resident mast cells (MCs) and presumably from peripheral blood basophils (*2*, *3*). However, in contrast with other forms of allergy such as asthma, allergic rhinitis, and atopic dermatitis, anaphylaxis is systemic and life-threatening owing to cardiovascular and respiratory symptoms (*2*). Allergic and anaphylactic reactions imply that allergens from the circulation cross epithelial and/or endothelial (the vascular lining) barriers to interact with IgE bound to MCs. However, the mechanisms were unknown. On page 656 of this issue, Choi *et al.* (*4*) elegantly demonstrate that such allergens are trapped by dendritic cells (DCs) that produce microvesicles (MVs) to interact with MC-IgE in the skin.

MCs are the cardinal cells of allergy, strategically present perivascularly throughout the body in both connective and mucosal tissues (*5*). Importantly, MCs are highly granulated, containing several preformed proinflammatory mediators [such as histamine, tryptase, and tumor necrosis factor (TNF)] and express FcεRI, the high affinity receptor for IgE. When MC-bound IgEs encounter a specific allergen, FcεRIs are cross-linked, leading to immediate cell activation, degranulation, and production and release of lipid metabolites. Subsequently, several cytokines and chemokines are released in a second wave, which, together with the activation of infiltrated inflammatory cells, cause symptoms similar to the immediate response.

In the model proposed by Choi *et al.*, conventional DCs (cDCs) of the cDC2 subset, and to a lesser extent cDC1s, are pivotal in anaphylaxis by shuttling allergens to IgE-bearing MCs. DCs are members of the professional antigen-presenting cell (APC) family, which also includes Langerhans cells and macrophages in the skin. Various phenotypical and functionally distinguishable dermal DC subsets have been identified in humans and mice (*6*), with cDC2 being the most abundant subset in the dermis. DCs originate from the bone marrow and engraft all nonlymphoid tissues, where they “sample” their environment by taking up foreign- or host-derived material (antigens). The ingested material is processed and presented as small peptides via major histocompatibility complex (MHC) class II molecules to $CD4^+$ T cells in lymphoid tissues, such as the lymph nodes.

Choi *et al.* found that allergen-sensitized, cDC2-deficient mice had significantly less-severe anaphylaxis when challenged intravenously with allergen. They discovered that cDC2s continuously probe adjacent blood vessels for allergens, with extensions projecting through the endothelial wall. Importantly, perivascular MCs were unable to probe the cutaneous circulation for allergens by themselves, as has been described for their acquisition of IgE (*7*). In contrast to the usual antigen-presentation process, a major portion of the allergens were not internalized by the cDC2s but were instead passed on to allergen-specific IgE bound to MCs via MVs (see the figure). MVs originate from the plasma membrane, range between 40 and 1000 nm in size, and are frequently shed from the membrane of cDC2s and other cells (*8*). In their experiments, the process of MV shedding was extremely rapid, reflecting the fast onset of symptoms during anaphylaxis. Direct cell-cell contact between MCs and DCs may also occur despite not being essential and thus not the primary means of MC activation in their models. Notably, MVs were also used to distribute allergens to neighboring macrophages and DCs, thereby spreading the antigens in the tissue. Choi *et al.* further showed that MV budding from the plasma membrane was dependent on the adenosine triphosphatase activity of vacuolar protein sorting–associated protein 4 (VPS4). Consequently, mice deficient for VPS4 in DCs exhibited much milder anaphylaxis when challenged with allergen. Choi *et al.* suggest that promiscuous scavenger or man-

---

[1] Pharmacology and Experimental Therapeutics Unit, School of Pharmacy, Institute for Drug Research, Faculty of Medicine, The Hebrew University of Jerusalem, Post Office Box 12065, Jerusalem 91120, Israel. [2] Department of Dermatology and Allergy, Charité Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. Email: francescal@ekmd.huji.ac.il

PHOTO: BRUNO CRESCIA PHOTOGRAPHY INC/GETTY IMAGES

nose receptors facilitate allergen binding to the plasma membrane of the cDC2s. In their model, blocking or depleting the mannose receptor in DCs reduced their capacity to activate MCs via MVs.

The global incidence of anaphylaxis is increasing, particularly in response to food-derived allergens. The study of Choi *et al.* strengthens the role of skin MCs in anaphylaxis, but the role of blood basophils (which are the first to encounter allergens) remains unclear (*2*, *3*). A number of other open questions also remain. Notably, the human and the mouse systems might work differently. In mice, IgG can also induce systemic anaphylaxis. In addition, the nature of the allergen might influence the mechanism. Moreover, besides mannose receptors, which other receptors facilitate allergen binding? Do DCs discriminate between internalization, processing, and MHC presentation versus antigen spreading via MVs? The described mechanism also suggests a high turnover rate of membrane lipids and proteins in DCs to sustain the permanent shedding of MVs from the plasma membrane. How do these cells keep up with the production of new plasma membrane? How specific are MVs in their cellular interactions? What happens to the allergen-loaded MVs in the tissue, do nonimmune cells interact with them, and what are the consequences? Much still remains unknown about MV biology.

The work of Choi *et al.* adds an important piece to the puzzle of how allergens become distributed to MCs in the tissues to start the dangerous chain reaction that leads to anaphylaxis. MCs are spread out in the body. Are similar pathways triggered in other tissues involving other DC subsets? Is this newly discovered mechanism a possible target for therapeutic intervention in anaphylaxis? Probably not. Sentinel functions of DCs and other APCs fulfill an important role in protecting the host from pathogens and internal threats (*9*). Therefore, interference bears the risk of serious side effects. Current options for anaphylaxis include avoidance, immunotherapy, and immediate epinephrine injection followed by antihistamines and corticosteroids (*10*). In the future, directly targeting MCs or IgE might be the best prophylactic strategy for anaphylaxis (*5*). ■

REFERENCES AND NOTES

1. C. Richet, Nobel lecture: Anaphylaxis, 11 December 1913; www.nobelprize.org/prizes/medicine/1913/richet/lecture/.
2. L. L. Reber *et al.*, *J. Allergy Clin. Immunol.* **140**, 335 (2017).
3. H. Karasuyama, K. Miyake, S. Yoshikawa, Y. Yamanishi, *J. Allergy Clin. Immunol.* **142**, 370 (2018).
4. H. W. Choi *et al.*, *Science* **362**, eaao0666 (2018).
5. R. S. Gangwar, N. Landolina, L. Arpinati, F. Levi-Schaffer, *Pharmacol. Ther.* **170**, 37 (2017).
6. S. W. Kashem, M. Haniffa, D. H. Kaplan, *Annu. Rev. Immunol.* **35**, 469 (2017).
7. L. E. Cheng *et al.*, *Immunity* **38**, 166 (2013).
8. G. Raposo, W. Stoorvogel, *J. Cell. Biol.* **200**, 373 (2013).
9. P. Matzinger, *Annu. Rev. Immunol.* **12**, 991 (1994).
10. F. E. Simons *et al.*, *World Allergy Organ J.* **8**, 32 (2015).

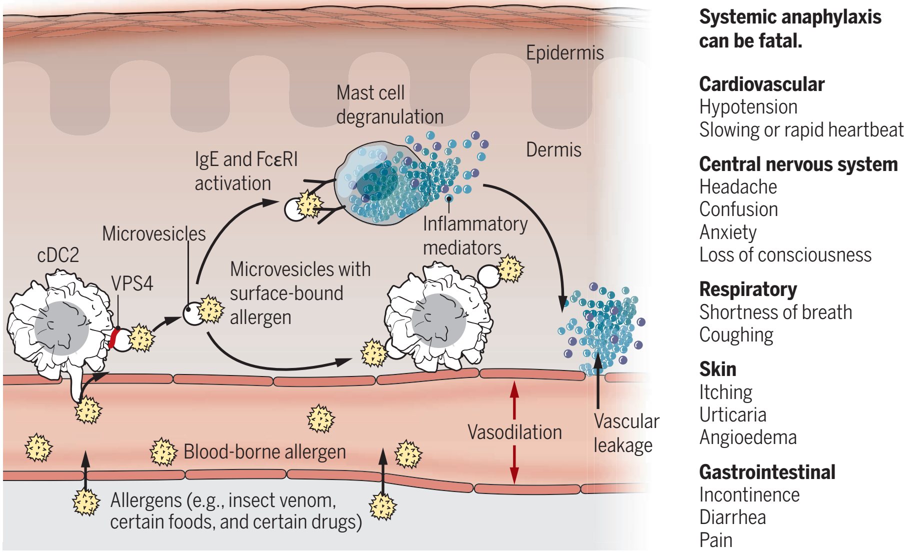
ACKNOWLEDGMENTS

F.L.-S. is funded by the Israel Science Foundation (grant 472/15), the Aimwell Charitable Trust (UK), Emalie Gutterman Memorial Endowed Fund for chronic obstructive pulmonary disease related research (USA), and the Rosetrees Trust (UK).

10.1126/science.aav4505

The allergen shuttle

cDC2s in the skin continuously sample allergens from the circulation and distribute them via microvesicles to APCs or IgE-coated mast cells. This results in mast cell degranulation and immediate release of various proinflammatory mediators. The simultaneous degranulation of mast cells in different body locations leads within minutes to systemic symptoms and may cause death of the individual.



GRAPHIC: N. DESAI/SCIENCE

IMMUNOLOGY

A key ingredient for priming killer T cells

WDFY4 protein controls the generation of antiviral and antitumor CD8⁺ T cell immunity

By **Gaëtan Barbet** *and* **J. Magarian Blander**

For the adaptive immune system to detect intracellular infection, microbial protein fragments (antigens) must be presented by major histocompatibility complex class I (MHC-I) molecules. This entails an elaborate intracellular folding of MHC-I molecules with peptides from the infecting pathogen, followed by their display on the cell surface for surveillance by $CD8^+$ T cells (*1*). In a variation of this classical pathway of MHC-I antigen presentation, a subset of professional antigen-presenting cells called dendritic cells (DCs) conduct cross-presentation, whereby MHC-I molecules are directed to present exogenous peptides from internalized microbes and dying infected cells or cancer cells (*2*). Because of the potential to prime cytotoxic $CD8^+$ T cells against intracellular pathogens and cancer cells, cross-presentation has preoccupied immunologists for decades. But how this process is orchestrated remains poorly understood. On page 694 of this issue, Theisen *et al.* (*3*) reveal a critical role for the protein WDFY4 (WD repeat- and FYVE domain–containing protein 4) in cross-presentation and not in classical MHC-I antigen presentation. The authors show that WDFY4 works specifically in the conventional DC1 (cDC1) subset and is responsible for their specialization at cross-presentation (*4*). This revelation identifies WDFY4 as a therapeutic target to mobilize cytotoxic $CD8^+$ T cells against tumors and infected cells.

WDFY4 is a large transmembrane protein and member of the BDCP (BEACH domain–

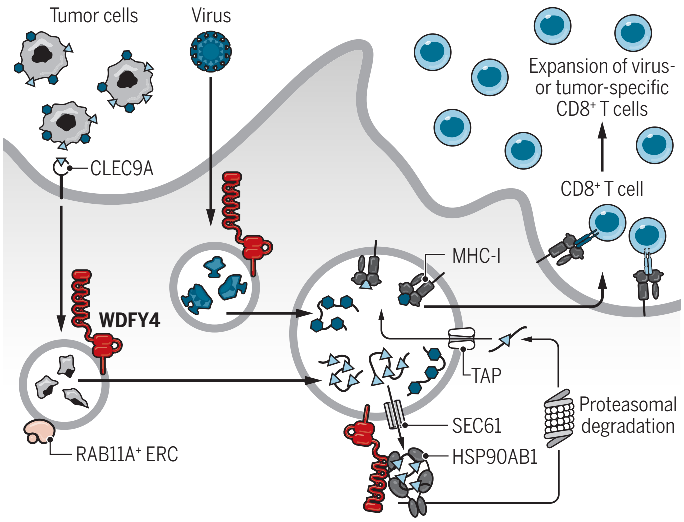
Jill Roberts Institute for Research in Inflammatory Bowel Disease; Gastroenterology and Hepatology Division, Joan and Sanford I. Weill Department of Medicine; Department of Microbiology and Immunology; Sandra and Edward Meyer Cancer Center; and Immunology and Microbial Pathogenesis Program, Weill Cornell Graduate School of Medical Sciences, Weill Cornell Medicine, Cornell University, New York, NY, USA. Email: jmblander@med.cornell.edu

containing protein) family. Theisen *et al.* identified a role for WDFY4 in cross-presentation, and in mice lacking WDFY4 (*Wdfy4*$^{-/-}$) they found that cDC1s developed normally, allowing investigation of the precise role of cross-presentation in immunity. All previous mouse models for studying cDC1 functions lack cDC1 cells because they rely on deletions of transcription factors or growth factors important for cDC1 development (*5*).

Wdfy4$^{-/-}$ mice were unable to mount an antiviral CD8$^+$ T cell response to cowpox or West Nile virus infection, but they survived an infection with the parasite *Toxoplasma gondii*, concordant with intact cDC1 production of interleukin-12 (IL-12) and tumor necrosis factor (TNF), two inflammatory cytokines that mediate *Toxoplasma* clearance. Moreover, *Wdfy4*$^{-/-}$ mice could not reject a highly immunogenic fibrosarcoma or mount antitumor CD8$^+$ T cell responses despite intact tumor infiltration by cDC1s. These observations provide direct experimental evidence for the importance of cross-presentation by cDC1s in antiviral and antitumor immunity. Continued investigations using *Wdfy4*$^{-/-}$ mice should cement the physiological relevance of cross-presentation, which was questioned mainly because of a lack of appropriate in vivo models.

cDC1 cross-presentation depends on WDFY4

Specialized receptors on cDC1s, such as CLEC9A, route antigen to RAB11A$^+$ ERC, where WDFY4 is highly expressed. WDFY4 anchors HSP90AB1 to endosomal membranes to facilitate endosome-to-cytosol translocation of antigen in preparation for its degradation, loading onto MHC-I molecules, and presentation to CD8$^+$ T cells.



Twenty single-nucleotide polymorphisms in *WDFY4* have been identified in Asians with the autoimmune disease systemic lupus erythematosus (*6*). Although mutations in other BDCPs have similarly been linked to human diseases, their precise functions have remained unknown (*7*). BDCPs are proposed to mediate membrane fission and fusion (*7*), events that are inherent in cross-presentation during antigen internalization (endocytosis), subcellular routing, and MHC-I trafficking (*8*). Through their BEACH domains, BDCPs are thought to attach to specific phospholipid or protein moieties on endolysosomal membranes to create large platforms for binding partner proteins, and to carry out specific membrane-associated functions (*7*).

WDFY4 bound to the cytosolic chaperone heat shock protein 90-β (HSP90AB1) and the protein transport protein SEC61A1, both of which have ascribed roles in cross-presentation (*8*). HSP90AB1 interacts with endocytosed proteins as they translocate across endosomal membranes into the cytosol to undergo proteasomal degradation, a step that is integral to cross-presentation (*9*). Potential interaction of SEC61A1 with WDFY4, although not validated by Theisen *et al.*, is also interesting. SEC61A1 is a subunit of the SEC61 complex that imports newly synthesized proteins into the endoplasmic reticulum (ER), but also mediates reverse ER-to-cytosol transport of misfolded proteins for proteasomal degradation (*8*). Enrichment of SEC61A1 with WDFY4 revives the debated involvement of the SEC61 complex in endosome-to-cytosol protein transport. These findings strengthen a model whereby WDFY4 serves as a scaffold to localize HSP90AB1 to the SEC61 complex during cytosolic translocation of internalized antigens (see the figure). Furthermore, these events are tightly coupled to the activity of the transporter for antigen processing (TAP), which mediates reentry of proteasome-degraded peptides into phagosomes and endosomes for loading onto MHC-I (*8*). Indeed, Theisen *et al.* show that cDC1 cross-presentation of cell-associated antigen is dependent on TAP.

High homology between mouse and human WDFY4 indicates a conserved cross-presentation pathway in cDC1s that is different from another DC subset, monocyte-derived DCs (moDCs). This difference may reflect the divergent programs of cross-presentation in cDC1s and moDCs (*10*), but may also relate to the "constitutive" nature of cross-presentation in cDC1s as opposed to its induction by inflammatory signals in other DC subsets (*11*). Through its ability to interact with vesicular trafficking proteins, WDFY4 provides cDC1 cross-presentation with a special pass that might also bypass help from inflammatory signals.

Theisen *et al.* noted that WDFY4 localizes to the plasma membrane and overlaps with the location of the small guanosine triphosphatase (GTPase) RAB11A; this finding suggests localization to endosomal recycling compartments (ERCs). In splenic cDC1s, ERCs have been noted to harbor large pools of MHC-I molecules (*11*). Coupled with the expression of specialized receptors such as CLEC9A, reported to route internalized antigen into ERCs (*12*), WDFY4 might also serve to route antigens to specialized intracellular compartments where several critical components of cross-presentation converge. Further studies are needed to pinpoint the subcellular localization of WDFY4 in cDC1s, follow it along antigen and MHC-I trafficking routes, and elucidate precisely how WDFY4 mediates superior cross-presentation by cDC1s. Understanding WDFY4 function should clarify the molecular underpinnings of CD8$^+$ T cell cross-priming against intracellular pathogens and tumors.

Mounting a potent CD8$^+$ T cell response that can kill infected cells and cancer cells remains one of the main challenges for vaccination. Most vaccines preferentially induce antibody responses that neutralize extracellular microbes (*13*). Present strategies to combat intracellular pathogens, such as malaria and HIV, target the extracellular phases of infections. For cancer therapeutics, success of one of the most promising drug types, checkpoint inhibitors (which reawaken antitumor CD8$^+$ T cells to attack cancer cells), is contingent on the presence of tumor-specific CD8$^+$ T cells (*14*). Thus, identifying ways to manipulate cross-presentation to generate CD8$^+$ T cells would inform effective vaccine design for cancer and also for infectious diseases such as tuberculosis, Ebola, malaria, and AIDS. ■

REFERENCES

1. J. S. Blum *et al.*, *Annu. Rev. Immunol.* **31**, 443 (2013).
2. J. M. den Haan *et al.*, *Curr. Opin. Immunol.* **13**, 437 (2001).
3. D. J. Theisen *et al.*, *Science* **362**, 694 (2018).
4. J. M. den Haan *et al.*, *J. Exp. Med.* **192**, 1685 (2000).
5. D. A. Anderson *et al.*, *Cold Spring Harb. Perspect. Biol.* a028613 (2017).
6. W. Yang *et al.*, *PLOS Genet.* **6**, e1000841 (2010).
7. A. R. Cullinane *et al.*, *Traffic* **14**, 749 (2013).
8. J. M. Blander, *Annu. Rev. Immunol.* **36**, 717 (2018).
9. R. J. Binder, *J. Immunol.* **193**, 5765 (2014).
10. C. G. Briseño *et al.*, *Cell Rep.* **15**, 2462 (2016).
11. P. Nair-Gupta *et al.*, *Cell* **158**, 506 (2014).
12. S. Zelenay *et al.*, *J. Clin. Invest.* **122**, 1615 (2012).
13. I. Delany *et al.*, *EMBO Mol. Med.* **6**, 708 (2014).
14. R. W. Jenkins *et al.*, *Br. J. Cancer* **118**, 9 (2018).

10.1126/science.aav3683

GRAPHIC: A. KITTERMAN/*SCIENCE*

Neonicotinoid exposure can affect bumblebee behavior in their nest.

PHOTO: NIGEL E. RAINE

ECOLOGY

Pesticide affects social behavior of bees

Neonicotinoid exposure impairs the social dynamics of bumblebees inside the nest

By **Nigel E. Raine**

Bees are critically important to agricultural crop production and to the reproduction of most flowering plant species on the planet (*1*). Yet, these essential ecosystem service providers are in decline around the world (*1*, *2*). Widespread pesticide use associated with increasingly intensive agriculture is one of several, likely interacting, factors that contribute to these concerning pollinator declines (*2*). Although insecticide applications are targeted at controlling pests, their use can have unintended impacts on beneficial insects, including bees. As the most widely used class of insecticides in the world, neonicotinoids have come under considerable scrutiny following concerns around their nontarget impacts on bees (*3*). On page 683 of this issue, Crall *et al.* (*4*) identify how exposure to these neurotoxic insecticides can adversely affect individual bumblebees and social dynamics within their colony.

Using an innovative automated, robotic platform, Crall *et al.* continuously monitored the behavior of all workers inside multiple bumblebee colonies, each housed in a specially constructed nest box attached to a foraging chamber (see the figure). They found that environmentally realistic exposure to the neonicotinoid imidacloprid, dissolved in artificial nectar collected from the foraging chamber, resulted in measurable behavioral changes in workers inside the nest: They were less active, less likely to feed and care for larvae (act as nurses), and more likely to be found toward the periphery of the nest, compared to workers in control colonies. Furthermore, these changes were stronger at night, potentially relating to daily patterns in pesticide consumption or detoxification; this requires further investigation.

In spring, bumblebee queens emerge from hibernation and seek nest sites. Once a site has been chosen, each queen tries to establish her colony. She secretes wax to form the nest structure, including pots to store nectar and pollen. The queen then lays eggs that will become workers. Until these hatch, the queen is a single mother performing all the foraging, incubation of eggs, larval feeding, and nest construction and cleaning. When adult workers emerge, they take over most of these tasks. Larger workers typically undertake foraging and other tasks outside the colony, whereas smaller workers nurse their sister larvae and perform other roles inside the nest. Although the sight of bees foraging on flowers is more familiar to us, these roles inside the nest are equally important to colony success.

Environmentally realistic exposure to neonicotinoids (imidacloprid and thiamethoxam) has been shown to affect multiple aspects of worker foraging behavior, including their flower choices, foraging trip duration, and ability to collect pollen (*5, 6*). Such impacts on foraging performance can delay or reduce colony growth by limiting the quantity of pollen entering the nest to support larval growth (*7*). However, studies exposing bumblebee colonies to the neonicotinoid clothianidin, or an alternative systemic insecticide (sulfoxaflor), report impacts on colony development (*8, 9*) and reproductive output (*9*), without observing any significant impacts on pollen foraging performance. A potential alternative mechanism for these results is provided by the observations of Crall *et al.*, suggesting that recording behavioral dynamics and interactions inside the nest could be an important component of future assessments of pesticide exposure on bumblebees. It will be interesting to see whether the automated monitoring system they developed can operate under less controlled conditions.

Impacts of neonicotinoid exposure on nursing behavior might also limit the flow of nutrients to some or all larvae, functionally acting on colony growth in the same way as limiting pollen flow into the nest through forager impairment. If this is the case, it might be possible to detect greater differentials in pollen storage between treated and control colonies, assuming pollen is coming into all colonies at similar rates but being fed to larvae less quickly in those exposed to insecticide. Differences in the spatial position of both larvae and workers within the nest also have important consequences for bumblebee colony dynamics. Larvae found further from the

---

School of Environmental Sciences, University of Guelph, Guelph, Ontario, N1G 2W1, Canada. Email: nraine@uoguelph.ca

center of the nest receive less food from workers, so they develop into smaller adults (*10*). If neonicotinoid-exposed workers move further away from the center of the nest and are less likely to engage in nursing behavior, this might substantially affect food intake rates of peripheral larvae. Future experiments could examine whether this leads to higher mortality of these larvae, or changes the body size distribution of workers, which can already vary 10-fold within a single colony that is not exposed to pesticides (*10*).

Rates of larval development and survival might be further affected by impaired thermoregulation in response to neonicotinoid exposure (*4*). This would likely have the greatest impact in early spring, when temperatures are lowest and thermoregulation relies on the queen and, if present, a few workers. Queens exposed to neonicotinoids under laboratory conditions have significantly lower rates of colony establishment (*11*); such impacts would likely be exacerbated if thermoregulation was affected under less benign conditions in the field.

The approach of Crall *et al.* offers exciting future opportunities to examine impacts of pesticide exposure on the earliest, and potentially most vulnerable, phase of the bumblebee life cycle—by monitoring the queen's behavior during nest establishment and as the colony grows. Understanding the impacts of pesticide exposure on bumblebee queens might be particularly important if this could bridge concerns around differences in chemical sensitivity between social and solitary bee species (*12*). Current ecotoxicology risk assessments for pesticide regulation use the highly social honey bee (*Apis mellifera*) to represent potential impacts for all insect pollinators. However, the highly social honey bee lifestyle is atypical as the majority of bee species are solitary, with a small number of bumblebee species (~250 worldwide) being social for part of their annual life cycle.

When considering relatively subtle behavioral impacts of long-term, chronic exposure to pesticides, the dynamics of large (honey bee) or small (bumblebee) colonies might act to buffer and conceal the appearance of adverse effects. However, females from solitary bee species, like early spring bumblebee queens, are overworked single mothers, making it more likely that a small behavioral change as a result of pesticide exposure might have a measurable impact on their ability to produce as many high-quality offspring. The work by Crall *et al.* also offers important potential avenues to improve risk assessments of impacts on bees for any pesticide. ■

REFERENCES AND NOTES

1. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Assessment Report on Pollinators, Pollination and Food Production; www.ipbes.net/assessment-reports/pollinators.
2. A. J. Vanbergen *et al.*, *Front. Ecol. Environ.* **11**, 251 (2013).
3. H. C. J. Godfray *et al.*, *Proc. R. Soc. B* **281**, 20140558 (2014).
4. J. D. Crall *et al.*, *Science* **362**, 683 (2018).
5. R. J. Gill, N. E. Raine, *Funct. Ecol.* **28**, 1459 (2014).
6. D. A. Stanley, *J. Appl. Ecol.* **53**, 1440 (2016).
7. R. J. Gill *et al.*, *Nature* **491**, 105 (2012).
8. A. N. Arce *et al.*, *J. Appl. Ecol.* **54**, 1199 (2017).
9. H. Siviter *et al.*, *Nature* **561**, 109 (2018).
10. M. J. Couvillon, A. Dornhaus, *Proc. R. Soc. B* **276**, 2411 (2009).
11. G. L. Baron *et al.*, *Nat. Ecol. Evol.* **1**, 1308 (2017).
12. M. Arena, F. Sgolastra, *Ecotoxicology* **23**, 324 (2014).

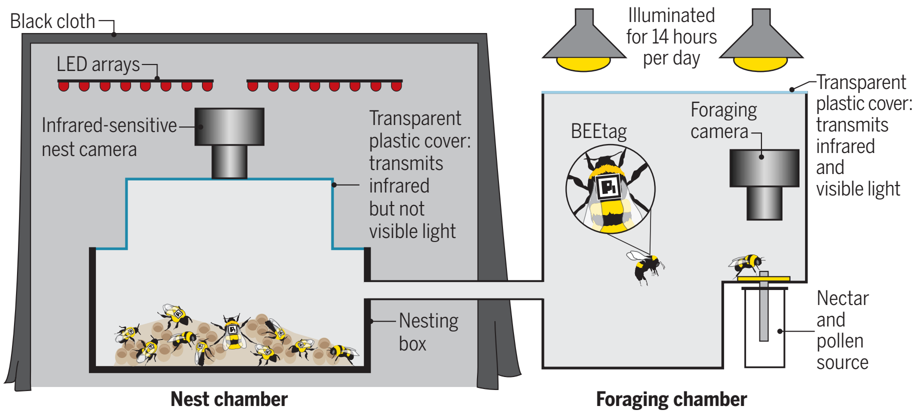
ACKNOWLEDGMENTS

N.E.R. is supported as the Rebanks Family Chair in Pollinator Conservation by the W. Garfield Weston Foundation, Natural Sciences and Engineering Research Council of Canada, Ontario Ministry of Environment and Climate Change, and Canada First Research Excellence Fund "Food from Thought" initiative.

10.1126/science.aav5273

Assessing behavior inside the colony

A computer-controlled, robotic observation platform with a mobile camera array enables automated periodic video monitoring of all bumblebees individually marked with BEEtags within 12 colonies. Schematic not to scale.



BIOCHEMISTRY

Recruiting more proteins to the RNA world

A primordial but still essential ribozyme co-opts proteins as it evolves

By **William G. Scott**[1] *and* **Kiyoshi Nagai**[2]

Ribonuclease P (RNase P) recognizes precursor transfer RNA (pre-tRNA) and processes it to generate mature tRNAs that are used for assembling proteins. Unlike almost all other enzymes, RNase P is a ribozyme, an enzyme with an active site that is composed of RNA, and it is present in every living organism. RNase P is among the most ancient of enzymes, a living molecular fossil from an "RNA world" in which life is thought to have originated. On page 657 of this issue, Lan *et al.* (*1*) present structures of the yeast RNase P enzyme by itself and bound to its pre-tRNA substrate. Additionally, the structure of the human form, by itself and bound to its tRNA product, is reported by Wu *et al.* (*2*). These reveal the detailed mechanism by which RNase P hydrolyzes pre-tRNA to produce the required 5′-phosphorylated tRNA of exactly the correct length. These structures unambiguously reveal how an assortment of proteins conspire to form a measuring device that ensures that the pre-tRNA substrate is correctly processed by the catalytic RNA subunit of this universal and essential enzyme.

RNase P was discovered by Robertson, Altman, and Smith in 1972 (*3*), who described an enzyme in the bacterium *Escherichia coli* that precisely removes the 5′ end of pre-tRNAs to produce mature tRNA products,

CREDITS: (GRAPHIC) N. CARY/*SCIENCE*; (PHOTO) MARK RALSTON/STAFF/GETTY IMAGES

Life is proposed to have begun with an RNA or RNA-like molecule.

the adaptor molecules that are required for translating the genetic code into proteins. A decade later, Altman, Pace, and colleagues published a remarkable discovery: The RNA subunit of bacterial RNase P is the active catalytic component (*4*). Before this discovery, and the nearly simultaneous discovery by Zaug and Cech of self-splicing RNA (*5*, *6*), it was assumed that only proteins could be enzymes. Altman and Cech soon shared the Nobel Prize, and added renewed traction to the "RNA world hypothesis" (*7*), which proposes that life began with RNA or an RNA-like molecule that can carry both genetic information and catalytic activity.

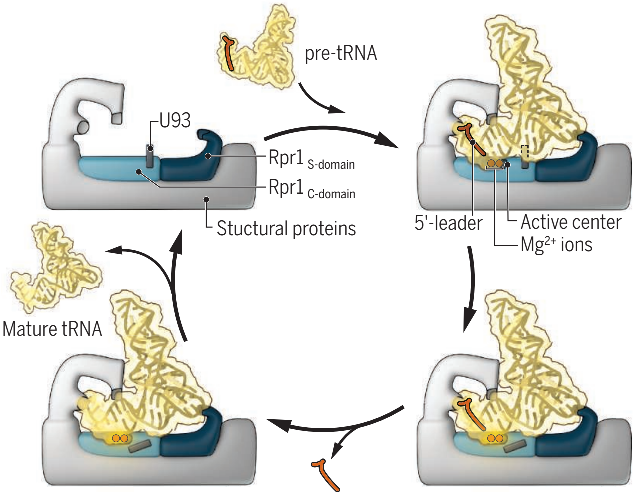
Although the RNA subunit in bacterial RNase P is catalytic, the enzyme also has a single protein subunit that appears to play an ancillary role in substrate recognition. More complex versions of RNase P are found in *Archaea* and *Eukarya*. Although they all possess similar RNA subunits, the archaeal and eukaryotic versions also contain multiple proteins. Are these more complicated RNase P enzymes also ribozymes? Or has evolution replaced the catalytic component with more conventional protein active sites? The new structures of yeast and human RNase P finally answer this question and in so doing provide a fascinating glimpse of the evolutionary transition between the RNA and protein worlds.

In addition to the RNA subunit, yeast RNase P has 10 protein subunits. The yeast RNase P RNA has a more elaborate structure compared with its more compact bacterial counterpart, but it shares a similar central region (*8*, *9*). In the yeast enzyme, the proteins form a "hook" that wraps around the RNA subunit, stabilizing the entire structure. The resulting protein-RNA complex forms a precise measuring device that targets pre-tRNAs for hydrolysis by the enzyme's active site, which is composed entirely of RNA (see the figure). The new structures show that yeast RNase P, despite being adorned with proteins, is indeed still a ribozyme.

Similar to bacterial RNase P, the active site of the yeast enzyme reveals two Mg^{2+} cations coordinated by adjacent nonbridging phosphates within the backbone of the RNA subunit, as well as the phosphate at the cleavage site of the pre-tRNA. One of the Mg^{2+} ions is also coordinated by a conserved nucleotide base in the RNA, as previously observed in the bacterial enzyme (*8*, *9*)—apparently the only nucleotide base that participates directly in catalysis. Many of the other known ribozymes also have active sites with two Mg^{2+} cations, which is consistent with an early but insightful prediction (*10*); they are observed in every ribozyme involved in RNA splicing, including the spliceosome (*5*, *6*, *11*–*14*). In the case of RNase P, one of the Mg^{2+} ions activates a water molecule to hydrolyze (break) the phosphodiester backbone of the pre-tRNA substrate. This geometrical arrangement of the reactants explains not only how the phosphate comes to reside on the 5′ end of the product tRNA as required but also how a more typical RNA degradation reaction, which could result from phosphodiester isomerization, is prevented. Lan *et al.* also provide a thorough and compelling set of computational reaction simulation results, demonstrating the plausibility of their proposed tRNA cleavage mechanism.

Eukaryotic RNase P tRNA processing

The yeast RNase P consists of ten structural proteins and one RNA subunit (Rpr1). The Rpr1 specificity (S) domain and the protein subunits form a measuring device that positions the pre-tRNA at the active center. The Rpr1 catalytic (C) domain contains a conserved nucleotide (U93) that coordinates one of two Mg^{2+} ions, which are involved in catalyzing pre-tRNA cleavage.



Like the yeast enzyme, human RNase P has one RNA (H1) and 10 homologous proteins. Therefore, the architecture of the holoenzyme is similar to that of yeast RNase P, and the mechanism of tRNA recognition is conserved. However, in the human structure, the 5′ end of the bound tRNA product is disordered, and the structure does not reveal bound catalytic metal cations (*2*).

Much of our prior knowledge of the recognition mechanism of RNase P was based on the crystal structure of bacterial RNase P, in complex with the mature tRNA reaction product (*9*). The single protein subunit in the bacterial RNase P guides binding of the 5′ leader sequence of pre-tRNA. The multiple protein components in yeast and human RNase P stabilize the catalytic RNA subunit and similarly guide pre-tRNA binding.

Many intriguing questions about the origin, evolution, and function of RNase P sharpen in focus with these new structures. Why has eukaryotic RNase P recruited many more noncatalytic protein subunits, given that bacterial RNase P, with a single protein subunit, is optimized for tRNA processing? Do these proteins extend the functionality of RNase P in processing other structural RNAs in eukaryotic cells? A multitude of other, entirely protein-based RNase enzymes exist, even including some forms of RNase P (*15*). Despite this, with one intriguing exception to the rule (*16*), the ribozyme version of RNase P is found in every cell of every organism, implying not only that this ribozyme is indeed a living fossil that dates back to the original life forms, but that the catalytic RNA subunit has been rigorously conserved during several billion years of evolution instead of being replaced with recruited proteins. What is it about this RNA that is so indispensable? The same type of hydrolysis reaction can be catalyzed by a variety of proteins that possess an RNase III catalytic domain. The new structures provide us with many intriguing hints. ■

REFERENCES

1. P. Lan *et al.*, *Science* **362**, eaat6678 (2018).
2. J. Wu *et al.*, *Cell* 10.1016/j.cell.2018.10.003 (2018).
3. H. D. Robertson *et al.*, *J. Biol. Chem.* **247**, 5243 (1972).
4. C. Guerrier-Takada *et al.*, *Cell* **35**, 849 (1983).
5. A. J. Zaug, T. R. Cech, *Cell* **19**, 331 (1980).
6. A. J. Zaug, T. R. Cech, *Science* **231**, 470 (1986).
7. W. Gilbert, *Nature* **319**, 618 (1986).
8. A. V. Kazantsev *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13392 (2005).
9. N. J. Reiter *et al.*, *Nature* **468**, 784 (2010).
10. T. A. Steitz, J. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6498 (1993).
11. N. Toor *et al.*, *Science* **320**, 77 (2008).
12. S. M. Fica *et al.*, *Nature* **503**, 229 (2013).
13. R. Wan *et al.*, *Science* **353**, 895 (2016).
14. W. P. Galej *et al.*, *Nature* **537**, 197 (2016).
15. J. Holzmann *et al.*, *Cell* **135**, 462 (2008).
16. A. I. Nickel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11121 (2017).

10.1126/science.aav4743

---

[1]*Department of Chemistry and Biochemistry and The Center for the Molecular Biology of RNA, University of California at Santa Cruz, Santa Cruz, CA 95064, USA.* [2]*MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK. Email: wgscott@ucsc.edu; kn@mrc-lmb.cam.ac.uk*

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

QUANTUM OPTICS

Scaling up solid-state quantum photonics

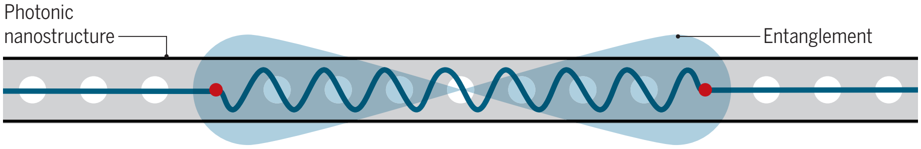
Radiative coupling creates an entangled state between two silicon vacancies in diamond

By **Peter Lodahl**

A deterministic interface between a single atom and a single optical photon is the essential building block underpinning many quantum applications of light for quantum communication, sensing, and simulations. Light and matter interact weakly with each other, so the challenge is to create conditions enabling strong interactions. Fortunately, solid-state quantum photonics has matured dramatically, and it is possible to create artificial photonic nanostructures that markedly enhance light-matter coupling. Moreover, single atoms, which are cumbersome to control experimentally because they need to be trapped and cooled, can be replaced by solid-state quantum emitters such as vacancy centers in diamond, molecules, or quantum dots (*1*, *2*). The high quality and purity of these systems now imply that coherent and near-deterministic photon-emitter interfaces are routinely constructed (*2*), but it is still challenging to scale up and deterministically couple multiple quantum emitters. On page 662 of this issue, Evans *et al*. (*3*) report on the successful coupling of two diamond silicon vacancy (SiV) quantum emitters mediated by their mutual coupling to a nanophotonic cavity. Radiative coupling leads to the formation of an entangled state between the two emitters (see the figure).

Radiative coupling with a photonic nanostructure

Two solid-state quantum emitters (red dots) are mutually entangled by their strong radiative coupling, mediated by a photonic nanostructure. The indicated oscillating wave packets illustrate the spatial extent of the electromagnetic field radiated by each emitter.



The essential figure of merit for a photon-emitter interface is the cooperativity *C*. It quantifies the probability *P* that a narrow-bandwidth single photon launched in a nanophotonic waveguide or cavity will interact with the emitter as $P = C^2/(1 + C)^2$. Cooperativity determines the obtainable fidelity of photon-photon quantum gates and quantum nonlinearities that can be mediated with a deterministically coupled emitter (*4*), and is maximized by decreasing the interaction volume, reducing loss of the resonator, or by suppressing spontaneous emission leakage or other decoherence processes. The limit of a deterministic photon-emitter interface occurs for $C \gg 1$, and steady progress of solid-state systems has resulted in extremely high *C* values, up to several hundred for self-assembled quantum dots. Evans *et al*. report an impressive improvement of cooperativity for SiV centers in a diamond nanocavity by more than an order of magnitude over previous work (*5*) to reach $C \sim 20$.

Such state-of-the-art photon-emitter interfaces can produce very long strings of photonic qubits, and by controlling a coherent ground-state spin, the photons can be created on demand in a multiphoton entangled state (*6*). However, scaling up these systems to couple multiple emitters and generate more than a single string of photonic qubits requires overcoming inhomogeneities, such as structural differences between emitters or internal strain. Evans *et al*. combated such intrinsic inhomogeneity with magnetic field tuning and preselection of near-identical emitters. They entangled two SiV color centers in a diamond nanocavity and controllably coupled the spin ground states. This work could be extended to a mature heterojunction semiconductor platform that would allow local tuning of each quantum emitter, e.g., with a local electric field (*7*).

Even a few efficiently coupled quantum emitters can have exciting applications. A single ground-state spin in an emitter can be used for a quantum gate operation on another coupled spin. Although the manipulation of a single spin in a quantum emitter enables multiphoton entanglement generation, the addition of a spin-spin gate operation could lead to the deterministic preparation of arrays of multiphoton entangled states, notably two-dimensional photonic cluster states (*8*). This constitutes a universal resource for measurement-based quantum computing, where a multiqubit entangled state is prepared and quantum algorithms can be implemented with just single-qubit measurements on the entangled state (*9*).

For quantum communication, an outstanding challenge is to construct a quantum repeater for distributing quantum information over long distances in the presence of unavoidable losses over optical fibers. All-photonic quantum repeaters have been proposed where quantum information is encoded and protected in multiphoton entangled states (*10*). Explicit protocols for generating such all-photonic quantum-repeater states on demand were put forward recently (*11*), where only two coupled quantum emitters are required.

Such approaches exploit the strengths of the solid-state quantum photonics platform ideally, because only a few matter qubits are required that rapidly can emit dozens of high-quality photonic qubits before the system decoheres. Alternative quantum-repeater architectures can be pursued if the solid-state quantum emitter is efficiently coupled to a long-lived quantum memory, such as nuclear spins (*12*). Solid-state quantum photonics systems can be tailored to these applications because each emitter can generate many photons. Thus, the ultimate dream of a quantum internet (*13*) may have come a step closer. ■

Hy-Q Center for Hybrid Quantum Networks, Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, DK-2100 Copenhagen, Denmark. Email: lodahl@nbi.ku.dk

REFERENCES

1. I. Aharonovich, D. Englund, M. Toth, *Nat. Photonics* **10**, 631 (2016).
2. P. Lodahl, S. Mahmoodian, S. Stobbe, *Rev. Mod. Phys.* **87**, 347 (2015).
3. R. E. Evans *et al.*, *Science* **362**, 662 (2018).
4. L.-M. Duan, H. J. Kimble, *Phys. Rev. Lett.* **92**, 127902 (2004).
5. A. Sipahigil *et al.*, *Science* **354**, 847 (2016).
6. N. H. Lindner, T. Rudolph, *Phys. Rev. Lett.* **103**, 113602 (2009).
7. R. J. Warburton, *Nat. Mater.* **12**, 483 (2013).
8. S. E. Economou, N. Lindner, T. Rudolph, *Phys. Rev. Lett.* **105**, 093601 (2010).
9. R. Raussendorf, H. J. Briegel, *Phys. Rev. Lett.* **86**, 5188 (2001).
10. K. Azuma *et al.*, *Nat. Commun.* **6**, 6787 (2015).
11. D. Buterakos *et al.*, *Phys. Rev. X* **7**, 041023 (2017).
12. N. Kalb *et al.*, *Science* **356**, 928 (2017).
13. H. J. Kimble, *Nature* **453**, 1023 (2008).

10.1126/science.aav3076

GRAPHIC: N. DESAI/*SCIENCE*

Republican president Theodore Roosevelt made conservation a priority, establishing five national parks, including Crater Lake, Oregon.

PHOTO: POSNOV/GETTY IMAGES

ENVIRONMENTAL POLICY

Why conservatives abandoned conservation

Today's Republicans see environmental problems, and those raising the alarm, differently

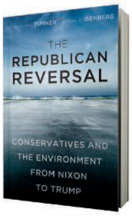
By **Neal D. Woods**

During the 2016 U.S. presidential campaign, Donald Trump suggested that, if elected, he would roll back environmental regulations, open public lands up for natural resource extraction, and pull the United States out of the Paris Accord, calling global warming a "hoax." Once in office, he followed through on this campaign rhetoric with a speed and intensity that was greeted with cheers by most Republicans in Congress.

The irony is that the major environmental laws that today's Republican Party seeks to weaken were widely championed by Republicans of an earlier era. Although not himself an environmentalist, President Richard Nixon established the Environmental Protection Agency in 1970 and led the push for several laws that form the foundation of the modern environmental state. Astonishing as it may seem today, the Clean Air Act (1970), Clean Water Act (1972), and Endangered Species Act (1973) were each passed without a single dissenting vote in the Senate. Even ardent conservatives supported a government role in environmental protection; Senator Barry Goldwater, for instance, cosponsored the bill that became the Clean Air Act.

The reviewer is at the Department of Political Science and the School of the Earth, Ocean and Environment, University of South Carolina, Columbia, SC 29208, USA. Email: woodsn@mailbox.sc.edu

What caused the shift? In *The Republican Reversal*, James Turner and Andrew Isenberg focus on three characteristics of the Republican Party that changed over time: (i) a shift from viewing environmental issues as urgent to viewing them as alarmist and exaggerated; (ii) a shift from relying on scientific research and expertise to viewing these entities with suspicion; and (iii) a shift from embracing a central role of government in addressing environmental problems to viewing regulations as a threat to economic growth, individual freedom, and free enterprise.



The Republican Reversal
James Morton Turner and Andrew C. Isenberg
Harvard University Press, 2018. 280 pp.

The most original portion of the book juxtaposes the conservative and environmental movements in the 1960s, offering new insights into their coevolution. The remainder traverses ground that has largely been covered elsewhere, most notably in Judith Layzer's excellent *Open for Business*. However, Turner and Isenberg deliver a sure-handed account, deftly distilling a complex array of political motives and actions into a lucid narrative.

The Republican Reversal posits three broad reasons for the sharp shift in Republican sentiment. The first is the ascendance of conservative ideology in the Republican Party, which, as it relates to the environment, marries the "dominion theology" of some evangelical Christian churches with technological optimism, a suspicion of scientific elites, and a strong pro–free market/antigovernment orientation. Over time, conservatives espousing this ideology have replaced moderate Republicans in Congress, many of whom had advanced environmental legislation.

Second, and relatedly, the authors argue that the Republican reversal reflects the growing importance of conservative interest groups in shaping the party's agenda. In addition to corporations and trade associations, these include law firms, think tanks, and other organizations that often act in concert to constrain the regulatory state.

The final factor pertains to the way environmental problems themselves have changed since the 1970s. This theme, which is less fully articulated than the other two, focuses on how environmental protection issues have been transformed from immediate, observable problems to problems that are more abstract and global and whose effects may not be fully felt for some time.

The Republican Reversal is a well-written account of the Republican Party's dramatic transformation on environmental policy over the past 40 years. Although there will inevitably be topics that readers will wish had received greater attention—my list would include the role of conservative ideas [such as the cornucopian ideas of Julian Simon (*1*) and the development of market-driven alternatives to traditional regulation]—the book does an excellent job of weaving the important facets into a clear, compelling narrative. ■

REFERENCE

1. J. Simon, *The Ultimate Resource* (Princeton Univ. Press, 1981).

10.1126/science.aav2324

SCIENCE & SOCIETY

What’s next for humanity?

With an emphasis on data science and biotechnology, a historian looks to the future

By **Dov Greenbaum**[1,2] *and* **Mark Gerstein**[2]

Historian Yuval Noah Harari’s new book is his third in a trilogy that examines the annals of humanity. Divided into five parts and 21 chapters, it discusses the near future, as extrapolated from our current trajectories.

Like his other bestsellers, *21 Lessons for the 21st Century* is meant for a broad audience, and Harari effortlessly jumps between diverse topics, from biology and information science to history, religion, and philosophy. But despite the wide-ranging subject matter, many scientists will also find substantial value in reading this book because it places recent scientific and technological developments within the broader arc of history.

Following a theme from his second work, Harari focuses on how current innovations in information science and biotechnology will affect humanity in the not-too-distant future. “When the biotech revolution merges with the infotech revolution, it will produce Big Data algorithms that can monitor and understand my feelings much better than I can,” he predicts, adding, worryingly: “and then authority will probably shift from humans to computers.” Harari claims that this phenomenon will eventually be so pervasive that we will be stripped of free will by machine-learning algorithms that will “analyze the biometric data streaming from sensors on and inside your body, determine your personality type and your changing moods, and calculate the emotional impact that a particular song—even a particular musical key—is likely to have on you.”

As biometric data collection becomes more pervasive, machine-learning algorithms are beginning to play a greater role in society.

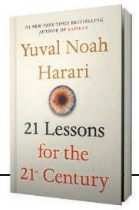
To reinforce just how important data already is in our lives, Harari explains how societal institutions such as government, education, and advertising can be conceptualized in terms of information processing. He describes how many autocratic regimes of the past century failed because their top-heavy, overly centralized structures were unable to handle the vast amounts of information from their populace. By contrast, he notes, many democratic regimes found success by using a more distributed structure and associated information-processing ability.

But Harari reminds the reader that advances in information science can change this balance: Large-scale data mining now favors highly centralized collection and processing of data, leveraging social structures typical of autocratic states and data monopolies. In an extreme example, he sketches out a futuristic autocracy in which pervasive sensors intimately monitor the population, giving both despots and entrenched corporate data platforms an unparalleled grip on power.

In this potential future, power is concentrated in the hands of a few elites, who use centralized data collection, artificial intelligence (AI) algorithms, and biometrics. This “data collective” will know us better than we do. And within a small amount of time, it will make arguably better decisions than we can individually. This need not be wholly nefarious, however; the benefits of data-driven decisions will be particularly evident in areas such as disease diagnosis, transportation, and personalized product recommendations.

Of course, the hyperefficiency of the collective will also leave the bulk of the population irrelevant as classical labor. But even this may not be all bad: In perhaps an unintentional nod to the gentlemen scientists of old, Harari argues that under the right circumstances, the future might allow displaced laborers to pursue scholarly and philosophical understanding rather than toiling in the pursuit of a livelihood.

21 Lessons for the 21st Century
Yuval Noah Harari
Spiegel & Grau, 2018.
395 pp.



One wonders whether individuals will still be necessary for producing data. Harari is unclear on this point. In some passages, he seems to think so. At other times, he seems unconvinced, such as when he asserts that AI chess masters no longer need external data to vanquish their opponents.

There is more good news for the digital future in the book: Modern technology will continue to enmesh us in an ever-bigger collective in which data shared will be immediately and thoroughly used to optimize everything from networked cars to AI doctors, which will be able to immediately use the entirety of all medical knowledge, as it becomes available in real time, to better treat our illnesses.

The scientific readership no doubt already appreciates the value of aggregating data because much of modern scientific research is effectively impossible to achieve by individuals alone. As we accumulate more scientific knowledge, the gap between the expansive knowledge of collective humanity and the knowledge of each individual grows larger.

In his eyes, only science can save us from an unpleasant outcome: “…the mark of science is the willingness to admit failure and try a different tack. That’s why scientists gradually learn how to grow better crops and make better medicines.”

But how do we use science to steer society toward the better future? Harari concludes with a call to action: Scientists need to learn to communicate better. We need to discard the notion that we can win over the public with tedious data. Rather, he argues, we must use narrative. “Humans think in stories rather than in facts, numbers, or equations, and the simpler the story, the better.” ■

10.1126/science.aav4297

PHOTO: ISTOCKPHOTO.COM/PETERHOWELL

---

[1] *Zvi Meitar Institute for Legal Implications of Emerging Technologies, Radzyner Law School, Interdisciplinary Center, Herzliya, Israel.* [2] *Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA. Email: dov.greenbaum@aya.yale.edu*

A freeze in military exercises could help to establish trust during nuclear negotiations with North Korea.

PHOTO: SEONGJOON CHO/BLOOMBERG/GETTY IMAGES

Edited by **Jennifer Sills**

Denuclearizing North Korea requires trust

In their Policy Forum "Denuclearizing North Korea: A verified, phased approach" (7 September, p. 981) A. Glaser and Z. Mian describe a pathway for verified denuclearization of North Korea. I agree that such an approach is necessary and, equally importantly, technically feasible. However, Glaser and Mian only highlight the disarmament side of the denuclearization agreement, without a plan to develop the mutual trust and the assurances on which such a deal depends. Incentivizing North Korea to reduce nuclear weapons and fissile materials will require confidence-building measures, ease of sanctions, and security guarantees. These elements are strongly related to the disarmament questions and must be regulated with similar precision.

Coordinating with the proposed phased approach, the involved parties could pair North Korea's freeze on weapon-related activities with a freeze of new nuclear-related sanctions or military exercises in the region. Such commitments would lay the foundation for an interim agreement, paving the way for long-term denuclearization. In a final step, the facilitation of humanitarian trade in areas such as health and nutrition would initiate the ease of sanctions and the establishment of credible security guarantees.

These measures need control and verification mechanisms, too. In case of nonfulfillment of such an agreement, it must be possible to swiftly reinstate the United Nations Security Council's sanctions. The structure of this contingency could be similar to the snapback mechanism in Article 37 of the Joint Comprehensive Plan of Action with Iran (*1*). Likewise, North Korea will insist on similar guarantees if it dismantles its nuclear weapons. It is always a challenge to create mechanisms that can credibly assure such guarantees for both parties, and this has become even more difficult after the U.S. withdrawal from the Iran nuclear agreement.

Tobias W. Langenegger
Chair of Negotiation and Conflict Management, ETH Zurich, 8092 Zurich, Switzerland.
Email: tlangenegger@ethz.ch

REFERENCE

1. United Nations Security Council Resolution 2231 (2015); https://undocs.org/S/RES/2231(2015).

10.1126/science.aav4636

Neutrino physics for Korean diplomacy

Continued diplomatic progress with North Korea will be a journey of many steps, as A. Glaser and Z. Mian describe in their Policy Forum "Denuclearizing North Korea: A verified, phased approach" (7 September, p. 981). Leaders in North Korea, South Korea, and the United States agree that one step could be dismantlement or civilian repurposing of the nuclear reactors at Yongbyon. We propose a cooperative method for verifying reactor shutdown or conversion. The key tools are meter-scale, field-deployable detectors that track neutrino emissions from reactor cores.

Neutrino detectors can track power levels and fuel evolution in nuclear reactors, as experiments in South Korea, China, Russia, the United States, and Europe have demonstrated (*1*–*7*). At Yongbyon, neutrino detectors could be deployed to verify reactor shutdown or civilian operations without the need for operational records or access inside reactor buildings. Shutdown of North Korea's main plutonium production reactor could be verified with a detector in a standard freight container parked outside the reactor building.

Existing neutrino technology may be attractive to all parties in the ongoing talks. North Korea may value a tool for demonstrating treaty compliance while maintaining custody of the reactor buildings. Other parties may value the tamper resistance of the neutrino signal and resilience of neutrino detectors, which require minimal on-site access and can reconstruct reactor operational history even after a data-taking pause. Neutrino projects are also a natural opportunity to strengthen relations between North and South Korea and to build international scientific ties. South Korea has an active neutrino community and could choose to deploy a counterpart to a Yongbyon-based detector at one of its own reactors. Resulting scientific collaboration could benefit Korea and the world. We encourage policy-makers to consider neutrino detectors as one step toward stability and security on the Korean Peninsula.

Rachel Carr,[1]* Jonathon Coleman,[2] Giorgio Gratta,[3] Karsten Heeger,[4] Patrick Huber,[5] YuenKeung Hor,[6] Takeo Kawasaki,[7] Soo-Bong Kim,[8] Yeongduk Kim,[9] John

Learned,[10] Manfred Lindner,[11] Kyohei Nakajima,[12] Seon-Hee Seo,[9] Fumihiko Suekane,[13] Antonin Vacheret,[14] Wei Wang,[6] Liang Zhan[15]

[1]Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139–4307, USA. [2]Department of Physics, University of Liverpool, Merseyside, UK. [3]Department of Physics, Stanford University, Stanford, CA 94305, USA. [4]Wright Laboratory, Department of Physics, Yale University, New Haven, CT 06520, USA. [5]Center for Neutrino Physics, Virginia Tech, Blacksburg, VA 24061, USA. [6]School of Physics, Sun Yat-Sen University, Guangzhou, China. [7]Department of Physics, Kitasato University, Sagamihara, Japan. [8]Department of Physics, Seoul National University, Seoul, Korea. [9]Center for Underground Physics, Institute for Basic Science, Daejeon, Korea. [10]University of Hawaii at Manoa, Honolulu, HI 96822, USA. [11]Max-Planck Institute for Nuclear Physics, Heidelberg, Germany. [12]Graduate School of Engineering, University of Fukui, Fukui, Japan. [13]Research Center for Neutrino Science, Tohoku University, Sendai, Japan. [14]Department of Physics, Imperial College London, London, UK. [15]Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, China.
*Corresponding author. Email: recarr@mit.edu

REFERENCES

1. Y. J. Ko *et al*. (NEOS Collaboration), *Phys. Rev. Lett*. **118**, 121802 (2017).
2. F. P. An *et al*. (Daya Bay Collaboration), *Phys. Rev. Lett*. **118**, 251801 (2017).
3. I. Alekseev *et al*. (DANSS Collaboration), arXiv:1804.04046 (2018).
4. G. Bak *et al*. (RENO Collaboration), arXiv:1806.00574 (2018).
5. J. Ashenfelter *et al*. (PROSPECT Collaboration), arXiv:1806.02784 (2018).
6. N.S. Bowden *et al*. (SONGS Collaboration), *J. Appl. Phys*. **105**, 064902 (2009).
7. G. Boireau *et al*. (Nucifer Collaboration), *Phys. Rev. D* **93**, 112006 (2016).

10.1126/science.aav8136

Improve traffic death statistics in China

Sustainable Development Goal 3.6 aims to "halve the number of global deaths and injuries from road traffic accidents" by 2020 (*1*). Accurate statistics are crucial to monitoring the progress toward this goal. According to current data, China contributed 23% of global road traffic deaths in 2016 (*2*). However, Chinese police departments have been criticized for underreporting road traffic deaths (*3*, *4*) and consistently attributing more than 99% of fatal crashes to road users' behavior rather than vehicle or environmental factors (*5*). Incomplete and distorted fatal crash data not only mislead policy-makers and researchers but also misdirect prevention efforts.

Police-reported data suggest that China witnessed a 46% decrease in road traffic deaths from 2002 to 2016 (*6*), but other data contradict this trend (*4*, *7*, *8*). Such discrepancies raise questions about the efficacy of national prevention efforts. The distorted cause spectrum statistics, which indicate that human error causes more than 99% of Chinese road traffic crashes (*6*) [compared with India, where human error cause rates hover in the 75 to 85% range (*9*)], mistakenly lead national prevention efforts to focus on safety education rather than multifaceted programs that incorporate legislation, engineering, and environmental approaches (*10*).

Improving the accuracy of Chinese road traffic injury statistics is achievable. The Chinese central government should establish and fund an independent body to collect crash data. China should also correct methodological oddities that yield inaccurate data. Improvements would include adopting the more commonly used window of 30 days postcrash for a mortality to be classified as a road traffic injury death (*11*), replacing exclusive-cause with multicause classification systems, and integrating other health and occupational surveillance data with traffic crash data to enable more accurate mortality estimation. Finally, China should develop a publicly available crash data database, modeled on other national systems such as the Fatality Analysis Reporting System in the United States (*12*), to allow researchers, practitioners, and the public to cite and use valid data for road traffic injury prevention.

Helai Huang,[1] Fangrong Chang,[1] David C. Schwebel,[2] Peishan Ning,[3] Peixia Cheng,[3] Guoqing Hu[3]*

[1]Urban Transport Research Center, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China. [2]Department of Psychology, College of Arts and Sciences, University of Alabama at Birmingham, Birmingham, AL 35294, USA. [3]Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha 410078, China.
*Corresponding author.
Email: huguoqing009@gmail.com

REFERENCES

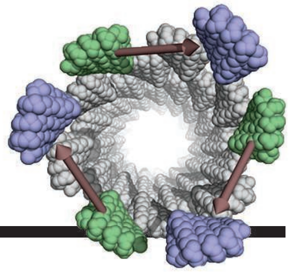
1. "Progress of Goal 3 in 2017," Sustainable Development Knowledge Platform (2017); https://sustainabledevelopment.un.org/sdg3.
2. The Global Burden of Disease (GBD) (2016); http://ghdx.healthdata.org/gbd-results-tool.
3. G. Hu *et al*., *Bull. World Health Org*. **89**, 41 (2011).
4. H. Huang *et al*., *Accid. Anal. Prev*. **108**, 220 (2017).
5. Traffic Management Bureau of the Ministry of the Public Security, *China Road Traffic Accident Statistics 2016* (Traffic Management Research Institute of the Ministry of Public Security, 2017).
6. "China Statistical Yearbook," National Bureau of Statistics of China (2016); http://data.stats.gov.cn/easyquery.htm?cn=C01&zb=A0SOD02&sj=2010-2016.
7. T. Alcorn, *Lancet* **378**, 305 (2011).
8. S. Ma, Q. Li, M. Zhou, L. Duan, D. Bishai, *Traffic Inj. Prev*. **13** (suppl. 1), 57 (2012).
9. "Road Accidents in India," Ministry of Road Transport and Highways for India (2010–2016); www.morth.nic.in/index2.asp?slid=297&sublinkid=139&lang=1.
10. F. P. McKenna, "Education in road safety: Are we getting it right?" (RAC Foundation Report 10/113, UK, 2010).
11. World Health Organization, "Global status report on road safety: Time for action" (2009); www.who.int/violence_injury_prevention/road_safety_status/2009.
12. National Highway Traffic Safety Administration, Fatality Analysis Reporting System (www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars).

10.1126/science.aav5117

RESEARCH

Self-assembling helical fragments by design

Shen et al., *p. 705*



IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

NEONICOTINOIDS

Trouble at the hive

Neonicotinoid pesticides cause mortality and decline in insect pollinators. One repeatedly noted effect is a reduction in bee colony size. However, the mechanism behind this reduction is unclear. Crall *et al*. performed complex real-time monitoring of bumblebee behavior within their nests (see the Perspective by Raine). Neonicotinoid exposure reduced nurse and caretaking behaviors, which affected productivity and harmed colony thermoregulation. These changes in behavior acted together to decrease colony viability, even when exposure was nonlethal. —SNV

Science, this issue p. 683; see also p. 643

Bumblebee (*Bombus impatiens*) worker foraging outdoors, fitted with a tracking device

QUANTUM OPTICS

Inducing interactions between quantum emitters

The development of scalable quantum systems will require the ability to control the interactions between the individual quantum building blocks of the system. Evans *et al*. used a pair of silicon vacancy centers embedded in a diamond nanocavity to show that interactions between the quantum emitters can be mediated optically (see the Perspective by Lodahl). Such optical control provides a speed advantage as well as the potential to develop an integrated platform for future quantum communication and quantum networking. —ISO

Science, this issue p. 662; see also p. 646

NEUROSCIENCE

Theta rhythm protects sleep replay

Hippocampal replay of place cell sequences during sleep is critical for memory consolidation in target cortical areas. How is the sequential organization of place cell assemblies maintained across different time scales? Drieu *et al*. compared periods when a rat either sat passively on a moving train or ran actively on a treadmill on the same train. During the passive movement, the slow behavioral sequence of place cells was still present, but the rapid generation of theta sequences was lost. Active running on the treadmill, however, maintained the theta rhythm. After passive transport, sequence replay during sleep was destroyed, whereas active running protected replay. —PRS

Science, this issue p. 675

ORGANIC CHEMISTRY

The staying power of electron-poor ligands

The venerable Suzuki coupling reaction originally used palladium to pair up unsaturated carbon centers. The protocol has been widely extended to chiral saturated alkyl carbons, but control over product stereochemistry is a pressing challenge. Zhao *et al*. systematically studied how the properties of the phosphine ligands that are coordinated to the catalyst influence the stereochemical outcome. Certain electron-withdrawing phosphines favored retention of the initial configuration in chiral alkyltrifluoroborate reactants. Conversely, bulky electron-rich phosphines lead to inverted configurations in the products. —JSY

Science, this issue p. 670

GENETIC PRIVACY

Detecting familial matches

Recent advances in DNA technology and companies that provide array-based testing have led to services that collect, share, and analyze volunteered genomic information. Privacy concerns have been raised, especially in light of the use of

IMAGE: SHEN *ET AL*.; PHOTO: JAMES CRALL

these services by law enforcement to identify suspects in criminal cases. Testing models of relatedness, Erlich *et al*. show that many individuals of European ancestry in the United States—even those that have not undergone genetic testing—can be identified on the basis of available genetic information. These results indicate a need for procedures to help maintain genetic privacy for individuals. —LMZ

Science, this issue p. 690

BATTERIES

Oil when not in use

For primary or nonrechargeable batteries, the overall energy density will be limited by any discharge or open-circuit corrosion that occurs during storage. For batteries based on aluminum and air, this longstanding problem has prevented their widespread use and has been challenging to overcome. Hopkins *et al*. used commercially available components to construct aluminum-air batteries. During standby periods, the electrolyte in the batteries was replaced with oil to protect the electrodes from corrosion, thus preventing energy loss. —MSL

Science, this issue p. 658

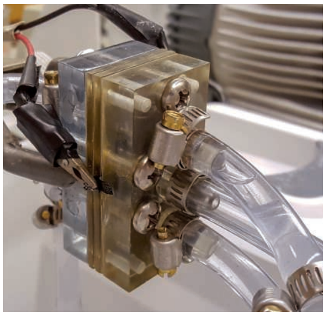


Photo of aluminum-air prototype battery

OSTEOPOROSIS

Building and rebuilding bone

WNT signaling is important for proper embryonic development, shaping cell fate and migration, stem cell renewal, and organ and tissue formation. Luther *et al*. investigated the role of WNT1 in osteoporosis. Patients with early-onset osteoporosis and mutations in the *WNT1* gene had low bone turnover and high fracture rates, and loss of WNT1 activity caused fracture and osteoporosis in mice. Inducing WNT1 in bone-forming cells increased bone mass in aged mice, and this process did not require LRP5, a co-receptor involved in WNT signaling. Thus, WNT1 acts as an anabolic (bone building) factor and might represent a therapeutic target for osteoporosis. —CC

Sci. Transl. Med. **10**, eaau7137 (2018).

SKIN BIOLOGY

A basic way to tan

Darker-skinned individuals have more melanin in their skin and a lower risk for skin cancers. The production of melanin in organelles called melanosomes is pH sensitive. Zhou *et al*. found that the enzymatic activity of soluble adenylyl cyclase (sAC) resulted in decreases in both melanosome pH and melanin synthesis. sAC deficiency or inhibitors increased melanosome pH and pigmentation in mice. This mechanism for rapidly regulating melanin synthesis could potentially be exploited to reduce skin cancer risk for fair-skinned individuals. —WW

Sci. Signal. **11**, eaau7987 (2018).

PREDATION

No longer a safe haven

Many biological patterns have a latitudinal component. One long-recognized pattern is that predation rates are higher at lower latitudes. This may explain why many migratory birds travel thousands of miles from the tropics to the poles to breed. Looking across thousands of records, Kubelka *et al*. found that climate change seems to have altered this fundamental pattern. In shorebirds, at least, predation rates on nests are now higher in the Arctic than in the tropics. —SNV

Science, this issue p. 680

IN OTHER JOURNALS

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

A handheld device should greatly expand the application of adaptive optics scanning laser ophthalmoscopy.

NEUROSCIENCE

Decisions and their future implications

We constantly make choices. It is often important to consider not only the short-term but also the longer-term implications of a choice, also known as its prospective value. In a combined decision-making and brain-scanning study, Kolling *et al*. found that when humans make sequential decisions, they do not only consider their immediately available options. Instead, they incorporate the average value of future options, their variability, and the time frame or search horizon, as well as search costs. Furthermore, humans even consider individual capacities and preferences for making decisions in the future. Brain activity in the dorsal anterior cingulate cortex represents prospective value, but when there is an increased need to overcome costs, activity in the ventral striatum and perigenual anterior cingulate cortex is more strongly coupled. —PRS

Neuron **99**, 1069 (2018).

CELL BIOLOGY

The intraflagellar transport train

Assembly of the cilium requires bidirectional intraflagellar transport (IFT) of building blocks along microtubules to and from the site of assembly at its tip. Dynein-1b motors are required to power retrograde transport and are believed to reach the ciliary tip by kinesin-2–driven anterograde IFT. It is unclear which mechanism prevents a tug-of-war between these oppositely directed microtubule motors. Jordan *et al*. used cryo–electron tomography to examine the architecture of IFT trains in *Chlamydomonas* cilia in situ. Their findings revealed the relative positions of IFT motors on anterograde versus retrograde trains. Dynein-1b in its autoinhibited form was an integral part of anterograde trains but is positioned to prevent premature engagement with the microtubules. Once at the cilia tip, the dynein converted into its activated form, engaged the microtubules, and then powered retrograde transport. —SMH

Nat. Cell Biol. **20**, 1250 (2018).

PHOTOS (LEFT TO RIGHT): BRANDON J. HOPKINS; CARMELO GERACI/GETTY IMAGES

OPHTHALMIC IMAGING

Making vision clearer

Adaptive optics scanning laser ophthalmoscopy (AOSLO) is now a routine tool used by eye care professionals to aid the detection and diagnosis of retinal disease. These machines, however, are large and not portable, so the patient needs to travel to where they are located. By combining computational algorithms with developments in miniaturized deformable mirrors and microelectromechanical technology, DuBose *et al.* have developed a handheld AOSLO device that weighs less than 200 grams. Such a technological development opens up the possibility of examination to a broader range of patient populations, such as children and the physically incapacitated. The light weight of the device also offers the opportunity to provide examinations to people in geographically remote areas. —ISO

Optica **9**, 1027 (2018).

CLIMATE CHANGE

Uneven results

What are likely to be the specific results of anthropogenic activities on climate, beyond higher temperatures? One may be change in the way rainfall is distributed in time. Rainfall is an uneven phenomenon: There are wet days and dry days, floods and droughts, and hard rains and gentle rains. Pendergrass and Knutti use observations and models to show that climate change should only exacerbate that unevenness. Whereas today it takes an average of 12 days each year for half of the rain to fall, by 2100, in a scenario with high greenhouse gas emissions, the same fraction should fall in the wettest 6 days each year. —HJS

Geophys. Res. Lett. 10.1029/2018GL080298 (2018).

IMMUNOMETABOLISM

A role for mastocytes in ketosis

Ketogenesis is a fundamental biochemical process occurring mainly in liver mitochondria. This pathway provides energy to vital organs when access to glucose is limited for prolonged periods, such as during fasting. Lipolysis-derived fatty acids normally initiate this process via the transcription factor PPAR-α. Misto *et al*. found that fasting stimulates mast cells to release histamine, which stimulates the biosynthesis of the high-affinity PPAR-α agonist oleoylethanolamide via G protein–coupled H_1 receptors. Thus, mast cells, recognized for their role in allergy and anaphylaxis, unexpectedly take part in the regulation of a major metabolic pathway. Future studies will be needed to uncover the underlying mechanisms of this process and understand how it may contribute to metabolic dysfunction. —STS

Cell Metab. 10.1016/j.cmet.2018.09.014 (2018).

SUPERCONDUCTIVITY

Thermal transport to the rescue

The pseudogap phase in cuprate superconductors remains one of the most puzzling aspects of these materials. To shed light on the nature of the pseudogap, Michon *et al*. studied thermal transport in the cuprate $La_{1.6-x}Nd_{0.4}Sr_xCuO_4$. By applying magnetic fields high enough to destroy superconductivity and approaching absolute zero temperature, the researchers found that the thermal and charge transport in the pseudogap phase were locked in step. This indicates that the ground state of the nonsuperconducting pseudogap phase has the character of a conventional metal. Comparison to the data at zero field further suggests that applying magnetic fields does not affect some of the signatures of the pseudogap phase. —JS

Phys. Rev. X **8**, 041010 (2018).

MOLECULAR IMAGING

Revealing chromosome features

Cryo–electron tomography (cryo-ET) is used to visualize cellular structures in the native environment without chemical fixation and dye labeling. Cai *et al*. used cryo-ET on both interphase and mitotic fission yeast cells to explore some interesting features of chromosome organization. Nucleosomes in situ do not appear to resemble the canonical conformation obtained by crystallography in vitro. In vivo, it appears that nucleosomes are partially unwrapped. Moreover, nucleosomes cluster irregularly, with the clusters being more condensed and less dynamic in mitotic cells. Yet, the condensation in mitotic chromosomes is uneven, and there are loosely packed regions where, possibly, mitotic transcription occurs. —SYM

Proc. Natl. Acad. Sci. U.S.A. **115**, 10977 (2018).

Climate change will exacerbate the unevenness of precipitation.

PHOTO: NITICHUYSAKUL PHOTOGRAPHY/GETTY IMAGES

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

NEUROSCIENCE

A framework for cognitive spaces

Ever since Tolman's proposal of cognitive maps in the 1940s, the question of how spatial representations support flexible behavior has been a contentious topic. Bellmund *et al*. review and combine concepts from cognitive science and philosophy with findings from neurophysiology of spatial navigation in rodents to propose a framework for cognitive neuroscience. They argue that spatial-processing principles in the hippocampal-entorhinal region provide a geometric code to map information domains of cognitive spaces for high-level cognition and discuss recent evidence for this proposal. —PRS

Science, this issue p. 654

EVOLUTION

Replaying the tape of life

The evolutionary biologist Stephen Jay Gould once dreamed about replaying the tape of life in order to identify whether evolution is more subject to deterministic or contingent forces. Greater influence of determinism would mean that outcomes are more repeatable and less subject to variations of history. Contingency, on the other hand, suggests that outcomes are contingent on specific events, making them less repeatable. Blount *et al*. review the numerous studies that have been done since Gould put forward this question, both experimental and observational, and find that many patterns of adaptation are convergent. Nevertheless, there is still much variation with regard to the mechanisms and forms that converge. —SNV

Science, this issue p. 655

IMMUNOLOGY

Dendritic cells give mast cells a nudge

Anaphylaxis is a life-threatening allergic reaction triggered after antigen-specific immunoglobulin E (IgE) antibodies bind to target allergens. These antibodies then cross-link IgE-specific Fc receptors on the surface of mast cells. The mast cells rapidly release inflammatory mediators, including histamine, resulting in smooth muscle contraction, vasodilation, and blood vessel leakage. Because mast cells are usually found in the perivascular abluminal surface of blood vessels, it has been unclear how blood-borne allergens can interact with them. Choi *et al*. used live intravital imaging of the mouse vasculature to show that a specialized subset of dendritic cells sample blood-borne antigens and relay them to mast cells on the surface of microvesicles (see the Perspective by Levi-Schaffer and Scheffel). IgE-bound mast cells then vigorously degranulate after contact with these microvesicles. —STS

Science, this issue p. 656;
see also p. 640

STRUCTURAL BIOLOGY

Structures of eukaryotic ribonuclease P

Ribonuclease P (RNase P) is a ribozyme that processes transfer RNA (tRNA) precursors and is found in all three kingdoms of life. Now, Lan *et al*. report the structures of yeast RNase P (see the Perspective by Scott and Nagai). The aporibozyme structure reveals how the protein components stabilize the RNA and explains how the structural roles of bacterial RNA elements have been delegated to the protein components in RNase P of higher organisms during evolution. The structure of yeast RNase P in complex with its natural substrate, a tRNA precursor, demonstrates the structural basis for substrate recognition and provides insights into its catalytic mechanism. —SYM

Science, this issue p. 657;
see also p. 644

NANOMATERIALS

Cleaving with a metal handle

Using adhesive tape to pull off monolayers of two-dimensional (2D) materials is now a well-established approach. However, the flakes tend to be micrometer scale, and the creation of multilayer stacks for device application can be challenging and time consuming. Shim *et al*. show that monolayers of a variety of 2D materials, including molybdenum disulfide and hexagonal boron nitride, can be cleaved from multilayers grown as 5-centimeter-diameter wafers. The multilayer is capped with a nickel layer, which can be used to pull off the entire grown stack. The bottom of the stack is again capped with nickel, and a second round of cleaving leaves the monolayer on the bottom nickel layer. The monolayers could be transferred to other surfaces, which allowed the authors to make field-effect transistors with high charge-carrier mobilities. —PDS

Science, this issue p. 665

ANTIBIOTIC RESISTANCE

Efflux pumps and mutation

Antibiotic resistance is an alarming and growing challenge. Bacteria show great heterogeneity in growth and mutation rates. Such variability allows some cells to persist during transient antibiotic exposure. During this persistent phase, mutations accumulate, which can result in selection for full-blown antibiotic resistance. El Meouche and Dunlop found that increased expression of efflux pumps on some cells affords them some relief from antibiotic toxicity. But up-regulating efflux pumps is costly for the bacteria, reducing growth rate and expression of MutS, a protein involved in DNA mismatch repair. These changes thus lift the lid on increased levels of bacterial mutation. —CA

Science, this issue p. 686

IMMUNOLOGY

An IgG1 SNP enhances autoimmunity

One common feature of autoimmune diseases like systemic lupus erythematosus (SLE) is the presence of high titers of self-reactive antibodies. These result in immune complexes, inflammation, and tissue pathology. Consequently, the checkpoints that normally keep immunoglobulin G (IgG)–positive autoreactive B cells in check are of intense interest. Chen *et al*. report the presence of a common IgG1 single-nucleotide polymorphism (SNP) in East Asian populations (hIgG1-G396R). This SNP was enriched in SLE patients and associated with increased disease severity. Humans with this SNP, as well as knockin mice, showed enhanced plasma cell accumulation and antibody production. This SNP enhanced IgG1 immunoglobulin tail tyrosine motif phosphorylation, triggering longer adaptor protein Grb2 dwell times in immunological synapses and hyper–Grb2–Bruton's tyrosine kinase signaling after antigen binding. —STS

Science, this issue p. 700

PROTEIN DESIGN

Built to be reversible

There has been some success in designing stable peptide filaments; however, mimicking the reversible assembly of many natural protein filaments is challenging. Dynamic filaments usually comprise independently folded and asymmetric proteins and using such building blocks requires the design of multiple intermonomer interfaces. Shen *et al*. report the design of self-assembling helical filaments based on previously designed stable repeat proteins. The filaments are micron scale, and their diameter can be tuned by varying the number of repeats in the monomer. Anchor and capping units, built from monomers that lack an interaction interface, can be used to control assembly and disassembly. —VV

Science, this issue p. 705

IMMUNOLOGY

Adding to the cross-presentation family

Immune responses to viral or tumor antigens are typically initiated by the process of cross-presentation. Cross-presentation is believed to be the major way that innate immune cells, such as the classical dendritic cell 1 (cDC1) subset, activate and prime immunological T cells. Theisen *et al*. used CRISPR-based screening to identify regulators of cross-presentation by cDC1s (see the Perspective by Barbet and Blander). One such regulator that was identified, WDFY4 (WD repeat- and FYVE domain–containing protein 4), was required for cross-presentation of cell- and bacterial-associated antigens. WDFY4 played a critical role in cDC1-mediated viral and tumor immunity yet did not seem necessary for major histocompatibility complex class II presentation or for cross-presentation by monocyte-derived DCs. —PNK

Science, this issue p. 694;
see also p. 641

CANCER IMMUNOLOGY

Tandem immunotherapy achieves synergy

Immune checkpoint–inhibitor therapies bolster the antitumor activity of $CD8^+$ T lymphocytes. Wang *et al*. performed single-cell analysis of tumor-infiltrating lymphocytes in mouse cancer models in which inhibitory anti–PD-1 (programmed cell death protein 1) and stimulatory anti–GITR (glucocorticoid-induced tumor necrosis factor receptor–related protein) antibodies together enhanced tumor control. This combination immunotherapy led to a synergistic increase in tumor antigen–specific memory precursor effector T cells dependent on the CD226 costimulatory pathway. Biochemical studies in liposomes identified CD226 as an additional target of dephosphorylation mediated by the PD-1–SHP2 (Src homology region 2) complex. Thus, further clinical trials could usefully test the efficacy of combined anti-GITR and anti–PD-1 immunotherapy in human cancer. —IW

Sci. Immunol. **3,** eaat7061 (2018).

REVIEW SUMMARY

NEUROSCIENCE

Navigating cognition: Spatial codes for human thinking

Jacob L. S. Bellmund*, **Peter Gärdenfors**, **Edvard I. Moser**, **Christian F. Doeller***

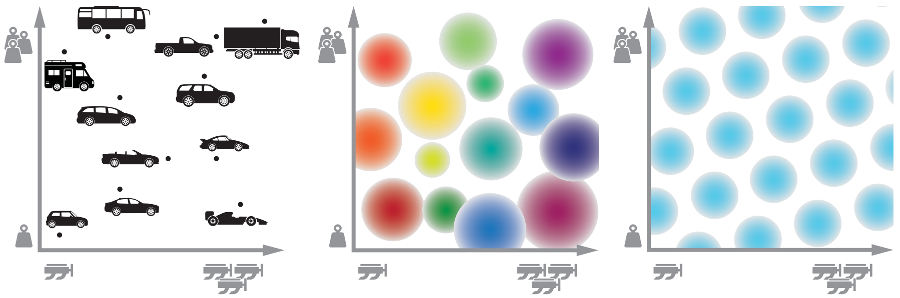
BACKGROUND: Ever since Edward Tolman's proposal that comprehensive cognitive maps underlie spatial navigation and, more generally, psychological functions, the question of how past experience guides behavior has been contentious. The discovery of place cells in rodents, signaling the animal's position in space, suggested that such cognitive maps reside in the hippocampus, a core brain region for human memory. Building on the description of place cells, several other functionally defined cell types were discovered in the hippocampal-entorhinal region. Among them are grid cells in the entorhinal cortex, whose characteristic periodic, six-fold symmetric firing patterns are thought to provide a spatial metric. These findings were complemented by insights into key coding principles of the hippocampal-entorhinal region: Spatial representations vary in scale along the hippocampal long axis, place cells remap to map different environments, and sequential hippocampal activity represents nonlocal trajectories through space. In humans, the existence of spatially tuned cells has been demonstrated in presurgical patients, and functional magnetic resonance imaging provides proxy measures for the noninvasive investigation of these processing mechanisms in human cognition. Intriguingly, recent advances indicate that place and grid cells can encode positions along dimensions of experience beyond Euclidean space for navigation, suggesting a more general role of hippocampal-entorhinal processing mechanisms in cognition.

ADVANCES: We combine hippocampal-entorhinal processing mechanisms identified in spatial navigation research with ideas from cognitive science describing a spatial representational format for cognition. Cognitive spaces are spanned by dimensions satisfying geometric constraints such as betweenness and equidistance, enabling the representation of properties and concepts as convex regions of cognitive space. We propose that the continuous population code of place and grid cells in the hippocampal-entorhinal region maps the dimensions of cognitive spaces. In these, each stimulus is located according to its feature values along the relevant dimensions, resulting in nearby positions for similar stimuli and larger distances between dissimilar stimuli. The low-dimensional, rigid firing properties of the entorhinal grid system make it a candidate to provide a metric or distance code for cognitive spaces, whereas hippocampal place cells flexibly represent positions in a given space. This mapping of cognitive spaces is complemented by the additional coding principles outlined above: Along the hippocampal long axis, cognitive spaces are mapped with varying spatial scale, supporting memory and knowledge representations at different levels of granularity. Via hippocampal remapping, spaces spanned by different dimensions can be flexibly mapped and established maps can be reinstated via attractor dynamics. The geometric definition of cognitive spaces allows flexible generalization and inference, and sequential hippocampal activity can simulate trajectories through cognitive spaces for adaptive decision-making and behavior.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat6766

OUTLOOK: Cognitive spaces provide a domain-general format for processing in the hippocampal-entorhinal region, in line with its involvement beyond navigation and memory. Spatial navigation serves as a model system to identify key coding principles governing cognitive spaces. An important question concerns the extent to which firing properties of spatially tuned cells are preserved in cognitive spaces. Technological advances such as calcium imaging will clarify coding principles on the population level and facilitate the translation to human cognitive neuroscience. Spatial navigation is mostly investigated in two dimensions and naturally limited to three dimensions; however, the processing of complex, multidimensional concepts is vital to high-level human cognition, and the representation of such high-dimensional spaces is an intriguing question for future research. Further, the role of brain networks acting in concert with the hippocampus, in navigation specifically and cognitive function in general, will provide insight into whether and how cognitive spaces are supported beyond the hippocampal-entorhinal region. Finally, the precise way in which cognitive spaces and trajectories through them are read out for behavior remains to be elucidated. ■



Place and grid cells map cognitive spaces. Cognitive spaces are defined by dimensions satisfying geometric constraints. The example space (left) is spanned by the dimensions of engine power and car weight. Black dots show different vehicles whose positions reflect their feature combinations. Place cells (center; colored circles represent firing fields of different cells) and grid cells (right; circles illustrate firing pattern of one cell) provide a continuous code for cognitive spaces.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: bellmund@cbs.mpg.de (J.L.S.B.); doeller@cbs.mpg.de (C.F.D.)
Cite this article as J. L. S. Bellmund *et al.*, *Science* **362**, eaat6766 (2018). DOI: 10.1126/science.aat6766

Bellmund *et al.*, *Science* **362**, 654 (2018) 9 November 2018 **1 of 1**

REVIEW

NEUROSCIENCE

Navigating cognition: Spatial codes for human thinking

Jacob L. S. Bellmund[1,2,3]*, Peter Gärdenfors[4,5], Edvard I. Moser[1], Christian F. Doeller[1,3]*

The hippocampal formation has long been suggested to underlie both memory formation and spatial navigation. We discuss how neural mechanisms identified in spatial navigation research operate across information domains to support a wide spectrum of cognitive functions. In our framework, place and grid cell population codes provide a representational format to map variable dimensions of cognitive spaces. This highly dynamic mapping system enables rapid reorganization of codes through remapping between orthogonal representations across behavioral contexts, yielding a multitude of stable cognitive spaces at different resolutions and hierarchical levels. Action sequences result in trajectories through cognitive space, which can be simulated via sequential coding in the hippocampus. In this way, the spatial representational format of the hippocampal formation has the capacity to support flexible cognition and behavior.

How past experience guides behavior has been a highly contested topic for decades. In 1948, Tolman (*1*) described evidence for learning beyond stimulus-response–driven behavior in rats, foreshadowing the cognitive revolution to follow. He proposed that rats learn comprehensive maps of their environments, which can guide flexible goal-directed behavior such as finding shortcuts. Tolman coined the term "cognitive map" and speculated how these maps might underlie psychological functions. Support for his controversial speculations was provided by the first reports of hippocampal place cells more than 20 years later (*2*). Place cells are generally only active when an animal occupies the cell's preferred location. These cells were proposed by O'Keefe and Nadel (*3*) as a potential neural substrate of Tolman's cognitive map. Incorporating the discovery of place cells and findings from lesion studies, O'Keefe and Nadel suggested that the hippocampus provides a map-like reference system, signaling both the position of the agent and features of the environment for navigational and mnemonic processing of events (*3*). The sensitivity of hippocampal coding to items or discrete events in the environment (*4*–*6*), differential coding of overlapping path segments based on task state (*7*), and widespread deficits in patients with hippocampal lesions (*8*) have been captured in the development of relational memory theory (*8*–*10*). This account builds on the capabilities of the hippocampus for associative binding to link event representations into relational networks and has offered a counterpole to an exclusively spatial-processing view of the hippocampus. Episodic memories can be formed by linking successive event representations and episodic trajectories into ensemble patterns stored in hippocampal networks for subsequent retrieval (*8*, *10*–*14*). Recent advances demonstrate a hippocampal involvement in flexible cognition beyond the domains of navigation and memory (*15*).

Space codes as a representational format for cognition

We describe domain-general core coding principles from spatial navigation research that have the potential to support a wider span of cognitive functions. Specifically, we propose that the hippocampal-entorhinal system represents experience in cognitive spaces [see conceptual spaces in (*16*, *17*)]. A cognitive space is thought to be spanned by a set of quality dimensions, which can be closely related to sensory inputs but also comprise abstract features (*16*). A given stimulus can be located in a cognitive space according to its set of feature values along a set of quality dimensions. Each dimension is equipped with an underlying metric and follows geometric constraints satisfying the mathematical notions of betweenness and equidistance (*16*).

Consider the following example: When planning to buy a new car, you might describe cars along two dimensions: their engine power and their weight (Fig. 1). Depending on the two features, vehicles can be located in this two-dimensional space; for instance, a "sports car" might be rather lightweight but have a strong engine. In this example, we might treat the dimensions as separable. However, stimuli can also vary along integral dimensions, on which a given stimulus cannot be described independently. For example, a car's color can only be fully described when assigning values on all three integral dimensions (hue, saturation, and brightness) that constitute the color domain (*16*). Different cognitive spaces might have different underlying metrics—for example, the Euclidean or the city-block metric—that allow distances to be assigned along a dimension reflecting the degree of similarity between locations in cognitive space with similar stimuli located close together (*16*, *18*).

Building on the geometric characterization of quality dimensions, a property is defined as a convex region in some domain, with convexity meaning that for all points x and y in the region, the points between x and y also fall into the region (*16*). In this framework, a property constitutes the simplest form of a concept based on only one domain—for example, describing a car as "heavy." Defining properties as convex regions enables generalization, in that a property of two stimuli x and y can be inferred to be shared by any stimulus z falling between x and y. Therefore, the geometric constraints on cognitive spaces allow inference about never-experienced stimuli. Hence, cognitive spaces afford great cognitive flexibility, going beyond associative and transitive inferences also permitted by relational networks (*8*–*10*).

More complex concepts comprise multiple domains and information about their interrelations. Thus, a concept is defined as a set of convex regions in a number of domains, with domains weighted on the basis of salience and additional information about how the regions in different domains are correlated (*16*). In our example, racing cars would occupy a region characterized by high power and low weight (Fig. 1). Following the spatial definition of concepts, different exemplars correspond to more or less central positions, with prototypical members (*19*) located centrally in the conceptual region. Using a Voronoi tessellation of a continuous space with a Euclidean metric, where all locations in a given space are discretized as belonging to the nearest prototype, convex regions emerge and allow classification of individual stimuli (*16*).

We propose that processing mechanisms in the hippocampal-entorhinal system are well suited to support cognitive spaces as a domain-general format for flexible, high-level cognition in humans. Studies of spatial navigation describe how space is mapped by a continuous, multiscale code of function-specific cells in the hippocampal-entorhinal system. This system enables flexible mapping of different environments and simulations of trajectories through space via temporally compressed sequences. These processing mechanisms might have developed from originally mapping navigable space to also representing cognitive spaces, consistent with an evolutionarily preserved circuitry of the hippocampal-entorhinal region across mammals (*20*), the translation of insights from spatial navigation research in rodents

---

[1]Kavli Institute for Systems Neuroscience, Centre for Neural Computation, The Egil and Pauline Braathen and Fred Kavli Centre for Cortical Microcircuits, NTNU, Norwegian University of Science and Technology, Trondheim, Norway. [2]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands. [3]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. [4]Department of Philosophy and Cognitive Science, Lund University, Lund, Sweden. [5]Centre for Artificial Intelligence, University of Technology Sydney, Sydney, Australia.
*Corresponding author. Email: bellmund@cbs.mpg.de (J.L.S.B.); doeller@cbs.mpg.de (C.F.D.)

to human navigation and beyond (3, 21, 22) (Box 1 and Fig. 2), and the notion of a core system of geometry (23).

From spatial navigation to cognitive spaces

In the mammalian brain, positional information is conveyed by the spatially constrained firing of place cells (2) and grid cells (24) during spatial navigation. Place cells in the hippocampus are preferentially active when the animal occupies a certain position within the environment: the cell's receptive field or place field. The firing fields of the population of place cells are thought to cover the entire environment, thereby providing a map-like representation of the animal's surroundings (2, 3). Although the firing of a place cell is usually restricted to one place in a small environment, grid cells in the medial entorhinal cortex (EC), one synapse from the hippocampus, exhibit multiple firing fields located at the vertices of equilateral triangles tiling the entire environment (24). These regular, six-fold symmetric firing patterns are assumed to support spatial navigation by providing a coordinate system of the environment (24, 25). The entorhinal grid system is assigned a key role in path integration and vector navigation (25–28).

Intracranial recordings in patients navigating virtual reality environments established the existence of place cells (29) and grid cells (30) in humans. The six-fold symmetry of grid cell firing has been translated to noninvasive functional magnetic resonance imaging (fMRI) in healthy volunteers, where grid-like hexadirectional signals have been observed in the EC during navigation (22, 31–37) (Box 1 and Fig. 2). The brain's spatial navigation system (25, 38) further includes head direction cells conveying information about the animal's head direction (39), goal and goal direction cells signaling egocentric directions to navigational goals (40, 41), speed cells sensitive to running speed (42), and border (43, 44) or boundary vector cells (45) responding to borders in the environment.

The firing of place and grid cells conveys positional information to navigate Euclidean space. We hypothesize that the spatially constrained firing of place cells and the metric provided by the entorhinal grid system might provide a domain-general mechanism to map dimensions of experience. In this framework, the activity of place cells can be conceived as indexing locations in a cognitive space spanned by the entorhinal grid system. Three further neural coding mechanisms identified in spatial navigation studies illustrate how the hippocampal-entorhinal system may support a core mechanism of mapping cognitive spaces. First, the firing fields of place and grid cells increase in size along the dorsoventral axis of the rodent hippocampus (24, 46–49), in line with mapping of cognitive spaces at different levels of granularity for multiscale representations of knowledge hierarchies or nested conceptual information. Second, the ability of place cells to undergo global remapping (50–52) allows the flexible formation of a multitude of uncorrelated maps for different cognitive spaces, which can be reinstated via attractor dynamics (11, 13, 53, 54). Third, sequential activity of place cells (55–57) and grid cells during replay (58–60) and theta oscillations (61) enables the simulation of trajectories (12, 62, 63) through different locations in a cognitive space for adaptive cognition and behavior.

A continuous map of experience

The spatially constrained firing of place and grid cells provides a continuous code for the dimensions of space, in which neighboring positions have similar representations due to partially overlapping firing fields across the population of cells (Fig. 3, A and B). We build on findings that the continuous code of spatially selective cells maps dimensions of experience beyond Euclidean space, which affords flexible cognition via the formation of cognitive spaces, whose dimensions are geometrically constrained as described above.

Unlike in rodents, the visual system has evolved to be the dominant source of sensory information in primates. During visual exploration of naturalistic images, neurons in the primate EC encode gaze position with six-fold symmetric firing patterns that are the hallmark of grid cell firing during navigation (64). Further, some entorhinal neurons preferentially discharged when monkeys explored the edges of the visual stimuli (64), akin to the firing of border cells in rodents (43–45).

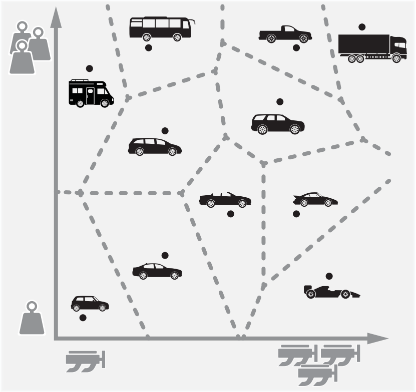


Fig. 1. Two-dimensional cognitive space. Schematic of a cognitive space spanned by the dimensions of car weight (y axis) and engine power (x axis). Feature values along the two axes define positions of different cars, resulting in stimuli with similar properties being located nearby. Concepts (car icons) are defined as convex regions of the cognitive space and are indicated by dashed lines obtained from a Voronoi tessellation of the space. Under the assumption of a Euclidean metric, this discretizes the space into convex regions by assigning each point in space to the region around the closest prototypical exemplar (black dots), with distances based on dissimilarity along the feature dimensions.

Box 1. Hexadirectional signals in fMRI.

Grid cells are defined by their six-fold symmetric firing patterns tiling environments in a highly regular fashion (24). Hexadirectional signals serve as a proxy measure for grid-like activity in BOLD-fMRI during trajectories through cognitive spaces to investigate the role of the grid system in higher-level cognition in the healthy human brain (22, 31–37). Three analysis approaches relying on a directional bias of activity have been used (Fig. 2). In the orientation-estimation approach (22), the data are partitioned to estimate the orientation of the hexadirectional signal in one part of the data. The prediction of increased levels of BOLD activity for trajectories aligned versus misaligned with the estimated orientation is then tested on an independent data partition. In the second analysis approach, which is based on fMRI repetition suppression, the hexadirectional grid code is reflected in correlations of BOLD activity during trajectories in a given direction with the time since the last trajectory at an angular offset of 60° (22). The third approach analyzes the similarity of multivoxel patterns as a function of angular differences between trajectory directions to test the assumption that activity of the grid system should be reflected in a 60° modulation of entorhinal activity patterns (33). Hexadirectional signals can serve as a showcase example for how insights from rodent electrophysiology might be translated to human navigation in fMRI and explored in human cognition and behavior more broadly (22, 31–37).

Although the BOLD response measured with fMRI does not reflect single-cell activity, it provides a hemodynamic proxy of population activity (159). How can the 60° symmetry of grid cell firing patterns be translated to a bias in population activity picked up by fMRI during trajectories through cognitive space? Although a single grid cell intuitively exhibits hexadirectional activity, a hexadirectional bias on the population level might result if orientations (78, 79) and spatial offsets to boundaries (160) of the firing patterns are clustered across cells. Further, this could potentially be reflected in directional biases in the local field potential in the EC; see also recent findings in MEG and intracranial electroencephalography (161–163). Additionally, conjunctive grid cells (25) also modulated by heading direction might contribute to the hexadirectional signal if the preferred directions of these cells align with the axes of the cells' grid pattern (22). Technological advances, such as two-photon microscopy (145), that enable imaging of cellular responses in larger portions of the rodent brain might allow future research to shed further light on the dynamics underlying hexadirectional signals on the population level.

Others were tuned to saccade direction (65), putatively signaling information similar to that provided by head direction cells (39). Together with two recent reports of eye movement–dependent hexadirectional signals in the human EC (35, 36), these findings support the notion that representations of visual space follow coding principles identified in spatial navigation research in rodents, suggesting a shared neural basis.

Next to its role in parsing sensory information, this spatial mapping mechanism might also encode a dimension inherent to all experience: time. During space-clamped running throughout a temporal delay, so-called time cells preferentially fire at specific time points (66, 67). The populations of these time cells in the hippocampus and EC overlap substantially with the populations of place and grid cells, respectively, which suggests that cells in these regions might exhibit mixed selectivity for space and time (68, 69). Although these hippocampal time cells map repeated intervals in the range of seconds, it remains elusive how time might be discriminated at larger scales. For this, an ongoing temporal signal slowly drifting across hours and days might be transmitted to the hippocampus from the lateral EC, enabling the tagging of time and order to events for mnemonic processing (70).

The mapping of dimensions by the hippocampal-entorhinal system might be a general mechanism to code for task-relevant information. In a sound modulation task, rats manipulated the frequency of an auditory stimulus to “navigate” to a target frequency (71). Cells in the hippocampus and EC exhibited discrete firing fields with elevated firing constrained to each cell’s preferred frequency range, thereby mapping the continuous dimension of sound frequency so that neighboring frequencies exhibited similar firing patterns across cells (Fig. 3C). A subset of the cells carrying frequency-specific information in the sound modulation task could be characterized as place and grid cells during random foraging (71). Hippocampal representations of continuously changing odor concentrations suggest that the directional selectivity of place cells on a linear track might be preserved when mapping a nonspatial dimension (72). This suggests a flexible recruitment of cells to map task-relevant dimensions of experience while maintaining spatial coding properties.

However, we are constantly faced with complex stimuli defined by not one but multiple feature dimensions. Spatial coding has been demonstrated for stimuli varying along two independent dimensions. In one recent study, human participants learned to associate target objects with silhouettes of birds (Fig. 3D) differing in the length of their necks and legs (32). While undergoing fMRI, participants watched trajectories through this two-dimensional feature space, and BOLD (blood oxygen level–dependent) signals were analyzed as a function of the angular orientation of these trajectories following the analysis developed to investigate grid coding during virtual navigation (Box 1 and Fig. 2) (22). Hexadirectional signals were observed in the EC, as well as in a network of regions (32) involved in human autobiographical memory (63, 73, 74) that also exhibited hexadirectional signals during navigation (22). Within participants, the orientation of the hexadirectional signal was consistent across frontal and medial temporal brain areas and stable over separate sessions more than a week apart (32), suggesting a role of the entorhinal grid system in storage and retrieval of consolidated conceptual knowledge.

In our framework, place and grid cells not only map Euclidean space for navigation, but also map cognitive spaces spanned by relevant feature dimensions (Fig. 3A). Building on the geometric description of cognitive spaces, this mapping provides an account for how multidimensional spaces of experience can be instantiated in the hippocampal formation. If stimuli are located in the space based on their characteristic values on the feature axes, place cell firing might encode stimuli at specific locations in cognitive space based on the respective values along the feature dimensions (16). Similar stimuli are located at nearby positions, whereas dissimilar stimuli might be located farther apart in cognitive space. Distances between positions are reflected in population vectors of place cell activity, which will be more similar for nearby positions because of the overlapping firing fields of cell ensembles. Therefore, the mapping of cognitive spaces by cell populations in the hippocampal-entorhinal region provides a mechanism to generate similarity between stimuli (16), a concept central to generalization and planning (75).

Applying this to the introductory example of a two-dimensional space of engine power and weight, positions encoded by constrained firing of place cells reflect specific combinations of values along the dimensions—for instance, high engine power and low weight. This might fall in a region representing the concept of racing cars, which would typically be characterized by this feature combination. The convexity of conceptual regions enables the generalization that a car whose engine power and weight fall between those of two known sports cars also belongs to this category (16). In this framework, the regular firing patterns of grid cells might provide a metric (24, 26–28) for the dimensions of cognitive spaces. This would allow not only for the encoding of positions but also their relations in a fashion similar to that proposed for spatial navigation and memory (21, 25). Thereby, the grid system might provide the coordinate system for the dimensions of cognitive spaces, in line with its suggested role in extracting dimensions from task states (76, 77). However, the precise nature of the metric remains elusive, because different metrics can underlie cognitive spaces (16) and because the impact of deviations from regular hexagonal firing patterns (47, 78, 79) is unclear.

Neural recordings in patients suggest the existence of cells in the hippocampus and EC that selectively respond to a narrow range of stimuli (80). Despite the limitations imposed by the necessity of stimulus selection, the claim that this phenomenon reflects coding of conceptual entities such as famous people or places is supported by the preservation of responses across stimulus modalities. Even though there are no experimentally defined feature axes in these studies, these cells sometimes respond similarly to entities that appear nearby in cognitive space. For example, a cell responding to pictures of the Tower of Pisa also exhibited increased firing in response to the Eiffel Tower (80). This is in line with the assumption that proximal positions in cognitive space are represented by overlapping cell assemblies, comparable to the representations of nearby locations in space by overlapping place cell populations.

The actions and positions of other agents in space are central to interacting with and learning from conspecifics (81). In bats and rodents, hippocampal place cells of an observer animal encode the location of a conspecific or a moving object, indicating the tracking of relevant agents in the environment (82, 83). The human hippocampus also signals the position of others along interpersonal dimensions. In one study, participants were asked to respond to statements of fictitious characters located in a two-dimensional space

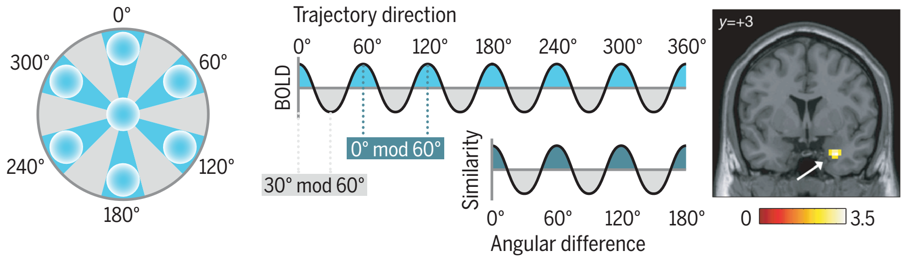


Fig. 2. Hexadirectional signals in fMRI. Left: The number of grid cell firing fields crossed depends on the direction of a trajectory through cognitive space. More fields are crossed during trajectories aligned with the grid (blue), translating to stronger entorhinal BOLD activity (top center; see text for details). Right: This effect was first observed during virtual navigation in the entorhinal cortex. [Adapted from (22)] The similarity structure of entorhinal multivoxel patterns exhibits a 60° modulation when comparing trajectories as a function of their angular difference as the grid is sampled at the same phase every 60° (bottom center), irrespective of its alignment.

of power and affiliation (84). Hippocampal activity tracked the position of the counterpart in this social space at times of interaction, exhibiting greater activity during interaction with counterparts with higher power and affiliation (84). Hippocampal encoding of the power dimension has also been demonstrated after learning of social hierarchies (85). These results suggest a hippocampal involvement in the representation of others in both navigable and interpersonal spaces.

It is thus conceivable that coding of fundamental dimensions of experience also underlies the involvement of the hippocampal-entorhinal system in episodic memory, where individual episodes are considered to be embedded in a spatiotemporal context (14). Because place and grid cells carry information about both space and time, they might thereby signal the context in which events are experienced. Evidence for this stems from a study in which participants encountered objects along a fixed route through a virtual city. Through learning, the similarity structure of multivoxel activity patterns in the hippocampus changed to reflect remembered spatial and temporal distances between object pairs (86). This is in line with the notion that the hippocampus encodes the dimensions of space and time along which episodic experiences are organized and the central role of the hippocampal formation in binding stimuli or events to a context in service of episodic memory (8, 10, 87).

Together, these findings suggest that the firing of functionally defined cell types in the hippocampal-entorhinal system prevails across task-relevant dimensions to map dimensions of experience in cognitive spaces. Stimuli are arranged in a spatial format where similarity between positions is reflected in the distance along the dimensions spanning the cognitive space. The representation of cognitive spaces allows not only associative or transitive inference, accounted for via overlapping relational networks in the realm of relational memory theory (8–10), but also generalization and inference to novel stimuli and situations. Below, we describe how key coding principles of the hippocampal-entorhinal system make it an ideal candidate for entertaining cognitive spaces.

Multiple scales of coding

One hallmark of abstract knowledge is the representation of information at different hierarchical levels (88, 89). For example, you might identify a particular vehicle characterized by low weight and high engine power as a Porsche (Fig. 4A). On a more general level, you might classify it as a sports car, allowing you to infer some properties such as high maximum speed, whereas on a more specific level you might wonder about the particular model and its associated characteristics. A likely mechanism for learning and representing information at different scales (88) is described in studies investigating the response properties of place and grid cells along the dorsoventral axis of the rodent hippocampal formation (Fig. 4B) (24, 46–49). On an 18-m-long linear track, the width of hippocampal place cell firing fields increased from less than 1 m in the dorsal hippocampus to approximately 10 m in the

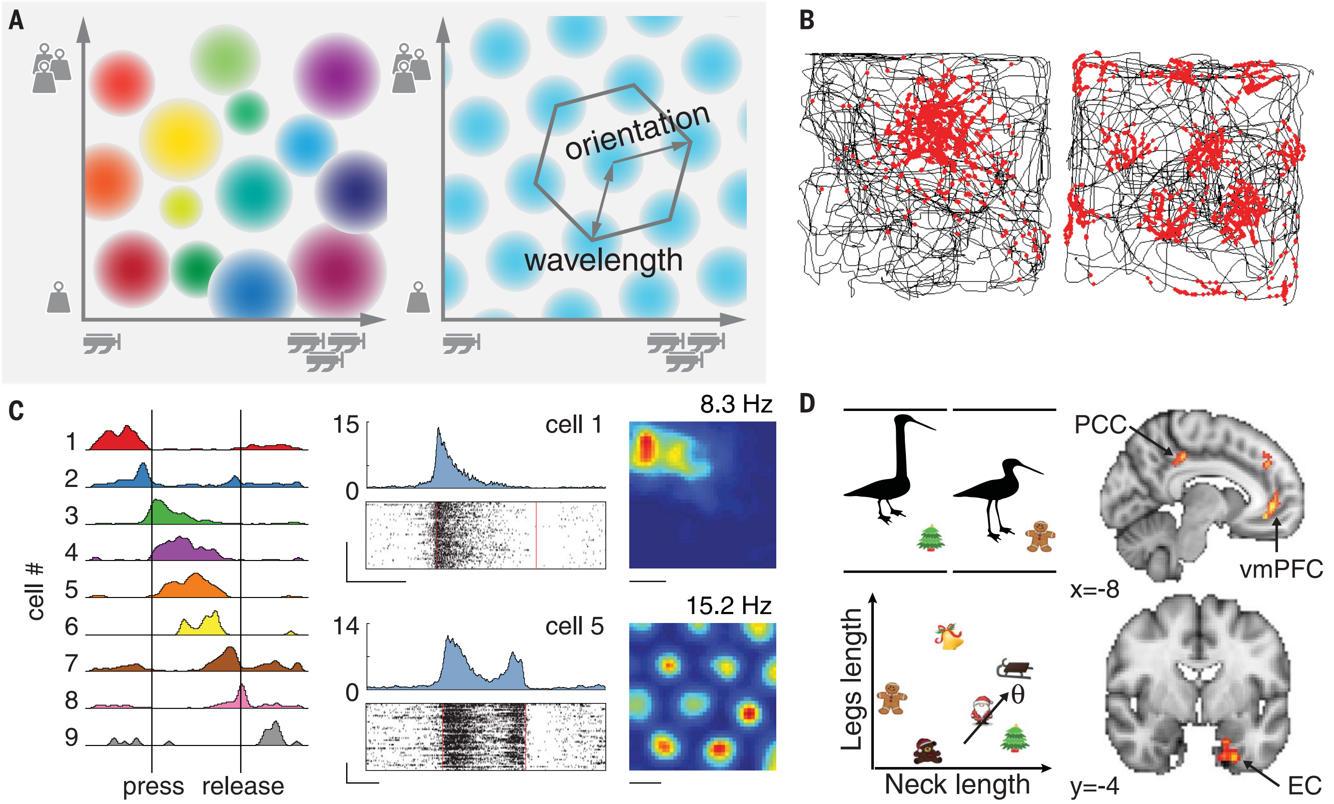


Fig. 3. Place and grid cells map dimensions of cognitive spaces. (**A**) Left: Colored circles illustrate firing fields of hippocampal place cells in a space spanned by the dimensions of car engine power and weight. Each color represents one individual place cell. Collectively, the firing fields map all locations of the space. Right: The hexagonally symmetric firing pattern of grid cells (one cell's pattern is shown) provides a metric for the dimensions of the space. (**B**) Firing of a place cell (left) and a grid cell (right) recorded from the hippocampus and entorhinal cortex, respectively. Spike locations (red dots) are shown on the animal's path (black line) through a square enclosure. [Adapted from (164)] (**C**) Frequency fields of place and grid cells. Left: While the animal presses a lever to manipulate the frequency of a tone, different cells fire at different frequencies during the frequency sweep. Center and right: Two example cells active in the sound-modulation task identified as a place cell and a grid cell, respectively, during navigation. Center panels show peristimulus histograms with firing rate in hertz (top) and raster plots displaying spikes as a function of sound frequency for different trials (bottom). Right panels show spatial firing rate maps with maximum firing rate. [Adapted from (71)] (**D**) Hexadirectional signals in two-dimensional bird space. Left: Participants learned associations of Christmas symbols and bird silhouettes with variable leg and neck length. Arrow illustrates example trajectory with angle θ. Right: During trajectories through this space defined by morphing birds, hexadirectional signals were observed in the entorhinal cortex (EC) and a network of brain regions implicated in mnemonic processing (PCC, posterior cingulate cortex; vmPFC, ventromedial prefrontal cortex). [Adapted from (32)]

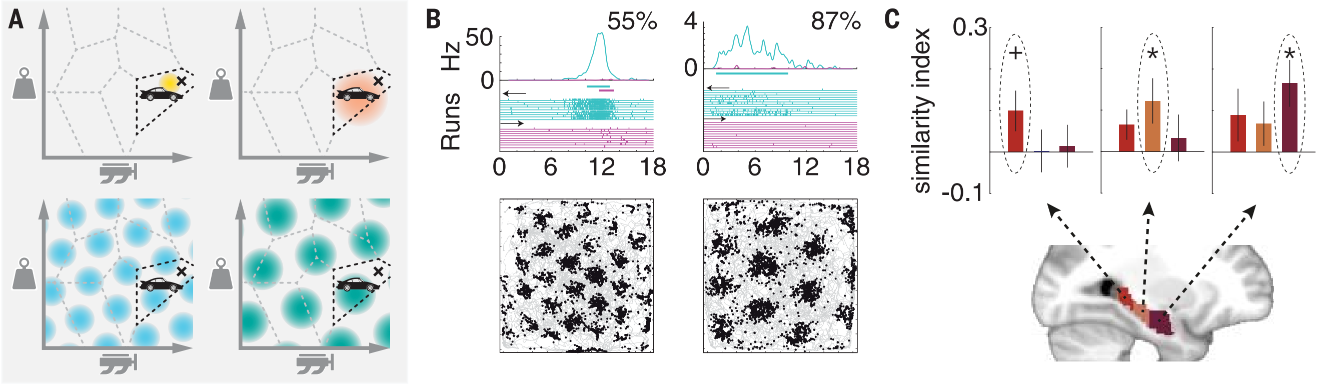


Fig. 4. Multiple levels of representation. (**A**) Representations of information at different levels of granularity can be supported by multiple spatial scales in the hippocampal formation. Narrower (left column) or broader (right column) positions in cognitive spaces can be encoded by place cells (top row) and grid cells (bottom row), respectively. Schematics illustrate firing fields of four different cells. Higher-level information (e.g., about the concept of sports cars) can be ascribed to a lower-level stimulus (black cross). For example, when learning that a specific car model is a sports car, we can infer that it is likely to have high engine power. (**B**) The firing field size of place cells (top) and the size and spacing of grid firing fields (bottom) increase from dorsal to ventral recording sites in the rodent hippocampal-entorhinal system. Top panels show firing rates in hertz as a function of position along the linear track. Percentages indicate location along the dorsoventral axis. Bottom panels show spike locations (black) overlaid on the animal's path (gray). [Adapted from (*48*, *49*)] (**C**) The granularity of mnemonic networks scales along the long axis of the human hippocampus with pairwise associations of elements in posterior and integrated networks in anterior hippocampus. Bars show model evidence (mean ± SEM) for mnemonic networks of small (red), medium (orange), and large (purple) scale. $^{*}P \leq 0.05$, $^{+}P = 0.053$. [Adapted from (*93*)]

ventral hippocampus (*48*). In human fMRI, this might be reflected in a voxel-similarity gradient along the hippocampal long axis (*90*). Similarly, the scale of entorhinal grid cells increases from the dorsal to the ventral parts of the medial EC, reflecting larger firing fields and larger spacing between fields (*24*, *47*, *49*, *91*). In contrast to place cell firing field width (*48*), the scale of grid patterns changes in discrete steps between modules of cells sharing a similar scale and orientation (*24*, *47*, *49*, *78*).

Different scales of information represented at distinct anatomical locations of the hippocampal formation might serve as a general mechanism across different stimulus domains. Positional decoding in cognitive spaces might benefit from the combination of multiple scales of representation analog to navigable space (*26*, *27*). The spacing of firing fields of grid cells responding to locations in visual space increases with anatomical distance to the rhinal sulcus (*64*), which approximately reflects the anterior-posterior axis of the primate hippocampus corresponding to the dorsal-ventral axes of the rodent hippocampus (*92*). This is further paralleled by broader tuning of saccade direction cells as anatomical distance to the rhinal sulcus increases (*65*).

Multiple scales of coding along the dorsoventral axis of the rodent hippocampus are in line with the scale increase of mnemonic networks represented along the anterior-posterior axis of the human hippocampus (Fig. 4C) in participants watching videos of lifelike events forming narratives (*93*, *94*). The scale at which these mnemonic networks were represented differed across the hippocampus: The posterior portion represented associations of the most recently linked pair of events, whereas the mid-portion held information about multiple event pairs and patterns in the anterior hippocampus were indicative of integrated networks of all events in a narrative (*93*). Multiscale event representations, interacting with mnemonic processing in the hippocampus, have been dissociated along the cortical hierarchy in humans (*95*). Representing cognitive spaces at different scales allows for the generalization of specific experience and the formation of contextual codes via more global representations. In rodents performing a context-dependent object discrimination task, ventral hippocampal neurons exhibited responses generalizing across events within a spatial context while strongly distinguishing between contexts after extended learning, whereas dorsal hippocampal neurons discriminated events within the same context with activity patterns reflecting the hierarchical task structure (*96*, *97*). The representation of integrated codes for overlapping memories might further enable inferential reasoning about related memories. In humans, transitive inference and generalization are supported by anterior portions of the hippocampus (*98*), whereas more posterior portions are associated with the retention of original memories and element segregation when associations overlap (*99*, *100*). These findings suggest a mnemonic gradient along the hippocampal long axis in humans, paralleling the differences in granularity of spatial representations along the dorsoventral axis of the rodent hippocampus. The most detailed representation might allow for fine-grained discrimination of locations in cognitive space, whereas the representation of larger areas might enable inference and the generalization of behavior to never-experienced stimuli and situations, not limited by the need for associations between nodes in relational mnemonic networks (*8*–*10*).

Representing knowledge at different levels of granularity in cognitive spaces requires a cognitive code that can signal positions in this space at different resolutions. Generalizing from the above findings to conceptual information, it appears plausible that the gradient of granularity plays an important role in learning and representing hierarchical knowledge structures (*88*, *89*). In these, overarching categories can be conceived as larger areas of cognitive space, putatively represented by place and grid cells with larger firing fields. Subcategories at a finer scale would then correspond to locations nested within these larger areas. Such a nested representation of cognitive spaces could enable inference via the transfer of knowledge from the superordinate category to new exemplars or subcategories (*16*).

Flexible formation of stable cognitive spaces via remapping and attractor dynamics

The hippocampus has been shown to contribute to a variety of cognitive domains (*15*). If the hippocampal-entorhinal system maps multitudes of cognitive spaces, this system needs to exert remarkable flexibility in terms of the dimensions it can represent, as well as an ability to rapidly switch between cognitive spaces (Fig. 5A). This flexibility is demonstrated by the capacity of hippocampal place cells to undergo global remapping (*50*–*52*). Different subsets of hippocampal cells will exhibit place fields in two different environments with spatial relationships among cells active in both environments not being maintained, rendering the two maps orthogonal to each other (Fig. 5B) (*51*). In contrast to place cells, entorhinal grid patterns have varying offsets to boundaries in different environments but maintain their spatial phase relative to each other, resulting in consistent relations between the firing fields of grid cells (*101*, *102*). These relationships are maintained not only between environmental contexts but also across

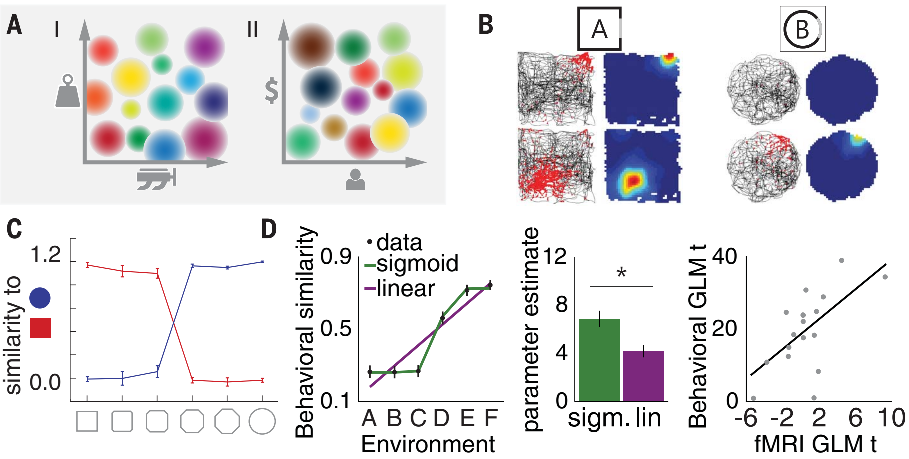


Fig. 5. Remapping and attractor dynamics for flexible cognitive spaces. (**A**) Different task-relevant dimensions (e.g., a car's passenger capacity and price instead of its engine power and weight) can be mapped by the recruitment of a different subset of place cells and the rearrangement of firing fields between spaces. (**B**) In navigating rodents, place cells remap to represent different environments. Spike locations (red) overlaid on the animal's path (black) and rate maps (warmer colors indicating increased firing) are shown for two place cells (rows) in two environments (columns). One place cell is active in the square environment but not in the circular environment, whereas the other exhibits a firing field in both environments but at unrelated positions. [Adapted from (*101*)] (**C**) Attractor dynamics enable rapid switches between established maps. The similarity to established maps of a square environment (red) and a circular environment (blue) is shown for a sequence of intermediate environments. The data show a sigmoidal rather than linear shift function. [Adapted from (*105*)] (**D**) Attractor dynamics in human spatial memory. After learning object locations in two base environments, A and F, participants were tested in a sequence of intermediate environments. Left: Spatial memory responses exhibited patterns better explained by a sigmoidal function than by a linear function. Center: Hippocampal multivoxel pattern similarity was better predicted by a sigmoidal model than by a linear model of behavior. Right: The fit of a canonical sigmoidal function to spatial memory responses was associated with hippocampal pattern similarity values across participants. T values from general linear modeling (GLM) reflect fit of sigmoidal model to hippocampal fMRI and behavioral data. Error bars in (C) and (D) reflect SEM. *$P < 0.05$. [Adapted from (*108*)]

behavioral states, with essentially identical cross-correlation patterns exhibited by populations of grid cells and other medial entorhinal cells during free foraging, slow-wave sleep (SWS), or rapid–eye movement (REM) sleep (*59*, *60*).

Remapping-like behavior of hippocampal cells has also been observed for time cells encoding temporal intervals during the delay of a memory-based discrimination task (*67*). When the length of the delay was altered, a subset of time cells remapped. Some cells ceased to be active, others shifted their activity to a different time during the delay, or previously silent cells became active, whereas other cells maintained their firing during absolute or relative times of the altered delay (*67*). The flexible recruitment of cells to map positions in cognitive spaces is further illustrated by hippocampal representations of conspecifics, where subsets of place cells exclusively encoded the location of the self, the conspecific, or an inanimate object, whereas other cells exhibited firing fields for both the self and other, but at different locations (*83*). The decorrelation of hippocampal representations has also been observed using multivariate pattern analyses of human episodic memory. Reconfigurations of hippocampal activity patterns reflect associations and narrative insight in increased similarity but also reduce overlap, resulting in decorrelation through experience (*103*, *104*).

If different cognitive spaces are represented by orthogonal subsets of hippocampal cells, how can already-formed representations of spaces be reinstated to provide stable maps over multiple encounters? In rodents, reinstatement of place cell firing patterns has been observed on a trial-by-trial level, indicating rapid switching between maps upon reexposure to a highly familiar circular and square enclosure (Fig. 5C), respectively (*105*). A likely candidate mechanism governing the reactivation of established cognitive spaces may involve attractor networks (*11*, *13*, *53*, *54*). In intermediate steps of a morph sequence transforming a square to a circular environment, hippocampal patterns resemble the original map until a switch point at which the representation is pulled toward exhibiting the other map (*105*), although progressive transformations of hippocampal codes have been observed under different experimental conditions (*106*). Neuroimaging research points toward a role for stored hippocampal representations in perceptual discrimination (*107*) and attractor dynamics specifically in memory-guided human behavior (Fig. 5D) (*108*). Participants learned positions of identical sets of objects in two virtual environments distinguished by background cues and were subsequently tested in a series of morph environments following a linear transformation between the original environments. Spatial memory responses in the intermediate environments followed the sinusoidal pattern predicted by the influence of attractor networks and were paralleled by corresponding nonlinear changes of hippocampal activity patterns (*108*).

These findings indicate that response properties of the hippocampal-entorhinal system enable the formation of independent maps for distinct cognitive spaces to map different dimensions. Consider again the car example: In a different context, weight and engine power might not be the most relevant feature dimensions; rather, a new space might be spanned by the dimensions of passenger capacity and price (Fig. 5A). Via hippocampal remapping, positions in the new space can be mapped and the positions of different cars are redefined based on their feature values on the new dimensions, resulting in new similarity relations. The more rigid firing properties of the entorhinal grid system might provide a stable metric through its intrinsic coherence across spaces. Once established, rapid switching between the maps of different cognitive spaces through remapping demonstrates the flexibility of the system and permits the representation of relevant information based on behavioral context. Switches between maps might be triggered by internal or external cues, attentional shifts, or changes in task demands rendering other dimensions relevant. Attractor dynamics might serve as anchors for stable representations, which enable learning over multiple encounters as well as the generalization of behavior across similar experiences.

Simulations and readout of trajectories for decisions

Recording studies in rodents have revealed structured activity of hippocampal place cells during rest in which cells fire sequentially, resulting in trajectories reflecting past experience on a maze (*55*, *56*). This so-called "replay" demonstrates the ability of spatially tuned cells to represent locations beyond the animal's current position (*57*). During replay sequences, place cells maintain their spatial relationships to each other (*55*, *109*), resulting in trajectories through space occurring both in forward and reverse order (*109*, *110*). Place cell sequences during replay are temporally compressed relative to trajectories during running (*56*, *57*, *111*) and might thereby allow fast simulations of trajectories through cognitive space (Fig. 6, A and B). Sequences replaying trajectories in reverse order might evaluate previous paths and associate the states visited with reward information for learning of adaptive behavior (*57*, *112*–*114*). Hippocampal sequences may extend to the simulation of potential future paths and their outcomes. In a spatial alternation task, replay of both correct and incorrect future trajectories supports

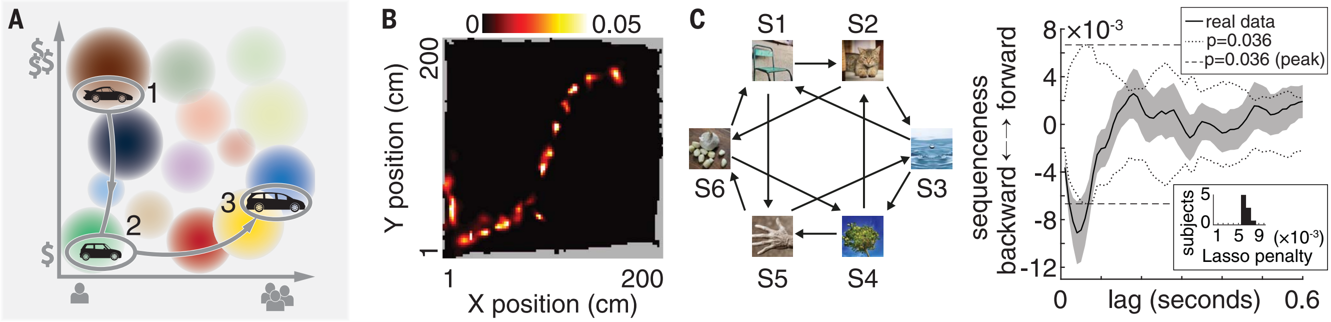


Fig. 6. Hippocampal simulations via sequential activity. (**A**) Sequential activity simulates different positions (1 to 3) in cognitive space, allowing the evaluation of different car types along the trajectory when deciding which car to buy. (**B**) Simulation of trajectory in a square enclosure. The posterior probabilities of positions from Bayesian decoding applied to different time frames during a sharp wave ripple event result in a trajectory through space. [Adapted from (*165*)] (**C**) Reverse trajectories through state space. Left: States are represented by images; arrows indicate possible transitions between states. Right: The solid black line shows a "sequenceness" measure indicating the probability of decoding a successive or preceding state (positive and negative values, respectively) at different time points after the decoding of a state during planning. Reverse sequences were observed with a lag of around 40 ms between state space positions. Inset shows histogram of lasso penalties from logistic regression. [Adapted from (*125*)]

learning and planning (*115*) and the disruption of sequences during sharp-wave ripples impairs successful performance (*116*). Further, place cell sequences reflect future paths during goal-directed behavior (*117*) and trajectories extending into the shock zone of a linear track prior to avoidance behavior (*118*). These findings dovetail with extensive research on place cell sequences during ongoing theta oscillations, where trajectories toward potential goals are represented within different theta cycles (*61*), suggesting simulations of possible trajectories through cognitive space.

What role do mental simulations of trajectories play for planning and decision-making in humans? During navigational planning on a circular track, hippocampal activity patterns carry information not only about the start and goal location, but also locations along the optimal rather than suboptimal path, in accordance with sequential simulations of trajectories through space (*119*). The entorhinal grid system has also been linked to imagining navigation (*34*) and snapshots from stationary viewpoints (*33*), implicating it in the planning of trajectories through space in line with replay in the EC (*58*–*60*, *120*) and the observation of grid cell firing during covert attentional trajectories through visual space (*121*). Beyond navigable space, the hippocampal-entorhinal system extracts statistical regularities of nonspatial sequences (*122*, *123*) and forms relational maps of the underlying structures from which the sequences were derived (*104*, *124*), potentially drawn upon to plan trajectories through task spaces. Indeed, in a nonspatial decision-making task, sequential reactivations of previously visited states reflected reverse trajectories (Fig. 6C) through a space of discrete states represented by objects (*125*). Although the nature of the magnetoencephalography (MEG) signal and the analysis approach focusing on visual responses make a direct hippocampal origin of these results unlikely, they might reflect sequential reinstatement of visual representations orchestrated by the hippocampus (*125*). Initial evidence suggests that sequential hippocampal activity can be observed using fMRI (*126*). Episodic cues can elicit memory-guided simulation of past experience influencing choice behavior (*127*), and prospective simulations of trajectories through a task's state space have been linked to model-based behavior in a two-stage decision-making task (*128*). These findings indicate that trajectories through cognitive spaces representing task states can be simulated on the basis of prior experience.

Cognitive spaces enable generalization and can reveal novel trajectories via the representation of positions along defined dimensions. The rodent hippocampus can recombine separate trajectories across segments of a two-choice T-maze, usually not experienced successively, to infrequently construct never-experienced place cell sequences (*113*). Similarly, hippocampal place cells constructed trajectories through previously nontraversed space after rats observed the placement of a reward on one T-maze arm while confined to the stem of the maze (*129*). In humans, hippocampal simulations have been linked to the construction of imaginary scenarios (*12*, *62*, *63*, *73*)—an ability impaired in patients with hippocampal lesions (*130*). Mental simulations can be conceived as putative trajectories through the space of episodic experience (*12*, *63*) and have been shown to influence decision-making in a delay-discounting task (*131*). Furthermore, in a task in which participants imagined novel compound goods consisting of two familiar foods, the hippocampus and medial prefrontal cortex flexibly combine past experiences (*132*). These findings show that the hippocampus can flexibly draw on past experience to form and simulate novel trajectories through cognitive space allowing adaptive decision-making and behavior.

Simulations of trajectories building on spatially tuned cells have been incorporated in models of episodic memory and consolidation (*12*, *63*, *133*). Replay-inspired simulation of experience has also been used to integrate reinforcement learning and deep neural networks in artificial intelligence (*134*). Here, we suggest that contemplating a number of stimuli can be conceived as a trajectory through cognitive space. In line with its proposed role in future anticipation and prediction (*62*, *63*, *73*, *77*, *130*), the hippocampal-entorhinal system supports these trajectories via sequential activity of spatially tuned cells. The entorhinal grid system might span a space based on a set of dimensions and thereby provide the framework for flexible simulations of positions and trajectories by the hippocampus. Relevant feature dimensions can range from locations in a maze to abstract state spaces in decision-making problems. Novel trajectories can be generated from past experience, and trajectories can be simulated via sequential hippocampal activity to guide future behavior. Building on the geometric characterization of cognitive spaces, experience can be generalized to the outcome of novel trajectories or actions. For example, if two prior actions undershot and overshot a goal, then an intermediate trajectory through cognitive space will approximate the goal more closely. Although it has been suggested that sequential activity might reflect inherent hippocampal dynamics (*135*, *136*), other accounts highlight hippocampal interactions with state-space representations in frontal regions in sequence generation and action selection, as well as the role of striatal regions in sequence evaluation (*131*, *137*–*140*).

Open questions and future directions

We used spatial navigation in rodents and humans (*25*, *38*) as a model system to identify key neural mechanisms and fused them with concepts from cognitive science to describe the central neural coding machinery underlying higher-level cognition in humans. Future research should help to elucidate the generation of cognitive spaces and their governing principles. For example, the conditions of the generation of the continuous code in the hippocampal-entorhinal system are still largely unclear. This concerns both the circuit-level interactions of brain regions (*141*) and their maturation (*142*), the role of spatial cognition during development (*143*), and the potential breakdown of spatial codes in aging and disease (*31*, *37*, *144*). Of particular interest is how precisely the entorhinal grid code emerges in its hexagonally symmetric form. Advances in techniques such as two-photon calcium imaging (*145*) will foster the population-level understanding of neural codes in rodents. In concert with the rise of high-resolution fMRI and optimized MEG protocols, this should further bridge the gap between systems and cognitive neuroscience to unravel neural mechanisms promoting refinements of processing in cognitive spaces, and to spur the mapping of brain structures to specific functions.

Cognitive spaces can be multidimensional. Therefore, the question of how a continuous code can be extended to map additional dimensions is fundamental. Research in humans (*146*), rats (*147*), and bats has shed light on how spatial coding of place cells (*148*) and head direction cells (*149*) can be extended to the third dimension. However, grid-like coding in three dimensions remains elusive, as does evidence for spatial coding in cognitive spaces of higher dimensionality. Related to this is the question of how cognitive spaces are spanned by multiple, potentially integral, dimensions. Further, the nature of the metric underlying cognitive spaces can be investigated using measures of neural and behavioral similarity between positional representations. The assumption of a Euclidean metric (*150*) might be most intuitive when comparing cognitive to navigable space, although there is evidence for topological representations of spaces in both rodents (*151*) and humans (*124*, *152*, *153*).

Grid cells have been implicated in representing dimensions of cognitive spaces, but the extent to which they retain their specific firing patterns remains to be explored. For instance, is the modular organization of the EC stable across cognitive spaces? One might expect that grid cells from the same module also show similar orientation and spacing when representing dimensions of a given cognitive space, but differ in their spatial phase from the boundaries of the space. Generalizing from this question, the role of other cell types encoding spatial information in the context of navigation remains to be elucidated. For example, one could envision a role for border cells in signaling event boundaries (*154*–*156*) or conceptual boundaries when learning to categorize stimuli drawn from a cognitive space as belonging to different concepts. Likewise, head direction cells (*39*), goal direction cells (*41*), or object vector cells (*157*) might be involved in representing relationships between stimuli located at different locations in a cognitive space. Similarly, the role played by networks of brain regions acting in concert with the hippocampus (*32*, *63*, *73*, *74*) for cognitive spaces and differences in the encoding of multidimensional stimulus spaces in other brain regions—for example, during face processing in monkey inferotemporal cortex (*158*)—should be explored.

An intriguing question concerns the measurement of behavioral benefits of cognitive space formation—for example, via the generation of shortcuts through cognitive space or the impaired ability to find shortcuts in lesion patients. Further, how does information encoded in different cognitive spaces interact? Can trajectories encoded in one space be transferred to another and be retrieved there to guide behavior? Conversely, it might be possible to bring codes from different cognitive spaces into conflict with one another, which might result in interference across spaces, or to investigate effects of deformations of firing patterns (*47*, *78*, *79*) across spaces to further elucidate how representations of different spaces are entertained by overlapping neural substrates.

Conclusion

In this theoretical article, we propose cognitive spaces as a primary representational format for information processing in the brain. Combining key mechanisms identified in systems neuroscience and concepts from cognitive science and philosophy, we developed a cognitive neuroscience framework for processing and representing information in cognitive spaces in the hippocampal-entorhinal system. Place and grid cells might have evolved to represent not only navigable space, but to also map dimensions of experience spanning cognitive spaces governed by geometric principles. In these cognitive spaces, stimuli can be located based on their values along the feature dimensions mapped by place and grid cells. These spatially specific cells provide a continuous code that allows similar stimuli to occupy neighboring positions in cognitive space, encoded by overlapping population responses. In this framework, concepts are represented by convex regions of similar stimuli. The multiscale spatial code along the long axis of the hippocampal formation enables representing stimuli at different granularities for both generalization and maintenance of fine details in hierarchical knowledge structures. Ever-changing demands requiring the flexible mapping of different dimensions of relevance are met by the capacity of the hippocampus to remap to flexibly form cognitive spaces for which the low-dimensional entorhinal grid code might provide a stable metric. An established mapping of a cognitive space might be reinstated via attractor dynamics and pattern completion to provide stable representations of familiar dimensions. Experiencing a sequence of stimuli results in a trajectory through cognitive space. We propose that sequential hippocampal activity in the form of replay and theta sequences allows simulations of temporally compressed trajectories through cognitive spaces for flexible cognition and adaptive behavior. In sum, we suggest cognitive spaces as a domain-general format for human thinking, thus providing an overarching framework, which can also help to elucidate cognitive breakdown in neurodegenerative diseases (*31*, *144*) and to inform novel architectures in artificial intelligence (*134*).

REFERENCES AND NOTES

1. E. C. Tolman, Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208 (1948). doi: 10.1037/h0061626; pmid: 18870876
2. J. O'Keefe, J. Dostrovsky, The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171–175 (1971). doi: 10.1016/0006-8993(71)90358-1; pmid: 5124915
3. J. O'Keefe, L. Nadel, *The Hippocampus as a Cognitive Map* (Clarendon, 1978).
4. K. M. Gothard, W. E. Skaggs, K. M. Moore, B. L. McNaughton, Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. *J. Neurosci.* **16**, 823–835 (1996). doi: 10.1523/JNEUROSCI.16-02-00823.1996; pmid: 8551362
5. E. R. Wood, P. A. Dudchenko, H. Eichenbaum, The global record of memory in hippocampal neuronal activity. *Nature* **397**, 613–616 (1999). doi: 10.1038/17605; pmid: 10050854
6. B. J. Young, G. D. Fox, H. Eichenbaum, Correlates of hippocampal complex-spike cell activity in rats performing a nonspatial radial maze task. *J. Neurosci.* **14**, 6553–6563 (1994). doi: 10.1523/JNEUROSCI.14-11-06553.1994; pmid: 7965059
7. E. R. Wood, P. A. Dudchenko, R. J. Robitsek, H. Eichenbaum, Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* **27**, 623–633 (2000). doi: 10.1016/S0896-6273(00)00071-4; pmid: 11055443
8. N. J. Cohen, H. Eichenbaum, *Memory, Amnesia, and the Hippocampal System* (MIT Press, 1993).
9. H. Eichenbaum, N. J. Cohen, Representation in the hippocampus: What do hippocampal neurons code? *Trends Neurosci.* **11**, 244–248 (1988). doi: 10.1016/0166-2236(88)90100-2; pmid: 2465617
10. H. Eichenbaum, N. J. Cohen, Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron* **83**, 764–770 (2014). doi: 10.1016/j.neuron.2014.07.032; pmid: 25144874
11. D. Marr, Simple memory: A theory for archicortex. *Philos. Trans. R. Soc. London Ser. B* **262**, 23–81 (1971). doi: 10.1098/rstb.1971.0078; pmid: 4399412
12. M. E. Hasselmo, *How We Remember: Brain Mechanisms of Episodic Memory* (MIT Press, 2011).
13. B. L. McNaughton, R. G. M. Morris, Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci.* **10**, 408–415 (1987). doi: 10.1016/0166-2236(87)90011-7
14. D. Schiller *et al.*, Memory and Space: Towards an Understanding of the Cognitive Map. *J. Neurosci.* **35**, 13904–13911 (2015). doi: 10.1523/JNEUROSCI.2618-15.2015; pmid: 26468191
15. D. E. Hannula, M. C. Duff, *The Hippocampus from Cells to Systems: Structure, Connectivity, and Functional Contributions to Memory and Flexible Cognition* (Springer, 2017).
16. P. Gärdenfors, *Conceptual Spaces: The Geometry of Thought* (MIT Press, 2000).
17. C. Balkenius, P. Gärdenfors, Spaces in the Brain: From Neurons to Meanings. *Front. Psychol.* **7**, 1820 (2016). doi: 10.3389/fpsyg.2016.01820; pmid: 27920740
18. F. Zenker, P. Gärdenfors, Eds., *Applications of Conceptual Spaces: The Case for Geometric Knowledge Representation* (Springer, 2015).
19. E. Rosch, Cognitive reference points. *Cognit. Psychol.* **7**, 532–547 (1975). doi: 10.1016/0010-0285(75)90021-3
20. R. E. Clark, L. R. Squire, Similarity in form and function of the hippocampus in rodents, monkeys, and humans. *Proc. Natl. Acad. Sci. U.S.A.* **110** (suppl. 2), 10365–10370 (2013). doi: 10.1073/pnas.1301225110; pmid: 23754372
21. G. Buzsáki, E. I. Moser, Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nat. Neurosci.* **16**, 130–138 (2013). doi: 10.1038/nn.3304; pmid: 23354386
22. C. F. Doeller, C. Barry, N. Burgess, Evidence for grid cells in a human memory network. *Nature* **463**, 657–661 (2010). doi: 10.1038/nature08704; pmid: 20090680
23. E. S. Spelke, S. A. Lee, Core systems of geometry in animal minds. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 2784–2793 (2012). doi: 10.1098/rstb.2012.0210; pmid: 22927577
24. T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, E. I. Moser, Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005). doi: 10.1038/nature03721; pmid: 15965463

25. E. I. Moser, M.-B. Moser, B. L. McNaughton, Spatial representation in the hippocampal formation: A history. *Nat. Neurosci.* **20**, 1448–1464 (2017). doi: 10.1038/nn.4653; pmid: 29073644
26. D. Bush, C. Barry, D. Manson, N. Burgess, Using Grid Cells for Navigation. *Neuron* **87**, 507–520 (2015). doi: 10.1016/j.neuron.2015.07.006; pmid: 26247860
27. A. V. Herz, A. Mathis, M. Stemmler, Periodic population codes: From a single circular variable to higher dimensions, multiple nested scales, and conceptual spaces. *Curr. Opin. Neurobiol.* **46**, 99–108 (2017). doi: 10.1016/j.conb.2017.07.005; pmid: 28888183
28. B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, M.-B. Moser, Path integration and the neural basis of the 'cognitive map'. *Nat. Rev. Neurosci.* **7**, 663–678 (2006). doi: 10.1038/nrn1932; pmid: 16858394
29. A. D. Ekstrom *et al*., Cellular networks underlying human spatial navigation. *Nature* **425**, 184–188 (2003). doi: 10.1038/nature01964; pmid: 12968182
30. J. Jacobs *et al*., Direct recordings of grid-like neuronal activity in human spatial navigation. *Nat. Neurosci.* **16**, 1188–1190 (2013). doi: 10.1038/nn.3466; pmid: 23912946
31. L. Kunz *et al*., Reduced grid-cell-like representations in adults at genetic risk for Alzheimer's disease. *Science* **350**, 430–433 (2015). doi: 10.1126/science.aac8128; pmid: 26494756
32. A. O. Constantinescu, J. X. O'Reilly, T. E. J. Behrens, Organizing conceptual knowledge in humans with a gridlike code. *Science* **352**, 1464–1468 (2016). doi: 10.1126/science.aaf0941; pmid: 27313047
33. J. L. S. Bellmund, L. Deuker, T. Navarro Schröder, C. F. Doeller, Grid-cell representations in mental simulation. *eLife* **5**, e17089 (2016). doi: 10.7554/eLife.17089; pmid: 27572056
34. A. J. Horner, J. A. Bisby, E. Zotow, D. Bush, N. Burgess, Grid-like Processing of Imagined Navigation. *Curr. Biol.* **26**, 842–847 (2016). doi: 10.1016/j.cub.2016.01.042; pmid: 26972318
35. M. Nau, T. Navarro Schröder, J. L. S. Bellmund, C. F. Doeller, Hexadirectional coding of visual space in human entorhinal cortex. *Nat. Neurosci.* **21**, 188–190 (2018). doi: 10.1038/s41593-017-0050-8; pmid: 29311746
36. J. B. Julian, A. T. Keinath, G. Frazzetta, R. A. Epstein, Human entorhinal cortex represents visual space using a boundary-anchored grid. *Nat. Neurosci.* **21**, 191–194 (2018). doi: 10.1038/s41593-017-0049-1; pmid: 29311745
37. M. Stangl *et al*., Compromised Grid-Cell-like Representations in Old Age as a Key Mechanism to Explain Age-Related Navigational Deficits. *Curr. Biol.* **28**, 1108–1115.e6 (2018). doi: 10.1016/j.cub.2018.02.038; pmid: 29551413
38. R. A. Epstein, E. Z. Patai, J. B. Julian, H. J. Spiers, The cognitive map in humans: Spatial navigation and beyond. *Nat. Neurosci.* **20**, 1504–1513 (2017). doi: 10.1038/nn.4656; pmid: 29073650
39. K. E. Cullen, J. S. Taube, Our sense of direction: Progress, controversies and challenges. *Nat. Neurosci.* **20**, 1465–1473 (2017). doi: 10.1038/nn.4658; pmid: 29073639
40. V. Hok, E. Save, P. P. Lenck-Santini, B. Poucet, Coding for spatial goals in the prelimbic/infralimbic area of the rat frontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4602–4607 (2005). doi: 10.1073/pnas.0407332102; pmid: 15761059
41. A. Sarel, A. Finkelstein, L. Las, N. Ulanovsky, Vectorial representation of spatial goals in the hippocampus of bats. *Science* **355**, 176–180 (2017). doi: 10.1126/science.aak9589; pmid: 28082589
42. E. Kropff, J. E. Carmichael, M.-B. Moser, E. I. Moser, Speed cells in the medial entorhinal cortex. *Nature* **523**, 419–424 (2015). doi: 10.1038/nature14622; pmid: 26176924
43. F. Savelli, D. Yoganarasimha, J. J. Knierim, Influence of boundary removal on the spatial representations of the medial entorhinal cortex. *Hippocampus* **18**, 1270–1282 (2008). doi: 10.1002/hipo.20511; pmid: 19021262
44. T. Solstad, C. N. Boccara, E. Kropff, M.-B. Moser, E. I. Moser, Representation of geometric borders in the entorhinal cortex. *Science* **322**, 1865–1868 (2008). doi: 10.1126/science.1166466; pmid: 19095945
45. C. Lever, S. Burton, A. Jeewajee, J. O'Keefe, N. Burgess, Boundary vector cells in the subiculum of the hippocampal formation. *J. Neurosci.* **29**, 9771–9777 (2009). doi: 10.1523/JNEUROSCI.1319-09.2009; pmid: 19657030
46. M. W. Jung, S. I. Wiener, B. L. McNaughton, Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *J. Neurosci.* **14**, 7347–7356 (1994). doi: 10.1523/JNEUROSCI.14-12-07347.1994; pmid: 7996180
47. C. Barry, R. Hayman, N. Burgess, K. J. Jeffery, Experience-dependent rescaling of entorhinal grids. *Nat. Neurosci.* **10**, 682–684 (2007). doi: 10.1038/nn1905; pmid: 17486102
48. K. B. Kjelstrup *et al*., Finite scale of spatial representation in the hippocampus. *Science* **321**, 140–143 (2008). doi: 10.1126/science.1157086; pmid: 18599792
49. H. Stensola *et al*., The entorhinal grid map is discretized. *Nature* **492**, 72–78 (2012). doi: 10.1038/nature11649; pmid: 23222610
50. E. Bostock, R. U. Muller, J. L. Kubie, Experience-dependent modifications of hippocampal place cell firing. *Hippocampus* **1**, 193–205 (1991). doi: 10.1002/hipo.450010207; pmid: 1669293
51. S. Leutgeb, J. K. Leutgeb, A. Treves, M.-B. Moser, E. I. Moser, Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science* **305**, 1295–1298 (2004). doi: 10.1126/science.1100265; pmid: 15272123
52. R. U. Muller, J. L. Kubie, The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *J. Neurosci.* **7**, 1951–1968 (1987). doi: 10.1523/JNEUROSCI.07-07-01951.1987; pmid: 3612226
53. J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554–2558 (1982). doi: 10.1073/pnas.79.8.2554; pmid: 6953413
54. J. L. McClelland, B. L. McNaughton, R. C. O'Reilly, Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* **102**, 419–457 (1995). doi: 10.1037/0033-295X.102.3.419; pmid: 7624455
55. M. A. Wilson, B. L. McNaughton, Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679 (1994). doi: 10.1126/science.8036517; pmid: 8036517
56. W. E. Skaggs, B. L. McNaughton, Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* **271**, 1870–1873 (1996). doi: 10.1126/science.271.5257.1870; pmid: 8596957
57. D. J. Foster, Replay Comes of Age. *Annu. Rev. Neurosci.* **40**, 581–602 (2017). doi: 10.1146/annurev-neuro-072116-031538; pmid: 28772098
58. H. F. Ólafsdóttir, F. Carpenter, C. Barry, Coordinated grid and place cell replay during rest. *Nat. Neurosci.* **19**, 792–794 (2016). doi: 10.1038/nn.4291; pmid: 27089021
59. R. J. Gardner, L. Lu, T. Wernle, M.-B. Moser, E. I. Moser, Correlation structure of grid cells is preserved during sleep. bioRxiv [Preprint]. 5 October 2017. pmid: 198499
60. S. G. Trettel, J. B. Trimper, E. Hwaun, I. R. Fiete, L. L. Colgin, Grid cell co-activity patterns during sleep reflect spatial overlap of grid fields during active behaviors. bioRxiv [Preprint]. 5 October 2017. pmid: 198671
61. A. Johnson, A. D. Redish, Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J. Neurosci.* **27**, 12176–12189 (2007). doi: 10.1523/JNEUROSCI.3761-07.2007; pmid: 17989284
62. R. L. Buckner, The role of the hippocampus in prediction and imagination. *Annu. Rev. Psychol.* **61**, 27–48 (2010). doi: 10.1146/annurev.psych.60.110707.163508; pmid: 19958178
63. P. Byrne, S. Becker, N. Burgess, Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychol. Rev.* **114**, 340–375 (2007). doi: 10.1037/0033-295X.114.2.340; pmid: 17500630
64. N. J. Killian, M. J. Jutras, E. A. Buffalo, A map of visual space in the primate entorhinal cortex. *Nature* **491**, 761–764 (2012). doi: 10.1038/nature11587; pmid: 23103863
65. N. J. Killian, S. M. Potter, E. A. Buffalo, Saccade direction encoding in the primate entorhinal cortex during visual exploration. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15743–15748 (2015). doi: 10.1073/pnas.1417059112; pmid: 26644558
66. E. Pastalkova, V. Itskov, A. Amarasingham, G. Buzsáki, Internally generated cell assembly sequences in the rat hippocampus. *Science* **321**, 1322–1327 (2008). doi: 10.1126/science.1159775; pmid: 18772431
67. C. J. MacDonald, K. Q. Lepage, U. T. Eden, H. Eichenbaum, Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron* **71**, 737–749 (2011). doi: 10.1016/j.neuron.2011.07.012; pmid: 21867888
68. B. J. Kraus, R. J. Robinson 2nd, J. A. White, H. Eichenbaum, M. E. Hasselmo, Hippocampal "time cells": Time versus path integration. *Neuron* **78**, 1090–1101 (2013). doi: 10.1016/j.neuron.2013.04.015; pmid: 23707613
69. B. J. Kraus *et al*., During Running in Place, Grid Cells Integrate Elapsed Time and Distance Run. *Neuron* **88**, 578–589 (2015). doi: 10.1016/j.neuron.2015.09.031; pmid: 26539893
70. A. Tsao *et al*., Integrating time from experience in the lateral entorhinal cortex. *Nature* **561**, 57–62 (2018). doi: 10.1038/s41586-018-0459-6
71. D. Aronov, R. Nevers, D. W. Tank, Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. *Nature* **543**, 719–722 (2017). doi: 10.1038/nature21692; pmid: 28358077
72. B. A. Radvansky, D. A. Dombeck, An olfactory virtual reality system for mice. *Nat. Commun.* **9**, 839 (2018). doi: 10.1038/s41467-018-03262-4; pmid: 29483530
73. D. L. Schacter, D. R. Addis, R. L. Buckner, Remembering the past to imagine the future: The prospective brain. *Nat. Rev. Neurosci.* **8**, 657–661 (2007). doi: 10.1038/nrn2213; pmid: 17700624
74. C. Ranganath, M. Ritchey, Two cortical systems for memory-guided behaviour. *Nat. Rev. Neurosci.* **13**, 713–726 (2012). doi: 10.1038/nrn3338; pmid: 22992647
75. R. N. Shepard, Toward a universal law of generalization for psychological science. *Science* **237**, 1317–1323 (1987). doi: 10.1126/science.3629243; pmid: 3629243
76. Y. Dordek, D. Soudry, R. Meir, D. Derdikman, Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. *eLife* **5**, e10094 (2016). doi: 10.7554/eLife.10094; pmid: 26952211
77. K. L. Stachenfeld, M. M. Botvinick, S. J. Gershman, The hippocampus as a predictive map. *Nat. Neurosci.* **20**, 1643–1653 (2017). doi: 10.1038/nn.4650; pmid: 28967910
78. J. Krupic, M. Bauza, S. Burton, C. Barry, J. O'Keefe, Grid cell symmetry is shaped by environmental geometry. *Nature* **518**, 232–235 (2015). doi: 10.1038/nature14153; pmid: 25673417
79. T. Stensola, H. Stensola, M.-B. Moser, E. I. Moser, Shearing-induced asymmetry in entorhinal grid cells. *Nature* **518**, 207–212 (2015). doi: 10.1038/nature14151; pmid: 25673414
80. R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, I. Fried, Invariant visual representation by single neurons in the human brain. *Nature* **435**, 1102–1107 (2005). doi: 10.1038/nature03687; pmid: 15973409
81. A. Stolk, L. Verhagen, I. Toni, Conceptual Alignment: How Brains Achieve Mutual Understanding. *Trends Cogn. Sci.* **20**, 180–191 (2016). doi: 10.1016/j.tics.2015.11.007; pmid: 26792458
82. T. Danjo, T. Toyoizumi, S. Fujisawa, Spatial representations of self and other in the hippocampus. *Science* **359**, 213–218 (2018). doi: 10.1126/science.aao3898; pmid: 29326273
83. D. B. Omer, S. R. Maimon, L. Las, N. Ulanovsky, Social place-cells in the bat hippocampus. *Science* **359**, 218–224 (2018). doi: 10.1126/science.aao3474; pmid: 29326274
84. R. M. Tavares *et al*., A Map for Social Navigation in the Human Brain. *Neuron* **87**, 231–243 (2015). doi: 10.1016/j.neuron.2015.06.011; pmid: 26139376
85. D. Kumaran, A. Banino, C. Blundell, D. Hassabis, P. Dayan, Computations Underlying Social Hierarchy Learning: Distinct Neural Mechanisms for Updating and Representing Self-Relevant Information. *Neuron* **92**, 1135–1147 (2016). doi: 10.1016/j.neuron.2016.10.052; pmid: 27930904
86. L. Deuker, J. L. S. Bellmund, T. Navarro Schröder, C. F. Doeller, An event map of memory space in the hippocampus. *eLife* **5**, e16534 (2016). doi: 10.7554/eLife.16534; pmid: 27710766
87. A. D. Ekstrom, C. Ranganath, Space, time, and episodic memory: The hippocampus is all over the cognitive map. *Hippocampus* 10.1002/hipo.22750 (2017). doi: 10.1002/hipo.22750
88. A. M. Collins, M. R. Quillian, Retrieval time from semantic memory. *J. Verbal Learn. Verbal Behav.* **8**, 240–247 (1969). doi: 10.1016/S0022-5371(69)80069-1
89. J. B. Tenenbaum, C. Kemp, T. L. Griffiths, N. D. Goodman, How to grow a mind: Statistics, structure, and abstraction. *Science* **331**, 1279–1285 (2011). doi: 10.1126/science.1192788; pmid: 21393536
90. I. K. Brunec *et al*., Multiple Scales of Representation along the Hippocampal Anteroposterior Axis in Humans. *Curr. Biol.* **28**, 2129–2135.e6 (2018). doi.org/10.1016/j.cub.2018.05.016
91. V. H. Brun *et al*., Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. *Hippocampus* **18**, 1200–1212 (2008). doi: 10.1002/hipo.20504; pmid: 19021257

92. B. A. Strange, M. P. Witter, E. S. Lein, E. I. Moser, Functional organization of the hippocampal longitudinal axis. *Nat. Rev. Neurosci.* **15**, 655–669 (2014). doi: 10.1038/nrn3785; pmid: 25234264
93. S. H. P. Collin, B. Milivojevic, C. F. Doeller, Memory hierarchies map onto the hippocampal long axis in humans. *Nat. Neurosci.* **18**, 1562–1564 (2015). doi: 10.1038/nn.4138; pmid: 26479587
94. B. Milivojevic, C. F. Doeller, Mnemonic networks in the hippocampal formation: From spatial maps to temporal and conceptual codes. *J. Exp. Psychol. Gen.* **142**, 1231–1241 (2013). doi: 10.1037/a0033746
95. C. Baldassano *et al*., Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron* **95**, 709–721.e5 (2017). doi: 10.1016/j.neuron.2017.06.041; pmid: 28772125
96. R. W. Komorowski *et al*., Ventral hippocampal neurons are shaped by experience to represent behaviorally relevant contexts. *J. Neurosci.* **33**, 8079–8087 (2013). doi: 10.1523/JNEUROSCI.5458-12.2013; pmid: 23637197
97. S. McKenzie *et al*., Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron* **83**, 202–215 (2014). doi: 10.1016/j.neuron.2014.05.019; pmid: 24910078
98. A. R. Preston, Y. Shrager, N. M. Dudukovic, J. D. E. Gabrieli, Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus* **14**, 148–152 (2004). doi: 10.1002/hipo.20009; pmid: 15098720
99. B. A. Kuhl, A. T. Shah, S. DuBrow, A. D. Wagner, Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nat. Neurosci.* **13**, 501–506 (2010). doi: 10.1038/nn.2498; pmid: 20190745
100. M. L. Schlichting, A. R. Preston, Memory integration: Neural mechanisms and implications for behavior. *Curr. Opin. Behav. Sci.* **1**, 1–8 (2015). doi: 10.1016/j.cobeha.2014.07.005; pmid: 25750931
101. M. Fyhn, T. Hafting, A. Treves, M.-B. Moser, E. I. Moser, Hippocampal remapping and grid realignment in entorhinal cortex. *Nature* **446**, 190–194 (2007). doi: 10.1038/nature05601; pmid: 17322902
102. K. Yoon *et al*., Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nat. Neurosci.* **16**, 1077–1084 (2013). doi: 10.1038/nn.3450; pmid: 23852111
103. B. Milivojevic, A. Vicente-Grabovetsky, C. F. Doeller, Insight reconfigures hippocampal-prefrontal memories. *Curr. Biol.* **25**, 821–830 (2015). doi: 10.1016/j.cub.2015.01.033; pmid: 25728693
104. A. C. Schapiro, L. V. Kustner, N. B. Turk-Browne, Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr. Biol.* **22**, 1622–1627 (2012). doi: 10.1016/j.cub.2012.06.056; pmid: 22885059
105. T. J. Wills, C. Lever, F. Cacucci, N. Burgess, J. O'Keefe, Attractor dynamics in the hippocampal representation of the local environment. *Science* **308**, 873–876 (2005). doi: 10.1126/science.1108905; pmid: 15879220
106. J. K. Leutgeb *et al*., Progressive transformation of hippocampal neuronal representations in "morphed" environments. *Neuron* **48**, 345–358 (2005). doi: 10.1016/j.neuron.2005.09.007; pmid: 16242413
107. H. M. Bonnici *et al*., Decoding representations of scenes in the medial temporal lobes. *Hippocampus* **22**, 1143–1153 (2012). doi: 10.1002/hipo.20960; pmid: 21656874
108. B. Steemers *et al*., Hippocampal Attractor Dynamics Predict Memory-Based Decision Making. *Curr. Biol.* **26**, 1750–1757 (2016). doi: 10.1016/j.cub.2016.04.063; pmid: 27345167
109. K. Diba, G. Buzsáki, Forward and reverse hippocampal place-cell sequences during ripples. *Nat. Neurosci.* **10**, 1241–1242 (2007). doi: 10.1038/nn1961; pmid: 17828259
110. T. J. Davidson, F. Kloosterman, M. A. Wilson, Hippocampal replay of extended experience. *Neuron* **63**, 497–507 (2009). doi: 10.1016/j.neuron.2009.07.027; pmid: 19709631
111. A. K. Lee, M. A. Wilson, Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* **36**, 1183–1194 (2002). doi: 10.1016/S0896-6273(02)01096-6; pmid: 12495631
112. D. J. Foster, M. A. Wilson, Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440**, 680–683 (2006). doi: 10.1038/nature04587; pmid: 16474382
113. A. S. Gupta, M. A. A. van der Meer, D. S. Touretzky, A. D. Redish, Hippocampal replay is not a simple function of experience. *Neuron* **65**, 695–705 (2010). doi: 10.1016/j.neuron.2010.01.034; pmid: 20223204
114. D. Dupret, J. O'Neill, B. Pleydell-Bouverie, J. Csicsvari, The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nat. Neurosci.* **13**, 995–1002 (2010). doi: 10.1038/nn.2599; pmid: 20639874
115. A. C. Singer, M. F. Carr, M. P. Karlsson, L. M. Frank, Hippocampal SWR activity predicts correct decisions during the initial learning of an alternation task. *Neuron* **77**, 1163–1173 (2013). doi: 10.1016/j.neuron.2013.01.027; pmid: 23522050
116. S. P. Jadhav, C. Kemere, P. W. German, L. M. Frank, Awake hippocampal sharp-wave ripples support spatial memory. *Science* **336**, 1454–1458 (2012). doi: 10.1126/science.1217230; pmid: 22555434
117. B. E. Pfeiffer, D. J. Foster, Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* **497**, 74–79 (2013). doi: 10.1038/nature12112; pmid: 23594744
118. C.-T. Wu, D. Haggerty, C. Kemere, D. Ji, Hippocampal awake replay in fear memory retrieval. *Nat. Neurosci.* **20**, 571–580 (2017). doi: 10.1038/nn.4507; pmid: 28218916
119. T. I. Brown *et al*., Prospective representation of navigational goals in the human hippocampus. *Science* **352**, 1323–1326 (2016). doi: 10.1126/science.aaf0784; pmid: 27284194
120. J. O'Neill, C. N. Boccara, F. Stella, P. Schoenenberger, J. Csicsvari, Superficial layers of the medial entorhinal cortex replay independently of the hippocampus. *Science* **355**, 184–188 (2017). doi: 10.1126/science.aag2787; pmid: 28082591
121. N. Wilming, P. König, S. König, E. A. Buffalo, Entorhinal cortex receptive fields are modulated by spatial attention, even without movement. *eLife* **7**, e31745 (2018). doi: 10.7554/eLife.31745; pmid: 29537964
122. D. Kumaran, E. A. Maguire, An unexpected sequence of events: Mismatch detection in the human hippocampus. *PLOS Biol.* **4**, e424 (2006). doi: 10.1371/journal.pbio.0040424; pmid: 17132050
123. L.-T. Hsieh, M. J. Gruber, L. J. Jenkins, C. Ranganath, Hippocampal activity patterns carry information about objects in temporal context. *Neuron* **81**, 1165–1178 (2014). doi: 10.1016/j.neuron.2014.01.015; pmid: 24607234
124. M. M. Garvert, R. J. Dolan, T. E. Behrens, A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. *eLife* **6**, e17086 (2017). doi: 10.7554/eLife.17086; pmid: 28448253
125. Z. Kurth-Nelson, M. Economides, R. J. Dolan, P. Dayan, Fast Sequences of Non-spatial State Representations in Humans. *Neuron* **91**, 194–204 (2016). doi: 10.1016/j.neuron.2016.05.028; pmid: 27321922
126. N. W. Schuck, Y. Niv, Sequential replay of non-spatial task states in the human hippocampus. bioRxiv [Preprint]. 7 May 2018. pmid: 315978
127. A. M. Bornstein, K. A. Norman, Reinstated episodic context guides sampling-based decisions for reward. *Nat. Neurosci.* **20**, 997–1003 (2017). doi: 10.1038/nn.4573; pmid: 28581478
128. B. B. Doll, K. D. Duncan, D. A. Simon, D. Shohamy, N. D. Daw, Model-based choices involve prospective neural activity. *Nat. Neurosci.* **18**, 767–772 (2015). doi: 10.1038/nn.3981; pmid: 25799041
129. H. F. Ólafsdóttir, C. Barry, A. B. Saleem, D. Hassabis, H. J. Spiers, Hippocampal place cells construct reward related sequences through unexplored space. *eLife* **4**, e06063 (2015). doi: 10.7554/eLife.06063; pmid: 26112828
130. D. Hassabis, D. Kumaran, S. D. Vann, E. A. Maguire, Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1726–1731 (2007). doi: 10.1073/pnas.0610561104; pmid: 17229836
131. J. Peters, C. Büchel, Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. *Neuron* **66**, 138–148 (2010). doi: 10.1016/j.neuron.2010.03.026; pmid: 20399735
132. H. C. Barron, R. J. Dolan, T. E. J. Behrens, Online evaluation of novel choices by simultaneous representation of multiple memories. *Nat. Neurosci.* **16**, 1492–1498 (2013). doi: 10.1038/nn.3515; pmid: 24013592
133. D. Kumaran, D. Hassabis, J. L. McClelland, What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends Cogn. Sci.* **20**, 512–534 (2016). doi: 10.1016/j.tics.2016.05.004; pmid: 27315762
134. V. Mnih *et al*., Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015). doi: 10.1038/nature14236; pmid: 25719670
135. G. Dragoi, S. Tonegawa, Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature* **469**, 397–401 (2011). doi: 10.1038/nature09633; pmid: 21179088
136. A. D. Grosmark, G. Buzsáki, Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. *Science* **351**, 1440–1443 (2016). doi: 10.1126/science.aad1935; pmid: 27013730
137. M. A. A. van der Meer, A. Johnson, N. C. Schmitzer-Torbert, A. D. Redish, Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task. *Neuron* **67**, 25–32 (2010). doi: 10.1016/j.neuron.2010.06.023; pmid: 20624589
138. N. W. Schuck, M. B. Cai, R. C. Wilson, Y. Niv, Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron* **91**, 1402–1412 (2016). doi: 10.1016/j.neuron.2016.08.019; pmid: 27657452
139. A. M. Wikenheiser, G. Schoenbaum, Over the river, through the woods: Cognitive maps in the hippocampus and orbitofrontal cortex. *Nat. Rev. Neurosci.* **17**, 513–523 (2016). doi: 10.1038/nrn.2016.56; pmid: 27256552
140. R. Kaplan, N. W. Schuck, C. F. Doeller, The Role of Mental Maps in Decision-Making. *Trends Neurosci.* **40**, 256–259 (2017). doi: 10.1016/j.tins.2017.03.002; pmid: 28365032
141. E. I. Moser *et al*., Grid cells and cortical representation. *Nat. Rev. Neurosci.* **15**, 466–481 (2014). doi: 10.1038/nrn3766; pmid: 24917300
142. F. Donato, R. I. Jacobsen, M.-B. Moser, E. I. Moser, Stellate cells drive maturation of the entorhinal-hippocampal circuit. *Science* **355**, eaai8178 (2017). doi: 10.1126/science.aad1935; pmid: 27013730
143. L. Hermer, E. S. Spelke, A geometric process for spatial reorientation in young children. *Nature* **370**, 57–59 (1994). doi: 10.1038/370057a0; pmid: 8015605
144. H. Fu *et al*., Tau Pathology Induces Excitatory Neuron Loss, Grid Cell Dysfunction, and Spatial Memory Deficits Reminiscent of Early Alzheimer's Disease. *Neuron* **93**, 533–541.e5 (2017). doi: 10.1016/j.neuron.2016.12.023; pmid: 28111080
145. D. A. Dombeck, A. N. Khabbaz, F. Collman, T. L. Adelman, D. W. Tank, Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* **56**, 43–57 (2007). doi: 10.1016/j.neuron.2007.08.003; pmid: 17920014
146. M. Kim, K. J. Jeffery, E. A. Maguire, Multivoxel Pattern Analysis Reveals 3D Place Information in the Human Hippocampus. *J. Neurosci.* **37**, 4270–4279 (2017). doi: 10.1523/JNEUROSCI.2703-16.2017; pmid: 28320847
147. R. Hayman, M. A. Verriotis, A. Jovalekic, A. A. Fenton, K. J. Jeffery, Anisotropic encoding of three-dimensional space by place cells and grid cells. *Nat. Neurosci.* **14**, 1182–1188 (2011). doi: 10.1038/nn.2892; pmid: 21822271
148. M. M. Yartsev, N. Ulanovsky, Representation of three-dimensional space in the hippocampus of flying bats. *Science* **340**, 367–372 (2013). doi: 10.1126/science.1235338; pmid: 23599496
149. A. Finkelstein *et al*., Three-dimensional head-direction coding in the bat brain. *Nature* **517**, 159–164 (2015). doi: 10.1038/nature14031; pmid: 25470055
150. S. M. Kosslyn, T. M. Ball, B. J. Reiser, Visual images preserve metric spatial information: Evidence from studies of image scanning. *J. Exp. Psychol. Hum. Percept. Perform.* **4**, 47–60 (1978). doi: 10.1037/0096-1523.4.1.47; pmid: 627850
151. Y. Dabaghian, V. L. Brandt, L. M. Frank, Reconceiving the hippocampal map as a topological template. *eLife* **3**, e03476 (2014). doi: 10.7554/eLife.03476; pmid: 25141375
152. Z. W. Pylyshyn, What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychol. Bull.* **80**, 1–24 (1973). doi: 10.1037/h0034650
153. W. H. Warren, D. B. Rothman, B. H. Schnapp, J. D. Ericson, Wormholes in virtual space: From cognitive maps to cognitive graphs. *Cognition* **166**, 152–163 (2017). doi: 10.1016/j.cognition.2017.05.020; pmid: 28577445
154. A. Ben-Yakov, N. Eshel, Y. Dudai, Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. *J. Exp. Psychol. Gen.* **142**, 1255–1263 (2013). doi: 10.1037/a0033558; pmid: 23815458
155. Y. Ezzyat, L. Davachi, Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron* **81**, 1179–1189 (2014). doi: 10.1016/j.neuron.2014.01.042; pmid: 24607235
156. I. K. Brunec, M. Moscovitch, M. D. Barense, Boundaries Shape Cognitive Representations of Spaces and Events. *Trends Cogn. Sci.* **22**, 637–650 (2018). doi: 10.1016/j.tics.2018.03.013; pmid: 29706557

157. Ø. A. Høydal, E. R. Skytøen, M.-B. Moser, E. I. Moser, Object-vector coding in the medial entorhinal cortex. bioRxiv [Preprint]. 22 March 2018. pmid: 286286
158. L. Chang, D. Y. Tsao, The Code for Facial Identity in the Primate Brain. *Cell* **169**, 1013–1028.e14 (2017). doi: 10.1016/j.cell.2017.05.011; pmid: 28575666
159. N. K. Logothetis, What we can do and what we cannot do with fMRI. *Nature* **453**, 869–878 (2008). doi: 10.1038/nature06976; pmid: 18548064
160. J. G. Heys, K. V. Rangarajan, D. A. Dombeck, The functional micro-organization of grid cells revealed by cellular-resolution imaging. *Neuron* **84**, 1079–1090 (2014). doi: 10.1016/j.neuron.2014.10.048; pmid: 25467986
161. T. Staudigl *et al*., Hexadirectional Modulation of High-Frequency Electrophysiological Activity in the Human Anterior Medial Temporal Lobe Maps Visual Space. *Curr. Biol.* **28**, 3325–3329.e4 (2018). doi: 10.1016/j.cub.2018.09.035; pmid: 30318353
162. S. Maidenbaum, J. Miller, J. M. Stein, J. Jacobs, Grid-like hexadirectional modulation of human entorhinal theta oscillations. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 10798–10803 (2018). doi: 10.1073/pnas.1805007115; pmid: 30282738
163. D. Chen *et al*., Hexadirectional Modulation of Theta Power in Human Entorhinal Cortex during Spatial Navigation. *Curr. Biol.* **28**, 3310–3315.e4 (2018). doi: 10.1016/j.cub.2018.08.029; pmid: 30318350
164. E. I. Moser, E. Kropff, M.-B. Moser, Place Cells, Grid Cells, and the Brain's Spatial Representation System. *Annu. Rev. Neurosci.* **31**, 69–89 (2008). doi: 10.1016/j.neuron.2014.10.048; pmid: 25467986
165. B. E. Pfeiffer, D. J. Foster, Autoassociative dynamics in the generation of sequences of hippocampal place cells. *Science* **349**, 180–183 (2015). doi: 10.1016/j.neuron.2014.10.048; pmid: 25467986

ACKNOWLEDGMENTS

Funding: C.F.D.'s research is supported by the Max Planck Society; the European Research Council (ERC-CoG GEOCOG 724836); the Kavli Foundation, the Centre of Excellence scheme of the Research Council of Norway–Centre for Neural Computation, the Egil and Pauline Braathen and Fred Kavli Centre for Cortical Microcircuits, the National Infrastructure scheme of the Research Council of Norway–NORBRAIN; and the Netherlands Organisation for Scientific Research (NWO-Vidi 452-12-009; NWO-Gravitation 024-001-006; NWO-MaGW 406-14-114; NWO-MaGW 406-15-291). **Author contributions:** J.L.S.B. and C.F.D. conceived the presented ideas; J.L.S.B. wrote the manuscript with input from C.F.D., P.G., and E.I.M. All authors contributed to the final manuscript. **Competing interests:** The authors declare no conflict of interest.

10.1126/science.aat6766

REVIEW SUMMARY

EVOLUTION

Contingency and determinism in evolution: Replaying life's tape

Zachary D. Blount*, **Richard E. Lenski**, **Jonathan B. Losos**

BACKGROUND: Evolution is a strongly historical process, and evolutionary biology is a field that combines history and science. How the historical nature of evolution affects the predictability of evolutionary outcomes has long been a major question in the field. The power of natural selection to find the limited set of high-fitness solutions to the challenges imposed by environments could, in principle, make those outcomes deterministic. However, the outcomes also may depend on idiosyncratic events that an evolving lineage experiences—such as the order of appearance of random mutations or rare environmental perturbations—making evolutionary outcomes unrepeatable. This sensitivity of outcomes to the details of history is called "historical contingency," which Stephen Jay Gould argued was an essential feature of evolution. Gould illustrated this view by proposing the thought experiment of replaying life's tape to see if the living world that we know would re-evolve. But, Gould wrote, "The bad news is that we can't possibly perform the experiment."

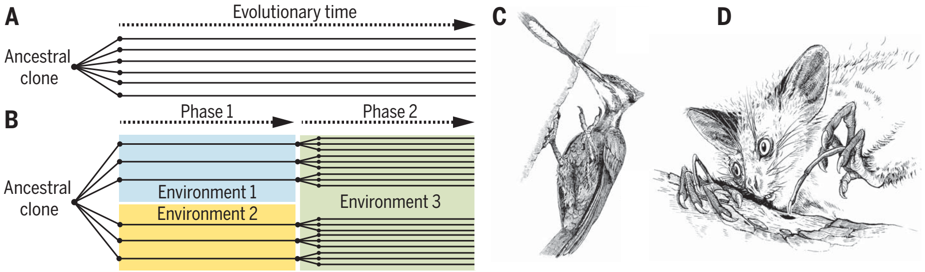
Gould's pessimistic assessment notwithstanding, experimental evolutionary biologists have now performed many replay experiments, albeit on a small scale, while comparative biologists are analyzing evolutionary outcomes in nature as though they were natural replay experiments. These studies provide new examples and insights into the interplay of historical contingency and natural selection that sits at the heart of evolution.

ADVANCES: Biologists have devised a variety of approaches to study the effects of history on the repeatability of evolutionary outcomes. On the experimental side, several designs have been employed, mostly using microbes, including "parallel replay experiments," in which initially identical populations are followed as they evolve in identical environments, and "historical difference experiments," in which previously diverged populations evolve under identical conditions (see the figure). Our review of many such experiments indicates that responses across replicate populations are often repeatable to some degree, although divergence increases as analyses move from overall fitness to underlying phenotypes and genetic changes. It is common for replicates with similar fitness under the conditions in which they evolved to vary more in their performance in other environments. Idiosyncratic outcomes also occur. For example, aerobic growth on citrate has evolved only once among 12 populations in an experiment with *Escherichia coli*, even after more than 65,000 generations. In that case, additional replays showed that the trait's evolution was dependent on the prior occurrence of particular mutations.

Meanwhile, comparative biologists have cataloged many notable examples of convergent evolution among species living in similar environments, illustrating the power of natural selection to produce similar phenotypic outcomes despite different evolutionary histories. Nonetheless, convergence is not inevitable—in many cases, lineages adapt phenotypically in different ways to the same environmental conditions. For example, the aye-aye (a lemur) and woodpeckers have evolved different morphological adaptations to similar ecological niches (see the figure). An emerging theme from comparative studies, tentatively supported by replay experiments, is that repeatability is common when the founding populations are closely related, perhaps resulting from shared genetics and developmental pathways, whereas different outcomes become more likely as historical divergences become greater.

ON OUR WEBSITE
Read the full article at http://dx.doi.org/10.1126/science.aam5979

OUTLOOK: Gould would be pleased that his thought experiment of replaying life's tape has been transformed into an empirical research program that explores the roles of historical contingency and natural selection at multiple levels. However, his view of historical influences as the central feature of evolution remains debatable. Laboratory replay experiments show that repeatable outcomes are common, at least when defined broadly (e.g., at the level of genes, not mutations). Moreover, convergence in nature is more common than many biologists would have wagered not long ago. On the other hand, as evolving lineages accumulate more differences, both experimental and comparative approaches suggest that the power of selection to drive convergence is reduced, and the contingent effects of history are amplified. Recognizing the joint contributions of contingency and natural selection raises interesting questions for further study, such as how the extent of prior genetic divergence affects the propensity for later convergence. Theory and experiments indicate that the "adaptive landscape"—that is, how specific phenotypes, and ultimately fitness, map onto the high dimensionality of genotypic space—plays a key role in these outcomes. Thus, a better understanding of these mappings will be important for a deeper appreciation of how fate and chance intertwine in the evolutionary pageant. ■



Replaying the tape of life. The tape of life is replayed on a small scale in evolution experiments of different designs. (**A**) In a parallel replay experiment, initially identical replicate populations evolve under the same conditions to see whether evolution is parallel or divergent. (**B**) A historical difference experiment explores the influence of earlier history in phase 1 on later evolution during phase 2. In nature, diverged lineages exposed to similar environmental conditions are similar to a historical difference experiment, in that the potential for convergence on the same adaptive response may depend on their earlier evolutionary histories. In the case of (**C**) the woodpecker and (**D**) the aye-aye, they have adapted to the same ecological niche (locating grubs, excavating through dead wood, and extracting them), but they evolved different anatomical traits to do so, reflecting the legacy of their evolutionary histories (e.g., primates lack beaks, birds lack fingers).

ILLUSTRATION: MARLIN PETERSON (PANELS C AND D)

The list of author affiliations is available in the full article online.
*Corresponding author. Email: blountza@msu.edu
Cite this article as Z. D. Blount *et al.*, *Science* **362**, eaam5979 (2018). DOI: 10.1126/science.aam5979

Blount *et al.*, *Science* **362**, 655 (2018) 9 November 2018 **1 of 1**

REVIEW

EVOLUTION

Contingency and determinism in evolution: Replaying life's tape

Zachary D. Blount[1,2]*, Richard E. Lenski[1], Jonathan B. Losos[3]

Historical processes display some degree of "contingency," meaning their outcomes are sensitive to seemingly inconsequential events that can fundamentally change the future. Contingency is what makes historical outcomes unpredictable. Unlike many other natural phenomena, evolution is a historical process. Evolutionary change is often driven by the deterministic force of natural selection, but natural selection works upon variation that arises unpredictably through time by random mutation, and even beneficial mutations can be lost by chance through genetic drift. Moreover, evolution has taken place within a planetary environment with a particular history of its own. This tension between determinism and contingency makes evolutionary biology a kind of hybrid between science and history. While philosophers of science examine the nuances of contingency, biologists have performed many empirical studies of evolutionary repeatability and contingency. Here, we review the experimental and comparative evidence from these studies. Replicate populations in evolutionary "replay" experiments often show parallel changes, especially in overall performance, although idiosyncratic outcomes show that the particulars of a lineage's history can affect which of several evolutionary paths is taken. Comparative biologists have found many notable examples of convergent adaptation to similar conditions, but quantification of how frequently such convergence occurs is difficult. On balance, the evidence indicates that evolution tends to be surprisingly repeatable among closely related lineages, but disparate outcomes become more likely as the footprint of history grows deeper. Ongoing research on the structure of adaptive landscapes is providing additional insight into the interplay of fate and chance in the evolutionary process.

The world in which we live—with all its splendor, tragedy, and strangeness—is the product of a vast, tangled web of events that form what we call history. Had history taken another route, the world of today would be different. Indeed, the historical record is filled with accidents and coincidences that shaped the course of events, critical twists of fate in which wrong turns and stalled cars helped start wars, dropped cigars changed military outcomes, and mutations contributed to toppling empires (*1*–*3*). These instances illustrate a property of history called "contingency," which makes outcomes sensitive to the details of the interacting events that led up to them. Contingency is why even though some trends may be predictable over the long-term and the past may be explicable, the future is unknowable.

Unlike many natural phenomena, evolution is a historical process, and evolutionary biology is a field in which science and history necessarily come together. Just as historians debate the extent to which certain historical events were inevitable, so too similar debates have raged in evolutionary biology. One person was especially influential in forcing biologists to grapple with the role of history in evolution: Stephen Jay Gould. In many of his writings, and most forcefully in his 1989 book *Wonderful Life* (*4*), Gould argued that historical contingency is central to evolution. He asserted that the living world is the product of a particular history, and had that history gone differently, the world of today would be utterly unlike the one we know.

In *Wonderful Life*, Gould illustrated his view with the now-famous gedankenexperiment of replaying life's tape and seeing whether the outcome would be at all like the original. Gould's conclusion was "Replay the tape a million times… and I doubt that anything like *Homo sapiens* would ever evolve again." But, Gould lamented, "The bad news is that we can't possibly perform the experiment." In recent years, however, evolutionary biologists have shown that Gould's experiment can, indeed, be conducted, at least on smaller scales. Indeed, a thriving subfield of experimental evolution has performed many replay experiments in both the lab and the field. Moreover, many paleontologists and comparative biologists contend that evolution in nature has conducted natural experiments that can be interpreted as replay experiments. These empirical studies are providing new insights into the interplay of contingency and determinism at the heart of evolution.

"Replaying life's tape" and the meaning of "contingency"

Any attempt to review the body of empirical research on contingency's role in evolution must first grapple with two sources of confusion that Gould himself introduced. The first comes from inconsistencies in how Gould described the replay metaphor. As pointed out by the philosopher John Beatty (*5*), in *Wonderful Life*, Gould first describes his gedankenexperiment as a strict replaying of the tape of life from identical earlier conditions (*6*), but later on Gould asks how slight variations at the outset would have altered the outcome (*7*). One can quibble about which idea Gould really favored, but a number of quotes from *Wonderful Life* suggest he was thinking more about the latter scenario (*8*). In any case, different researchers have designed tests of the replay hypothesis based on Gould's alternative versions, which both complicates and enriches the synthesis of their findings.

Gould also introduced confusion about the concept of contingency itself. Despite its centrality to his thinking, Gould never formally defined "contingency." He gave various informal descriptions, but these tended to be unfulfilling and circular. Moreover, he often conflated the two common meanings of the word "contingency": "dependence on something else" and "an accidental or chance event." Other writers have attempted to define contingency based on their interpretations of Gould's works, and different researchers have, again, designed work based on different notions of contingency (*9*–*13*). The definitions largely boil down to two alternatives that correspond to the different versions of the replay metaphor (*5*): unpredictability in outcomes from identical starting conditions, and causal dependence on the history leading to an outcome.

Philosophers of science have worked to clarify and formalize the concept of contingency. Beatty (*14*, *15*) points out that contingency ultimately means that an outcome depends on a history that did not necessarily have to happen. Desjardins (*16*–*18*) has further identified this property as intrinsic to path-dependent systems in which there are multiple possible paths from an initial state, multiple possible outcomes, and "probabilistic causal dependence" that links the two. These characteristics make path-dependent systems sensitive to differences over their entire history, including initial conditions, as well as later events that may cause paths to diverge even when starting from identical conditions (*16*, *17*). Thus, Gould's two alternative notions of contingency are just facets of the same thing. These characteristics also mean that a system's historical sensitivity will vary. In extreme cases, certain events along a historical path might completely preclude a given outcome, or render another outcome inevitable.

Desjardins' identification of contingency as a property of path-dependent systems is important because evolution inevitably has characteristics of path dependency. In particular, the stochastic processes of mutation and genetic drift virtually guarantee that different histories

---

[1]Department of Microbiology and Molecular Genetics and BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI 48824, USA. [2]Department of Biology, Kenyon College, Gambier, OH 43022, USA. [3]Department of Biology, Washington University, St. Louis, MO 63130, USA.
*Corresponding author. Email: blountza@msu.edu

will occur even when populations start from the same state and evolve under identical conditions (Box 1) (*19*). Such differences, in turn, constitute the sort of unpredictable antecedent events that might preclude populations from evolving the same solutions when confronting the same selective circumstances or, at least, change the relative likelihoods of different outcomes (*5*). These effects arise from how mutations and the order in which they occur affect later evolution. Indeed, the particular mutations that occur, their effects, and their fates can alter the rates of occurrence, phenotypic and developmental effects, and fates of later mutations, thereby shifting the probabilities of alternative evolutionary paths (*20*). These differences may be further amplified or dampened by environmental perturbations that may themselves be stochastic. In short, past genetic changes that originate stochastically through mutations can become the contingencies that shape subsequent evolution. Therefore, just like human history, evolution permits different historical paths, the instantiation of which is governed by probabilistic causal dependence. The central question that remains is whether, and under what conditions, those different paths lead to meaningfully different outcomes. Evolution involves the strongly deterministic force of natural selection, which has no clear analog in human history. Is evolution still meaningfully contingent, despite this deterministic element?

One reason that evolution might be meaningfully contingent, even with the deterministic force of natural selection, is the extraordinarily complex relationship of genotype to fitness. This relation is often described using the metaphor of an "adaptive landscape" (*21*). The metaphor is often drawn as a vista or topographical map, in which genotypes are arranged according to their mutational distance, while the elevation represents each genotype's fitness in a given environment. As a population evolves, new genotypes arise and their relative abundances shift, and the population thereby moves through the landscape. Absent any changes in conditions, natural selection tends to push the population uphill to higher average fitness, whereas the stochastic processes of mutation and drift tend to increase dispersion. If a landscape is smooth, with a single peak, then selection will eventually drive a population to that peak. If the landscape is rugged, with multiple peaks, then not all possible paths will lead to the highest peak, and evolutionary outcomes will be more sensitive to the population's initial state. Moreover, environmental changes may alter the shape of the adaptive landscape, potentially moving peaks or even turning hills into valleys and vice versa. Of course, this analogy of the adaptive landscape to a physical landscape is flawed, in part because the extreme high-dimensionality of potentially relevant genotypic states makes it impossible to identify and represent the possible paths that an evolving population might take. Moreover, the adaptive landscape metaphor as usually put forth implicitly ignores the role of developmental processes in translating genotypes into phenotypes. Nonetheless, while imperfect, the adaptive landscape metaphor remains widely used and is helpful when discussing the role of history in evolution.

Approaches to "replaying the tape" in evolutionary biology

Gould's writings have inspired many studies of evolutionary contingency using a variety of approaches. Some comparative and paleontological analyses have used "macroevolutionary" data to examine contingency and convergence in key innovations and other phenotypic features (*22*–*26*). Others have reconstructed ancestral genes to examine contingency in the historical transitions in protein function (*27*–*29*). However, the main approach has been to perform Gould's replay experiment, albeit on a smaller scale. In some studies, this approach has been used to evolve replicate populations of digital organisms—programs that replicate, mutate, compete, and evolve—in which all parameters can be controlled and histories reconstructed perfectly (*30*–*32*). More often, however, replay studies have employed three other approaches: (i) experiments in the laboratory with fast-evolving organisms; (ii) experiments in nature; and (iii) comparative studies of lineages that have experienced similar environments.

A note on the issue of development

The field of evolutionary developmental biology, or "evo-devo," has shown that development is a key aspect of the evolution of multicellular life, affecting the relationship between genotype,

Box 1. Contingency, determinism, and related words in an evolutionary context.

The vocabulary of evolution includes many words used both in ordinary language and to convey specific scientific ideas. Some of them also have different technical definitions in different scholarly contexts. Here we clarify what we mean by some of these words. To do so, we will build up from the basic processes that govern evolution to the conceptual issues that are the focus of this review.

At its core, evolution occurs by four fundamental processes: mutation, recombination, natural selection, and genetic drift. The first two produce genetic variation, whereas the last two govern the fate of variants. (Gene flow, interspecific hybridization, and horizontal gene transfer are special forms of recombination. The first describes the movement of genes across a spatial landscape; the second and third involve genes moving between species and microbial lineages, respectively.) Three of the processes—all except natural selection—are stochastic, in the sense that the specific variants produced or lost in a given generation are (or appear to be) a matter of chance. Chance is a tricky concept, however. There may well be some underlying cause for a chance event, such as a UVB (ultraviolet B) photon hitting DNA to produce a particular mutation or an asteroid striking Earth at a particular moment, but whether any specific event happens is unknowable or, at the least, impossible to incorporate into a mathematically efficient and useful theory of evolution. By contrast, natural selection is a deterministic process that reflects systematic differences in the propensity of alternative genotypes to survive and reproduce, depending on their fit to the environment. Thus, the "determinism" in our paper's title makes reference to the systematic effects of natural selection that promote repeatable outcomes in evolution. Of course, natural selection can act only on variation that exists within the realm of physical and biological constraints, which might thus be viewed as also contributing to that determinism.

Determinism implies inevitability in some philosophical contexts, but it does not in an evolutionary context because of the interplay between natural selection and the various stochastic processes. For example, a deleterious mutation might reach fixation in a small population by genetic drift, and a beneficial mutation may go extinct by drift, even in a large population, because the number of individuals initially carrying the mutation is small. Thus, our paper attempts to review studies that provide evidence about the repeatability of evolution, rather than to resolve conflicting philosophical positions.

To be sure, evolutionary theory involves higher-level processes, such as speciation and extinction, but they emerge from these four fundamental processes playing out in time and space. This situation is comparable to that in physics, in which a few fundamental forces—gravity, electromagnetism, and the weak and strong nuclear forces (the second and third of these are now unified as the electroweak force)—together gave rise to chemical elements and galaxies.

The words "parallel" and "convergent" are widely used to describe repeatable evolutionary outcomes. If two lineages are ancestrally similar or identical, and if they evolve similar adaptations, then that is often called parallel evolution (although several other definitions of parallel evolution are sometimes used as well). By contrast, if they diverged substantially in the past, but subsequently evolve similar structures or functions, then that is called convergence. However, the distinction is often unclear, especially for organisms in nature and even sometimes in long-running experiments. For this reason, we follow Arendt and Reznick (*134*) in referring to all cases of independently derived similarity as convergent evolution.

phenotype, and fitness (33–35). Indeed, the evolution of developmental systems can introduce the various constraints and biases that preclude or predispose subsequent evolutionary outcomes, making development an important factor in evolutionary contingency (36, 37). In this review, we couch our discussion in terms of genetic changes and gloss over the details of how development affects the contingency of evolution. However, this approach is not intended to discount the role of development. Rather, development is generally encoded by genes (including developmental responses to environmental perturbations), so although our presentation emphasizes genetic changes, we recognize that genes produce phenotypes in multicellular organisms via the developmental process. Moreover, our review places substantial emphasis on experiments with unicellular microbes, for which development is less relevant. Although we discuss studies with multicellular plants and animals with complex developmental programs, we aim to present a view that integrates them with the microbial work, and thus have focused on genetics. For these reasons, we do not dwell on the manner in which the evolution of developmental systems can produce the historical contingencies that are the subject of this essay. Such a topic provides excellent material for dissecting the role of evolutionary contingency, but is beyond the scope of this review.

Laboratory evolution experiments

In these experiments, replicate populations of a given species (or sometimes a community of two or more species) are propagated under controlled conditions, and their evolution monitored (38). History can play out repeatedly in these experiments, with initial and ongoing conditions that are either kept as identical as feasible or subtly changed, depending on the experiment, providing a valuable tool with which researchers can probe and even quantify the effects of contingency. Candidate events upon which particular outcomes are putatively contingent can then be identified, and their effects tested in further experiments. Although these experiments take place in laboratories, their results illuminate the potential role of contingency in the natural world.

The experiments have been performed with a variety of organisms. Microbes have been particularly useful because they are easy to handle and manipulate, they have fast generation times and large populations, and their (typically) asexual reproduction allows researchers to found replicate populations from the same clonal genotype. Moreover, some microbes can be frozen and later revived, allowing the preservation of living "frozen fossil records" of evolving populations (39). These fossil records provide direct access to population histories, making them particularly useful in contingency studies (40).

Alternative experimental designs

Three basic designs have been used to examine contingency and repeatability in laboratory evolution experiments (40) (Fig. 1). The simplest and most common is the "parallel replay experiment" in which initially identical replicate populations evolve under identical conditions, thus effectively playing the same tape several times simultaneously (Fig. 1A). In parallel replay experiments with frozen fossil records, the contingency of a particular outcome can later be tested with "analytic replay experiments," which are often called simply "replay" or "re-evolution" experiments (Fig. 1B). These experiments highlight the probabilistic nature of evolution and contingency. In an analytic replay experiment, archived samples are used to restart a population from multiple time points in its history. The resurrected populations are then allowed to evolve, and the patterns of recurrence of the outcome of interest examined (41, 42). Researchers use this design to probe for critical historical points at which the probability of a particular eventual outcome shifted to become more or less likely to occur than beforehand. These points can then be examined to identify the critical mutations or other events upon which the outcome's occurrence or nonoccurrence was contingent. Analytic replay experiments come closest to representing Gould's thought experiment, as they involve rerunning evolution from a previous point in history and seeing whether (and when and how often) the outcome is the same as the original.

Finally, "historical difference experiments" use a two-phase design to examine the effect of divergent evolutionary histories on subsequent evolution (40). In the simplest design, initially identical populations evolve in a single condition, just as in a parallel replay experiment. During this phase, each replicate acquires a unique history. In the second phase, the replicates are moved to a new environment where they evolve for another period (43) (Fig. 1C). Typically, the purpose of the second phase is to see whether the replicates adapt in the same way despite the differences accumulated during the first phase. There are several variations on this design; in all cases, the object of the first phase is for replicate populations to accumulate different histories, whereas the effect of those different histories on subsequent evolution is assessed in the second phase. In one variation, the populations evolve under multiple conditions in the first phase, before being shifted to a single condition in the second phase (Fig. 1D). In another variation, populations are founded from natural isolates and then evolved in a common laboratory environment; in this case, their prior evolution in the wild constitutes the first phase (Fig. 1E).

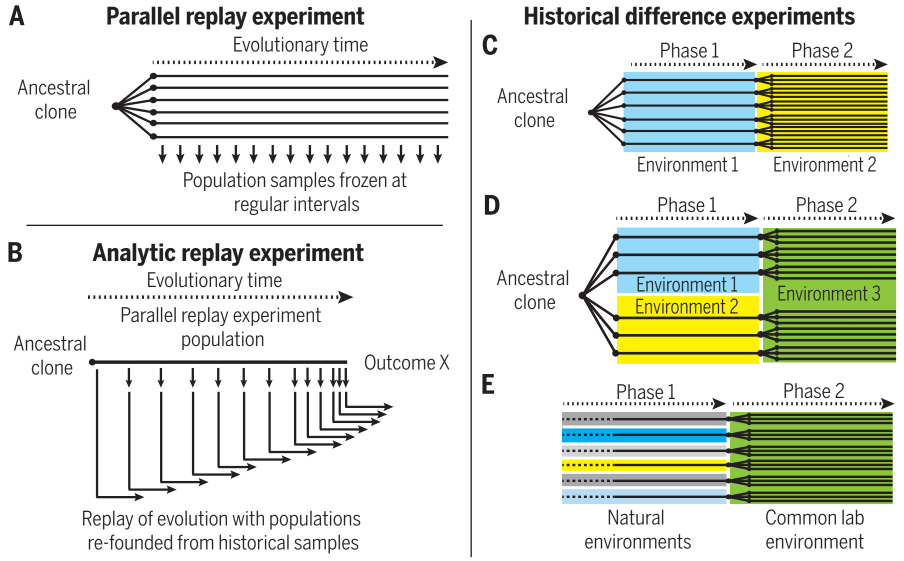


Fig. 1. Designs of microbial evolution experiments to explore historical contingency in parallel replay experiments. (**A**) Initially identical replicate populations are evolved under the same conditions to see whether evolution is parallel or divergent. Analytic replay experiments (**B**) are used to assess the contingency of a given outcome observed in a parallel replay experiment by replaying the population's evolution from various points in its history to see whether the likelihood of that outcome changes over time. Historical difference experiments explore the influence of differences caused by earlier history in phase 1 on later evolution during phase 2. In the simplest historical difference experiment design (**C**), initially identical populations evolve under one condition for a period of time. They are then shifted to a second condition, in which they evolve for another period, typically to see whether they evolve convergently despite differences accumulated in the first period. In one variant historical difference experiment design (**D**), the first phase of evolution is carried out under multiple conditions before the populations are shifted to a single, common condition. In another (**E**), wild isolates are used to found populations that evolve under a common, laboratory environmental condition. In this case, prior evolution in the wild constitutes phase 1.

Survey of findings

In recent years, the number of laboratory evolution experiments relevant to historical contingency has increased greatly. Both the parallel replay and historical difference experimental designs have often been used to address various questions other than contingency. Indeed, the parallel replay experiment is effectively the default design for replicated evolution experiments. Consequently, many studies can be evaluated for what they say about evolutionary contingency, even when they were not explicitly designed for that purpose. A formal meta-analysis of the full body of experiments would be difficult because of their heterogeneity, and it is beyond the scope of this review. Instead, we surveyed 51 studies chosen for their variety of designs and organisms. These studies include 35 that used a parallel replay experiment design, 5 that involved some type of analytic replay experiment, and 14 with variations of the historical difference experiment design (these sum to more than 51 because some studies used multiple designs). Altogether, they involved 17 different species, including bacteria, viruses, and unicellular and multicellular eukaryotes (table S1). For each study, we noted the experimental design, organisms used, specific questions asked, and the sources of any historical differences either among the founding populations or that arose during the experiments. We then evaluated whether and how history affected the measured outcomes. Collectively, the studies present a complex, and sometimes contradictory, picture that suggests a more nuanced role for contingency in evolution than Gould envisioned.

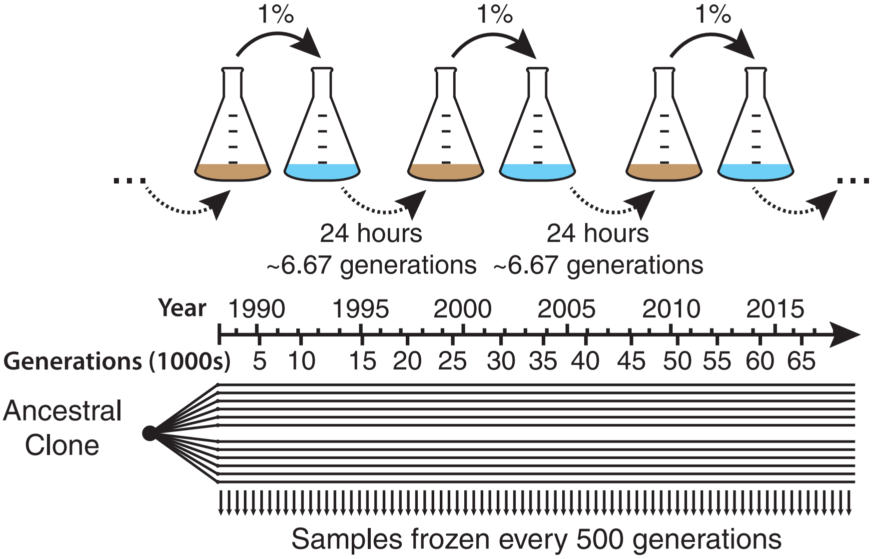


Fig. 2. The Long-Term Evolution Experiment with *E. coli* (LTEE). The LTEE is a paradigmatic parallel replay experiment that has studied 12 initially identical populations of *E. coli* for more than 65,000 generations of laboratory evolution under conditions of serial batch culture with daily 100-fold dilution into fresh medium. Samples of each population are frozen every 500 generations to provide a fossil record of viable bacteria.

The Long-Term Evolution Experiment with *Escherichia coli* (LTEE) is the most extensively studied example of a parallel replay experiment. The LTEE has followed 12 populations for over 65,000 generations since they were founded from a single clone in 1988 (*44*) (Fig. 2). The populations have been serially propagated in a glucose-limited medium that is considerably different from their natural environment, providing substantial opportunities for adaptation. Evolution in the LTEE occurs by de novo mutations, drift, and natural selection, making it a good model for investigating the contributions of these core processes to contingency. The populations have evolved in parallel (i.e., repeatedly) in several ways (*45*, *46*). All have evolved much higher fitness, faster growth, and larger cells than the ancestor. Also, beneficial mutations have accumulated in many of the same genes across some or all of the populations, although the mutations are rarely the same at the nucleotide level. The populations have also diverged in various ways (*45*, *46*). Each has accumulated a unique suite of mutations. Half evolved much higher mutation rates, causing the number of mutations accrued in each population to vary greatly. Most populations have evolved very similar fitness levels under the conditions of the experiment, but even so there are persistent differences in fitness between them, suggesting that they are ascending different peaks on the adaptive landscape. Moreover, the evolved populations vary considerably in their fitness under other conditions, including on different resources (*47*). Finally, many of the populations have evolved simple ecosystems in which two or more lineages stably coexist (*48*–*50*), although it remains to be seen whether coexistence typically involves the same ecological and genetic mechanisms. Overall, the LTEE populations seem to be following subtly different evolutionary paths, albeit in the same general direction, with one major exception that we will address later.

Broadly speaking, other parallel replay experiments, although much shorter in duration, show a similar pattern of generally consistent evolutionary responses across replicate populations under a variety of conditions. In some instances, these responses have been markedly parallel (*51*–*57*). However, heterogeneity in evolutionary responses across replicates is not uncommon (*58*–*60*). Such divergence is often more evident as analyses move from fitness per se to underlying phenotypic and genotypic responses (*61*). For instance, phenotypic parallelism often involves more variable genotypic changes, although instances of phenotypic variability with genotypic parallelism, at least at the level of genes mutated, have also been reported (*54*, *62*–*64*). Similarly, as in the LTEE, it is not unusual for replicates with similar fitness under the conditions in which they evolved to have genetic differences that cause significant variation in fitness and phenotype under other conditions (*65*). This cross-condition variability makes it difficult to compare levels of divergence among experiments. This difficulty is exacerbated by logistical differences in obtaining genetic and phenotypic information. Modern genome sequencing and bioinformatics make the detection and comparison of evolved genetic changes easy and cost-effective. By contrast, measuring phenotypes is difficult, costly, and time consuming, so most studies have examined relatively few phenotypic changes under a restricted set of conditions.

Divergence among replicates, when it occurs, is not always subtle (*66*–*71*). Collins and Bell (*66*), for example, observed two starkly different adaptive responses among five replicate algal populations that evolved under an elevated CO$_2$ level. Another notable example comes from the LTEE. After more than 31,000 generations, one population evolved the capacity to grow aerobically on citrate (Cit$^+$), which was included in the culture medium as a chelating agent. Although many bacteria are Cit$^+$, *E. coli* has been historically defined as a species in part by its inability to grow aerobically on citrate (Cit$^-$). Occasional environmental isolates of *E. coli* have been found to be Cit$^+$, but as the result of the acquisition of foreign plasmids, not chromosomal mutations. The Cit$^+$ mutant that arose in the LTEE was only the second case ever reported (*72*), despite decades of study of this organism in hundreds of laboratories. A recent study found additional spontaneous Cit$^+$ mutants, but their isolation required prolonged, intense, and focused selection (*73*). When this new ability arose in the LTEE, it changed the population's ecological circumstances and evolutionary direction in several important ways—allowing the cell numbers to increase several-fold, causing metabolic by-products to accumulate, changing the bacteria's stoichiometric evolution, and perhaps even setting the Cit$^+$ lineage on a path toward incipient speciation (*41*, *74*–*76*).

The ability to grow on citrate is highly beneficial in the LTEE environment, yet the Cit$^+$ trait has evolved in only 1 of 12 populations, even after more than 65,000 generations. There are two plausible explanations for this seeming paradox. The trait might have been caused by a single extremely rare mutation that could have occurred at any time in any of the populations. Alternatively, the ability to grow on citrate might have required multiple mutations. If so, selection for the Cit$^+$ trait per se would not have facilitated spread of the earlier mutations that, nonetheless, were required for the evolution of the Cit$^+$ trait under the experimental conditions. According to that second hypothesis, the evolution of the Cit$^+$ trait was therefore contingent on a particular history during which one or more required mutations happened to accumulate, "potentiating" the trait's appearance. To test these ideas, Blount *et al.* (*41*) devised the analytic replay experiment design, recognizing that a contingent outcome should be more likely after the potentiating event (or events). In several experiments, they restarted the population with clones isolated at 16 time points in its frozen fossil record, replayed evolution thousands of times, and examined the outcomes. The Cit$^+$ trait re-evolved only in populations founded by clones from 20,000 generations onward, implying that some potentiating mutation had arisen by then.

Subsequent work has revealed the complex evolutionary history that led to the Cit$^+$ trait. Leon *et al.* (*77*) found that the trait was slightly beneficial in the ancestral genetic background. However, early evolution in the population was dominated by high-fitness, glucose-adapted mutations against which any rare Cit$^+$ mutants could not effectively compete. This adaptation led to a genetic background in which the Cit$^+$ trait had become detrimental. Further mutations, some of which seem to have been involved in adaptation to growth on acetate (a by-product of glucose metabolism), accumulated between 10,000 and 29,000 generations. The Cit$^+$ trait was slightly beneficial again on this new background (*78*). At this point, high-fitness mutations were no longer sweeping through the population, and the weakly beneficial Cit$^+$ cells were able to persist long enough to accumulate refining mutations that made the trait highly beneficial (*74*, *75*, *79*).

The analytic replay experiment design has since been used to test the contingency of other outcomes seen in parallel replay experiments. Using four closely related clones isolated very early from another LTEE population, Woods *et al.* (*80*) performed an analytic replay experiment to investigate why one lineage had eventually prevailed over another, even though the clones representing the eventual winner had demonstrably lower fitness than clones from the lineage that later went extinct. Replays showed that the eventual winners prevailed because they were more evolvable; that is, they were more likely to generate beneficial mutations of large effect. Genome sequencing and genetic manipulations showed that this difference reflected a strong epistatic interaction between mutations at two specific loci. Meyer *et al.* (*81*) performed a multispecies analytic replay experiment, which showed that the evolution of a phage λ variant able to infect *E. coli* via an alternative receptor was contingent on mutations in the coevolving host population. This work highlights how evolutionary contingency can play a key role in community dynamics that are more typically addressed in purely ecological terms.

The analytic replay design is relatively new, and few have been performed to date. However, they show that particular outcomes can hinge on small historical differences between populations that can then lead to substantial divergence even under identical conditions. They also indicate that genetic and ecological interactions can play critical roles in generating the events that drive such divergence. Altogether, analytic replay experiments provide compelling examples of how evolutionary outcomes can hinge on the particulars of history.

Parallel replay experiments show that differences can arise among initially identical populations evolving under identical conditions, and analytic replay experiments show that those differences can alter evolutionary potentials in important ways, even in the absence of environmental change. By contrast, historical difference experiments examine how different histories can affect subsequent evolution when the environment is changed. Forerunners to this design included experiments in which bacteria were challenged to grow in different environments to see whether the sequence of challenges affected the propensity to acquire an altered metabolic or resistance phenotype (*82*, *83*). In the first historical difference experiment to explicitly quantify the effect of history, Travisano *et al.* (*43*) isolated clones from each LTEE population after 2000 generations of adaptation to the glucose-limited medium. They then founded three replicate populations from each clone, which evolved for 1000 generations in the same medium except with glucose replaced by maltose. Owing to their different histories, the clones varied greatly in their initial fitness in the maltose environment. However, they rapidly converged in their fitness on maltose during evolution in that new environment (Fig. 3). Several later historical difference experiments have also shown that adaptation to new conditions can drive convergence at the level of fitness, despite initial differences, although the mark of history often lingers at the genetic level (*84*–*89*).

Some historical difference experiments, however, have shown stronger historical effects that preclude complete convergence, although those effects can vary with the environment used for the second phase of adaptation (*71*). Burch and Chao (*89*) found that two closely related phage ϕ6 genotypes had distinctly different capacities for further adaptation after prior evolution left them in different regions of the adaptive landscape, and Flores-Moya *et al.* (*90*) found that history strongly affected the evolution of two dinoflagellate strains. Moore and Woods (*91*) also found that *E. coli* strains isolated from different hosts varied significantly in the rate at which they adapted to a glucose-limited medium; this variation did not simply reflect differences in their initial adaptation to the laboratory environment, but instead indicated more idiosyncratic effects of prior history. Similarly, a study of 230 *Saccharomyces cerevisiae* strains (produced by crossing two highly diverged parental strains) showed a strong tendency for later, convergent adaptation to erase prior history, although the degree of erasure was subtly affected by specific genotypes (*92*). Taken together, historical difference experiments indicate that the capacity of selection to overcome historical differences has limits. Specifically, the historical difference experiments suggest that adaptation's ability to drive convergence declines as populations have spent more time diverging from one another, and when that divergence occurred in more distinct environments.

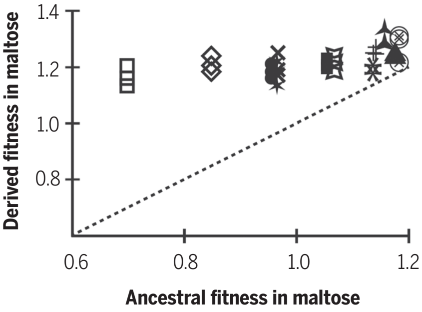


Fig. 3. Rapid convergence in a historical difference experiment. Single clones of *E. coli* were isolated from each of the 12 LTEE populations after 2000 generations of evolution in glucose-limited medium. Three replicate populations of each were founded and then evolved for 1000 generations in a maltose medium. Despite substantial initial variation due to their independent histories of adaptation to glucose, the replicate populations rapidly converged in their fitness on maltose. [Redrawn from Travisano *et al.* (*43*)]

Synopsis of laboratory studies

These replay experiments present a rich and complex picture of the repeatability and contingency of evolutionary outcomes. The direction of evolutionary change typically seems to be broadly consistent in a given condition, regardless of history, and phenotypic and genetic parallelisms are often striking (*45*). Even so, there remains scope for history to drive substantially divergent outcomes. These divergences are often subtle, such as differences in genotype that nonetheless lead to parallel evolution in phenotypes, including especially fitness itself. But subtlety of immediate effects does not necessarily negate the importance of long-term effects, as differences can build on one another. The evolution of the Cit$^+$ trait in the LTEE is a case in point, illustrating how seemingly minor changes can shift the potential for further evolution in ways that lead to marked divergence (*41*). Moreover, subtle divergences that matter little in the environment where they emerge can have major effects when conditions change, as a consequence of mutations that have not been tested under the new conditions (*45*, *88*). On the other hand, historical difference experiments show that selection in the new environment can sometimes overcome those previously evolved differences. The deeper the imprint of history, however, the less likely it becomes that evolution can reverse the prior divergence.

One interpretation of the results of the laboratory replay experiments is that the potential for contingency to matter is determined, in part, by the structure of the adaptive landscape encountered by the replicate evolving populations. As might be expected, a rugged landscape that presents multiple adaptive peaks makes distinct outcomes possible, and starting conditions, as well as the form and strength of interactions between mutations (epistasis), will affect the probabilities of those outcomes. Alternatively, a smooth landscape will tend to yield more repeatability if the time scale examined allows replicates to find the peak (*67*). However, these inferences are potentially circular, because our knowledge of adaptive landscapes typically comes from such experimental outcomes. This issue highlights the need for further investigation into landscape parameters. One factor that may affect ruggedness is environmental complexity; an environment with spatial structure or multiple resources, for instance, may often provide more opportunities for divergent adaptive responses (*61*, *68*). Exogenous events and how organisms modify their environments complicate things further by changing the structure of the landscape in ways that can affect opportunities for subsequent divergence (*69*, *70*). However, a genotype may have multiple distinct paths to higher fitness even in a homogeneous, single-resource environment (*71*).

Altogether, laboratory experiments on contingency support a nuanced view. Evolution is more likely to be historically insensitive and repeatable if the adaptive landscape offers few alternative paths or many that lead to similar outcomes. If, however, the landscape is rugged, with multiple avenues available that lead to dissimilar adaptations, then outcomes are likely to be more variable and more sensitive to historical contingencies. Evolutionary repeatability varies because the degree to which outcomes are contingent varies.

Experimental evolution in nature

Although most replay experiments have been conducted in the laboratory, an ambitious new direction involves replicated evolution experiments in natural settings. The realization that natural selection can produce rapid evolutionary change (*93*–*96*) opened the door to evolution experiments in nature. To date, results are available from only a few such experiments, but many more are now under way (*97*). Some of these studies take advantage of long-running ecological experiments, including the Park Grass Experiment, which was started in 1856 (*98*, *99*).

These studies have focused on hypotheses about adaptation in the wild. However, they often also constitute de facto replay experiments, as replicate populations can be compared to examine variation in evolutionary responses. Several differences should be kept in mind when comparing these studies to laboratory experiments. In particular, the experiments in nature often involve vertebrate animals, rather than the microorganisms and invertebrates typically used in laboratory experiments; therefore, populations are smaller, generations are longer, and founding populations are genetically heterogeneous. These factors make it more likely that evolutionary responses in field experiments rely on standing genetic variation present at the outset, rather than on de novo variation generated during the experiment. They therefore increase the opportunity for parallel responses based on shared variation, on the one hand, and the scope for differences in initial conditions between replicate populations to produce contingent evolutionary responses, on the other hand. Furthermore, in some experiments, such as those on color and life histories in guppies, different populations were used to establish the experimental populations, making these studies more akin to historical difference experiments than to parallel replay experiments (*97*).

It is perhaps too early to generalize from the field evolution experiments reported to date. Nonetheless, the results so far—including guppies evolving slower life histories in the absence of predators (*100*) and lizards evolving shorter limbs when forced to use narrow substrates (*101*)—tend to indicate a high degree of repeatability in evolutionary responses (*97*).

Comparative studies: Evolutionary replays across space and time

The ideal experiment for characterizing repeatability and contingency in evolution would be to expose initially identical populations to the same conditions in nature and allow them to evolve not for a few years or tens of years, but for thousands and even millions of years. Even if funding were available for such studies, we would have to wait a long time to get the results. But fortuitously, nature has already conducted such experiments for us, albeit not quite as precisely as those performed in the laboratory.

Convergent evolution is broadly defined (Box 1) as the independent evolution of similar features in multiple species or clades (*102*). Convergent evolution can occur for many reasons. For instance, shared developmental programs may predispose species to evolve in the same way for reasons unrelated to natural selection (*103*, *104*). However, convergence occurring in distinct lineages living in similar environments has long been considered strong evidence of the operation of natural selection (*102*, *105*, *106*). For example, both the C_4 and CAM (crassulacean acid metabolism) photosynthetic pathways have evolved independently many times in plants, almost always in lineages that now occur in arid or semiarid regions; this evolutionary correlation suggests that the lower rates of water loss and other physiological features of these pathways are advantageous under these conditions (*107*, *108*). Similarly, strikingly convergent carnivorous pitcher plants have evolved in several unrelated genera as an adaptation to waterlogged soils with low nutrient availability and high light (*109*, *110*). Until fairly recently, such cases of convergence were considered relatively rare exceptions. In recent years, however, myriad examples of adaptive convergence have been reported (*23*, *111*, *112*). Particularly impressive are cases in which convergence involves not just two (or more) lineages adapting to the same niche, but entire multispecies assemblages evolving similarly, such as evolutionary radiations of Caribbean lizards and Pacific Ocean snails on multiple islands, and frog and bird faunas on different continents (*113*).

The extent of convergence has led some to argue that the repeated evolution of the same feature under similar circumstances means that evolution is predictable and that contingencies of history hold little sway in directing evolution. More specifically, they argue that the ubiquity of convergence indicates that optimal solutions exist to problems posed by the environment and that lineages have repeatedly, almost deterministically, found these solutions (*23*, *111*, *112*).

This argument assumes that the same selective conditions occur repeatedly, that there are a limited number of high-fitness phenotypic solutions ("adaptive peaks") to these challenges, and that populations inevitably evolve these phenotypes. According to Conway Morris (*23*), "the evolutionary routes are many, but the destinations are limited." McGhee (*114*) put it this way: "Convergent evolution is the result of the fact that there are limited numbers of ways to solve a functional problem within the constraints imposed by the laws of physics and geometry."

One prerequisite for adaptive convergence is that species respond to similar selective pressures by adopting the same ecological role [i.e., the same niche in the original Grinnellian sense (*115*)]. This need not be the case, however, because communities of species do not necessarily partition resources in similar ways. Moreover, even when species converge upon the same ecological role, they may evolve distinct nonconvergent phenotypic adaptations. For example, considering the aye-aye (a lemur) and the woodpecker to be convergent misses the point that they evolved very different phenotypic means to accomplish the same task of locating and extracting grubs from inside wood. They occupy the same niche but adapted in divergent, rather than convergent, ways.

Assuming that multiple lineages independently adopt the same ecological niche, how might contingency lead them to adapt in different ways to the same environmental challenge? We see three main possibilities. First, populations might evolve different solutions to the same challenge. For example, some plants may adapt to the presence of a herbivore by evolving physical defenses such as thorns, others by acquiring chemical defenses, and yet others by becoming cryptic. Second, populations may evolve the same function, but by means of different phenotypic changes. For example, the hammering beak and long bristly tongue of the woodpecker accomplish the same ends as the chiseling teeth and long, flexible finger of the aye-aye [more generally, the "many-to-one" phenomenon in biomechanics (116)]. Third, some populations may get stuck on a lower adaptive peak (local optimum) and be unable to evolve the best possible phenotype (global optimum). In all three cases, historical contingencies may predispose a lineage to adapt one way or another (birds lack teeth and hands, and primates lack beaks, explaining the different routes taken by the aye-aye and woodpecker). Their different histories thus may explain why two lineages fail to converge despite experiencing the same selective conditions for millions of years.

Fig. 4. Evolutionary one-offs. Evolutionary one-offs are species or clades that evolved unique adaptations to their ecological circumstances that have not been convergently evolved by other lineages. Clockwise from top left: African elephant, *Welwitschia*, Moyer's pygmy chameleon, red octopus. (Note that similarity among, for example, species of elephants or chameleons is not convergent; rather, their shared features are the result of inheritance from a common ancestral species that evolved their trademark features a single time.) [Photo credits: African elephant: Jonathan Losos. *Welwitschia*: Thomas Schoch, CC BY-SA 3.0 license; original at https://en.wikipedia.org/wiki/Welwitschia#/media/File:Welwitschia_mirabilis(2).jpg. Moyer's pygmy chameleon: Martin Neilsen, CC BY-SA 4.0 license; original at https://commons.wikimedia.org/wiki/Category:Rhampholeon_moyeri#/media/File:Stump-tailed_chameleons_Rhampholeon_moyeri.jpg. Red octopus: Jerry Kirkhart, CC BY-SA 2.0 license; original at https://commons.wikimedia.org/wiki/Category:Octopus_rubescens#/media/File:Red_Octopus_rescued.jpg]

In evaluating the extent to which convergence is evidence of evolutionary determinism, several points must be considered. Most generally, we need to ask what constitutes convergence. Birds, bats, and insects all fly, but their wings are constructed differently and their aerodynamics also differ. Are these convergent adaptations, or divergent adaptations accomplishing the same task? At some level, drawing a line becomes arbitrary. Another difficulty is that convergence is identified after the fact. The saber-toothed condition evolved at least three times in the Carnivora, as well as once each in creodonts and South American marsupials, presumably as an adaptation to a particular predatory strategy (117). But how many other taxa, faced with the same selective conditions, failed to evolve this adaptation? Knowing the denominator is key to determining how repeatable a convergent trend is (45), but rarely does one know how many other lineages experienced similar circumstances, yet failed to evolve the trait in question. Moreover, although recent compilations of convergence (23, 111, 112) are impressive, one could just as easily compile lists of adaptive types lacking a convergent doppelgänger: the two-leaved *Welwitschia mirabilis,* the platypus, chameleons, kiwis, elephants, octopuses, and hominins—all adaptive types that have evolved just once—to name a few (Fig. 4).

Finally, the occurrence of convergent evolution is not necessarily inconsistent with the evolutionary importance of contingency. Genetic changes can become the contingencies that shape subsequent evolution. To the extent that shared genetic and developmental systems predispose species to evolve in similar ways (103, 104, 106, 118), then adaptive convergence may often be shaped by the particular history that sculpted the genetics and development of their shared ancestors (119). In such cases, evolution may be deterministic within a clade but contingent at deeper phylogenetic levels when comparing species across clades (104, 112, 119). Moreover, the shared regulatory mechanisms and sometimes cryptic genetic similarities that underlie deep homologies indicate that contingent historical events can shape convergence even among distant relatives (36). The evolutionary reactivation of previously silenced, but still functional, developmental programs is another example of how distant relatives can exhibit evolutionarily derived phenotypic similarity as a result of contingent genetic events (120–122).

Some convergence proponents go so far as to say that if life has evolved on Earth-like exoplanets, it will look much like what we see here (23). But we need not look to the stars to test that hypothesis: All we need to do is go to New Zealand, an island lacking any native terrestrial mammals. In their absence, New Zealand's flora and fauna evolved to bear little resemblance to any other ecosystem in the world. In addition to kiwis, there are both carnivorous and flightless parrots, adzebills, moas, giant eagles, and flightless wrens, as well as a semiterrestrial bat ["the bat family's attempt to make a mouse" (123)], giant snails and orthopterans, and divaricating shrubs with leaves that grow in the interior of the bush. And going back in time, one would be hard-pressed to find many similarities between the Mesozoic world of the dinosaurs and today's faunas.

In short, lineages adapting to similar environmental conditions in nature can be thought of as evolutionary replays, even if these "natural experiments" are not as precise as carefully designed and controlled laboratory experiments. Because the lineages will have different genetic constitutions and will have experienced different histories, these cases are analogous to the historical difference experiments in laboratory studies. Unfortunately, however, the evidence boils down to one list of cases in which convergence occurred and another where it did not, rendering quantitative conclusions unsatisfactory. Nonetheless, the many impressive cases of convergence show that repeated outcomes can arise from similar environmental challenges. Conversely, the many cases in which convergence did not occur suggest that contingent effects can play a strong role in shaping divergent adaptive responses.

Against that murky conclusion, one trend stands out (despite some exceptions): Conspecific populations and closely related species seem to evolve in similar ways more often than distantly related taxa (124). Such a trend is expected in part because closely related species tend to interact with the environment in similar ways. Moreover, they share more of their history, and thus share more of the past changes in their genetic and development systems that can shape later evolution. Closely related lineages are thus predisposed to evolve in the same way. Indeed, some cases of parallel evolution have occurred by selection on shared variation that was present in a common ancestral population (125, 126). By contrast, convergence between distantly related lineages is less likely to result from selection on shared variation. A related finding is that when convergence occurs, the extent to which the response involves the same gene is greater when the taxa are closely related (*127*). This pattern accords with the tentative conclusion from laboratory studies that parallel replay experiments (with replicate populations founded by the same ancestor) tend to produce parallel outcomes more often than historical difference experiments (with populations founded by different ancestral strains or species).

Conclusions and future prospects

Gould's gedankenexperiment that "we can't possibly perform" has been transformed into a real experimental program, one in which increasingly sophisticated and audacious studies are exploring the roles of contingency and determinism at ever deeper levels. Although Gould's ideas on contingency have stimulated a great deal of productive work, his view that contingent effects were pervasive throughout evolution remains debatable. The laboratory replays performed to date typically show that repeatable outcomes are common, at least when the founding populations are similar, when repeatability is defined broadly (e.g., at the level of affected genes and pathways, as opposed to precise mutational changes), and over the time scales accessible to experiments. Moreover, evolutionary convergence across lineages that share similar natural environments has proven more common than most biologists would have wagered even two decades ago—its prevalence attests to the power of natural selection to repeatedly sculpt the same adaptive solutions. That it does so more often among closely related taxa, which share similar genetics and developmental programs, illustrates the yin and yang of contingency and determinism.

Where to now? Clearly, evolution can be both contingent and deterministic, and often in complicated and fascinating ways. Recognizing this mixed nature will allow future research to investigate how contingency and determinism interact. Many questions remain to be addressed; for example, what circumstances promote contingent and deterministic outcomes, how does the extent of prior genetic divergence affect the propensity for future parallelism versus contingency, what types of divergence—say, a few mutations of large effect versus the accumulation of minor variants over long periods—lead to which outcomes, and what circumstances allow convergence even in distantly related taxa? Theory and experiments show that the structure of the adaptive landscape plays a critical role in determining the potential for contingent outcomes. Therefore, a deeper understanding of adaptive landscapes will be important for understanding evolutionary contingency (*89*, *128*–*133*). In short, there's no shortage of work to do, and interesting outcomes to be discovered and quantified. Gould would be pleased that the field he inspired has such bright prospects, as the tape of life plays on.

REFERENCES AND NOTES

1. G. J. Meyer, *A World Undone: The Story of the Great War 1914–1918* (Bantam, 2007).
2. J. M. McPherson, *Crossroads of Freedom: Antietam* (Oxford Univ. Press, 2002).
3. E. I. Rogaev, A. P. Grigorenko, G. Faskhutdinova, E. L. Kittler, Y. K. Moliaka, Genotype analysis identifies the cause of the "royal disease". *Science* **326**, 817 (2009). doi: 10.1126/science.1180660; pmid: 19815722
4. S. J. Gould, *Wonderful Life* (Norton, 1989).
5. J. H. Beatty, Replaying life's tape. *J. Philos.* **103**, 336–362 (2006). doi: 10.5840/jphil2006103716
6. "I call this experiment 'replaying life's tape.' You press the rewind button and, making sure you thoroughly erase everything that actually happened, go back to any time and place in the past…Then let the tape run again and see if the repetition looks at all like the original" [p. 48 in (*4*)].
7. "[A]ny replay, altered by an apparently insignificant jot or tittle at the outset, would have yielded an…outcome of entirely different form…" [p. 248 in (*4*)].
8. "Alter any early event, ever so slightly and without apparent importance at the time, and evolution cascades into a radically different channel" [p. 51 in (*4*)]. "A historical explanation…[rests]…on an unpredictable sequence of antecedent states, where any major change in any step of the sequence would have altered the final result. This final result is therefore dependent, or contingent, upon everything that came before—the unerasable and determining signature of history" [p. 283 in (*4*)]. "Historical explanations take the form of narrative: E, the phenomenon to be explained, arose because D came before, preceded by C, B and A. If any of these earlier stages had not occurred, or had transpired in a different way, then E would not exist" [p. 283 in (*4*)].
9. Y. Ben–Menahem, Historical contingency. *Ratio* **10**, 99–107 (1997). doi: 10.1111/1467-9329.00032
10. Y. Ben–Menahem, "Historical necessity and contingency" in *A Companion to the Philosophy of History and Historiography*, A. Tucker, Ed. (Blackwell, 2009), chapter 9.
11. D. D. Turner, Gould's replay revisited. *Biol. Philos.* **26**, 65–79 (2011). doi: 10.1007/s10539-010-9228-0
12. R. Inkpen, D. D. Turner, The topography of historical contingency. *J. Philos. Hist.* **6**, 1–19 (2012). doi: 10.1163/187226312X625573
13. D. H. Erwin, in *Chance in Evolution*, G. Ramsey, C.H. Pence, Eds. (Univ. of Chicago Press, 2016), chapter 12.
14. J. H. Beatty, I. Carrera, When what had to happen was not bound to happen: History, chance, narrative, evolution. *J. Philos. Hist.* **5**, 471–495 (2011). doi: 10.1163/187226311X599916
15. J. Beatty, Narrative possibility and narrative explanation. *Stud. Hist. Philos. Sci. A* **62**, 31–41 (2017). doi: 10.1016/j.shpsa.2017.03.001; pmid: 28583357
16. E. C. Desjardins, *Historicity in Biology* (Univ. of British Columbia, 2009).
17. E. C. Desjardins, Historicity and experimental evolution. *Biol. Philos.* **26**, 339–364 (2011). doi: 10.1007/s10539-011-9256-4
18. E. C. Desjardins, in *Chance in Evolution*, C. Ramsey, C.H. Pence, Eds. (Univ. of Chicago Press, 2016), chapter 9.
19. J. B. S. Haldane, *The Causes of Evolution* (Longmans, Green, and Co., 1932).
20. G. S. Mani, B. C. Clarke, Mutational order: A major stochastic process in evolution. *Proc. R. Soc. Lond. B Biol. Sci.* **240**, 29–37 (1990). doi: 10.1098/rspb.1990.0025; pmid: 1972992
21. S. Wright, The roles of mutation, inbreeding, crossbreeding and selection in evolution, in *Proceedings of the Sixth International Congress of Genetics* (1932), pp. 356–366.
22. S. B. Emerson, A macroevolutionary study of historical contingency in the fanged frogs of Southeast Asia. *Biol. J. Linn. Soc. Lond.* **73**, 139–151 (2001). doi: 10.1111/j.1095-8312.2001.tb01352.x
23. S. Conway Morris, *Life's Solution* (Cambridge Univ. Press, 2003).
24. S. C. Morris, Evolution: Like any other science it is predictable. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 133–145 (2010). doi: 10.1098/rstb.2009.0154; pmid: 20008391
25. M. H. Dick, S. Lidgard, D. P. Gordon, S. F. Mawatari, The origin of ascophoran bryozoans was historically contingent but likely. *Proc. Biol. Sci.* **276**, 3141–3148 (2009). doi: 10.1098/rspb.2009.0704; pmid: 19515659
26. G. J. Vermeij, Historical contingency and the purported uniqueness of evolutionary innovations. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1804–1809 (2006). doi: 10.1073/pnas.0508724103; pmid: 16443685
27. E. A. Ortlund, J. T. Bridgham, M. R. Redinbo, J. W. Thornton, Crystal structure of an ancient protein: Evolution by conformational epistasis. *Science* **317**, 1544–1548 (2007). doi: 10.1126/science.1142819; pmid: 17702911
28. M. J. Harms, J. W. Thornton, Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* **512**, 203–207 (2014). doi: 10.1038/nature13410; pmid: 24930765
29. T. N. Starr, L. K. Picton, J. W. Thornton, Alternative evolutionary histories in the sequence space of an ancient protein. *Nature* **549**, 409–413 (2017). doi: 10.1038/nature23902; pmid: 28902834
30. T. Taylor, J. Hallam, Replaying the tape: An investigation into the role of contingency in evolution, in *Proceedings of Alife VI*, C. Adami, R. K. Below, H. Kitano, C. E. Taylor, Eds. (MIT Press, 1998), pp 256–265.
31. D. A. Wagenaar, C. Adami, Influence of chance, history, and adaptation on evolution of Digitalia, in *Proceedings of Artificial Life VII*, M. Bedau, J. McCaskill, N. Packard, S. Rasmussen, Eds. (MIT Press, 2000), pp. 216–220.
32. G. Yedid, C. A. Ofria, R. E. Lenski, Historical and contingent factors affect re-evolution of a complex feature lost during mass extinction in communities of digital organisms. *J. Evol. Biol.* **21**, 1335–1357 (2008). doi: 10.1111/j.1420-9101.2008.01564.x; pmid: 18564346
33. B. K. Hall, Evo-Devo: Evolutionary developmental mechanisms. *Int. J. Dev. Biol.* **47**, 491–495 (2003). pmid: 14756324
34. B. K. Hall, Evolutionary developmental biology (evo-devo): Past, present, and future. *Evol. Ed. Outreach* **5**, 184–193 (2012). doi: 10.1007/s12052-012-0418-x
35. M.-J. Favé, E. Abouheif, Integrating ecological genomics and eco-evo-devo reveals multiple adaptive peaks in ant populations of the Arizona Sky Islands. *bioRxiv* 045419 [Preprint]. 29 March 2016. https://doi.org/10.1101/045419.
36. N. Shubin, C. Tabin, S. Carroll, Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818–823 (2009). doi: 10.1038/nature07891; pmid: 19212399
37. P. M. Brakefield, Evo-devo and accounting for Darwin's endless forms. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2069–2075 (2011). doi: 10.1098/rstb.2011.0007; pmid: 21690125
38. T. J. Kawecki *et al*., Experimental evolution. *Trends Ecol. Evol.* **27**, 547–560 (2012). doi: 10.1016/j.tree.2012.06.001; pmid: 22819306
39. R. E. Lenski, M. Travisano, Dynamics of adaptation and diversification: A 10,000-generation experiment with bacterial populations. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6808–6814 (1994). doi: 10.1073/pnas.91.15.6808; pmid: 8041701
40. Z. D. Blount, in *Chance in Evolution*, G. Ramsey, C. H. Pence, Eds. (Univ. of Chicago Press, 2016), chapter 10.
41. Z. D. Blount, C. Z. Borland, R. E. Lenski, Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7899–7906 (2008). doi: 10.1073/pnas.0803151105; pmid: 18524956
42. C. B. Turner, Z. D. Blount, R. E. Lenski, Replaying evolution to test the cause of extinction of one ecotype in an experimentally evolved population. *PLOS ONE* **10**, e0142050 (2015). doi: 10.1371/journal.pone.0142050; pmid: 26581098
43. M. Travisano, J. A. Mongold, A. F. Bennett, R. E. Lenski, Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* **267**, 87–90 (1995). doi: 10.1126/science.7809610; pmid: 7809610
44. R. E. Lenski, M. R. Rose, S. C. Simpson, S. C. Tadler, Long–term experimental evolution in *Escherichia coli*. I. adaptation and divergence during 2,000 generations. *Am. Nat.* **138**, 1315–1341 (1991). doi: 10.1086/285289
45. R. E. Lenski, Convergence and divergence in a long–term experiment with bacteria. *Am. Nat.* **190** (S1), S57–S68 (2017). doi: 10.1086/691209; pmid: 28731830
46. R. E. Lenski, Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J.* **11**, 2181–2194 (2017). doi: 10.1038/ismej.2017.69; pmid: 28509909
47. N. Leiby, C. J. Marx, Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*. *PLOS Biol.* **12**, e1001789 (2014). doi: 10.1371/journal.pbio.1001789; pmid: 24558347
48. D. E. Rozen, R. E. Lenski, Long–term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. *Am. Nat.* **155**, 24–35 (2000). pmid: 10657174
49. J. Plucain *et al*., Epistasis and allele specificity in the emergence of a stable polymorphism in *Escherichia coli*. *Science* **343**, 1366–1369 (2014). doi: 10.1126/science.1248688; pmid: 24603152

50. B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, M. M. Desai, The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (2017). doi: 10.1038/nature24287; pmid: 29045390
51. J. J. Bull *et al*., Exceptional convergent evolution in a virus. *Genetics* **147**, 1497–1507 (1997). pmid: 9409816
52. L. Notley-McRobb, T. Ferenci, Adaptive *mgl*-regulatory mutations and genetic diversity evolving in glucose-limited *Escherichia coli* populations. *Environ. Microbiol.* **1**, 33–43 (1999). doi: 10.1046/j.1462-2920.1999.00002.x; pmid: 11207716
53. H. A. Wichman, M. R. Badgett, L. A. Scott, C. M. Boulianne, J. J. Bull, Different trajectories of parallel evolution during viral adaptation. *Science* **285**, 422–424 (1999). doi: 10.1126/science.285.5426.422; pmid: 10411508
54. S. S. Fong, A. R. Joyce, B. Ø. Palsson, Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res.* **15**, 1365–1372 (2005). doi: 10.1101/gr.3832305; pmid: 16204189
55. A. J. Betancourt, Genomewide patterns of substitution in adaptively evolving populations of the RNA bacteriophage MS2. *Genetics* **181**, 1535–1544 (2009). doi: 10.1534/genetics.107.085837; pmid: 19189959
56. D. van Ditmarsch *et al*., Convergent evolution of hyperswarming leads to impaired biofilm formation in pathogenic bacteria. *Cell Rep.* **4**, 697–708 (2013). doi: 10.1016/j.celrep.2013.07.026; pmid: 23954787
57. D. E. Deatherage, J. L. Kepner, A. F. Bennett, R. E. Lenski, J. E. Barrick, Specificity of genome evolution in experimental populations of *Escherichia coli* evolved at different temperatures. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1904–E1912 (2017). doi: 10.1073/pnas.1616132114; pmid: 28202733
58. H. J. E. Beaumont, J. Gallie, C. Kost, G. C. Ferguson, P. B. Rainey, Experimental evolution of bet hedging. *Nature* **462**, 90–93 (2009). doi: 10.1038/nature08504; pmid: 19890329
59. K. L. Hillesland, G. J. Velicer, R. E. Lenski, Experimental evolution of a microbial predator's ability to find prey. *Proc. Biol. Sci.* **276**, 459–467 (2009). doi: 10.1098/rspb.2008.1098; pmid: 18832061
60. W. C. Ratcliff *et al*., Experimental evolution of an alternating uni- and multicellular life cycle in *Chlamydomonas reinhardtii*. *Nat. Commun.* **4**, 2742 (2013). doi: 10.1038/ncomms3742; pmid: 24193369
61. S. F. Bailey, N. Rodrigue, R. Kassen, The effect of selection environment on the probability of parallel evolution. *Mol. Biol. Evol.* **32**, 1436–1448 (2015). doi: 10.1093/molbev/msv033; pmid: 25761765
62. G. J. Velicer, L. Kroos, R. E. Lenski, Loss of social behaviors by *myxococcus xanthus* during evolution in an unstructured habitat. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12376–12380 (1998). doi: 10.1073/pnas.95.21.12376; pmid: 9770494
63. M. Le Gac, T. F. Cooper, S. Cruveiller, C. Médigue, D. Schneider, Evolutionary history and genetic parallelism affect correlated responses to evolution. *Mol. Ecol.* **22**, 3292–3303 (2013). doi: 10.1111/mec.12312; pmid: 24624420
64. J. Plucain *et al*., Contrasting effects of historical contingency on phenotypic and genomic trajectories during a two-step evolution experiment with bacteria. *BMC Evol. Biol.* **16**, 86 (2016). doi: 10.1186/s12862-016-0662-8; pmid: 27108090
65. R. C. MacLean, G. Bell, Experimental adaptive radiation in *Pseudomonas*. *Am. Nat.* **160**, 569–581 (2002). doi: 10.1086/342816; pmid: 18707508
66. S. Collins, G. Bell, Phenotypic consequences of 1,000 generations of selection at elevated CO$_2$ in a green alga. *Nature* **431**, 566–569 (2004). doi: 10.1038/nature02945; pmid: 15457260
67. M. L. Friesen, G. Saxer, M. Travisano, M. Doebeli, Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in *Escherichia coli*. *Evolution* **58**, 245–260 (2004). doi: 10.1111/j.0014-3820.2004.tb01642.x; pmid: 15068343
68. R. Korona, C. H. Nakatsu, L. J. Forney, R. E. Lenski, Evidence for multiple adaptive peaks from populations of bacteria evolving in a structured habitat. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9037–9041 (1994). doi: 10.1073/pnas.91.19.9037; pmid: 8090765
69. P. B. Rainey, M. Travisano, Adaptive radiation in a heterogeneous environment. *Nature* **394**, 69–72 (1998). doi: 10.1038/27900; pmid: 9665128
70. H. A. Lindsey, J. Gallie, S. Taylor, B. Kerr, Evolutionary rescue from extinction is contingent on a lower rate of environmental change. *Nature* **494**, 463–467 (2013). doi: 10.1038/nature11879; pmid: 23395960
71. A. H. Melnyk, R. Kassen, Adaptive landscapes in evolving populations of *Pseudomonas fluorescens*. *Evolution* **65**, 3048–3059 (2011). doi: 10.1111/j.1558-5646.2011.01333.x; pmid: 22023573
72. B. G. Hall, Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. *J. Bacteriol.* **151**, 269–273 (1982). pmid: 7045076
73. D. J. Van Hofwegen, C. J. Hovde, S. A. Minnich, Rapid evolution of citrate utilization by *Escherichia coli* by direct selection requires *citT* and *dctA*. *J. Bacteriol.* **198**, 1022–1034 (2016). doi: 10.1128/JB.00831-15; pmid: 26833416
74. Z. D. Blount, J. E. Barrick, C. J. Davidson, R. E. Lenski, Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489**, 513–518 (2012). doi: 10.1038/nature11514; pmid: 22992527
75. C. B. Turner, Z. D. Blount, D. H. Mitchell, R. E. Lenski, Evolution and coexistence in response to a key innovation in a long–term evolution experiment with *Escherichia coli*. *bioRxiv* 020958 [Preprint]. 17 June 2015. https://doi.org/10.1101/020958.
76. C. B. Turner, B. D. Wade, J. R. Meyer, B. A. Sommerfeld, R. E. Lenski, Evolution of organismal stoichiometry in a long-term experiment with *Escherichia coli*. *R. Soc. Open Sci.* **4**, 170497 (2017). doi: 10.1098/rsos.170497; pmid: 28791173
77. D. Leon, S. D'Alton, E. M. Quandt, J. E. Barrick, Innovation in an *E. coli* evolution experiment is contingent on maintaining adaptive potential until competition subsides. *PLOS Genet.* **14**, e1007348 (2018). doi: 10.1371/journal.pgen.1007348; pmid: 29649242
78. E. M. Quandt *et al*., Fine-tuning citrate synthase flux potentiates and refines metabolic innovation in the Lenski evolution experiment. *eLife* **4**, e09696 (2015). doi: 10.7554/eLife.09696; pmid: 26465114
79. E. M. Quandt, D. E. Deatherage, A. D. Ellington, G. Georgiou, J. E. Barrick, Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2217–2222 (2014). doi: 10.1073/pnas.1314561111; pmid: 24379390
80. R. J. Woods *et al*., Second-order selection for evolvability in a large *Escherichia coli* population. *Science* **331**, 1433–1436 (2011). doi: 10.1126/science.1198914; pmid: 21415350
81. J. R. Meyer *et al*., Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* **335**, 428–432 (2012). doi: 10.1126/science.1214449; pmid: 22282803
82. B. G. Hall, Evolution of new metabolic functions in laboratory organisms, in *Evolution of Genes and Proteins*, M. Nei, R. K. Koehn, Eds. (Sinauer, 1983), pp. 234 – 257.
83. R. E. Lenski, Two-step resistance by *Escherichia coli* B to bacteriophage T2. *Genetics* **107**, 1–7 (1984). pmid: 6373494
84. H. Teotónio, I. M. Chelo, M. Bradić, M. R. Rose, A. D. Long, Experimental evolution reveals natural selection on standing genetic variation. *Nat. Genet.* **41**, 251–257 (2009). doi: 10.1038/ng.289; pmid: 19136954
85. S. Bedhomme, G. Lafforgue, S. F. Elena, Genotypic but not phenotypic historical contingency revealed by viral experimental evolution. *BMC Evol. Biol.* **13**, 46 (2013). doi: 10.1186/1471-2148-13-46; pmid: 23421472
86. S. Kryazhimskiy, D. P. Rice, E. R. Jerison, M. M. Desai, Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519–1522 (2014). doi: 10.1126/science.1250939; pmid: 24970088
87. A. Spor *et al*., Phenotypic and genotypic convergences are influenced by historical contingency and environment in yeast. *Evolution* **68**, 772–790 (2014). doi: 10.1111/evo.12302; pmid: 24164389
88. P. Simões *et al*., Predictable phenotypic, but not karyotypic, evolution of populations with contrasting initial history. *Sci. Rep.* **7**, 913 (2017). doi: 10.1038/s41598-017-00968-1; pmid: 28424494
89. C. L. Burch, L. Chao, Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature* **406**, 625–628 (2000). doi: 10.1038/35020564; pmid: 10949302
90. A. Flores-Moya *et al*., Effects of adaptation, chance, and history on the evolution of the toxic dinoflagellate *Alexandrium minutum* under selection of increased temperature and acidification. *Ecol. Evol.* **2**, 1251–1259 (2012). doi: 10.1002/ece3.198; pmid: 22833798
91. F. B.-G. Moore, R. Woods, Tempo and constraint of adaptive evolution in *Escherichia coli* (Enterobacteriaceae, Enterobacteriales). *Biol. J. Linn. Soc. Lond.* **88**, 403–411 (2006). doi: 10.1111/j.1095-8312.2006.00629.x
92. E. R. Jerison *et al*., Genetic variation in adaptability and pleiotropy in budding yeast. *eLife* **6**, e27167 (2017). doi: 10.7554/eLife.27167; pmid: 28826486
93. H. B. D. Kettlewell, *The Evolution of Melanism: The Study of a Recurring Necessity with Special Reference to Industrial Melanism in the Lepidoptera* (Oxford Univ. Press, 1973).
94. B. R. Grant, P. R. Grant, Darwin's finches: Population variation and sympatric speciation. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 2359–2363 (1979). doi: 10.1073/pnas.76.5.2359; pmid: 16592654
95. P. R. Grant, B. R. Grant, *40 Years of Evolution: Darwin's Finches on Daphne Major Island* (Princeton Univ. Press, 2014).
96. J. A. Endler, *Natural Selection in the Wild* (Princeton Univ. Press, 1986).
97. J. B. Losos, *Improbable Destinies* (Riverhead Books, 2017).
98. J. Silvertown *et al*., The Park Grass Experiment 1856–2006: Its contribution to ecology. *J. Ecol.* **94**, 801–814 (2006). doi: 10.1111/j.1365-2745.2006.01145.x
99. S. Y. Strauss, J. A. Lau, T. W. Schoener, P. Tiffin, Evolution in ecological field experiments: Implications for effect size. *Ecol. Lett.* **11**, 199–207 (2008). doi: 10.1111/j.1461-0248.2007.01128.x; pmid: 18031555
100. D. N. Reznick, H. Bryga, Life–history evolution in guppies (*Poecilia reticulate*) I. Phenotypic and genetic changes in an introduction experiment. *Evolution* **41**, 1370–1385 (1987). pmid: 28563598
101. J. J. Kolbe, J. C. Ehrenberger, H. A. Moniz, M. J. Angilletta Jr., Physiological variation among invasive populations of the brown anole (*Anolis sagrei*). *Physiol. Biochem. Zool.* **87**, 92–104 (2014). doi: 10.1086/672157; pmid: 24457924
102. J. B. Losos, Convergence, adaptation, and constraint. *Evolution* **65**, 1827–1840 (2011). doi: 10.1111/j.1558-5646.2011.01289.x; pmid: 21729041
103. D. B. Wake, Homoplasy – the result of natural selection, or evidence of design limitations. *Am. Nat.* **138**, 543–567 (1991). doi: 10.1086/285234
104. D. B. Wake, M. H. Wake, C. D. Specht, Homoplasy: From detecting pattern to determining process and mechanism of evolution. *Science* **331**, 1032–1035 (2011). doi: 10.1126/science.1188545; pmid: 21350170
105. P. H. Harvey, M. D. Pagel, *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, 1991).
106. A. A. Agrawal, Toward a predictive framework for convergent evolution: Integrating natural history, genetic mechanisms, and consequences for the diversity of life. *Am. Nat.* **190**, S1–S12 (2017). doi: 10.1086/692111; pmid: 28731831
107. R. F. Sage, P.-A. Christin, E. J. Edwards, The C$_4$ plant lineages of planet Earth. *J. Exp. Bot.* **62**, 3155–3169 (2011). doi: 10.1093/jxb/err048; pmid: 21414957
108. X. Yang *et al*., The *Kalanchoë* genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nat. Commun.* **8**, 1899 (2017). doi: 10.1038/s41467-017-01491-7; pmid: 29196618
109. T. J. Givnish, New evidence on the origin of carnivorous plants. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10–11 (2015). doi: 10.1073/pnas.1422278112; pmid: 25538295
110. C. J. Thorogood, U. Bauer, S. J. Hiscock, Convergent and divergent evolution in carnivorous pitcher plant traps. *New Phytol.* **217**, 1035–1041 (2018). doi: 10.1111/nph.14879; pmid: 29131340
111. S. Conway Morris, *The Runes of Evolution: How the Universe Became Self–Aware* (Templeton Press, 2015).
112. G. McGhee, *Convergent Evolution: Limited Forms Most Beautiful* (MIT Press, 2011).
113. J. B. Losos, Adaptive radiation, ecological opportunity, and evolutionary determinism. American Society of Naturalists E. O. Wilson award address. *Am. Nat.* **175**, 623–639 (2010). doi: 10.1086/652433; pmid: 20412015
114. G. R. McGhee Jr., Can evolution be directional without being teleological? *Stud. Hist. Philos. Biol. Biomed. Sci.* **58**, 93–99 (2016). pmid: 26754619
115. T. W. Schoener, in *Ecological Concepts: The Contribution of Ecology to an Understanding of the Natural World*, J. M. Cherrett, Ed. (Blackwell, 1989).

116. P. C. Wainwright, M. E. Alfaro, D. I. Bolnick, C. D. Hulsey, Many–to–one mapping of form to function: A general principle in organismal design? *Integr. Comp. Biol.* **45**, 256–262 (2005). doi: 10.1093/icb/45.2.256; pmid: 21676769
117. J. A. Meachen-Samuels, Morphological convergence of the prey–killing arsenal of sabertooth predators. *Paleobiology* **38**, 1–14 (2012). doi: 10.1017/S0094837300000373
118. K. Schwenk, G. P. Wagner, in *Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes*, M. Pigliucci, K. Preston, Eds. (Oxford Univ. Press, Oxford, UK 2004), pp. 390–408.
119. R. Powell, C. Mariscal, Convergent evolution as natural experiment: The tape of life reconsidered. *Interface Focus* **5**, 20150040 (2015). doi: 10.1098/rsfs.2015.0040; pmid: 26640647
120. N. Shubin, D. B. Wake, A. J. Crawford, Morphological variation in the limbs of *Taricha granulosa* (Caudata: Salamandridae): Evolutionary and phylogenetic implications. *Evolution* **49**, 874–884 (1995). doi: 10.1111/j.1558-5646.1995.tb02323.x; pmid: 28564880
121. M.-J. West-Eberhard, *Developmental Plasticity and Evolution* (Yale Univ. Press, 2008).
122. M. W. Kirschner, J. C. Gerhart, *The Plausibility of Life* (Yale Univ. Press, 2008).
123. J. Diamond, in *Ecological Restoration of New Zealand Islands*, D. R. Towns, C. H. Daugherty, I. A. E. Atkinson, Eds. (New Zealand Department of Conservation, 1990), pp. 3–8.
124. T. J. Ord, T. C. Summers, Repeated evolution and the impact of evolutionary history on adaptation. *BMC Evol. Biol.* **15**, 137 (2015). doi: 10.1186/s12862-015-0424-z; pmid: 26156849
125. P. F. Colosimo *et al*., Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**, 1928–1933 (2005). doi: 10.1126/science.1107239; pmid: 15790847
126. D. L. Stern, The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764 (2013). doi: 10.1038/nrg3483; pmid: 24105273
127. G. L. Conte, M. E. Arnegard, C. L. Peichel, D. Schluter, The probability of genetic parallelism and convergence in natural populations. *Proc. Biol. Sci.* **279**, 5039–5047 (2012). doi: 10.1098/rspb.2012.2146; pmid: 23075840
128. S. J. Gould, *The Structure of Evolutionary Theory* (Harvard Univ. Press, 2002).
129. D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006). doi: 10.1126/science.1123539; pmid: 16601193
130. A. E. Lobkovsky, Y. I. Wolf, E. V. Koonin, Predictability of evolutionary trajectories in fitness landscapes. *PLOS Comput. Biol.* **7**, e1002302 (2011). doi: 10.1371/journal.pcbi.1002302; pmid: 22194675
131. J. A. G. M. de Visser, J. Krug, Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* **15**, 480–490 (2014). doi: 10.1038/nrg3744; pmid: 24913663
132. S. Kryazhimskiy, D. P. Rice, M. M. Desai, Population subdivision and adaptation in asexual populations of *Saccharomyces cerevisiae*. *Evolution* **66**, 1931–1941 (2012). doi: 10.1111/j.1558-5646.2011.01569.x; pmid: 22671557
133. J. R. Nahum *et al*., A tortoise-hare pattern seen in adapting structured and unstructured populations suggests a rugged fitness landscape in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7530–7535 (2015). doi: 10.1073/pnas.1410631112; pmid: 25964348
134. J. Arendt, D. Reznick, Convergence and parallelism reconsidered: What have we learned about the genetics of adaptation? *Trends Ecol. Evol.* **23**, 26–32 (2008). doi: 10.1016/j.tree.2007.09.011; pmid: 18022278

ACKNOWLEDGMENTS

We thank E. Desjardins for helpful discussion and comments on an earlier draft of this paper. **Funding:** This work was supported in part by a grant from the National Science Foundation (DEB-1451740), the BEACON Center for the Study of Evolution in Action (DBI-0939454), Michigan State University, and Harvard University. **Competing interests:** We declare no competing interests.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/eaam5979/suppl/DC1
Supplementary Text
References (*135*–*150*)

10.1126/science.aam5979

RESEARCH ARTICLE SUMMARY

IMMUNOLOGY

Perivascular dendritic cells elicit anaphylaxis by relaying allergens to mast cells via microvesicles

Hae Woong Choi*, Jutamas Suwanpradid, Il Hwan Kim, Herman F. Staats, Muzlifah Haniffa, Amanda S. MacLeod, Soman N. Abraham

INTRODUCTION: An increasing number of individuals suffer from acute anaphylaxis, particularly food-associated reactions. The symptoms of anaphylaxis are triggered soon after allergens (such as peanut antigens, insect venom, or certain medications) enter the circulation of sensitized subjects who have elevated levels of antigen-specific immunoglobulin E (IgE) antibodies in their circulation. Mast cells (MCs) are the primary effectors of anaphylaxis and become activated when allergens make contact with IgE antibodies bound to the surfaces of MCs. Upon activation, MCs rapidly release prestored inflammatory mediators. When this occurs simultaneously throughout the body, anaphylaxis and shock result. As MCs are typically found in the perivascular abluminal surface of relatively impregnable blood vessels, how these cells perceive and interact with blood-borne allergens is unclear.

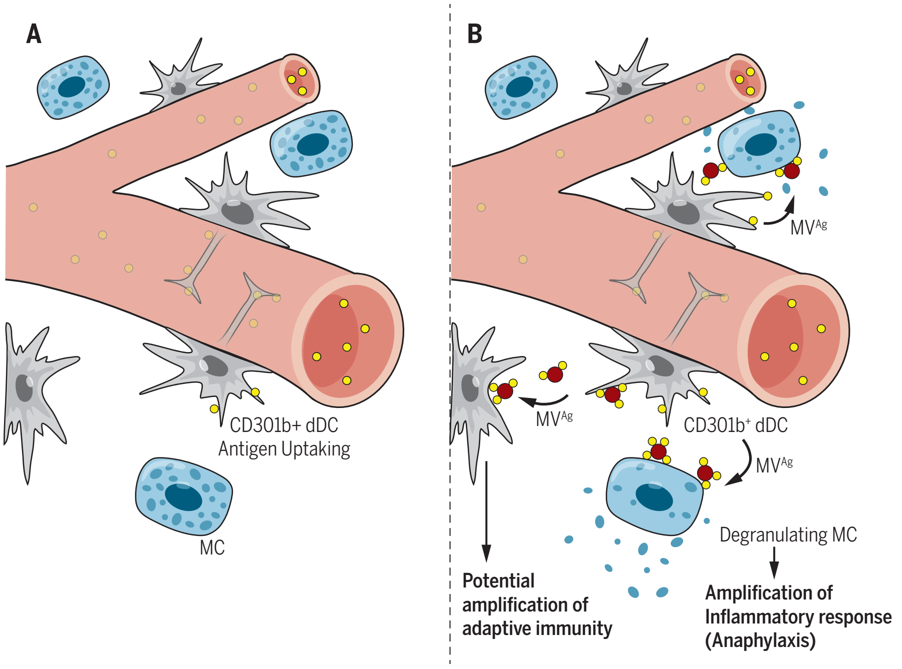
RATIONALE: Extravascular MCs were recently reported to be capable of extending cellular protrusions to directly sample vascular contents. Therefore, we reasoned that if MCs are directly sampling the blood, we should find that MCs are the first cells to acquire allergens upon systemic infusion of labeled allergen. We systemically infused mice with labeled dextran and examined which immune cells in the proximity of the vasculature had acquired the label within 30 min of antigen administration.

RESULTS: Surprisingly, MCs acquired only a small number of the dextran particles. Instead, most of the labeled antigen was acquired by a subset of dermal dendritic cells (DCs). These DCs, defined as CD301b^{+} dermal cDC2, were found in close proximity to MCs in the perivascular abluminal surface lining dermal blood vessels. Furthermore, they appeared to directly sample the blood through the extrusion of dendrites. These DCs were critical to anaphylaxis, as IgE-sensitized mice deficient in CD301b^{+} dermal cDC2 failed to undergo anaphylaxis. Live microscopy of the mouse vasculature revealed numerous perivascular DCs routinely sampling blood-borne allergens via dendrites, which penetrated the endothelial wall into the vascular lumen (left panel in the figure). Blood-borne antigens captured in this manner were relayed on the surfaces of 0.5- to 1.0-µm microvesicles (MVs) to neighboring MCs and DCs in the perivascular space. A panoply of receptors can be found on the surfaces of DCs, including the mannose receptor, which enable these cells to capture a broad range of foreign antigens in the blood. Knocking down mannose receptors markedly reduced the capacity of DCs to bind antigen and convey antigen to MCs. Furthermore, large numbers of MVs were constitutively shed by DCs regardless of whether contact with antigens was made. MVs were products of specific budding and pinching-off activities occurring on the plasma membranes of DCs. Additionally, MV shedding was abrogated when vacuolar protein sorting–associated protein 4 (VPS4), which mediates scission of MVs, was knocked down. DCs impaired in their ability to express VPS4 were found to be incapable of activating MCs or causing anaphylaxis. Noticeably, distribution of antigens by DCs on budding MVs occurred relatively quickly and was not preceded by internalization of antigen. Upon contact with allergens on the surfaces of MVs, IgE-bound MCs were found to degranulate vigorously, releasing their payload of inflammatory mediators (right panel in the figure) and triggering anaphylaxis. Thus, allergen-induced MC degranulation and anaphylaxis require the critical contribution of perivascular DCs in capturing and relaying circulating allergens.

CONCLUSION: We demonstrate how IgE-sensitized MCs are indirectly activated by blood-borne allergens. Additionally, our study reveals how perivascular DCs continuously sample blood and initiate and markedly enhance inflammatory and immune responses by rapidly discharging antigen-bearing MVs to surrounding immune cells. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aao0666



Transfer of blood-borne allergens by DCs to MCs during anaphylaxis. (**A**) Sampling of blood-borne allergens by perivascular DCs. (**B**) Relaying of allergens borne on vesicles to neighboring MCs and DCs. Ag, antigen; dDC, dermal DC.

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

*Corresponding author. Email: haewoong.choi@duke.edu

Cite this article as H. W. Choi *et al.*, *Science* **362**, eaao0666 (2018). DOI: 10.1126/science.aao0666

Choi *et al.*, *Science* **362**, 656 (2018) 9 November 2018

RESEARCH ARTICLE

IMMUNOLOGY

Perivascular dendritic cells elicit anaphylaxis by relaying allergens to mast cells via microvesicles

Hae Woong Choi[1]*, Jutamas Suwanpradid[2], Il Hwan Kim[3], Herman F. Staats[1,4,5], Muzlifah Haniffa[6,7], Amanda S. MacLeod[2,4], Soman N. Abraham[1,4,8,9]

Anaphylactic reactions are triggered when allergens enter the blood circulation and activate immunoglobulin E (IgE)–sensitized mast cells (MCs), causing systemic discharge of prestored proinflammatory mediators. As MCs are extravascular, how they perceive circulating allergens remains a conundrum. Here, we describe the existence of a CD301b$^+$ perivascular dendritic cell (DC) subset that continuously samples blood and relays antigens to neighboring MCs, which vigorously degranulate and trigger anaphylaxis. DC antigen transfer involves the active discharge of surface-associated antigens on 0.5- to 1.0-micrometer microvesicles (MVs) generated by vacuolar protein sorting 4 (VPS4). Antigen sharing by DCs is not limited to MCs, as neighboring DCs also acquire antigen-bearing MVs. This capacity of DCs to distribute antigen-bearing MVs to various immune cells in the perivascular space potentiates inflammatory and immune responses to blood-borne antigens.

Currently, 4 to 5 persons per 100,000 suffer from anaphylaxis annually. These numbers continue to grow, particularly for food-associated reactions (*1*, *2*). These reactions are especially frequent in the young, most of whom also present atopic diseases such as asthma, eczema, or allergic rhinitis (*3*). Acute anaphylaxis is associated with severe pathophysiological symptoms, such as hives, loss in blood pressure, vasculature leakage, and a drop in body temperature, which can be fatal (*4*). These symptoms are triggered soon after allergens such as peanut antigens, insect venom, and certain medications enter the circulation of antigen-specific immunoglobulin E (IgE)–sensitized individuals (*2*). Mast cells (MCs) are primary effectors of anaphylaxis because of their distinct ability to release large amounts of cytoplasmic granules enriched in inflammatory chemicals upon allergen activation of their surface IgE. MCs are typically found lining blood vessels, so when allergens enter the circulation, widespread MC degranulation is triggered, resulting in rapid and systemic onset of anaphylaxis. As MCs are located in the perivascular abluminal surface of relatively impregnable endothelial cells, how blood-borne allergens contact MCs is unclear.

MCs have the capacity to directly probe blood vessels with cellular protrusions to acquire IgE antibodies from the circulation (*5*). Dendritic cells (DCs) are often observed alongside MCs at many sites. DCs are primarily immune surveillance cells with the distinct capacity to extrude dendrites between cells that are connected via tight junctions (*6*). These probing dendrites allow DCs lying underneath gut and respiratory epithelial tracts to sample luminal contents (*6*, *7*). Additionally, epidermal Langerhans cells of the skin can penetrate the stratum corneum to sample external antigens (*8*). In this study, we investigated how abluminal perivascular MCs detect circulatory antigens through cooperation with adjacent DCs to trigger anaphylaxis.

Results

MCs and CD11c$^+$ cells are critical mediators of anaphylaxis

We investigated whether MCs were able to bind and detect blood-borne antigens in a manner similar to circulating IgE antibodies (*5*). We intravenously injected tetramethyl rhodamine isothiocyanate (TRITC)–conjugated dextran, which is unable to enter the extravascular space. We then used flow cytometry 30 min after injection to evaluate the ability of dermal abluminal MCs to acquire dextran-TRITC. Only ~1% of cKit$^+$FcεRI$^+$ skin MCs were positive for dextran-TRITC (Fig. 1A and fig. S1). Unexpectedly, up to 5% of CD11c$^+$ cells in the skin cell preparation were positive for dextran-TRITC (Fig. 1A and fig. S1). To locate these CD11c$^+$ cells within the tissue, we prepared whole mounts of the mouse ear and probed them for MCs and CD11c$^+$ cells. Numerous CD11c$^+$ cells were in close proximity to blood vessels and often in direct contact with both endothelial cells and MCs (Fig. 1B).

To investigate whether these CD11c$^+$ cells had any functional role in anaphylactic responses to allergens, we compared anaphylactic reactions in CD11c$^+$ cell–depleted mice and nondepleted mice by using two classical assays: passive systemic anaphylaxis (PSA) and passive cutaneous anaphylaxis (PCA). We administered repeated injections of diphtheria toxin (DT) to CD11c–DT receptor (DTR) mice depleted of CD11c$^+$ cells (*9*). Sixteen hours after locally sensitizing both wild-type (WT) and CD11c$^+$ cell–depleted mice with trinitrophenyl (TNP)–specific IgE, we challenged the mice intravenously with TNP-ovalbumin (TNP-OVA). To visualize the vascular leakage associated with PCA in the IgE-sensitized mice, we coinjected Evans blue dye and TNP-OVA intravenously. Strong vascular leakage occurred in the nondepleted mice, whereas the CD11c$^+$ cell–depleted mice had significantly less leakage (Fig. 1C). CD11c$^+$ cell depletion also protected mice from a severe PSA reaction upon systemic IgE sensitization and TNP-OVA challenge. A sharp drop in body temperature occurred in IgE-sensitized saline-treated CD11c-DTR mice but not in CD11c$^+$ cell–depleted mice after intravenous administration of TNP-OVA (Fig. 1D). CD11c$^+$ cell–depleted mice exhibited only a modest drop in body temperature, with no associated mortality (Fig. 1D). Thus, CD11c$^+$ cells contribute substantially to the development of PCA and PSA in IgE-sensitized mice.

To confirm the role of MCs in our anaphylaxis model, we compared the PCA and PSA responses in WT and MC-depleted mice (MC depletion was achieved by repeated administration of DT to Mcpt5-CreiDTR mice) (*10*). As expected, PCA challenge showed that TNP-specific IgE–sensitized control littermate mice experienced significantly increased blood vessel leakage compared with MC-depleted mice (Fig. 1E). To visualize the consequence of MC activation and vascular leakage at a local site, we examined the ears of TNP-specific IgE–sensitized control littermate and MC-depleted mice after intravenous injection of TNP-OVA along with dextran-TRITC. Whole-mount imaging of the mouse ears showed extensive MC degranulation and vascular leakage in control littermate mice but not in MC-depleted mice (Fig. 1F). When PSA was induced, TNP-specific IgE–sensitized control littermate mice showed a significant drop in body temperature. By contrast, MC-depleted mice exhibited only limited responses with no associated mortality after TNP-OVA challenge (Fig. 1G). Thus, in addition to MCs, CD11c$^+$ cells are important modulators of PCA and PSA in mice.

CD301b$^+$ dermal DCs are critical mediators of anaphylaxis

We next established the identity of the perivascular CD11c$^+$ cells that were collaborating with MCs to

---

[1]Department of Pathology, Duke University Medical Center, Durham, NC 27710, USA. [2]Department of Dermatology, Duke University Medical Center, Durham, NC 27710, USA. [3]Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC 27710, USA. [4]Department of Immunology, Duke University Medical Center, Durham, NC 27710, USA. [5]Duke Human Vaccine Institute, Duke University Medical Center, Durham, NC 27710, USA. [6]Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. [7]Department of Dermatology, Newcastle upon Tyne NHS Trust, Newcastle upon Tyne NE1 4LP, UK. [8]Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710, USA. [9]Program in Emerging Infectious Diseases, Duke–National University of Singapore, Singapore 169857, Singapore.
*Corresponding author. Email: haewoong.choi@duke.edu

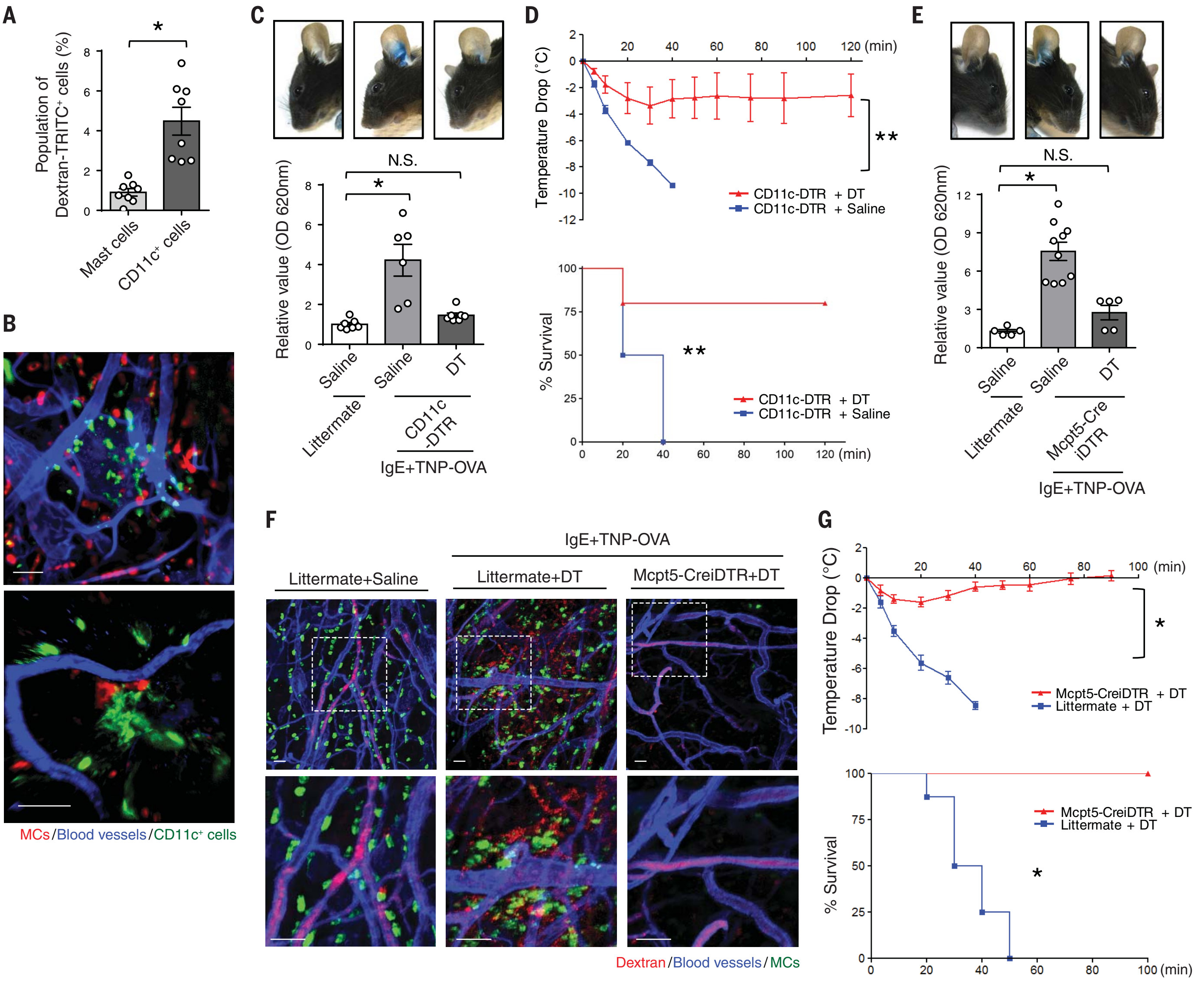


Fig. 1. PCA and PSA are mediated by MCs and CD11c$^+$ cells. (**A**) Efficient antigen uptake by CD11c$^+$ cells. CD11c-GFP mice were intravenously injected with dextran-TRITC. After 30 min, mice were sacrificed and their ears were dissected to generate a single-cell suspension. The dextran-positive populations among CD45$^+$FcεRI$^+$cKit$^+$ MCs and CD45$^+$CD11c$^+$ cells were compared. n = 8 mice per group. Data are presented as the mean ± SEM. *$P < 0.001$, unpaired Student's t test. (**B**) CD11c$^+$ cells lie in close proximity to blood vessels. The ears of mice expressing CD11c-GFP (green) were dissected. A whole mount was then prepared and stained for MCs (avidin, red) and blood vessels (CD31, blue). Scale bars, 50 μm (left) and 15 μm (right). (**C**) CD11c$^+$ cells mediate vascular leakage. CD11c-DTR or littermate mice were intraperitoneally and intravenously injected with DT or vehicle every other day twice to deplete the population of CD11c$^+$ cells. The day before antigen challenge, the ears of CD11c-DTR or WT mice were sensitized with TNP-specific IgE. TNP-OVA was intravenously injected along with Evans blue dye into both groups. One hour postinjection, mouse ears were imaged and dissected to extract dye for optical density (OD) measurements. n = 6 to 7 mice per group. Data are presented as the mean ± SEM. *$P < 0.001$, one-way ANOVA with Tukey's multiple comparisons test. N.S., not significant. (**D**) CD11c$^+$ cells mediate anaphylaxis. After CD11c$^+$ cells were depleted, IgE-sensitized CD11c-DTR mice were intravenously injected with TNP-OVA. Temperature changes were then monitored (top). Survival during PSA was recorded, and mice with temperature changes greater than 10°C were sacrificed (bottom). n = 4 to 5 mice per group. Data are presented as the mean ± SEM. **$P < 0.05$, two-way ANOVA (top) or log-rank test (bottom). (**E** and **F**) MC activation causes vascular leakage and anaphylaxis. Both Mcpt5-CreiDTR and littermate mice were intravenously injected with DT to deplete the population of MCs. The day before antigen challenge, the ears of each mouse were sensitized with TNP-specific IgE. TNP-OVA was intravenously injected into both groups of mice, along with (E) Evans blue dye or (F) TRITC-conjugated dextran. One hour postinjection, (E) mouse ears were imaged and dissected to extract dye for OD measurements. Data are presented as the mean ± SEM. *$P < 0.001$, one-way ANOVA with Tukey's multiple comparisons test. (F) Mouse ears were dissected and fixed for whole-mount imaging. Confocal microscopy was utilized to observe ears stained for MCs (avidin, green), blood vessels (CD31 antibody, blue), and dextran (red). The regions indicated by dashed squares in the upper panels are magnified in the corresponding lower panels. Scale bars, 50 μm. (**G**) MC activation triggers a sharp drop in body temperature. After MCs were depleted, littermate or Mcpt5-CreiDTR mice were injected intraperitoneally with TNP-specific IgE to sensitize them. The following day, mice were intravenously injected with TNP-OVA, and temperature changes were monitored (top). Survival during PSA was recorded (bottom). n = 5 to 8 mice per group. Data are presented as the mean ± SEM. *$P < 0.001$, two-way ANOVA (top) or log-rank test (bottom). Two to three separate experiments were performed for each individual panel.

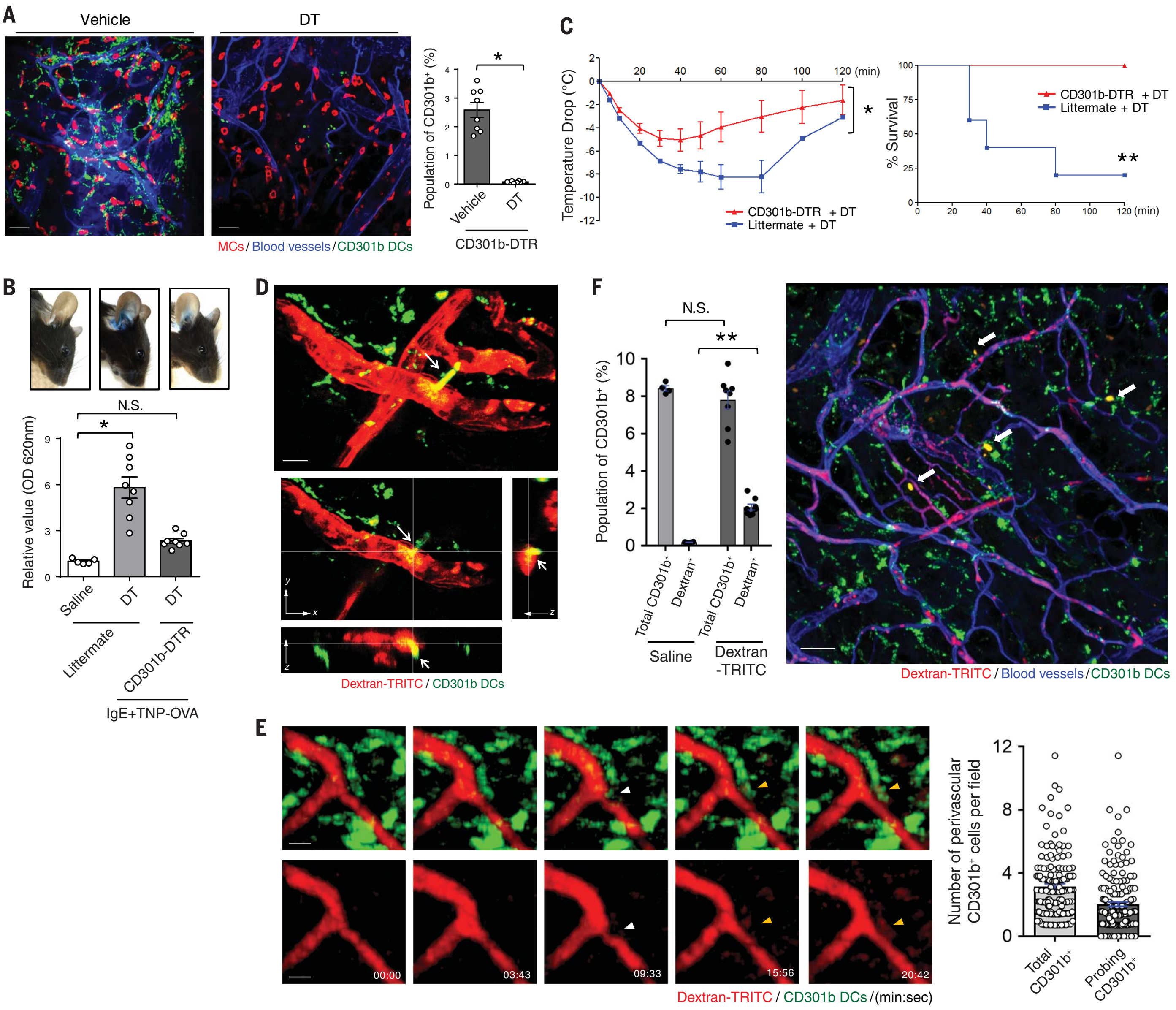


Fig. 2. CD301b$^+$ DCs mediate anaphylaxis after uptake of blood-borne antigen. (**A**) Treatment of CD301b-DTR-GFP mice with DT results in the depletion of CD301b$^+$ cells. CD301b-DTR-GFP mice were injected intraperitoneally once with DT or vehicle to deplete CD301b$^+$ cells. Depletion of CD301b$^+$ cells was quantified by flow cytometry. The ears of CD301b-DTR-GFP mice were then dissected, and a whole mount was prepared and examined by confocal microscopy. MCs (avidin), red; blood vessels (CD31), blue; and CD301b, green. $n = 8$ mice per group. Data are presented as the mean ± SEM. $^*P < 0.001$, unpaired Student's t test. Scale bars, 50 μm. (**B** and **C**) CD301b$^+$ cells mediate PCA and PSA. (B) After CD301b$^+$ cell depletion, the ears of CD301b-DTR mice or littermate controls were sensitized with TNP-specific IgE. TNP-OVA was injected intravenously, along with Evans blue dye. After 1 hour, mouse ears were imaged and then dissected to extract dye for OD measurements. $n = 5$ to 8 mice per group. Data are presented as the mean ± SEM. $^*P < 0.001$, one-way ANOVA with Tukey's multiple comparisons test. (C) After CD301b$^+$ cell depletion, CD301b-DTR or littermate mice were intraperitoneally injected with TNP-OVA, and then temperature changes were monitored (left). Mice with temperature changes greater than 10°C were sacrificed. The survival rate is shown (right). $n = 5$ mice per group. Data are presented as the mean ± SEM. $^*P < 0.001$, two-way ANOVA (left); $^{**}P < 0.01$ (survival during PSA was recorded and analyzed by a log-rank test) (right). (**D** and **E**) Three-dimensional visualization of CD301b$^+$ DCs probing the vasculature. CD301b-GFP mice were intravenously injected with dextran-TRITC. (D) After 30 min, mouse ears were dissected. Whole mounts were prepared and imaged by using two-photon microscopy. Arrows indicate a lamellipodium-like structure protruding into the vasculature from CD301b$^+$ cDC2. The top panel is an xy-plane projection from a Z-stack view, whereas the lower panels are cross-sectional views of selected xy, yz, and xz planes. Scale bar, 10 μm. (E) Immediately after injection, intravital two-photon microscopy of mouse ears was performed. Time-series events were captured and displayed. White arrowheads point to a dendrite protrusion into the vasculature from CD301b$^+$ cDC2. Yellow arrowheads point to CD301b$^+$ cDC2 taking up blood-borne dextran-TRITC. The probing activity of perivascular CD301b$^+$ cDC2 was quantified from multiple images taken over a 20-min period (right). Scale bars, 10 μm. (**F**) A substantial population of CD301b$^+$ cDC2 takes up blood-borne antigens. CD301b-GFP mice were intravenously injected with dextran-TRITC. After 30 min, the mouse ears were dissected and processed as a single-cell suspension for flow cytometry in which the dextran-TRITC$^+$ population was identified (left) or processed as whole mounts for microscopy in which antigen-sampling cells were readily detectable (arrows, right). $n = 4$ to 8 mice per group. Data are presented as the mean ± SEM. $^{**}P < 0.01$, unpaired Student's t test. Scale bar, 50 μm. Two to three separate experiments were performed for each individual panel.

mediate both PCA and PSA responses. CD11c is expressed on both DCs and macrophages (11). A recent comprehensive analysis of skin immune cells showed high CD11c expression on the dermal cDC1 subset (previously known as CD8α^{+} DCs) as well as the cDC2 subset (previously known as CD11b^{hi} dermal DCs). However, low levels of expression are also found on monocyte-derived DCs and some macrophages (11–13). The cDC2 subset constitutes the major population of DCs in the dermis. These DCs are defined as CD207^{-}CD11b^{hi}CD103^{-}EpCAM^{-}CD24$^{-/lo}$Sirp^{+}CD11c^{+} cells (13, 14) and function to promote the differentiation of Th2 and Th17 cells (15, 16).

Recently, a subset of CD301b-expressing CD11b^{hi} cDC2 was reported to be especially proficient in antigen uptake at the periphery after antigen challenge and in the mediation of subsequent antigen presentation in the draining lymph nodes (DLNs) (17–19). Although certain monocyte-derived DCs and even macrophages reportedly share the CD301b marker (11, 13, 20), the subset of CD301b-expressing CD11b^{hi} cells was designated dermal DCs because in addition to being positive for CD301b and CD11c, these cells have the phenotype CD11b^{hi}MHCII^{hi}CD64^{lo}Ly6C^{lo} (17, 21). Additionally, they express *Zbtb46*, a conventional DC (cDC)–specific transcription factor gene expressed by cDCs and their committed progenitors but not by macrophages or monocytes (22–24).

In view of their capacity to acquire antigen, we investigated whether the population of CD11c^{+} CD301b^{+} cells expressed markers characteristic of cDC2 and whether they participated in PCA and PSA. Using transgenic CD301b–green fluorescent protein (GFP)–DTR mice (18), we found that the GFP^{+}CD301b^{+} cells in the mice had the phenotype CD103^{-}CD11b^{hi}Sirpa^{+}EpCAM^{-}CD24^{+} and also expressed *Zbtb46* (fig. S2). Whole-mount–imaged ears from CD301b-GFP-DTR mice harbored numerous GFP^{+} cells in the dermal perivascular space adjacent to both the vasculature and MCs (Fig. 2A, left). Moreover, DT injection into these mice selectively and almost completely depleted the GFP^{+} cells (Fig. 2A, middle and right). Numbers of CD64^{+}CD11c^{-} cells (dermal perivascular macrophages) (11, 25) were not altered after the depletion of CD301b^{+} cells (fig. S3).

To investigate the functional contribution of CD301b^{+} DCs to anaphylaxis, we performed PCA and PSA challenges in these mice. The depletion of CD301b^{+} cDC2 markedly reduced vascular leakage in the ears (Fig. 2B) and alleviated the drop in body temperature (Fig. 2C) in CD301b-GFP-DTR mice compared with littermate controls, indicating that CD301b^{+} cDC2 contributed substantially to the development of anaphylaxis. A small subset of cDC1 (CD103^{+}CD8α^{+}CD11b^{-}) is known to be present in the dermis (26). *Batf3*$^{-/-}$ mice, which lack cDC1, also showed significantly reduced anaphylaxis (~50% of the effect observed after the removal of CD301b^{+} cDC2) (fig. S4). Thus, this capacity to sample blood-borne antigens is not confined to a particular DC subset.

To address whether this response involved the direct uptake of blood-borne antigens by perivascular CD301b^{+} cDC2, we injected intravenous dextran-TRITC into CD301b-GFP-DTR transgenic mice. Thirty minutes later, we generated whole-mount preparations of their ears, which were subjected to two-photon microscopy. We could visualize several of the fluorescent CD301b^{+} cDC2 cells adjacent to blood vessels in the process of taking up dextran-TRITC (Fig. 2D, fig. S5, and movie S1). In some cases, the CD301b^{+} cDC2 appeared to extrude long lamellipodia with multiple dendrites into the luminal surface of the blood vessel (Fig. 2D, top). Stacked images deployed in *x*-, *y*-, and *z*-axis planes revealed the movement of dextran-TRITC into CD301b^{+} cDC2. As shown in the *yz*- or *xz*-axis cross-sectioned planes, tips

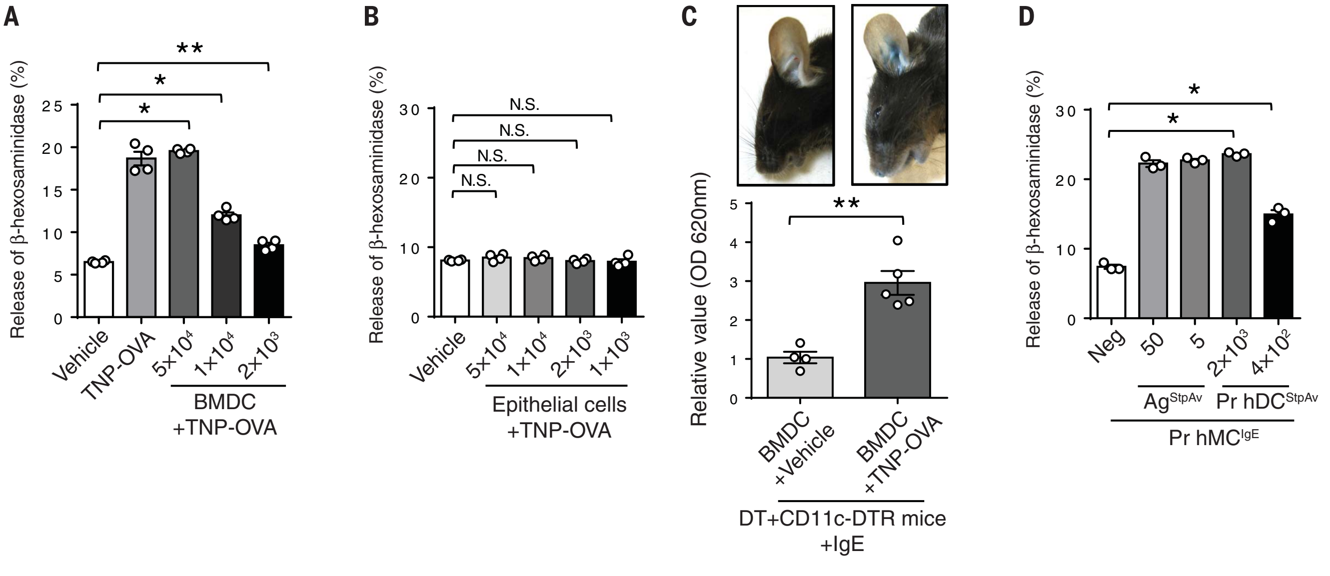


Fig. 3. Antigen-primed DCs induce MC degranulation in vitro and in vivo. (**A**) Antigen-primed BMDCs induce MC degranulation in a dose-dependent manner. BMDCs were primed with TNP-OVA for 15 min, rinsed three times, and coincubated with TNP-specific IgE–sensitized RBL-2H3 cells for 1 hour, after which β-hexosaminidase release was measured. Data are presented as the mean ± SEM. $^{*}P < 0.001$, $^{**}P < 0.05$, one-way ANOVA with Tukey's multiple comparisons test. (**B**) Antigen-primed epithelial cells fail to induce MC degranulation. The same procedure as for (A) was followed except that we replaced BMDCs with human bladder epithelial 5637 cells. Data are presented as the mean ± SEM. N.S., not significant by one-way ANOVA with Tukey's multiple comparisons test. (**C**) Antigen-primed BMDCs induce vascular leakage in vivo. CD11c^{+} cells were depleted in CD11c-DTR mice with intraperitoneal and intravenous injections of DT 2 days before challenge with antigen-primed BMDCs. The ears of these mice were sensitized with TNP-specific IgE 1 day before challenge. For the challenge, BMDCs were primed with TNP-OVA or vehicle for 20 min and then washed before they were injected intradermally into the ears of the IgE-sensitized mice. Evans blue dye was simultaneously injected intravenously into the same mice. After 1 hour, the mouse ears were imaged and dissected to extract dye for OD measurements. n = 4 to 5 mice per group. Data are presented as the mean ± SEM. $^{**}P < 0.05$, unpaired Student's t test. (**D**) Antigen-primed primary human skin DCs induce the degranulation of primary human MCs. Primary human MCs (Pr hMCs) cultured from peripheral blood were sensitized with biotinylated human IgE overnight. Primary human skin DCs (Pr hDCs) were isolated from skin tissue as described in Materials and methods, primed with streptavidin (StpAv) for 20 min, and rinsed three times. MCs and DCs were coincubated for 1.5 hours. β-Hexosaminidase released into the extracellular medium was then measured. No β-hexosaminidase release was detected in control cultures containing only DCs. Data are presented as the mean ± SEM. $^{*}P < 0.001$, one-way ANOVA with Tukey's multiple comparisons test. Neg, negative control; Ag, antigen. Two to three separate experiments were performed for each individual panel.

of protruding dendrites appeared to colocalize with dextran within the mouse blood vessel lumen (Fig. 2D, bottom). Dextran-TRITC uptake by CD301b$^+$ cDC2 in the ear was also captured by intravital two-photon microscopy (Fig. 2E, fig. S6, and movie S2). Dendrites (indicated by white arrowheads in Fig. 2E) protruding from GFP-labeled CD301b$^+$ cDC2 appeared to penetrate the vasculature within minutes and promptly became labeled with dextran-TRITC (yellow arrowheads in Fig. 2E). Thus, the uptake of blood-borne dextran-TRITC by CD301b$^+$ cDC2 appeared to be a rapid event. Additional intravital confocal microscopy of the ears of CD301b-GFP/Mcpt5-CretdTomato mice was undertaken. In these mice, CD301b$^+$ cDC2 and MCs were fluorescently labeled with GFP and tdTomato, respectively. We were able to confirm that within 15 min of intravenous administration of Alexa 647–conjugated OVA (OVA-A647), CD301b cDC2 but not MCs acquired antigen (fig. S7 and movie S3). To quantify antigen uptake by CD301b$^+$ cells in vivo, we generated a single-cell preparation from the ears of CD301b-GFP-DTR mice 30 min after intravenous injection of dextran-TRITC and examined the preparation by flow cytometry. Up to 30% of CD301b$^+$ cDC2 cells were positive for dextran-TRITC (Fig. 2F and fig. S8), suggesting that antigen uptake by CD301b$^+$ cDC2 occurs quickly, also corresponding to the rapid onset of anaphylaxis.

In vitro activation of IgE-sensitized MCs by antigen-primed DCs

Although MCs are the primary effector cells of PCA and PSA, our data suggest a surprisingly critical role for CD301b$^+$ DCs. These cells appear to be primarily responsible for acquiring antigens from the blood. We hypothesized that if this was true, the role of DCs in promoting anaphylaxis would be in relaying acquired antigens to adjacent IgE-sensitized MCs so as to activate the MCs. Thus, we performed in vitro assays to investigate whether mouse bone marrow–derived CD301b$^+$ DCs (BMDCs) are capable of activating IgE-sensitized MCs after the acquisition of allergen. We first confirmed that our generated BMDCs expressed CD301b but not markers of macrophages such as CD115 (fig. S9). BMDCs were pretreated with TNP-OVA for 20 min and then thoroughly washed to remove unbound antigen from the extracellular medium. Thereafter, the primed DCs were added in increasing number to cultures of the MC line (RBL-2H3) that had previously been sensitized with TNP-OVA–specific IgE (Fig. 3A). We observed strong and dose-dependent MC degranulation, indicating that antigens "presented" without any processing by the DCs were capable of activating MCs. When we replaced DCs with mouse epithelial cells in this assay, we failed to detect any MC activation, indicating that this activity was specific to DCs (Fig. 3B). To confirm that primed DCs have the capacity to activate sensitized MCs in vivo, we intradermally injected antigen-primed BMDCs into DT-treated CD11c-DTR mice. Before reconstitution, MCs from these mice were sensitized with TNP-specific IgE directed at TNP-OVA. As shown in Fig. 3C, intradermal delivery of antigen-primed BMDCs evoked a strong PCA reaction in the mice. This confirmed that antigen-primed DCs can activate sensitized MCs in vivo. We also examined whether human DCs can activate sensitized human MCs after being primed with antigen. We used a human skin–derived DC preparation obtained from skin explants (27) along with human peripheral blood–derived MCs (hPBMCs). The human DC preparation was exposed to streptavidin, and then the primed DCs were coincubated with hPBMCs sensitized with biotinylated IgE. As before, we observed dose-dependent MC degranulation (Fig. 3D), indicating that antigens bound onto human DCs have the capacity to cross-link and degranulate IgE-sensitized human MCs.

DCs activate MCs by discharging microvesicles with surface-bound allergens

Next, we sought to investigate how antigen-primed DCs subsequently relay antigen to MCs. Live confocal imaging of blood-borne antigen uptake by perivascular DCs in CD301b-GFP/Mcpt5-CretdTomato mice revealed the trafficking of "allergen particles" (~1 μm in diameter) between different perivascular CD301b$^+$ cDC2 cells and between CD301b$^+$ cDC2 and MCs (Fig. 4A, fig. S10, and movie S4). This raised the possibility that antigen-primed DCs shed antigen-bearing microvesicles (MVs), which traffic to neighboring MCs and DCs. Because the secretion of extracellular vesicles occurred relatively quickly (~15 min), these vesicles were unlikely to be multivesicular body–generated exosomes, which require markedly more time for their generation (28). We transiently exposed a readily transfectable DC line (29) to OVA-A647 to examine whether DCs have the capacity to shed antigen-bearing MVs after exposure to antigen. Consistent with our hypothesis, we observed that within minutes of washing off of unbound antigen, ~1-μm MVs budded off from the plasma membrane (Fig. 4B, top; fig. S11A; and movie S5). Shed MVs traveled randomly in the extracellular medium and bound neighboring immune cells (Fig. 4B, bottom; figs. S11B and S12; and movies S6 and S7). To examine whether these MVs were capable of relaying antigens, we harvested shed MVs from OVA–fluorescein isothiocyanate (FITC)–treated DCs by differential ultracentrifugation and viewed the MVs by microscopy. These MVs contained fluorescent OVA, and most of this OVA appeared to be localized on the surfaces of the vesicles (Fig. 4C).

Because antigens appeared to be loaded on the surfaces of MVs, isolated MVs from antigen-primed DCs may have the potential to activate MCs. Thus, we examined whether MVs harvested from TNP-OVA–treated BMDCs could induce the degranulation of TNP-specific IgE–sensitized MCs. MVs isolated from TNP-OVA–treated BMDCs by differential ultracentrifugation as described previously (30) were capable of activating bone marrow–derived MCs (BMMCs) in a dose-dependent manner (Fig. 4D). Notably, MVs from vehicle-treated DCs failed to activate MCs (fig. S13).

We developed a flow cytometric assay to quantitate DC-derived MVs. First, we established a gating area that corresponded to the reported size of MVs (31) by using latex beads ranging from 0.5 to 1.0 μm [x axis, forward scatter–height (FCS-H); y axis, forward scatter–width (FCS-W)] (fig. S14). Thereafter, we quantitated MVs generated by in vitro–cultured primary BMDCs (Fig. 4E) or by a DC line after exposure to fluorescent OVA-FITC (fig. S14). In each case, two distinct populations of MVs were shed by OVA-FITC–primed BMDCs. One population of MVs (~14%) was positive for FITC, whereas the remainder were FITC negative (Fig. 4F), indicating that MV production by DCs was not dependent on bound antigen. As the total number of MVs generated by vehicle-treated BMDCs during the same period of time was comparable to the total number of MVs generated after exposure to antigen (Fig. 4F), MV shedding by DCs appears to be a constitutive activity and not antigen induced.

Molecular determinants mediating DC shedding of MVs and acquisition of antigen

We investigated the molecular mechanism underlying antigen sharing and antigen acquisition by DCs. Specifically, we were interested in identifying critical cellular components mediating the shedding of antigen-bearing MVs, as well as components involved in the initial acquisition of antigen. As MCs become activated within minutes of the allergen's entering the blood, the time for DCs to acquire and then present antigens to MCs is short. It is unlikely that there would be sufficient time for DCs to endocytose antigens and then export these antigens as exosomes within vesicles originating from multivesicular bodies (28). Even if this did occur, it is unlikely that the internalized antigen would still be bound on the outer surface of the vesicle membrane, which is necessary to activate MCs. Immune cells have the innate capacity to outwardly bud off plasma membrane fragments in the form of 0.2- to 1.0-μm-diameter MVs (32). Although these MVs lack cytosolic organelles and nuclear fragments (33), they are selectively enriched in various plasma membrane proteins, including major histocompatibility complex class I, β1 integrin receptors, and vesicle-associated membrane protein 3 (VAMP3) (30, 34), and presumably, any molecules that these receptors are bound to when budding occurs. Mechanistically, outward membrane budding requires the endosomal sorting complexes required for transport (ESCRT) machinery to assemble a spiral structure at the neck of the budding vesicle, which promotes MV budding (32). This budding event is specifically catalyzed by vacuolar protein sorting–associated protein 4 (VPS4) through its specific AAA-type adenosine triphosphatase activity (35). We investigated whether VPS4 was involved in relaying allergens from DCs to MCs by using both microscopy and flow cytometry to examine MV shedding from antigen-primed DCs after their VPS4A expression was markedly reduced [by transfection with VPS4A small interfering RNA (siRNA)] (fig. S15) compared with that

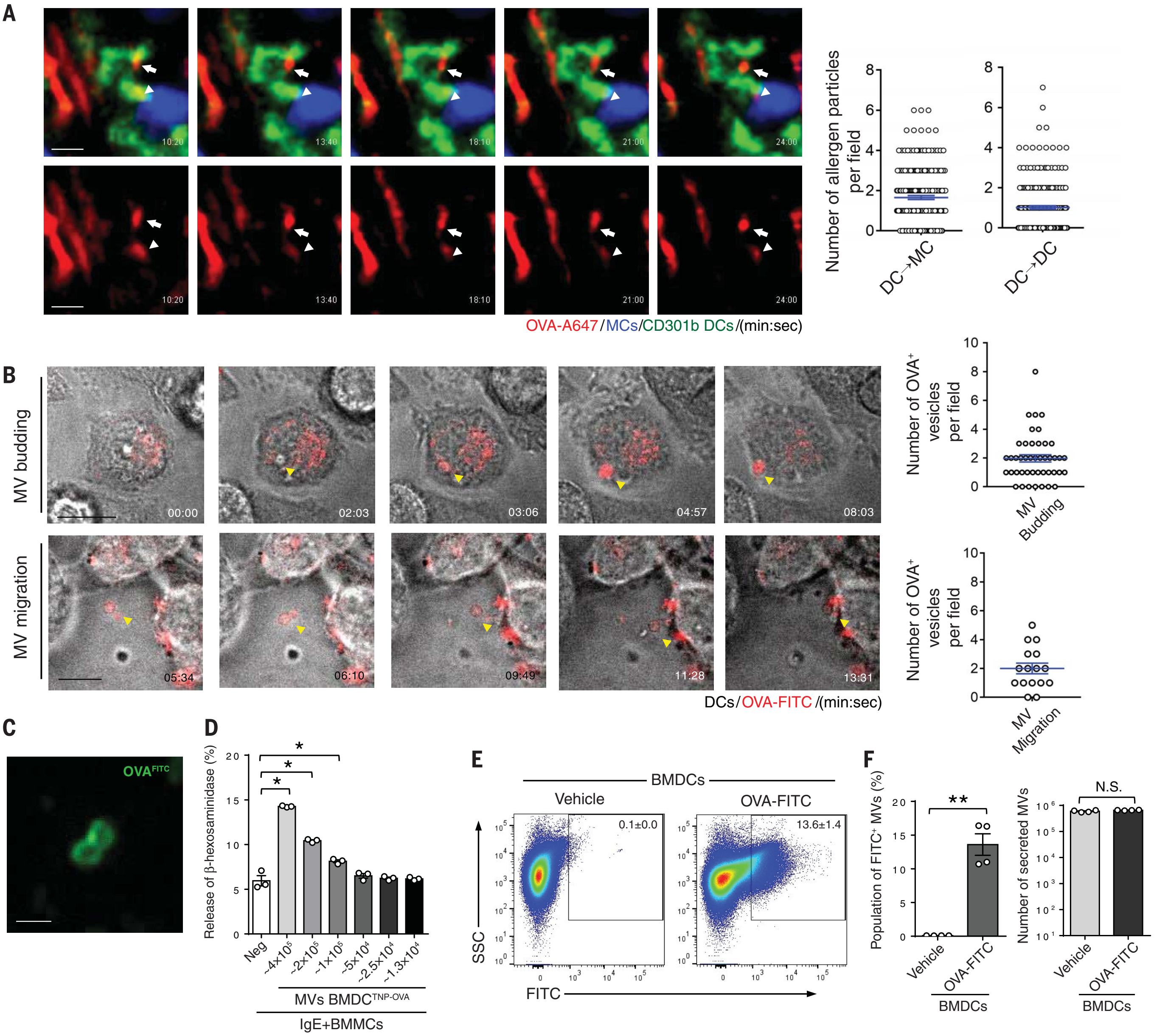


Fig. 4. Release of MVs by DCs triggers degranulation in neighboring MCs. (**A**) Time-lapse release of antigen-loaded vesicles from perivascular CD301b$^+$ cDC2 to neighboring MCs and DCs in vivo. CD301b-GFP/Mcpt5-CretdTomato mice were injected intravenously with OVA-A647. The ears of live mice were then imaged by using intravital confocal microscopy. MCs, blue; CD301b cDC2, green; A647-OVA, red. The arrowhead indicates the transfer of an antigen-bearing vesicle from a CD301b$^+$ DC (green) to an MC (blue). The arrow denotes the transfer of antigen-bearing vesicles from a CD301b$^+$ DC to another CD301b$^+$ DC. Quantification of fluorescent particles was achieved by viewing multiple images and fields. The movement of particles from DCs to MCs (DC→MC) and from DCs to other DCs (DC→DC) for a 25-min period was quantitated (right). Scale bars, 10 μm. (**B**) MV formation and secretion from a DC line (JAWSII). After the DC line was primed with antigen OVA-FITC (red), the cells were imaged by using differential interference contrast (DIC). Arrowheads denote the formation and release of antigen-loaded MVs. The budding MVs or their migration over a 20-min period was quantified by viewing multiple images and fields (right). Scale bars, 10 μm. (**C**) Fluorescent antigens are enriched on the membranes of recently formed MVs. After the DC line was primed with OVA-FITC (green), MVs released from these cells were collected through differential centrifugation and subjected to confocal microscopy. Scale bar, 1 μm. (**D**) TNP-OVA–loaded MVs induce MC degranulation. After BMDCs were treated with TNP-OVA for 15 min, the MVs were collected and MV numbers were calculated as described in Materials and methods. Collected MVs were applied to IgE-sensitized BMMCs in a dose-dependent manner (twofold dilution) for 1 hour, after which a β-hexosaminidase release assay was performed. Data are presented as the mean ± SEM. $^*P < 0.001$, one-way ANOVA with Tukey's multiple comparisons test. (**E**) Flow cytometric analysis of OVA-FITC–loaded MVs. BMDCs were primed with OVA-FITC for 15 min, and after thorough washing to remove unbound OVA-FITC, fresh medium was added and the cells were incubated for 2 hours to release MVs. These particles were collected by differential centrifugation and analyzed by flow cytometry. Gating on 0.5- to 1.0-μm MVs was performed as described in the legend to fig. S14. Pseudocolored dot plots of OVA-FITC–treated or vehicle-treated MVs were displayed with FITC on the x axis and side scatter (SSC) on the y axis. (**F**) Bar graphs demonstrating quantification of the FITC$^+$ population (left) or total MVs (right) for OVA-FITC–treated or vehicle-treated MVs. Data are presented as the mean ± SEM. $^{**}P < 0.05$, unpaired Student's t test. Two to three separate experiments were performed for each individual panel.

of control DCs [transfected with nonspecific (NS) siRNA] (Fig. 5, A and B). To visualize MV shedding, we stained DCs with fluorescent CellTracker CM-DiI dye, which selectively accumulates in the plasma membrane of the cell. We observed many fluorescent MVs shed from control antigen-primed DCs (Fig. 5A, top, and movie S8). By contrast, MV secretion was completely abrogated in VPS4A-silenced DCs (Fig. 5A, bottom, and movie S9). The surfaces of these cells appeared to be decorated with numerous vesicles in the process of budding but incapable of being released, consistent with the role of VPS4A in vesicle scission. We complemented these observations by quantitating MV shedding in control DCs (NS siRNA transfected) and an OVA-FITC–primed murine DC line with VPS4A silenced (VPS4A siRNA transfected) by using flow cytometry. We observed that the population of MVs from VPS4A-silenced DCs was reduced by up to 90% compared with the controls (Fig. 5B). Thus, the rapid relay of allergens by DCs after uptake was achieved via the budding of antigen-bound MVs from the plasma membrane in a VPS4A-dependent manner. To further support these observations, we sought to reproduce some of these observations by using primary DCs in our in vitro and in vivo studies. Using CRISPR-Cas9 technology (*36*), we knocked out *Vps4a* in primary DCs (figs. S16 and S17). These *Vps4a*$^{-/-}$ BMDCs or mock-transfected BMDCs were then exposed to OVA-FITC antigen and processed for transmission electron microscopy (TEM). Cross sections of these DCs revealed numerous 0.5- to 1-µm extracellular MVs in close association with the mock-transfected BMDCs (Fig. 5C and fig. S18). By contrast, *Vps4a*$^{-/-}$ BMDCs exhibited a limited capacity to shed MVs (Fig. 5C). Instead, numerous vesicular structures incapable of detaching were observed to form on the plasma membrane of these BMDCs (Fig. 5C), consistent with the phenotype observed via live imaging (Fig. 5A). Next, we sought to demonstrate the in vivo contribution of MVs emanating from DCs by comparing cutaneous anaphylaxis reactions in CD11c$^+$ cell–depleted mice reconstituted with *Vps4a*$^{-/-}$ BMDCs or mock-transfected BMDCs. Compared with the extensive cutaneous anaphylaxis observed in the mice reconstituted with mock-transfected BMDCs, the anaphylaxis in mice with *Vps4a*$^{-/-}$ BMDCs was greatly limited (Fig. 5D). These in vivo studies support our in vitro observations and strongly suggest that VPS4A-mediated MV secretion from DCs relays blood-borne antigen to MCs, leading to anaphylaxis.

Lastly, we investigated how antigens were bound on the surfaces of MVs. Several receptors identified on the surfaces of DCs are capable of binding a wide range of antigens. These promiscuous receptors include scavenger and mannose receptors, which could potentially bind the OVA antigen used in this study (*37*, *38*). The mannose receptor (MR) on DCs, which binds a wide range of mannosylated compounds, has previously been shown to bind OVA (*39*). To investigate the possible role of the MR in antigen uptake, we examined whether the pretreatment of mouse DCs with mannan, a linear polymer of mannose, would competitively block the binding of antigen (OVA-TNP), resulting in a loss of the ability of DCs to trigger MC degranulation. Mannan pretreatment significantly blocked the binding of antigen (OVA-TNP), as these cells subsequently failed to induce MC degranulation in a dose-dependent manner (Fig. 5E). To confirm that MRs on DCs were responsible for binding OVA-TNP, we knocked down the expression of MR in rodent DCs by using siRNA transfection (fig. S20). This significantly reduced the capacity of DCs to activate MCs after being primed with antigen (Fig. 5F). Immunoblots of isolated MVs generated by these DCs also revealed the presence of MR, confirming the role of MR in antigen transfer (fig. S21). Thus, the MRs on DCs and MVs are important for the acquisition of OVA antigens, which are subsequently presented to MCs via MVs.

Discussion

After the entry of allergens into the circulation, MCs are the primary effectors of the ensuing rapid and systemic allergic reactions. This is attributable, at least in part, to the large number of allergen-specific IgE molecules coating MCs. These cells prestore a panoply of powerful inflammatory mediators, which, upon release, can work in concert to trigger various symptoms of anaphylaxis, including vascular leakage, itchy rash, swelling of afflicted tissue, and a drop in blood pressure and body temperature. MCs can directly probe blood contents with protoplasmic protrusions to acquire circulatory IgE (*5*). This behavior was initially hypothesized to explain how these perivascular cells became so quickly activated after the entry of allergens into the circulation. However, we have shown that only a minority of MCs acquire blood-borne allergens in this manner. Rather, most perivascular MCs are activated in an indirect fashion by CD301b$^+$ dermal DCs that are continuously probing blood contents. Live and fixed-tissue microscopy of the mouse ear vasculature revealed that blood-borne antigens were adsorbed by lamellipodia of CD301b$^+$ cells that protrude through the endothelial wall into blood vessels. After the acquisition of antigens by the MR, CD301b$^+$ DCs spontaneously shed antigen-bound MVs to neighboring IgE-bearing MCs in the perivascular region. Upon degranulation, the many granule-associated MC mediators triggered both vascular leakage and a drop in body temperature. The functional importance of both MCs and CD301b$^+$ DCs to anaphylaxis was demonstrated by the abrogation of anaphylaxis when either one of these cell populations was depleted in mice. Although CD301b$^+$ cDC2 has previously been reported to predominate in the dermis (*14*, *17*, *18*), we have noticed the preferential location of this subset in the abluminal surface of the dermal vasculature. We do not exclude a role for other DC subsets (*26*, *40*) in anaphylaxis because CD103$^+$CD8α$^+$CD11b$^-$ cDC1 also contributes to this process, albeit to a lesser extent than CD301b$^+$ cDC2.

That CD301b$^+$ DCs and not MCs were primarily responsible for probing the vasculature for allergens is not surprising, as DCs are the professional immune sampling cells in the body. Subepithelial DCs have been implicated in sampling various body sites that are exposed to the environment, such as the airways, gut, and skin (*6*–*8*). Many environmental antigens can enter the circulation either when the vasculature is ruptured or when antigens, including microorganisms, are introduced into the blood (e.g., via insect bites) (*41*, *42*). Thus, there is a need to continuously sample the blood for these extrinsic agents. Dermal CD301b$^+$ DCs appear to perform this vital function. Although perivascular macrophages also appear to be capable of transendothelial sampling many hours after exposure to blood-borne antigens (*25*, *43*), their functional relevance, particularly in the context of anaphylaxis, remains unknown. It is currently unclear how allergic reactions in the skin and mucosal surfaces are triggered. As MCs are also found in the same subepithelial region where DCs probing these locations lie, it is conceivable that environmental allergens are also initially acquired by DCs and then delivered to adjacent sensitized MCs to trigger local allergic reactions. This could be the explanation for the observation made more than a decade ago showing that the depletion of lung CD11c$^+$ cells during allergen challenge abrogates the characteristic features of asthma in mice (*44*).

Antigen-presenting cells, including dermal CD301b$^+$ cells, ingest extrinsic antigens in the periphery and then traffic to proximal DLNs, where they present processed antigens to T cells, resulting in the development of antigen-specific immune responses (*15*). Here, we report a distinct form of antigen presentation by the CD301b$^+$ DC subset, where antigens adsorbed onto the DC surfaces are promptly relayed to neighboring MCs in the skin to cause systemic anaphylaxis. Although some of the antigens obtained by DCs were internalized, a substantial amount remained on the plasma membrane and were released to the extracellular medium associated with MVs. These vesicles are the products of specific budding and pinching-off activities occurring on the plasma membranes of DCs, because the secretion of MVs was abrogated when VPS4, which mediates the scission of MVs from the DC-derived plasma membrane, was specifically knocked down. We confirmed the functional role played by MVs in anaphylaxis by showing the attenuated ability of *Vps4a*-deficient primary DCs to induce cutaneous vascular leakage. Because MCs are activated within minutes of allergens entering the blood, there is unlikely to be sufficient time for the internalization of antigens and for any packaging of antigens in vesicles analogous to exosomes. Rather, the budding of membrane vesicles containing surface-bound antigens appears to be a possible mechanism for the rapid transfer of these antigens.

During immune surveillance, DCs regularly utilize the broad binding specificity of the cell surface MR to capture and clear a wide range of extrinsic and endogenous antigens that the DCs encounter. Although antigens bound by MR can be endocytosed (37), we show here that at least some OVA antigens bound by MRs are held onto the plasma membrane and subsequently presented to neighboring immune cells via MVs. The factors that determine the endocytosis of antigens by the MR or redistribution to neighboring cells via MVs remain unclear.

Our findings provide a mechanism for how blood-borne allergens trigger tissue MC degranulation through allergen acquisition from the circulation by perivascular CD301b^{+} cells and direct transfer to MCs through MVs. The capacity of CD301b^{+} DCs to sample extrinsic antigens in the dermis and traffic to the DLNs (15, 18) suggests that CD301b^{+} DCs are key mediators of tissue MC degranulation in response to blood-borne allergens, in addition to playing a recognized role as migratory DCs capable of presenting processed antigen in DLNs. The ability of CD301b^{+} DCs to acquire blood-borne extrinsic antigens and relay them to neighboring MCs and other DCs may be

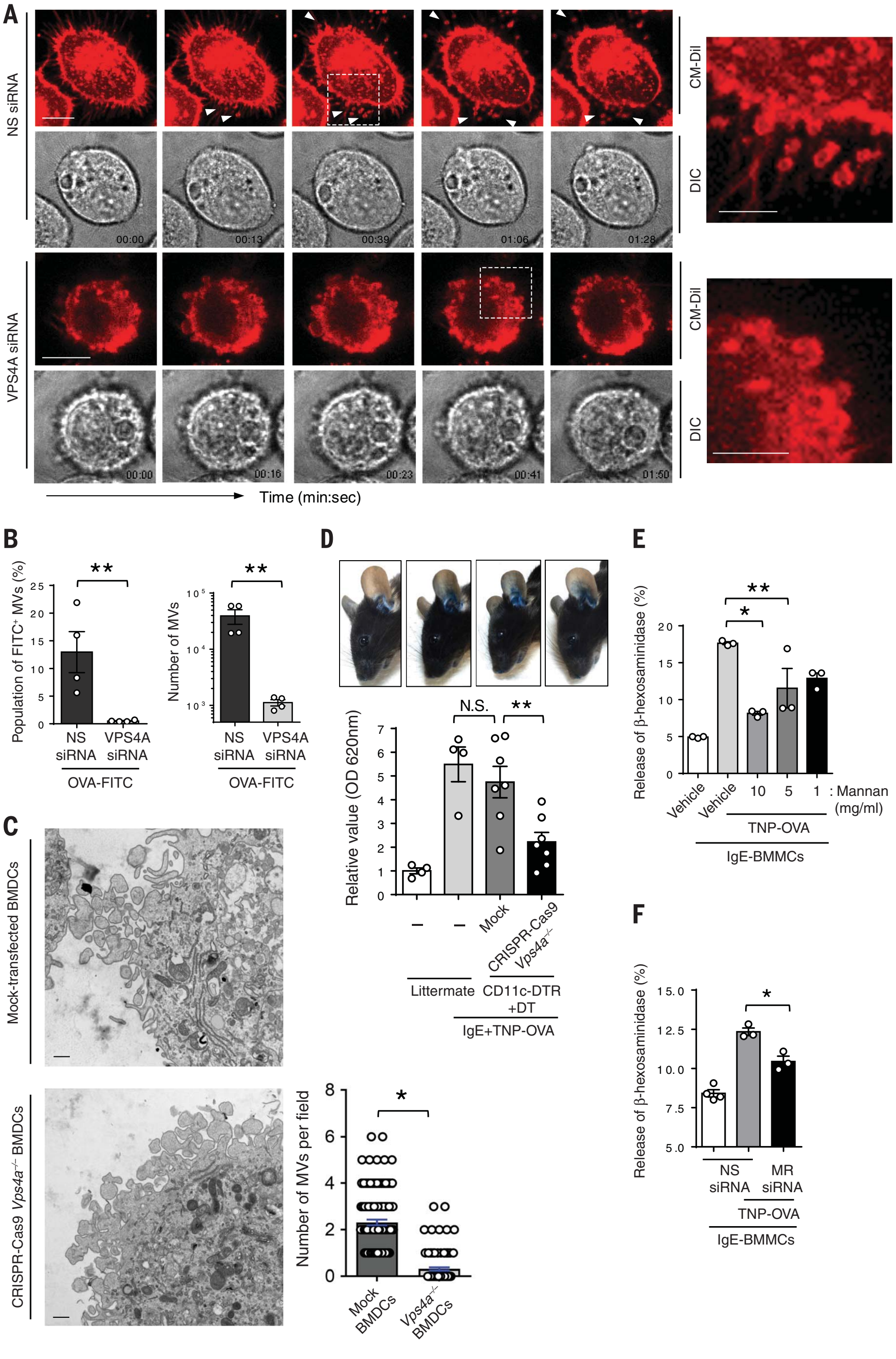


Fig. 5. MV shedding by DCs is VPS4 dependent, and antigens borne by MVs are bound by the MR. (**A**) VPS4 mediates MV shedding. MV shedding by DC lines transfected with VPS4A siRNA (bottom rows) or NS siRNA (top rows) was monitored. Forty-eight hours after transfection, the DCs were stained with CM-DiI dye (red), which stains membranes, and then exposed to antigen (OVA) and examined by time-lapse microscopy. In each group, the upper panels are fluorescence images and the lower panels are the corresponding DIC images. The arrowheads indicate sites of MV shedding. Representative enlarged images of the regions indicated by dashed squares demonstrate the shedding of MVs (top right) and the inhibition of shedding of formed vesicles (bottom right). Scale bars, 10 μm (time-lapse images) and 5 μm (magnified images). (**B**) MV shedding requires the VPS4A subunit. MVs produced by control and VPS4A-knockdown DC lines were quantitated by flow cytometry. Data are presented as the mean ± SEM. $^{**}P < 0.05$, unpaired Student's *t* test. (**C**) Ultrastructure of MV formation by BMDCs. CRISPR-Cas9–generated $Vps4a^{-/-}$ BMDCs or mock-transfected BMDCs were pretreated with OVA for 15 min and immediately fixed with 4% PFA and processed for TEM. MVs shed by mock-transfected BMDCs or CRISPR-Cas9 $Vps4a^{-/-}$ BMDCs were quantified by viewing multiple images and fields (right). Scale bars, 500 nm. Data are presented as the mean ± SEM. $^{*}P < 0.01$, unpaired Student's *t* test. (**D**) MC-dependent anaphylaxis requires the secretion of MVs from nearby DCs. CD11c cell–depleted mice (CD11c-DTR) were reconstituted with CRISPR-Cas9–generated $Vps4a^{-/-}$ BMDCs or mock-transfected BMDCs. (Confirmation of the specificity of $Vps4a^{-/-}$ is shown in figs. S16 and S17; the successful reconstitution of BMDCs is depicted in fig. S19.) MCs in the reconstituted mice or littermate control mice were sensitized by intradermal injection of TNP-specific IgE antibody. On the following day, mice were intravenously challenged with TNP-OVA, along with Evans blue dye. After 30 min, the ears of mice were imaged (top) and dissected for dye extraction and quantitation as shown in the bar graph (bottom). $n = 4$ to 7 mice per group. Data are presented as the mean ± SEM. $^{**}P < 0.05$, one-way ANOVA with Tukey's multiple comparisons test. (**E**) The MR mediates OVA antigen acquisition by DCs. The DC line was preincubated with various concentrations of mannan (1, 5, and 10 mg/ml) and then primed with TNP-OVA (0.1 μg/ml). This cell line was then added to IgE-sensitized BMMCs, and MC degranulation was assessed. Data are presented as the mean ± SEM. $^{*}P < 0.01$, $^{**}P < 0.05$, one-way ANOVA with Tukey's multiple comparisons test. (**F**) The MR on DCs is responsible for binding OVA antigen. A DC line transfected with MR siRNA or NS siRNA was primed with TNP-OVA (0.1 μg/ml) and exposed to IgE-sensitized BMMCs. The MC degranulation response was then evaluated. Data are presented as the mean ± SEM. $^{*}P < 0.01$, one-way ANOVA with Tukey's multiple comparisons test. Two to three separate experiments were performed for each individual panel.

a powerful but underappreciated mechanism for enhancing the magnitude of inflammatory and immune responses to blood-borne antigens. As effective treatments for anaphylaxis remain elusive, recognizing the role of CD301b$^+$ DCs in allergen sampling may serve as a target for future therapeutic strategies.

Materials and methods

Mouse strains

Eight- to twelve-week-old female and male mice were used for most of these studies. The following mouse strains were used: C57BL/6 WT, CD11c-DTR-GFP, CD301b-DTR-GFP, iDTR, AI14, and *Batf3*$^{-/-}$ mice (Jackson Laboratories) and Mcpt5-Cre$^+$ mice (a gift from A. Roers, University of Technology, Dresden, Germany). CD11c$^+$ cells were depleted as follows: 8- to 12-week-old CD11c-DTR-GFP mice and their littermates were given intraperitoneal and intravenous injections of 500 ng of DT per mouse every other day, twice. To deplete CD11c$^+$ cells in ear skin, CD11c-DTR mice were administered DT at 500 ng intraperitoneally and 50 ng intradermally in both ears. Two days after DT injection, 1×10^6 BMDCs, which were electroporated with a *Vps4a*-specific CRISPR-Cas9–expressing plasmid, mock plasmid (PX458), or control vector pmaxGFP (Lonza), were intradermally injected into the ears. CD301b$^+$ cells were depleted as follows: 8- to 12-week-old CD301b-DTR-GFP mice and their littermates were administered intraperitoneal and intravenous injections of 250 ng of DT per mouse every other day for 1 week or intraperitoneal and intravenous injections of 500 ng of DT per mouse once. MCs in Mcpt5-CreiDTR mice were conditionally depleted as follows: 8-week-old Mcpt5-CreiDTR mice and Mcpt5-Cre/+ littermates were given five intravenous injections of 200 ng of DT per mouse within 1 week. Mcpt5-CreAI14 mice and CD301b-GFP mice were crossed to generate Mcpt5-CreAI14:CD301b-GFP mice for intravital imaging. All procedures related to mice were performed in strict accordance with the animal protocol approved by the of Duke University Institutional Animal Care and Use Committee.

Cell culture

For BMMC culture, WT bone marrow was obtained from WT C57BL/6 mice in ice-cold Hanks' balanced salt solution (HBSS) (Gibco) and cultured in complete RPMI containing 10% fetal bovine serum (FBS) (Hyclone), 1 mM nonessential amino acids, 25 mM HEPES [4-(2-hydroxyethyl)-1-piperazineethane-sulfonic acid], 2 mM L-glutamine, 1 mM sodium pyruvate, 1× Antibiotic-Antimycotic solution (all reagents from Gibco), 10-ng/ml stem cell factor (SCF), and 5-ng/ml interleukin-3 (IL-3) (BioLegend) for 8 to 12 weeks. For BMDC culture, bone marrow was obtained from WT C57BL/6 or CD301b-DTR-GFP mice and cultured in RPMI containing 10% FBS, 1× GlutaMAX (Gibco), and 20-ng/ml granulocyte-macrophage colony-stimulating factor (GM-CSF) (BioLegend) or 20-ng/ml GM-CSF and 50-ng/ml IL-4 for 6 to 7 days. For bone marrow–derived macrophage (BMM) culture, the same procedures were followed but with 20-ng/ml macrophage colony-stimulating factor (M-CSF) replacing the growth factors. The rat MC RBL-2H3 cells (ATCC) were cultured in minimum essential medium (MEM) (Gibco) containing 15% FBS and antibiotics. JAWSII cells (ATCC) were maintained in MEMα (Gibco) supplemented with 20% FBS, 1 mM sodium pyruvate, 100-U/ml penicillin, and 100-μg/ml streptomycin. All cells were cultured at 37°C in a humidified water-jacketed incubator under a 5% CO_2–95% air atmosphere.

Primary human MC culture

Detailed procedures for primary human MC culture have been explained in detail elsewhere (45). For the culture of primary MCs, Iscove's modified Dulbecco's medium (IMDM) (Gibco) containing 200 μM bovine serum albumin (BSA) (Sigma), 1× insulin-transferrin-selenium (ITS) supplement (Gibco), 75 μM β-mercaptoethanol (Sigma), 100-U/ml penicillin, and 100-μg/ml streptomycin was used. In brief, purified CD34$^+$ peripheral blood cells (Allcells) were suspended at 5×10^5 cells/ml in supplemented IMDM containing 100-ng/ml recombinant human SCF (rhSCF), 10-ng/ml recombinant human IL-6 (rhIL-6), and 1-ng/ml rhIL-3 for 0 to 3 weeks. For the subsequent 3 to 6 weeks, the cells were suspended at 5×10^5 cells/ml in supplemented IMDM containing 100-ng/ml rhSCF and 50-ng/ml rhIL-6. After 6 weeks of differentiation, cells were finally suspended at 5×10^5 cells/ml in supplemented IMDM containing 100-ng/ml rhSCF, 50-ng/ml rhIL-6, and 10% FBS. After differentiation of primary human MCs, we evaluated the maturation of primary human MCs via flow cytometry (>95% were double positive for c-kit and FcεRI) and toluidine blue staining (dark purple staining).

β-Hexosaminidase assay

The day before the experiment, RBL-2H3 cells or BMMCs were sensitized with TNP-specific IgE antibody (BD Biosciences). For IgE antigen stimulation, JAWSII cells or BMDCs were pretreated with TNP-OVA (Biosearch Technologies) for 15 min and then thoroughly washed three times with phosphate-buffered saline (PBS). DCs and MCs were then coincubated for 1 hour, and supernatant was collected for β-hexosaminidase release assay. Briefly, collected supernatant was incubated with *p*-nitrophenyl-*N*-acetyl-β-D-glucosaminide (3.4 mg/ml) dissolved in 0.1 M citrate buffer (pH 4.5) for 1 hour at 37°C. Carbonate buffer (0.1 M, pH 10) was then applied to the reaction wells to stop the reaction and develop color. The colorimetric measurement was performed at 405 nm by using a microplate reader.

PCA and PSA

To elicit PCA, mice were injected intradermally in the ears with TNP-specific IgE antibody (100 ng/20 μl of PBS) (BD Biosciences). After 16 hours, mice were injected intravenously with 200 μg of TNP-OVA diluted in saline. For some experiments, 0.5% Evans blue dye or 1 mg of dextran-TRITC was injected along with 200 μg of TNP-OVA solution. To elicit PSA, mice were first sensitized with TNP-specific IgE (20 μg in saline) (BD Biosciences) by intravenous injection. After 24 hours, mice were injected intravenously with 2 μg of TNP-OVA. After antigen challenge, body temperature was measured with a rectal microprobe thermometer every 10 min to monitor systemic anaphylaxis.

Skin explant preparation and culture

Detailed procedures for skin explant preparation and culture have been described previously (27). Skin explants were obtained from biopsy specimens. Cut samples of approximately 1 cm by 1 cm by 0.5 cm were placed in 1 ml of Dulbecco's modified Eagle's medium (DMEM) (Gibco) containing 10% FBS. Two to three days after culture, migrated cells were harvested and used for further experiments. Human skin specimens were collected from healthy patients undergoing plastic surgery at Duke University Medical Center and used in an anonymized fashion. All human samples for this study were obtained according to the protocols approved by the Institutional Review Board at Duke University.

Flow cytometry

To produce single-cell suspensions for flow cytometry, dissected mouse ears were split open and the only dermis side lacking cartilage was digested with a digestion buffer (15% collagenase, 1% deoxyribonuclease I, 10 mM HEPES, and 1.5% FBS in HBSS). Single-cell suspensions were washed and stained with the fluorescently labeled antibodies against CD11b (clone M1/70), CD11c (clone N418), CD24 (clone M1/69), CD45 (clone 30-F11), CD64 (clone X54-5/7.1), CD103 (2E7), CD115 (clone AFS98), CD117 (clone 2B8), B220 (clone RA3-6B2), IA/IE (clone M5/114.15.2), FcεRI (clone MAR-1), Sirpα (clone P84), EpCAM (clone G8.8), and isotype controls (all antibodies from BioLegend) for 30 min at 4°C. These antibodies were used to compare expression levels. Flow cytometry was performed on a FACSCanto II or LSR II (BD) and was analyzed by using FlowJo software (Tree Star, Ashland, OR).

Vps4a-specific CRISPR-Cas9

To delete *Vps4a* in BMDCs and JAWSII cells, we used a genome-editing approach targeting the endogenous *Vps4a* gene through clustered regularly interspaced short palindromic repeats (CRISPR) and the endonuclease Cas9 (36). A CRISPR single guide RNA (sgRNA) targeting exon 2 of the *Vps4a* gene was designed by using the CRISPR design database (46) and was cloned into the pSpCas9 (BB)-2A-GFP (PX458) plasmid (Addgene). The oligonucleotide sequence for *Vps4a* sgRNAs was GGATAAAGCCAAGAACTACG [20-nt (nucleotide) sequence, 5′ to 3′]-AGG (protospacer adjacent motif). The CRISPR-Cas9 plasmid (PX458) with guide RNA targeting *Vps4a* exon 2 was electroporated with Amaxa Nucleofector (Lonza) into JAWSII cells or BMDCs. Forty-eight hours after electroporation into the JAWSII cells, the efficiency in generating site-specific double-strand breaks was measured through the SURVEYOR mutation detection assay (Integrated DNA Technologies) from purified genomic DNA derived from the JAWSII cells. After confirmation of the construction of CRISPR-Cas9 with sgRNA targeting *Vps4a*, BMDCs, which were matured with GM-CSF (20 ng/ml) and IL-4 (50 ng/ml) for 6 days, were electroporated with the constructed plasmid. Forty-eight hours after electroporation, immunoblotting with anti-VPS4A antibody (Santa Cruz Biotechnology) was performed to confirm the knockout of endogenous VPS4A protein.

Electroporation

BMDCs (2 × 10^6) were resuspended into 100 μl of Nucleofactor solution V (Lonza), mixed with *Vps4a*-specific CRIPSR-Cas9 or mock plasmid, and electroporated by using Amaxa Nucleofector (Lonza) with program Y-001. Electroporated cells were cultured in medium with 20-ng/ml GM-CSF and 50-ng/ml IL-4 for 48 or 72 hours.

Immunofluorescent staining and microscopy

For coculture visualization of MCs and DCs, BMMCs and BMDCs were fixed in 4% paraformaldehyde (PFA) and permeabilized in 0.1% saponin (Sigma) per 1% BSA in PBS. For whole-mount ear visualization, ears were excised and split open and immediately fixed in 4% PFA at room temperature for 2 hours. To block and permeabilize the tissue, buffer containing 0.3% Trion X-100 and 2.5% normal goat serum (Gibco) in 1% BSA–PBS was used.

To visualize blood vessels, 0.5-μg/ml rat anti-mouse CD31 antibody (clone MEC 13.3; BD Biosciences) was incubated with mouse ears overnight. A647-conjugated F(ab′)2 fragment anti-rat IgG (Jackson ImmunoResearch) was used as a secondary antibody. To visualize MCs, FITC-avidin (BD Biosciences) was used. The samples were visualized by confocal microscopy or two-photon microscopy.

For live imaging of MV secretion from DCs, JAWSII cells were grown on MatTek plates (MatTek Corporation) and treated with CellTracker CM-DiI dye (Molecular Probes) for 20 min in serum-free MEMα at 37°C and 5% CO$_2$ with humidification. After thorough washing with PBS, the prestained cells were treated with OVA or PBS for 15 min in culture medium. After thorough washing with PBS, cells were maintained at 37°C while a Nikon ECLIPSE TE200 confocal microscope equipped with a 100×/1.49NA oil immersion objective lens was used to capture live moments for 15 min.

For intravital imaging by confocal microscopy, Mcpt5-CreAI14:CD301b-GFP mice were generated as described above. Mice were anesthetized by using a mixture of ketamine and xylaxine and kept on a heating pad maintained at 38°C. The inner sides of mouse ears were taped onto a glass slide and immobilized with tape. One hundred microliters of 5-mg/ml A647-OVA (Molecular Probes) was injected intravenously, and ears were immediately imaged by a Nikon ECLIPSE TE200 confocal microscope equipped with a 20×/0.75NA multi-immersion objective lens.

For intravital imaging by two-photon microscopy, CD301b-eGFP-DTR mice were anesthetized with a mixture of isoflurane and O$_2$ supplied through an isoflurane vaporizer to a mask covering the snout. To prevent hypothermia of mice during intravital imaging, the imaging platform was heated and maintained at 38°C. Ear pinnae were taped onto the imaging platform to secure their position and minimize movement during imaging. Immediately after the injection of 200 μl of 10-mg/ml dextran-TRITC (150 kDa; Molecular Probes), time-lapse *Z*-stack imaging was initiated by using an Olympus FV1000 multiphoton upright microscope, characterizing in vivo movement of CD301b cDC2 near blood vessels at 30-s intervals for a total of 30 min. An Olympus FV1000 multiphoton microscope equipped with a 25×/1.05 NA water immersion objective lens was used for two-photon microscopy. Collected series of images were reconstructed in three dimensions by using Imaris software (Bitplane).

Preparation of MVs

JAWSII cells or BMDCs plated on 100-cm^2 culture dishes were treated with antigens for 15 min in culture medium and then thoroughly washed with PBS to clear unbound antigens. After culture medium was resupplied, cells were incubated under culture conditions for 2 hours. Collected culture supernatant was spun at 400 × *g* for 5 min, and the resulting supernatant was re-spun under the same conditions to remove cell debris. Thereafter, a final high-speed spin at 100,000 × *g* for 30 min was performed. An Optima L-90K ultracentrifuge (Beckman Coulter) was used for centrifugation. The collected MV pellet was resuspended in PBS and analyzed by flow cytometry, where particles ranging in size between 0.5 and 1.0 μm were counted. In our studies, we observed that DCs cultured in 100-cm^2 culture dishes with 70% confluency typically generated ~1 × 10^6 MVs.

Immunoblots

Samples were mixed with radio immunoprecipitation assay buffer and then vortexed for 30 s, three times. After centrifugation at 13,000 × *g*, the lysates were mixed with sample buffer and separated on 4 to 20% SDS–polyacrylamide gel electrophoresis gels. Proteins on gels were transferred onto polyvinylidene difluoride membranes (Bio-Rad) and blocked with 3% BSA in Tris-buffered saline with Tween-20 for 2 hours. Membranes were incubated with mouse anti-VPS4A monoclonal antibody (clone A-11, 0.2 μg/ml; Santa Cruz Biotechnology) or rabbit anti-MR1 polyclonal antibody (1 μg/ml; R&D Systems) overnight at 4°C. Anti-mouse IgG antibody (1:2500; Bio-Rad) or anti-rabbit IgG antibody (1:1000; Bio-Rad) conjugated with horseradish peroxidase was used as a secondary antibody.

siRNA sequences

siRNA sequences were as follows: for VPS4A, 5′-GUGGAAUGAUGUAGCUGGA[dT][dT]-3′, and for MR, C type 1, 5′-CUCCUUACUGGGCAAUGCA[dT][dT]-3′.

Statistical analyses

Statistical analyses were performed by using GraphPad Prism v.6 (GraphPad Software). Unpaired Student's *t* test, two-way analysis of variance (ANOVA), and one-way ANOVA with Tukey's multiple comparisons tests were used to calculate statistical significance. Log-rank tests were used to analyze survival during PSA. A *P* value of <0.05 was considered statistically significant. Data are presented as the mean ± SEM.

REFERENCES AND NOTES

1. M. A. Tejedor-Alonso, M. Moro-Moro, M. V. Múgica-García, Epidemiology of anaphylaxis: Contributions from the last 10 years. *J. Investig. Allergol. Clin. Immunol.* **25**, 163–175 (2015). pmid: 26182682
2. S. R. Boden, A. Wesley Burks, Anaphylaxis: A history with emphasis on food allergy. *Immunol. Rev.* **242**, 247–257 (2011). doi: 10.1111/j.1600-065X.2011.01028.x; pmid: 21682750
3. F. E. R. Simons, World Allergy Organization survey on global availability of essentials for the assessment and management of anaphylaxis by allergy-immunology specialists in health care settings. *Ann. Allergy Asthma Immunol.* **104**, 405–412 (2010). doi: 10.1016/j.anai.2010.01.023; pmid: 20486330
4. H. A. Sampson *et al.*, Symposium on the definition and management of anaphylaxis: Summary report. *J. Allergy Clin. Immunol.* **115**, 584–591 (2005). doi: 10.1016/j.jaci.2005.01.009; pmid: 15753908
5. L. E. Cheng, K. Hartmann, A. Roers, M. F. Krummel, R. M. Locksley, Perivascular mast cells dynamically probe cutaneous blood vessels to capture immunoglobulin E. *Immunity* **38**, 166–175 (2013). doi: 10.1016/j.immuni.2012.09.022; pmid: 23290520
6. M. Rescigno *et al.*, Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. *Nat. Immunol.* **2**, 361–367 (2001). doi: 10.1038/86373; pmid: 11276208
7. S. S. Sung *et al.*, A major lung CD103 (α_E)-β_7 integrin-positive epithelial dendritic cell population expressing Langerin and tight junction proteins. *J. Immunol.* **176**, 2161–2172 (2006). doi: 10.4049/jimmunol.176.4.2161; pmid: 16455972
8. A. Kubo, K. Nagao, M. Yokouchi, H. Sasaki, M. Amagai, External antigen uptake by Langerhans cells with reorganization of epidermal tight junction barriers. *J. Exp. Med.* **206**, 2937–2946 (2009). doi: 10.1084/jem.20091527; pmid: 19995951
9. S. Jung *et al.*, In vivo depletion of CD11c$^+$ dendritic cells abrogates priming of CD8$^+$ T cells by exogenous cell-associated antigens. *Immunity* **17**, 211–220 (2002). doi: 10.1016/S1074-7613(02)00365-5; pmid: 12196292
10. A. Dudeck *et al.*, Mast cells are key promoters of contact allergy that mediate the adjuvant effects of haptens. *Immunity* **34**, 973–984 (2011). doi: 10.1016/j.immuni.2011.03.028; pmid: 21703544
11. S. Tamoutounour *et al.*, Origins and functional specialization of macrophages and of conventional and monocyte-derived dendritic cells in mouse skin. *Immunity* **39**, 925–938 (2013). doi: 10.1016/j.immuni.2013.10.004; pmid: 24184057
12. M. Merad, P. Sathe, J. Helft, J. Miller, A. Mortha, The dendritic cell lineage: Ontogeny and function of dendritic cells and their subsets in the steady state and the inflamed setting. *Annu. Rev. Immunol.* **31**, 563–604 (2013). doi: 10.1146/annurev-immunol-020711-074950; pmid: 23516985
13. B. Yang *et al.*, IL-27 facilitates skin wound healing through induction of epidermal proliferation and host defense. *J. Invest. Dermatol.* **137**, 1166–1175 (2017). doi: 10.1016/j.jid.2017.01.010; pmid: 28132857
14. S. Henri *et al.*, CD207$^+$ CD103$^+$ dermal dendritic cells cross-present keratinocyte-derived antigens irrespective of the presence of Langerhans cells. *J. Exp. Med.* **207**, 189–206 (2010). doi: 10.1084/jem.20091964; pmid: 20038600
15. H. Tang *et al.*, The T helper type 2 response to cysteine proteases requires dendritic cell-basophil cooperation via ROS-mediated signaling. *Nat. Immunol.* **11**, 608–617 (2010). doi: 10.1038/ni.1883; pmid: 20495560
16. A. Schlitzer *et al.*, Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow. *Nat. Immunol.* **16**, 718–728 (2015). doi: 10.1038/ni.3200; pmid: 26054720
17. Y. Kumamoto, K. Denda-Nagai, S. Aida, N. Higashi, T. Irimura, MGL2 dermal dendritic cells are sufficient to initiate contact hypersensitivity in vivo. *PLOS ONE* **4**, e5619 (2009). doi: 10.1371/journal.pone.0005619; pmid: 19440334
18. Y. Kumamoto *et al.*, CD301b$^+$ dermal dendritic cells drive T helper 2 cell-mediated immunity. *Immunity* **39**, 733–743 (2013). doi: 10.1016/j.immuni.2013.08.029; pmid: 24076051

19. R. Murakami *et al*., A unique dermal dendritic cell subset that skews the immune response toward Th2. *PLOS ONE* **8**, e73270 (2013). doi: 10.1371/journal.pone.0073270; pmid: 24039898
20. B. Malissen, S. Tamoutounour, S. Henri, The origins and functions of dendritic cells and macrophages in the skin. *Nat. Rev. Immunol.* **14**, 417–428 (2014). doi: 10.1038/nri3683; pmid: 24854591
21. K. Denda-Nagai *et al*., Distribution and function of macrophage galactose-type C-type lectin 2 (MGL2/CD301b): Efficient uptake and presentation of glycosylated antigens by dendritic cells. *J. Biol. Chem.* **285**, 19193–19204 (2010). doi: 10.1074/jbc.M110.113613; pmid: 20304916
22. Y. Kumamoto, T. Hirai, P. W. Wong, D. H. Kaplan, A. Iwasaki, CD301b$^+$ dendritic cells suppress T follicular helper cells and antibody responses to protein antigens. *eLife* **5**, e17979 (2016). doi: 10.7554/eLife.17979; pmid: 27657168
23. M. M. Meredith *et al*., Expression of the zinc finger transcription factor zDC (Zbtb46, Btbd4) defines the classical dendritic cell lineage. *J. Exp. Med.* **209**, 1153–1165 (2012). doi: 10.1084/jem.20112675; pmid: 22615130
24. A. T. Satpathy *et al*., Zbtb46 expression distinguishes classical dendritic cells and their committed progenitors from other immune lineages. *J. Exp. Med.* **209**, 1135–1152 (2012). doi: 10.1084/jem.20120030; pmid: 22615127
25. O. Barreiro *et al*., Pivotal role for skin transendothelial radio-resistant anti-inflammatory macrophages in tissue repair. *eLife* **5**, e15251 (2016). doi: 10.7554/eLife.15251; pmid: 27304075
26. B. T. Edelson *et al*., Peripheral CD103$^+$ dendritic cells form a unified subset developmentally related to CD8α^+ conventional dendritic cells. *J. Exp. Med.* **207**, 823–836 (2010). doi: 10.1084/jem.20091627; pmid: 20351058
27. J. J. Lindenberg *et al*., IL-10 conditioning of human skin affects the distribution of migratory dendritic cell subsets and functional T cell differentiation. *PLOS ONE* **8**, e70237 (2013). doi: 10.1371/journal.pone.0070237; pmid: 23875023
28. K. Denzer, M. J. Kleijmeer, H. F. Heijnen, W. Stoorvogel, H. J. Geuze, Exosome: From internal vesicle of the multivesicular body to intercellular signaling device. *J. Cell Sci.* **113**, 3365–3374 (2000). pmid: 10984428
29. P. Brossart, A. W. Goldrath, E. A. Butz, S. Martin, M. J. Bevan, Virus-mediated delivery of antigenic epitopes into dendritic cells as a means to induce CTL. *J. Immunol.* **158**, 3270–3276 (1997). pmid: 9120283
30. V. Muralidharan-Chari *et al*., ARF6-regulated shedding of tumor cell-derived plasma membrane microvesicles. *Curr. Biol.* **19**, 1875–1885 (2009). doi: 10.1016/j.cub.2009.09.059 pmid: 19896381
31. E. Crompot *et al*., Avoiding false positive antigen detection by flow cytometry on blood cell derived microparticles: The importance of an appropriate negative control. *PLOS ONE* **10**, e0127209 (2015). doi: 10.1371/journal.pone.0127209 pmid: 25978814
32. K. Choudhuri *et al*., Polarized release of T-cell-receptor-enriched microvesicles at the immunological synapse. *Nature* **507**, 118–123 (2014). doi: 10.1038/nature12951 pmid: 24487619
33. K. McDaniel *et al*., Functional role of microvesicles in gastrointestinal malignancies. *Ann. Transl. Med.* **1**, 4 (2013). pmid: 24432300
34. V. Dolo *et al*., Selective localization of matrix metalloproteinase 9, beta1 integrins, and human lymphocyte antigen class I molecules on membrane vesicles shed by 8701-BC breast carcinoma cells. *Cancer Res.* **58**, 4468–4474 (1998). pmid: 9766680
35. S. Lata *et al*., Helical structures of ESCRT-III are disassembled by VPS4. *Science* **321**, 1354–1357 (2008). doi: 10.1126/science.1161070; pmid: 18687924
36. F. A. Ran *et al*., Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013). doi: 10.1038/nprot.2013.143; pmid: 24157548
37. S. Burgdorf, V. Lukacs-Kornek, C. Kurts, The mannose receptor mediates uptake of soluble but not of cell-associated antigen for cross-presentation. *J. Immunol.* **176**, 6770–6776 (2006). doi: 10.4049/jimmunol.176.11.6770; pmid: 16709836
38. E. P. McGreal, J. L. Miller, S. Gordon, Ligand recognition by antigen-presenting cell C-type lectin receptors. *Curr. Opin. Immunol.* **17**, 18–24 (2005). doi: 10.1016/j.coi.2004.12.001; pmid: 15653305
39. C. G. Figdor, Y. van Kooyk, G. J. Adema, C-type lectin receptors on dendritic cells and Langerhans cells. *Nat. Rev. Immunol.* **2**, 77–84 (2002). doi: 10.1038/nri723; pmid: 11910898
40. S. A. Mollah *et al*., Flt3L dependence helps define an uncharacterized subset of murine cutaneous dendritic cells. *J. Invest. Dermatol.* **134**, 1265–1275 (2014). doi: 10.1038/jid.2013.515; pmid: 24288007
41. S. F. Rodrigues, D. N. Granger, Blood cells and endothelial barrier function. *Tissue Barriers* **3**, e978720 (2015). doi: 10.4161/21688370.2014.978720; pmid: 25838983
42. C. Handfield, J. Kwock, A. S. MacLeod, Innate antiviral immunity in the skin. *Trends Immunol.* **39**, 328–340 (2018). doi: 10.1016/j.it.2018.02.003; pmid: 29526487
43. A. Abtin *et al*., Perivascular macrophages mediate neutrophil recruitment during bacterial skin infection. *Nat. Immunol.* **15**, 45–53 (2014). doi: 10.1038/ni.2769; pmid: 24270515
44. L. S. van Rijt *et al*., In vivo depletion of lung CD11c$^+$ dendritic cells during allergen challenge abrogates the characteristic features of asthma. *J. Exp. Med.* **201**, 981–991 (2005). doi: 10.1084/jem.20042311; pmid: 15781587
45. M. Holm *et al*., Seven week culture of functional human mast cells from buffy coat preparations. *J. Immunol. Methods* **336**, 213–221 (2008). doi: 10.1016/j.jim.2008.04.019; pmid: 18547584
46. Zhang Lab, CRISPR design platform, http://tools.genome-engineering.org.

ACKNOWLEDGMENTS

We thank A. Roers for the Mcpt5-Cre mice. B. Hayes is gratefully acknowledged for critical review of this manuscript. We thank Y. Jiao and B. Chen for assistance with two-photon microscopy. We thank R. Vancini and S. Miller for assistance with electron microscopy. We gratefully acknowledge the Duke Cancer Institute and Duke Human Vaccine Institute Flow Cytometry Facility for the use of FACSCanto II and LSR II. **Funding:** This work was funded by U.S. National Institutes of Health grants (U01-AI082107, R01-AI096305, and R56-DK095198). A.S.M. receives funding from U.S. National Institutes of Health grant R21-AI128727, the Duke Physician-Scientist Strong Start Award, the Duke University Medical Center Department of Dermatology, and the Dermatology Foundation. **Author contributions:** H.W.C., J.S., I.H.K., M.H., A.S.M., and S.N.A. designed experiments. H.W.C and J.S. performed the experiments. H.W.C., J.S., I.H.K., H.F.S., A.S.M., and S.N.A. contributed to data analysis. H.W.C., M.H., A.S.M., and S.N.A. wrote the manuscript. All authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare no competing interests associated with this work. **Data and materials availability:** Plasmids and reagents described in this study are available to the scientific community upon request to S.N.A. All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/eaao0666/suppl/DC1
Figs. S1 to S21
Movies S1 to S9

10 June 2017; resubmitted 20 April 2018
Accepted 12 September 2018
10.1126/science.aao0666

RESEARCH ARTICLE SUMMARY

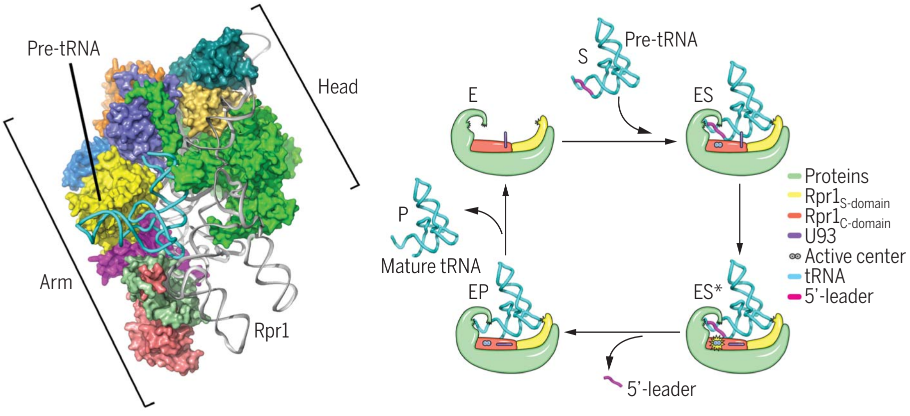
STRUCTURAL BIOLOGY

Structural insight into precursor tRNA processing by yeast ribonuclease P

Pengfei Lan*, Ming Tan*, Yuebin Zhang*, Shuangshuang Niu, Juan Chen, Shaohua Shi, Shuwan Qiu, Xuejuan Wang, Xiangda Peng, Gang Cai, Hong Cheng, Jian Wu†, Guohui Li†, Ming Lei†

INTRODUCTION: Ribonuclease P (RNase P), a universal ribozyme that has been found in organisms from all three domains of life, processes the 5′ end of transfer RNA (tRNA). RNase P is a ribonucleoprotein complex, composed of a single catalytic RNA component and a variable number of proteins. Unlike bacterial RNase P, which contains only one small protein cofactor, archaeal and eukaryotic nuclear RNase Ps have evolved considerably more complex protein subunits: five in archaea and 9 to 10 in eukarya. The pre-tRNA processing reaction can be described by a kinetic mechanism that includes four distinct events: (i) rapid and irreversible binding of RNase P (E) to pre-tRNA (S) to form the initial RNase P-pre-tRNA complex (ES); (ii) a conformational change isomerizing the ES complex to a catalytically competent conformer (ES*) in a magnesium ion (Mg^{2+})–dependent manner; (iii) the cleavage of the phosphodiester bond; and (iv) rapid dissociation of the 5′ leader and slow, rate-limiting release of the mature tRNA (see the figure, right).

RATIONALE: Despite extensive biochemical and genetic studies, however, the role of protein components and the reason for the increased complexity of the protein moieties in eukaryotic nuclear RNase P are still poorly understood. It is still enigmatic how the pre-tRNA substrate, especially the 5′-leader, is recognized by eukaryotic RNase P; how the catalytically important metal ions are coordinated in the active site; and what the chemical mechanism is of pre-tRNA 5′ cleavage. High-resolution structures of eukaryotic RNase Ps are required to answer these key questions.



Catalytic mechanism of pre-tRNA processing catalyzed by yeast RNase P. (**Left**) The overall structure of yeast RNase P holoenzyme in complex with pre-tRNA^{Phe}. The protein hook and the RNAs [the Rpr1 RNA (gray) and the pre-tRNA (cyan)] are in surface and ribbon representations, respectively. (**Right**) Pre-tRNA is cleaved by yeast RNase P by means of a kinetic mechanism that includes four distinct events. First, pre-tRNA is recognized by RNase P through a double-anchor mechanism to form the initial ES complex, which induces a local conformational change in the catalytic center of RNase P. In particular, nucleotide U93 of the Rpr1 RNA undergoes a dramatic conformational change to mediate an inner-sphere coordination of the catalytically important Mg^{2+} ion, so that the ES complex is isomerized to the active ES* state. Next, the activated ES* complex catalyzes the phosphodiester bond cleavage of pre-tRNA through a two-metal-ion SN2 mechanism to release the 5′-leader of pre-tRNA. Last, the mature tRNA dissociates from the holoenzyme in a slow, rate-limiting step, and RNase P is ready for the next round of catalysis.

RESULTS: Here, we report the 3.5-Å cryo–electron microscopy structures of *Saccharomyces cerevisiae* RNase P holoenzyme alone and in complex with pre-tRNA^{Phe}. The yeast RNase P holoenzyme consists of one catalytic RNA Rpr1 and nine protein components. The Rpr1 RNA adopts an extended single-layered conformation that maintains a central helical core but lacks most of the long-range RNA-RNA interactions that are essential for structural stability in bacterial RNase P. The protein components form an interconnected hook-shaped architecture that tightly wraps around the RNA and stabilizes yeast RNase P into a "measuring device," with two fixed anchors that recognize the L-shaped structure rather than specific sequences of pre-tRNA substrates (see the figure, left). This "measuring device" mediates the initial engagement with pre-tRNA to form the low-affinity ES complex. The recognition of the 5′-leader of pre-tRNA involves both the Rpr1 RNA and the protein subunit Pop5. Two catalytically important Mg^{2+} ions are coordinated in the catalytic center by highly conserved uridine U93 and the phosphate backbone of Rpr1, together with the scissile phosphate and the O3′ leaving group of pre-tRNA (see the figure, right). The configuration of this RNA-based catalytic center is universally conserved in all RNase Ps, from bacteria to eukarya. Pre-tRNA binding induces a dramatic conformational change in the catalytic center, corresponding to the isomerization step to the ES* state. Moreover, our simulation analysis visualized the mechanistic details of phosphodiester bond hydrolysis of pre-tRNA, which is a two-Mg^{2+}-ion–mediated S_N2 reaction (see the figure, right).

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat6678

CONCLUSION: The structures presented here represent a major step forward for mechanistic understanding of the function of eukaryotic RNase P. Our data support that all RNase P ribozymes share an RNA-based, substrate-induced catalytic mechanism of pre-RNA processing. Whereas bacterial RNase P RNA is catalytically active by itself, eukaryotic RNase P is a protein-controlled ribozyme; its protein components not only directly participate in substrate recognition but also stabilize the catalytic RNA in a conformation optimal for pre-tRNA binding and cleavage reaction. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.

†Corresponding author. Email: leim@shsmu.edu.cn (M.L.); ghli@dicp.ac.cn (G.L.); wujian@shsmu.edu.cn (J.W.)

Cite this article as P. Lan *et al.*, Science 362, eaat6678 (2018). DOI: 10.1126/science.aat6678

Lan *et al.*, *Science* **362**, 657 (2018) 9 November 2018 **1 of 1**

RESEARCH ARTICLE

STRUCTURAL BIOLOGY

Structural insight into precursor tRNA processing by yeast ribonuclease P

Pengfei Lan[1]*, Ming Tan[2,3]*, Yuebin Zhang[4]*, Shuangshuang Niu[2,3,5], Juan Chen[1], Shaohua Shi[1], Shuwan Qiu[6], Xuejuan Wang[6], Xiangda Peng[4], Gang Cai[6], Hong Cheng[2], Jian Wu[1]†, Guohui Li[4]†, Ming Lei[1,7,8,9]†

Ribonuclease P (RNase P) is a universal ribozyme responsible for processing the 5′-leader of pre–transfer RNA (pre-tRNA). Here, we report the 3.5-angstrom cryo–electron microscopy structures of *Saccharomyces cerevisiae* RNase P alone and in complex with pre-tRNA^{Phe}. The protein components form a hook-shaped architecture that wraps around the RNA and stabilizes RNase P into a "measuring device" with two fixed anchors that recognize the L-shaped pre-tRNA. A universally conserved uridine nucleobase and phosphate backbone in the catalytic center together with the scissile phosphate and the O3′ leaving group of pre-tRNA jointly coordinate two catalytic magnesium ions. Binding of pre-tRNA induces a conformational change in the catalytic center that is required for catalysis. Moreover, simulation analysis suggests a two-metal-ion S_N2 reaction pathway of pre-tRNA cleavage. These results not only reveal the architecture of yeast RNase P but also provide a molecular basis of how the 5′-leader of pre-tRNA is processed by eukaryotic RNase P.

Ribonuclease P (RNase P), one of only two universal ribozymes that have been found in organisms from all three domains of life, is responsible for the maturation of the 5′ end of transfer RNA (tRNA) (*1*). RNase P is ribonucleoprotein complex, composed of a single catalytic RNA component and a variable number of proteins (*2*–*4*). The RNA component of RNase P from all organisms shows marked similarities at the primary and secondary structure level, which is suggestive of the presence of a universally conserved catalytic RNA core (*2*, *4*). All RNase P RNAs can be divided into two independent folded domains, the catalytic domain (C domain) and the specificity domain (S domain), which play key roles in substrate cleavage and substrate binding, respectively (*5*, *6*). Bacterial RNase P contains a single small protein (Rpp) that is essential for substrate recognition and cleavage under physiological conditions (*7*–*9*). By contrast, archaeal and eukaryotic nuclear RNase Ps have evolved considerably more complex subunit compositions with an increased number of protein components, five in archaea and 9 to 10 in eukarya (*2*). Unlike bacterial RNAs, the RNAs themselves in archaeal and eukaryotic RNase Ps are not generally catalytically active, and protein subunits are required to enhance the pre-tRNA substrate binding affinity and cleavage efficiency (*10*–*13*).

Very little is known about the structure of eukaryotic RNase Ps. Only crystal structures of human Pop6-Pop7 subcomplex and *Saccharomyces cerevisiae* Pop6-Pop7 complexed with the P3 element of RNase MRP (mitochondrial RNA processing) RNA are available (*14*, *15*). In addition, a low-resolution structure of *S. cerevisiae* nuclear RNase P has been solved by means of cryo–negative staining electron microscopy (EM) (*16*). A high-resolution structure of eukaryotic nuclear RNase P holoenzyme still has yet to be determined, hindering our understanding of the structural organization and mechanism of action of RNase P from higher organisms.

RNase P is a multiple turnover ribozyme that recognizes its substrates in trans (*17*). Previous studies suggest that the cleavage of pre-tRNA by RNase P can be described by a kinetic mechanism that includes at least four distinct events: (i) rapid and irreversible binding of RNase P (E) to pre-tRNA (S) to form the initial RNase P–pre-tRNA complex (ES); (ii) the ES complex then undergoes a conformational change and is isomerized to a catalytically competent conformer (ES) in a magnesium ion(Mg^{2+})–dependent manner; (iii) the cleavage of the phosphodiester bond; and (iv) rapid dissociation of the 5′-leader and slow, rate-limiting release of the mature tRNA (*18*–*21*). Although this kinetic model was proposed about a decade ago, how RNase P facilitates such a reaction remains largely enigmatic.

The first step of this kinetic scheme represents the formation of the initial low-affinity ES complex. RNase P recognizes the three-dimensional structural feature rather than specific sequences of pre-tRNAs (*1*, *22*, *23*). It has been proposed that RNase P contains a "measuring device" that recognizes a common structural feature of all pre-tRNAs; the coaxially stacked acceptor and T stems have a fixed length of 12 base pairs of nucleotides (*1*). The crystal structure of the bacteria *Thermotoga maritima* RNase P–tRNA complex with a soaked 5′-leader provides the structural basis of a simple, RNA-based "measuring device" and some insight into the binding of pre-tRNA 5′-leader (*23*). However, pre-tRNA recognition by the much more complex and indispensable protein components of eukaryotic RNase Ps is still poorly understood. Kinetic evidences suggested that a conformational change occurs after the formation of the initial ES complex and before 5′-leader cleavage, transforming the ES complex into an active ES* state (*7*, *9*, *18*, *24*, *25*). The ES* complex is stabilized by at least two Mg^{2+} ions during the transition from ES to ES*, and the functional groups and metal ions in the active site are believed to be repositioned to coordinate the substrate for catalysis (*19*, *26*). However, there is still no structural evidence of the existence of this conformational change. The phosphodiester bond cleavage is an S_N2-type transesterification reaction, which uses a two-Mg^{2+}-ion mechanism that is also used by the group I and group II introns as well as the spliceosome (*27*–*30*). Despite extensive studies, the positions of the catalytic Mg^{2+} ions and the composition of the catalytic center in RNase P are still enigmatic because of the lack of structural information of a pre-tRNA substrate–bound RNase P (*24*, *31*–*39*). Consequently, the chemical mechanism of this two-Mg^{2+}-ion–mediated reaction still awaits to be revealed.

Here, we present the 3.5-Å cryo-EM structures of *S. cerevisiae* nuclear RNase P holoenzyme alone and in complex with a pre-tRNA substrate. The structures unveil the arrangement and function of all the subunits within yeast RNase P and provide an integrated model that depicts how

---

[1]Shanghai Institute of Precision Medicine, Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200125, China. [2]State Key Laboratory of Molecular Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences (CAS), Shanghai 200031, China. [3]University of Chinese Academy of Sciences, CAS, Shanghai 200031, China. [4]Laboratory of Molecular Modeling and Design, State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, CAS, Dalian 116023, China. [5]School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China. [6]Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China, Hefei 230027, China. [7]Key laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China. [8]National Facility for Protein Science in Shanghai, Zhangjiang Laboratory, Shanghai, 201210, China. [9]Shanghai Science Research Center, CAS, Shanghai, 201204, China.
*These authors contributed equally to this work.
†Corresponding author. Email: leim@shsmu.edu.cn (M.L.); ghli@dicp.ac.cn (G.L.); wujian@shsmu.edu.cn (J.W.)

the pre-tRNA substrate is recognized and how the hydrolysis of the 5′-leader of pre-tRNA is catalyzed by eukaryotic RNase P.

Overall structure of yeast RNase P

We used a two-step affinity purification scheme to obtain the endogenous *S. cerevisiae* RNase P complex. The highly purified RNase P was analyzed by means of SDS–polyacrylamide gel electrophoresis followed by mass spectrometry confirmation (fig. S1, A and B, and supplementary materials, materials and methods). Yeast RNase P exhibited a robust pre-tRNA cleavage activity in the presence of magnesium ions, indicating the recovery of a fully functional ribozyme (fig. S1C). Single-particle EM analysis of yeast RNase P yielded a well-defined EM density map at an overall resolution of 3.5 Å (figs. S1, D to J, and S2). We combined de novo model building and homologous modeling to generate an atomic structure for the RNase P complex (Fig. 1, A and B; fig. S3; and table S1). The final refined model of yeast RNase P contains all the previously identified components, one catalytic RNA—Rpr1—and nine proteins (Fig. 1B).

The protein components of yeast RNase P form an intimately interconnected hook-shaped architecture, with Pop1, Pop6, and Pop7 being the head and Pop3, Pop4, Pop5, Pop8, Rpp1, and Rpr2 being the arm (Fig. 1C). Structurally, the protein hook can be considered as the assembly of Pop1, Pop6-Pop7 heterodimer, Pop5-Pop8-$(\text{Rpp1})_2$ heterotetramer, and Pop4-Rpr2-Pop3 heterotrimer (Fig. 1C). The Rpr1 RNA adopts an extended and slightly curved single-layered configuration, with the C and S domains packing against each other (Fig. 2, A and B). Three coaxially stacked helical stems, P2-P19 and P4-P1 of the C domain and P8-P9 of the S domain, form the core of the RNA, which is covered by Pop1 on one side and by the Pop5-Pop8-$(\text{Rpp1})_2$ heterotetramer on the other (Fig. 1B). In the head module of the hook, the Pop6-Pop7 heterodimer together with Pop1 encircles the P3 branch of Rpr1 (Fig. 1B). The arm of the hook packs along one side of Rpr1 (Fig. 1B). At the end of the hook, the Pop4-Rpr2-Pop3 heterotrimer functions as a bridge between the C and S domains of Rpr1 (Fig. 1B). Together, the protein hook tightly wraps around the Rpr1 RNA, burying a total of ~14,715 Å^2 surface area between the RNA and the proteins (Fig. 1B).

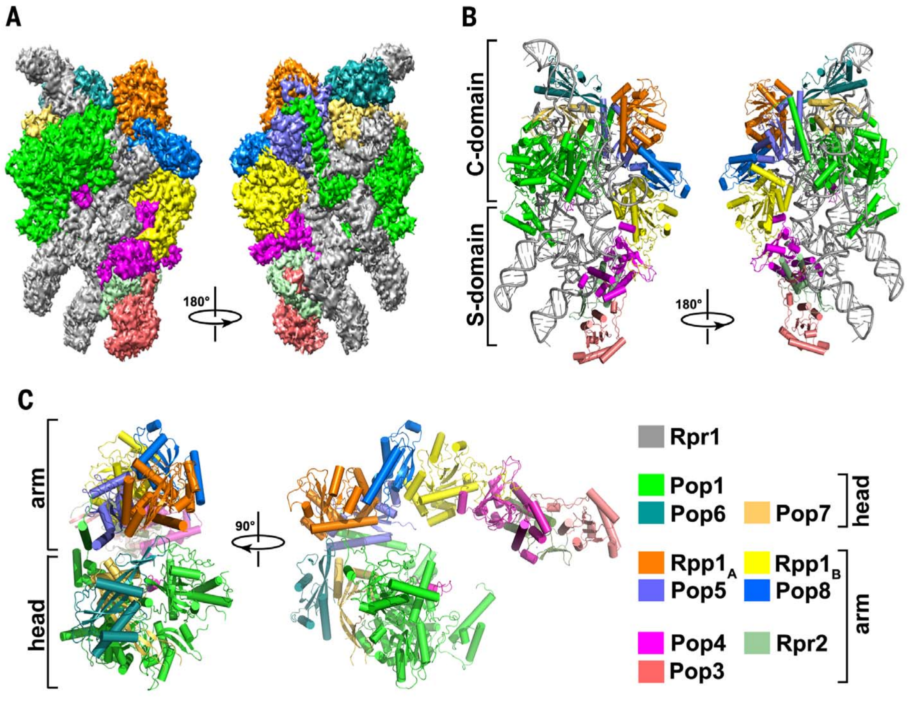


Fig. 1. EM Structure of the *S. cerevisiae* RNase P holoenzyme. (**A**) The EM density map of *S. cerevisiae* RNase P at an average resolution of 3.5 Å. The individual protein and RNA subunits of RNase P are colored according to the scheme shown at the bottom right of the figure. (**B**) Overall structure of the *S. cerevisiae* RNase P complex. (**C**) Two orthogonal views of the overall structure of the protein hook.

Structure of the Rpr1 RNA

In the coaxial helical core of Rpr1, three highly conserved regions—CR-I, CR-IV, and CR-V—fold together into a distinct pseudo-knot structure (Fig. 2, A and C). This pseudo-knot structure brings the P2-P19 stem into close vicinity to stem P4-P1 so that the two long stems pack parallel to each other (Fig. 2A). By contrast, stem P8-P9 in the S domain is not close to stem P4-P1 in primary sequence (Fig. 2B). Both the terminal loops of stems P8 and P9 adopt a tetraloop conformation and respectively interact with the minor grooves of stems P4 and P1 through tetraloop-tetraloop receptor interactions (Fig. 2, A and D). Beside CR-I, CR-IV, and CR-V, there are two additional universally conserved regions CR-II and CR-III, corresponding to two single-stranded junctions J11/12 and J12/11, between stems P10/11 and P12 in the S domain (Fig. 2B). J11/12 and J12/11 fold into two interleaving T-loop motifs that are stabilized by an intricate network of non–Watson-Crick interactions among the conserved nucleotides (Fig. 2E).

The most striking feature of yeast RNA is that six stems (P3, P7, P8′, P10/11, P12, and P15) are loosely connected to the helical core, expanding outward from the RNA center (Fig. 2A). This expanded conformation of Rpr1 yields three big open holes in the RNA structure (Fig. 2A). In sharp contrast, RNAs of bacterial RNase Ps adopt much more compact, two-layered configurations, with auxiliary elements that mediate long-range interactions among different RNA regions to ensure the correct fold and the stability of the RNA (fig. S5, A and B) (*23*, *40*, *41*).

Another difference between bacterial and yeast RNase P RNAs is from the P3 branch. In bacteria RNase P RNA, the P3 branch is a continuous base-paired stem (fig. S4, A and B). By contrast, the P3 branch of yeast Rpr1 contains a large unpaired region that separates the P3 branch into two helical stems (Fig. 2, A and B). Hereafter, we will refer to the proximal and distal helical stems as P3 and P3′, respectively, and the two unpaired single-stranded regions between stems P3 and P3′ as junctions J3/3′ and J3′/3, respectively (Fig. 2, A and B).

Head module of the protein hook and its interaction with Rpr1

The largest protein subunit Pop1 is specific to eukaryotic RNase Ps (Fig. 1, A to C) (*42*). From the N to C termini, Pop1 contains three motifs: an N-terminal motif (NTM), an internal motif (INM), and a large C-terminal globular domain (CTD) (Fig. 3A). A long helix α1 of Pop1_{NTM} sticks out and mediates the interaction with Pop5 (Fig. 3A). The rest of Pop1_{NTM} folds into a small helical and highly basic structure, plugging into the open junction between CR-IV and stems P4, P7, and P15 of Rpr1 (Fig. 3B). The side chains of multiple arginine residues in Pop1_{NTM} make extensive stacking and electrostatic interactions with nucleotides in stems P4, P7 and P15, and CR-IV from four different directions (Fig. 3B and fig. S5A). These arginine residues are highly conserved from yeast to humans (fig. S5B), suggesting that Pop1_{NTM} plays an important role in stabilizing the RNA structure in all eukaryotic RNase Ps. Consistent with this notion, mutation of conserved $\text{Arg}^{97}\text{Arg}^{98}\text{Arg}^{99}$ in yeast Pop1 resulted in severe defect in pre-tRNA processing (*43*). The INM of Pop1 contains three helices and a long loop, fitting into the big open hole among stems P7, P8′, P8, P9, and P11 and stabilizing the conformation of the S domain of Rpr1 (Fig. 3C and fig. S5C). Pop1_{CTD} is located at the concave side of the Rpr1 RNA opposite to Pop1_{NTM} (Fig. 3D).

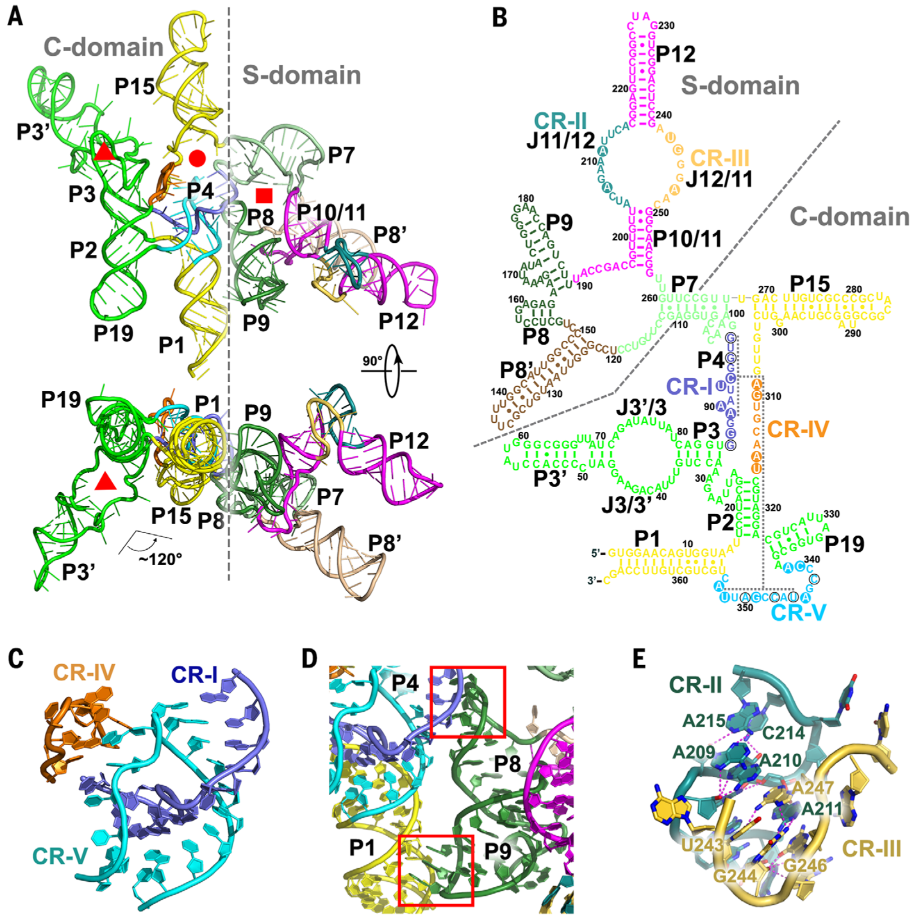


Fig. 2. Structure of the catalytic RNA subunit Rpr1. (**A**) Two orthogonal views of the overall structure of the Rpr1 RNA. The three open holes are highlighted by the red triangle, circle, and rectangle, respectively. (**B**) Secondary structure diagram of the yeast Rpr1 RNA. RNA elements are colored as in (A). The C and S domains are denoted. The canonical Watson-Crick and noncanonical base-pairing interactions are shown as solid lines and dots, respectively. Nucleotides in CRs that are universally conserved in bacterial, archaeal, and eukaryotic RNase P RNAs are highlighted with solid circles, whereas those that are invariant only within eukaryotic RNAs are highlighted with open circles. (**C**) Structure of the universally conserved pseudo-knot formed by CR-I, R-IV, and CR-V. (**D**) The tetraloop-tetraloop receptor interactions between stems P4 and P8 and between P1 and P9. The interacting regions are highlighted in red boxes. (**E**) A close-up cartoon representation of the CR-II/CR-III region in the S domain.

Three conserved basic patches in Pop1$_{CTD}$ mediate interactions with different regions of Rpr1, covering a large surface area of the helical core of Rpr1 and stabilizing the single-layered architecture of the RNA (Fig. 3D and fig. S5D). Consistent with this notion, mutations of key residues on all three patches cause defects in pre-tRNA processing (fig. S5B) (*43*).

The P3 branch of Rpr1 binds to the concaved surface of the saddle-shaped Pop6-Pop7 heterodimer, with junctions J3/3′ and J3′/3 sitting on the basic center of the saddle (fig. S6, A and B). In contrast to Pop6, which only interacts with the P3 branch, Pop7 also contacts stem P15 and is sandwiched in the wedge between stems P3 and P15 of Rpr1 (fig. S6C). Binding with Pop6-Pop7 induces an ~120° bend between the P3 branch and the single-layered core of Rpr1, so that Pop7 and Pop1 form an extensive interface and the P3 branch of Rpr1 is encircled by Pop1, Pop6, and Pop7 (fig. S6D). Taken together, the head module of the protein hook wraps around the C domain and part of the S domain of Rpr1, stabilizing the helical RNA core as well as loosely connected stems P3′ and P15 (Fig. 3E).

Arm module of the protein hook and its interaction with Rpr1

The arm module of the protein hook contains two subcomplexes, the Pop5-Pop8-(Rpp1)$_2$ heterotetramer and the Pop4-Rpr2-Pop3 heterotrimer (Fig. 1C). Pop5, Pop8, and two copies of Rpp1 form a heterotetramer with a pseudo-twofold symmetry (Fig. 3F). The heterotetramer is connected with the head module through two discrete, reciprocal interactions with Pop1 involving Pop5 and one Rpp1 molecule (Fig. 1C). Hereafter, we will refer to the Rpp1 molecule that interacts with Pop1 as Rpp1$_A$ and the other as Rpp1$_B$ (Fig. 1C). Although there is very limited sequence similarity between Pop5 and Pop8, their structures highly resemble each other, with a root mean square deviation of 2.5 Å (fig. S7A). The Pop5-Pop8-(Rpp1)$_2$ heterotetramer sits on one side of the C domain of Rpr1 opposite to Pop1$_{CTD}$ (Figs. 1B and 3G and fig. S7, B and C). The surface of one side of Pop5 is highly basic, forming a deep cleft that tightly holds the zig-zagged CR-IV of Rpr1 (Fig. 3H). The highly conserved basic N terminus of Pop5 embraces CR-IV, with the side chains of Arg^3 and Lys^5 pointing into the narrow space between CR-I, CR-IV, and CR-V (fig. S7D). In addition to Pop5, Rpp1$_B$ also coordinates electrostatic interactions with both CR-I and CR-V (fig. S7B), so that Pop5 and Rpp1$_B$ together stabilize the pseudo-knot structure of Rpr1 by means of a large extended complementary interface (Fig. 3G).

This Pop4-Rpr2-Pop3 heterotrimer is connected to the Pop5-Pop8-(Rpp1)$_2$ heterotetramer through extensive interactions between the β-barrel of Pop4 and the C-terminal long coiled coil of Rpp1$_B$, so that the two subcomplexes together form the rod-shaped arm module of the protein hook (Figs. 1C and 3I and fig. S8A). The β-barrel of Pop4 interacts with the minor groove of stem P1 of Rpr1 through electrostatic interactions (Fig. 3J and fig. S8B). The two ends of long helix α3 of Pop4 respectively pack on the termini of stems P1 and P9, stabilizing the tetraloop-tetraloop receptor interaction between stems P1 and P9 (Fig. 3J and fig. S8C). Rpr2 mainly interacts with the S domain of Rpr1; helices α2 and α3, one edge of the β-sheet, and peripheral loops of Rpr2 constitute a positively charged depression that holds the U-shaped T-loop of J12/11 of Rpr1 (Fig. 3J and fig. S8D). In contrast to Pop4 and Rpr2, Pop3 makes limited direct interactions with the RNA; only the extended β-sheet of Pop3 contacts J11/12 of Rpr1 (Fig. 3J and fig. S8E). Taken together, the Pop4-Rpr2-Pop3 heterotrimer functions as a bridge between the C and S domains of Rpr1, stabilizing the position of J11/12 and J12/11 relative to the helical core of the RNA (Fig. 3K).

The protein hook stabilizes the Rpr1 RNA

To corroborate our structural study, we carried out molecular dynamics (MD) simulations to investigate how the protein components affect the structural stability of the Rpr1 RNA. MD simulation trajectories indicated that Rpr1 alone exhibited a very large fluctuation in the backbone phosphate atoms (Fig. 4A). Association of the protein hook with Rpr1 greatly suppressed this fluctuation except for stems P8′ and P12, which make no direct contact with the proteins (Fig. 4A). Binding with Pop1 alone was able to substantially reduce the fluctuation of Rpr1, especially the P3 branch and stem P15 (Fig. 4A). Other than the structural variation of individual nucleotides, MD simulation also showed that except for stems P8′ and P12, local cross-correlation in Rpr1 was suppressed upon protein hook association (fig. S9), suggesting that binding with proteins reduced the relative movements of individual structural elements within Rpr1 so that the RNA behaves more as a whole in the RNase P complex.

To obtain an overall picture of the atomic motions in Rpr1, we performed principle component analysis (PCA) using Cartesian coordinates of the phosphate atoms of Rpr1. The projection of trajectories onto the first two principal components, PC1 and PC2, accounted for a substantial amount of overall motion of Rpr1 in phase space. Consistent with the cross-correlation analysis, except for stems P8′ and P12, the overall motion of Rpr1 along PC1 and PC2 was greatly constrained by proteins in the RNase P complex (Fig. 4B and Movie 1). Furthermore, Gibbs free-energy landscape (FEL) analysis showed that for Rpr1 in the RNase P complex, a single global energy minimum was observed, indicating that the conformational state of Rpr1 is well restricted by the protein components (Fig. 4C). In sharp contrast, many local minimum-energy basins were observed for Rpr1 alone, which is suggestive of thermodynamically less stable conformations (Fig. 4C). Collectively, MD simulation revealed an obvious stabilizing effect on the Rpr1 RNA by the protein hook.

pre-tRNA recognition

To gain insights into the pre-tRNA recognition and 5′-leader processing mechanism, we determined the cryo-EM structure of yeast RNase P in complex with a yeast pre-tRNA^{Phe} at a resolution of 3.5 Å (figs. S10, A to H, and S11, A to E, and table S1). In this pre-tRNA substrate bound structure, the pre-tRNA substrate sits in a large open pocket on one side of the holoenzyme (Fig. 5, A to C). The coaxially stacked acceptor and T stems of the pre-tRNA substrate make extensive intermolecular interactions with both the Rpr1 RNA and protein components, positioning the scissile phosphate of pre-tRNA right at the active site of the ribozyme (Fig. 6A and fig. S11, A to C). In addition, one side of the anticodon arm of pre-tRNA also mediates extensive contacts with Rpp1$_B$ through complementary interface, further stabilizing the pre-tRNA in the binding pocket (Fig. 5, A to C).

The cryo-EM structure of yeast RNase P–pre-tRNA complex provided an atomic model of the "measuring device" of eukaryotic nuclear RNase Ps. The "measuring device" for pre-tRNA recognition is mainly composed of two anchors in RNase P (Fig. 6A). At the first anchor site, two unpaired nucleotides—A204 and G245, from the conserved T-loops in the S domain of Rpr1—form π-π stacking interactions with two unstacked bases, C56 from the TψC loop and G19 from the D loop of the pre-tRNA, respectively (Fig. 6B and fig. S11D).

At the other end of the acceptor stem, where the scissile phosphate resides, the highly conserved N-terminal motif of Pop1 (Pop1$_{NTM}$) functions as the second anchor to stabilize the pre-tRNA at the catalytic center in the C domain of RNase P (Fig. 6C). The N-terminal long helix α1 in Pop1$_{NTM}$ that packs with Pop5 sits right on the acceptor stem of pre-tRNA so that the loop between helices α1 and η1 in Pop1$_{NTM}$ (L$_{1\text{-}1}$) packs on the G1-C72 base pair of pre-tRNA, forcing the 3′-tailor of pre-tRNA to adopt a sharp turn at A73 and fold back to the 5′-leader (Fig. 6C). The base of pre-tRNA A73 flips out away from the acceptor stem and is sandwiched between the base of Rpr1 G300 and the aliphatic side chain of Arg^{115} in Pop1$_{NTM}$ (Fig. 6D). The helical core of Pop1$_{NTM}$ also mediates extensive hydrogen-bonding and electrostatic interactions with the first three base pairs of the pre-tRNA (Fig. 6E).

MD simulation analysis revealed a broad distribution of the distance between the two anchor points of Rpr1, which is suggestive of a very dynamic behavior of the RNA (Fig. 6F). In the presence of the protein hook, this distance

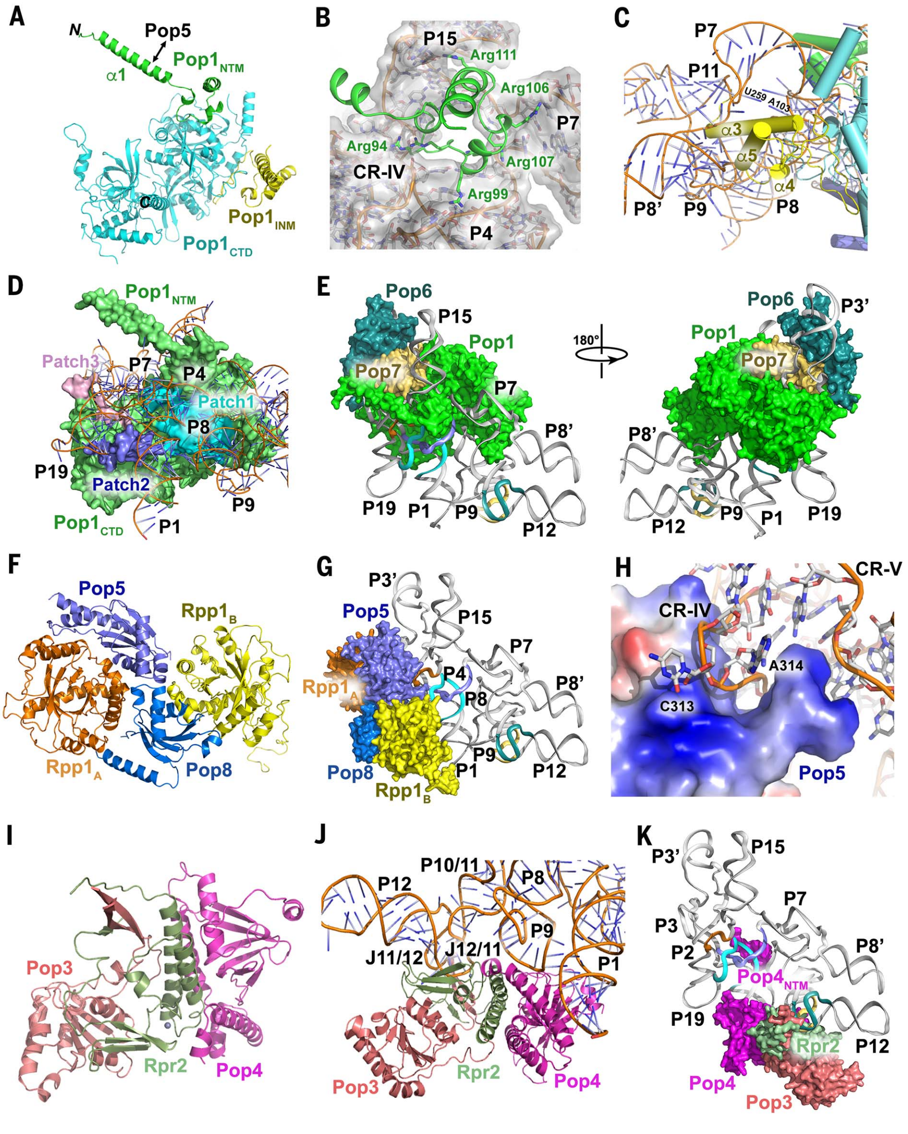


Fig. 3. The protein hook and its interaction with the Rpr1 RNA. (**A**) Overall structure of Pop1. (**B**) An overall view of the interaction between Pop1$_{NTM}$ and the Rpr1 RNA. The RNA is shown as cartoon with semitransparent surface. CR-IV and stems P4, P7, and P15 that forms the big open hole are denoted. (**C**) An overall view of Pop1$_{INM}$ (colored in yellow) that fits into the big open hole among stems P7, P8′, P8, P9, and P11. (**D**) Interactions between Pop1$_{CTD}$ and Rpr1. (**E**) The head module of the protein hook wraps around the C domain and part of the S domain of Rpr1, stabilizing stems P3′ and P15. (**F**) Overall structure of the Pop5-Pop8-(Rpp1)$_2$ heterotetramer. (**G**) The Pop5-Pop8-(Rpp1)$_2$ heterotetramer sits on one side of the C domain of Rpr1, mediating extensive interactions with the pseudo-knot. (**H**) Analysis of the electrostatic surface potential of Pop5 reveals a highly basic deep cleft that holds CR-IV tightly (red, negative; blue, positive). (**I**) Overall structure of the Pop4-Rpr2-Pop3 heterotrimer. (**J**) Overall view of the interaction between the Pop4-Rpr2-Pop3 heterotrimer and the Rpr1 RNA. (**K**) The Pop4-Rpr2-Pop3 heterotrimer functions as a bridge between the C and S domains of Rpr1.

is tightly constrained to ~51 Å and is optimal to accommodate the coaxially stacked acceptor stem and T-stem of pre-tRNAs, which have a fixed length of ~45 Å from the cleavage site to the TψC loop (Fig. 6G). The Pop4-Rpr2-Pop3 heterotrimer by itself already had a strong stabilizing effect (Fig. 6F), which is consistent with this trimer functioning as a bridge between the C and S domains, stabilizing the position of the T-loops relative to the catalytic center (Fig. 5, B and C). Taken together, our structural and MD simulation results suggested that the protein hook stabilizes yeast RNase P into a "measuring device," with two fixed anchors that recognize the L-shaped structure rather than specific sequences of pre-tRNA substrates (*1*).

5′-leader recognition

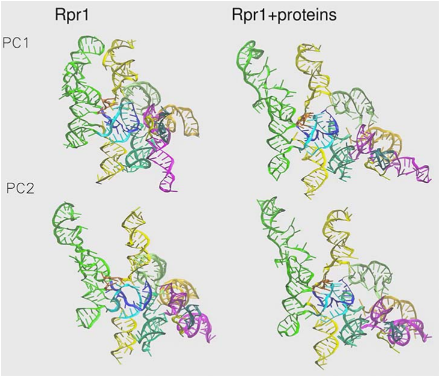
In the complex structure, at least four nucleotides in the 5′-leader of pre-tRNA were visible in the cryo-EM density map (fig. S12A). The recognition of the 5′-leader involves both the Rpr1 RNA and protein subunits. Pop5 and Rpp1$_B$ form a continuous basic surface that holds the 5′-leader (Fig. 6H). The 5′ terminal nucleotide A(–4) of pre-tRNA sticks into a hydrophobic pocket between Pop5 and Rpp1$_B$, with its base stacking with the side chain of Phe^{140} of Rpp1$_B$ (Fig. 6I). Nucleotides at the –1 and –2 positions of pre-tRNA—A(–2) and A(–1), respectively—sequentially stack on G1 following the double-stranded trajectory of the acceptor stem (Fig. 6I). This continuous base-stacking is extended into nucleotide A314 in the zig-zagged turn of CR-IV in Rpr1 (Fig. 6I). Nucleotide G(–3) of the 5′-leader diverges from this base stack and packs on C313 of Rpr1 (Fig. 6I). In addition to these stacking interactions, G(–3) and A(–2) also make multiple hydrogen-bonding interactions with C313 and A314 of Rpr1 (fig. S12B), further strengthening the connection between the 5′-leader of pre-tRNA and CR-IV of Rpr1. The zig-zagged turn of CR-IV is tightly grabbed by a highly basic cleft of Pop5, emphasizing that one of the major functions of Pop5 is to stabilize CR-IV of Rpr1 for 5′-leader recognition. In addition to this indirect recognition of the 5′-leader via CR-IV of Rpr1, Pop5 also directly contacts the backbone of nucleotides A(–1) and A(–2) of pre-tRNA through hydrogen-bonding interactions (Fig. 6I).

Active site

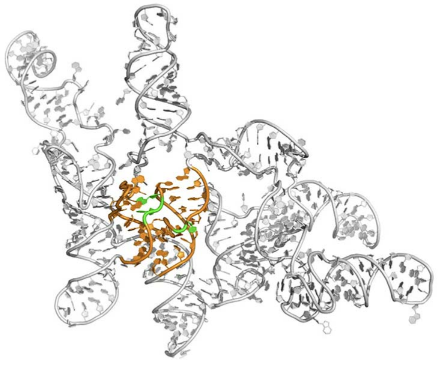
The location of the catalytic center is inferred from the scissile phosphate of the pre-tRNA substrate, which resides right on the P4 stem of the Rpr1 RNA (Fig. 7A). The distance between the scissile phosphate and the closest protein subunit Pop5 is ~7 Å, arguing against protein components playing any direct role in the catalysis (fig. S13A). By contrast, the scissile phosphate of pre-tRNA is closely surrounded by a panel of Rpr1 nucleotides: A91, U92, and U93 from CR-I and G343 and A344 from CR-V (Fig. 7A). Equivalent nucleotides of A91 and U92 in *Bacillus subtilis* and *Escherichia coli* RNase P RNAs have been implicated in metal ion coordination and catalysis (*24*, *35*, *36*, *39*, *44*, *45*).

A putative Mg^{2+} ion (M1) was identified in the vicinity of U92 and U93 in the EM density map (fig. S13B). This site coincides with a metal site observed in the crystal structures of *T. maritima* and *Bacillus stearothermophilus* RNase Ps (*23*, *41*). The M1 Mg^{2+} ion is coordinated by inner-sphere contacts with the hydroxyl of the conserved U93 in Rpr1 and three nonbridging phosphoryl oxygens, *pro*-S_p of Rpr1 A91, *pro*-R_p of Rpr1 U92, and *pro*-R_p of the scissile phosphate of pre-tRNA (Fig. 7A). Several lines of evidence from previous nucleotide substitution studies of bacterial RNase P support this geometry of the M1 Mg^{2+} ion and its environment. First, phosphorothioate-rescue experiments show that the *pro*-R_p but not *Pro*-S_p oxygen of G50 in *B. subtilis* RNase P RNA (equivalent to yeast Rpr1 U92) binds one metal ion (*36*). Second, the 4-thiouridine (4SU) modification of U51 of *B. subtilis* RNase P RNA (equivalent to yeast Rpr1 U93) greatly decreases phosphodiester bond cleavage of pre-tRNA (*24*).

A hallmark of the two–metal ion catalysis is that the two metal ions maintain a distance of ~4 Å and are located roughly in line with the phosphor-sugar backbone on the opposite sides of the scissile phosphate through interactions with the same nonbridging phosphoryl oxygen (*27*, *46*). Although the local cryo-EM density map did not unambiguously reveal the position of the second metal ion, we could model a putative Mg^{2+} (M2) ion that is within the density map and is in accordance with the aforementioned constraints (fig. S13B). The M2 Mg^{2+} ion is coordinated by nonbridging phosphoryl oxygens of A91, G343, and A344 of Rpr1 and the O3′ leaving group of A(–1) in the pre-tRNA substrate (Fig. 7A). Consistently, previous biochemical studies have implicated major constituents of the M2 site in *B. subtilis* RNase P RNA, *pro*-S_p oxygen of A50, and *pro*-R_p of A390 (equivalent to yeast Rpr1 U92 and G344, respectively), in direct inner-sphere coordination of a Mg^{2+} ion important for RNase P activity (*32*, *36*, *38*, *39*). Taken together, the structure of the catalytic center in yeast RNase P is in perfect agreement with previous biochemical data of bacterial RNase P, suggesting that the configuration of the RNA-based catalytic center is universally conserved in all RNase Ps, from bacteria to eukarya.



Movie 1. The first two essential modes of Rpr1 obtained from PCA using MD simulation trajectories. (Left) Absent of the protein components. (Right) Present of the protein components.



Movie 2. Local conformational change in the active center of RNase P upon pre-tRNA binding. The morphed trajectory was generated through linearly interpolating the conformations of the Rpr1 RNA between the apo and pre-tRNA–bound states of yeast RNase P based on the cryo-EM structures. Although the intermediates are unphysical, the movie clearly demonstrates the major conformational differences of Rpr1 upon pre-tRNA binding.

Substrate-induced activation of RNase P

Previous kinetic studies suggested that a conformational change may occur after the formation of the initial RNase P–pre-tRNA complex and before the cleavage step to form the product (*18*, *19*, *25*, *47*, *48*). To examine this conformational change, we superimposed the aporibozyme (apo) and pre-tRNA–bound yeast RNase P structures and found a dramatic conformational change in the active center of Rpr1, although the overall structure of the RNA remains largely unchanged (Fig. 7, B and C; fig. S14, A and B; and Movie 2). In the presence of pre-tRNA, the N-terminal tail of Pop5 rotates toward the catalytic center to form two hydrogen bonds with nucleotide A(–2) in pre-tRNA, helping stabilize the 5′-leader in the active site (Fig. 7B and fig. S14, A and B). Consequently, the conformation of nucleotides G343 and A344 in Rpr1 is adjusted to be optimal for coordinating a catalytic Mg^{2+} ion (M2) (Fig. 7B and fig. S14, A and B). The most prominent conformational change is from nucleotide U93 of Rpr1. Upon binding of pre-tRNA, U93 is forced to rotate a large angle from a position outside of stem P4 to point into the catalytic center and coordinate the other catalytic Mg^{2+} ion (M1) (Fig. 7, B and C, and fig. S14, A and B). The position of U93 in the active center is further stabilized by the C94-G349 base pair in stem P4 of Rpr1 through hydrogen-bonding interactions (Fig. 7D). MD simulation analysis revealed that in the apo-ribozyme, U93 bulges out of the P4 stem of Rpr1 and exhibits a dynamic conformation (fig. S15). In sharp contrast, binding of pre-tRNA tightly constrains U93 in a fix orientation that precisely matches the conformation observed in the complex structure (fig. S15). Taken together, comparative analysis of the apo and pre-tRNA–bound structures of yeast RNase P revealed a Mg^{2+} ion–dependent substrate induced conformational change in the catalytic center, which transforms the ribozyme from an inactive into an active state.

Catalytic mechanism

On the basis of the structure of the active site, we proposed a catalytic mechanism of pre-tRNA cleavage by RNase P. In this model, the M1 Mg^{2+} ion coordinates the *pro-S*$_p$ oxygen of the scissile phosphate and an attacking nucleophilic water molecule (Fig. 8A). The catalytic role for the M1 Mg^{2+} ion is to lower the pK_a value of the nucleophilic water (where K_a is the acid dissociation

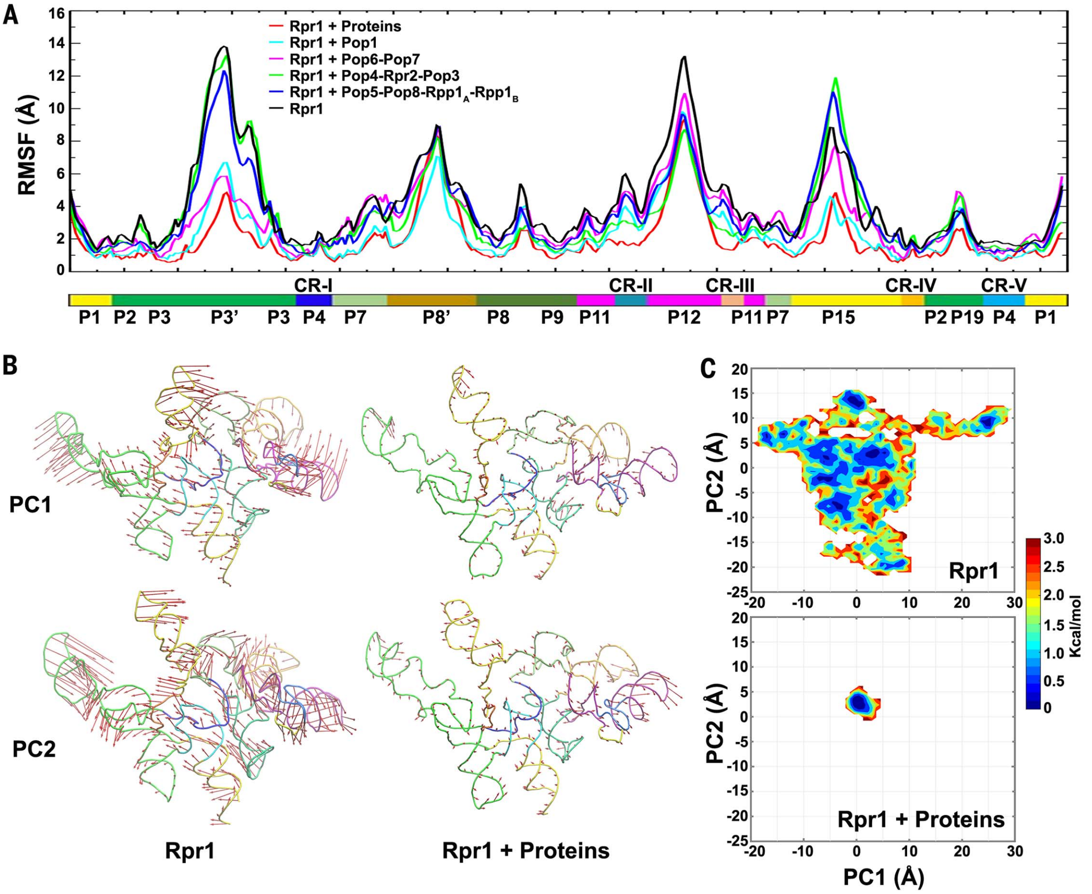


Fig. 4. The protein hook stabilizes the Rpr1 RNA. (**A**) The root mean square fluctuation (RMSF) of backbone phosphorous atoms of the Rpr1 RNA alone (black), in the presence of all protein components (red), or only Pop1 (cyan), the Pop6-Pop7 heterodimer (magenta), the Pop5-Pop8-$(Rpp1)_2$ heterotetramer (blue), or the Pop4-Rpr2-Pop3 heterotrimer (green). (**B**) Porcupine plots of the first two eigenvectors generated by means of PCA. The vectors represented as red arrows illustrate the tendency of the movement of the Rpr1 RNA in the absence (left) or in the presence (right) of the protein components. (**C**) Free-energy landscape calculated by projecting the conformational space onto the two principal components (PC1 and PC2) in the absence (top) or in the presence (bottom) of the protein components.

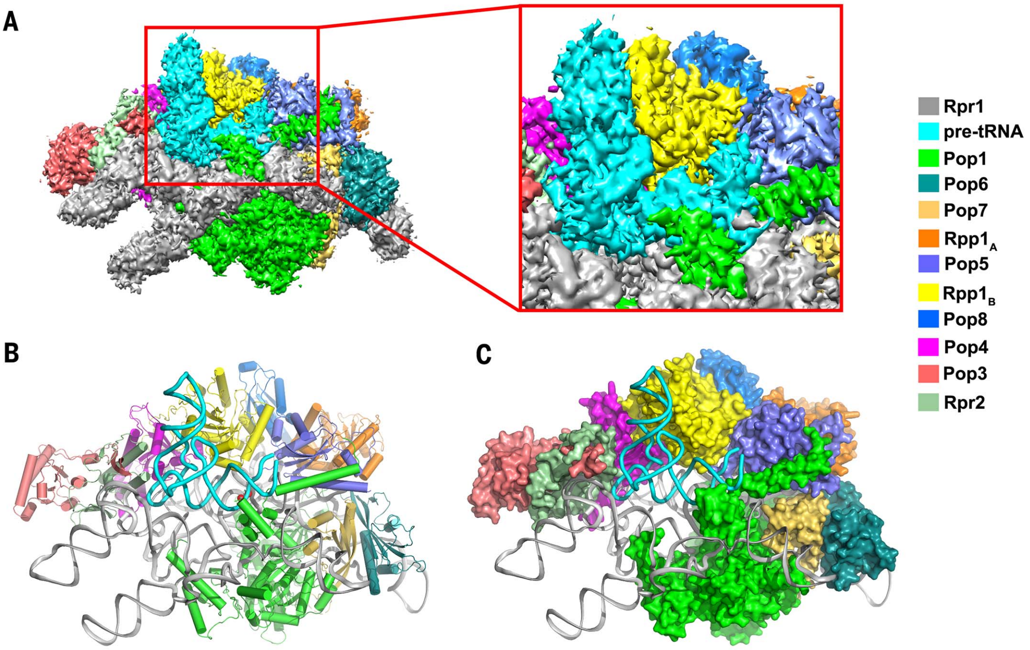


Fig. 5. Overall structure of the *S. cerevisiae* RNase P holoenzyme complexed with pre-tRNA^{Phe}. (**A**) Overall view of the EM density map of *S. cerevisiae* RNase P complexed with pre-tRNA^{Phe}. The individual protein and RNA subunits of RNase P and the pre-tRNA substrate are colored according to the scheme shown at the right of the figure. A close-up view of the active center bound with the pre-tRNA substrate is shown on the right. (**B** to **C**) Overall structure of the *S. cerevisiae* RNase P–pre-tRNA complex. The protein components are in cartoon (B) or surface (C) representation. The RNA component and pre-tRNA are shown in cartoon.

constant and pK_a = –log10K_a), thus facilitating the deprotonation of the water molecule to generate a hydroxide ion that performs an S_N2 in-line attack on the phosphodiester bond with a trigonal bipyramidal transition state (Fig. 8A). The M2 Mg^{2+} ion coordinates the same *pro*-S_p oxygen of the scissile phosphate as well as the upstream ribose O3′ oxygen atom, playing a stabilizing role in maintaining the geometry of the transition state and the leaving 5′-leader (Fig. 8A).

We used multiscale quantum mechanical/molecular mechanical (QM/MM) free-energy simulations to unveil the underlying catalytic mechanism of yeast RNase P. First, we performed 250-ns classical MD simulations to evaluate the structural arrangement of the catalytic center, where a putative nucleophilic water molecule is proposed to coordinate the M1 Mg^{2+} ion (fig. S16). The key distance profiles (d1 to d7) of the active site obtained from classical MD simulations are quantitatively consistent with the structural architecture of the catalytic site observed in the cryo-EM structure (fig. S16).

Next, we carried out QM/MM simulations, combining umbrella samplings and path collective variables (CVs), to obtain the free-energy profile of the phosphodiester cleavage reaction of pre-tRNA. A path connecting the reactant and product in a seven-CV space (distance d1 to d7) was used to explicitly define the progress of the reaction. The free-energy profile along the converged minimum free-energy path (MFEP) revealed a metastable intermediate (INT) region separated by two transition states (TS1 and TS2) (Fig. 8B). After crossing the TS1 by overcoming a free-energy barrier of ~18.56 kcal/mol, the free-energy profile exhibits an intermediate region, corresponding to the inversion of stereo configuration of the pentacovalent phosphorane intermediate (Fig. 8B). Then, the reaction proceeds to cross the rate-limiting step TS2 with a free-energy barrier of ~18.75 kcal/mol (Fig. 8B). The calculated free-energy barrier is in excellent agreement with the experimentally measured rate constants of 0.35 ± 0.03 s^{-1}, corresponding to a Gibbs free energy change ($\Delta G^{\ddagger}$) value of ~18.19 kcal/mol (*25*).

On the basis of the simulations, we depicted the details of the catalytic mechanism. (i) The early stage of the MFEP starts with a decrease of the distance between the scissile phosphate and the nucleophile water (d2), indicating the beginning of the nucleophilic water to attack the scissile phosphate (Fig. 8, B and C). The distance between the scissile phosphate and the leaving O3′ atom (d1) exhibits a perceptible increase after d2 is less than 2.7 Å (path node = 3) (Fig. 8B). The distance between the M2 Mg^{2+} and the leaving O3′ group (d7) also decreases substantially when the nucleophilic water approaches the scissile phosphate before the TS is reached, suggesting stabilization of the leaving O3′ by the

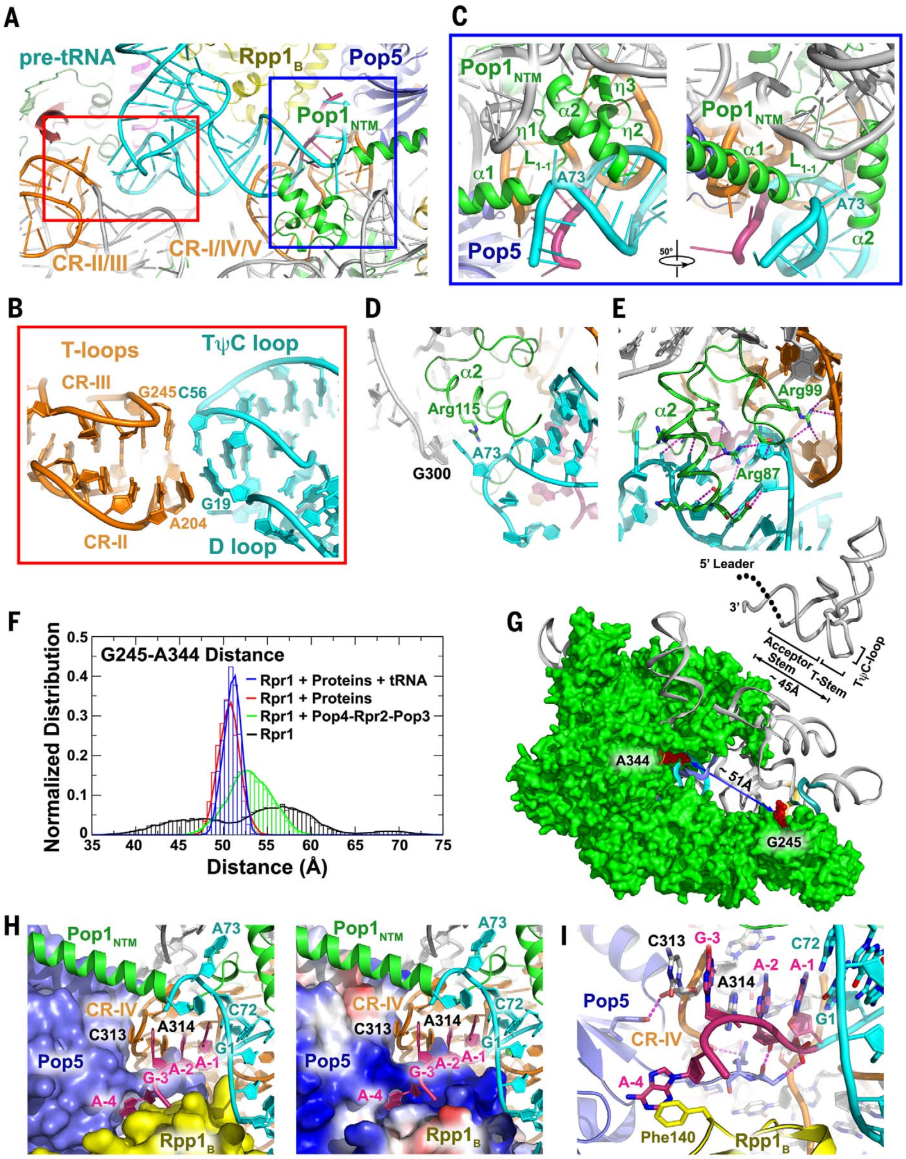


Fig. 6. The pre-tRNA^{Phe} substrate recognition. (**A**) Overview of the interactions between the yeast RNase P holoenzyme and the pre-tRNA substrate. Interactions at the two anchor sites are highlighted in red and blue boxes, respectively. (**B**) A close-up view of the stacking interactions between the TψC and D-loops of pre-tRNA and the T-loops in CR-II and CR-III of the Rpr1 RNA. (**C**) Two related close-up views of the interactions between Pop1$_{NTM}$ and pre-tRNA. The pre-tRNA, Pop1$_{NTM}$, and the pseudoknot of Rpr1 were colored in cyan, green, and orange, respectively. (**D**) A close-up view of the interaction between the 3′-tailor of pre-tRNA and Pop1$_{NTM}$. (**E**) A close-up view of the interactions between Pop1$_{NTM}$ and the first three base pairs in the acceptor stem of pre-tRNA. (**F**) Distribution of the distance between the backbone phosphorus atom of A344 and the C1′ atom of G245 of Rpr1, sampled from a total of ~1.8 μs integrated accelerated molecular dynamics simulation trajectories of each system. (**G**) Overall view of the distance between the conserved A344 and G245 located in C and S domains of Rpr1, respectively. A pre-tRNA in ribbon representation is also shown. The distance between the cleavage site and the TΨC-loop is denoted. (**H**) Recognition of the 5′-leader of pre-tRNA by yeast RNase P. The protein subunits Pop5 and Rpp1$_B$ are shown in surface (left) or electrostatic surface (right) representation, respectively. Four nucleotides in the 5′-leader were colored in dark pink and shown in cartoon. The rest of pre-tRNA and the conserved pseudoknot of Rpr1 were shown in cartoon and colored in cyan and orange, respectively. (**I**) A close-up view of the interactions between the 5′-leader and RNase P. The individual components of RNase P are colored as in (H).

M2 Mg^{2+} ion. (ii) At the TS1 point, where d1 equals d2, the free-energy profile reaches the first maximum value, and the scissile phosphate evolves into a pentavalent phosphorane-like geometry (Fig. 8, B and C). Simultaneously, the nucleophilic water's proton is released, and a hydroxyl group is formed after the TS1 (Fig. 8, B and C). We found that a bulk water molecule, stabilized by the nonbridging oxygen atoms of G1 in pre-tRNA and U93 in Rpr1, serves as a general base to accept the nucleophile's proton (Fig. 8, A to C). A transient hydronium ion (H_3O^+) is identified to facilitate the proton transfer to the *pro*-S_p oxygen of the scissile phosphate, forming a mono-anionic phosphorane intermediate (Fig. 8, A to C). (iii) The product state is accompanied by the formation of the new phosphoryl bond between the nucleophile and the scissile phosphate (d2 ≈ 1.64 Å) and the departure of the leaving O3′ group (d1 ≈ 3.00 Å). (iv) During the late stage of the MFEP, the free-energy profile exhibits a downhill process, and the stereo configuration at the phosphorus is inverted, indicating a typical in-line SN_2 nucleophilic substitution reaction mechanism (Fig. 8, B and C). We observed that the proton that is transferred onto the *pro*-S_p oxygen of the scissile phosphate at TS2 is spontaneously shuttled toward the leaving O3′ group after the inversion of the configuration (Fig. 8, B and C).

Discussion

The catalytic mechanism of how RNase P facilitates the cleavage of pre-tRNA is a central question in the RNase P field. Our structural and simulation data reported here provide an integrated mechanistic insight into the RNase P–catalyzed pre-tRNA processing (Fig. 8D). Our data reveal that the RNase P holoenzyme is largely preassembled to engage the pre-tRNA substrate through a two-anchor mechanism (Fig. 6A). One of the anchors, the T-loops in the S domain of Rpr1, that recognizes the TψC and D loops of pre-tRNA is universally conserved, whereas the other that recognizes the pre-tRNA cleavage site has evolved from an RNA-based apparatus in bacteria to a protein-based one in eukaryotic RNase Ps (Fig. 6C) (*23*). The distance between the two anchors is optimal for accommodating the coaxially stacked acceptor and T stems of pre-tRNA substrates (Fig. 6F), suggesting that these anchor sites function as the "measuring device" to recognize the pre-tRNA substrate. We propose that it is this "measuring device" of all RNase P complexes that mediates the initial engagement with pre-tRNA to form the low-affinity ES complex (Fig. 8D).

The pre-tRNA–bound RNase P structure reveals that the recognition of the 5′-leader of pre-tRNA involves both Rpr1 and the protein subunit Pop5. A highly basic cleft on the surface of Pop5 tightly holds a zig-zagged turn in CR-IV of Rpr1 that mediates both stacking and hydrogen-bonding interactions with the 5′-leader (Fig. 6, H and I, and fig. S12B). Bacterial sole protein subunit Rpp also contains a similar highly basic cleft that holds CR-IV of the bacterial RNA in the same fashion (fig. S17) (*23*). Therefore, it is likely that the 5′-leader of pre-tRNA is recognized by bacterial RNase P through the same mechanism as in yeast RNase Ps. Unlike the 5′-leader, the 3′-tailor of pre-tRNA only makes a limited contribution to the recognition by yeast RNase P (Fig. 6D), which is consistent with the fact that the 3′-CCA sequence in eukaryotic tRNA is added posttranscriptionally (*49*). This is in sharp contrast to bacterial tRNA, in which the 3′-CCA is transcriptionally encoded and serves as a key element recognized by bacterial RNase P through base-pairing with nucleotides in the L_{15} loop of RNase P RNA (*23*).

A comparison of the apo and pre-tRNA–bound structures of yeast RNase P indicates that after the initial recognition of pre-tRNA by the "measuring device" of RNase P, the pre-tRNA substrate induces a local conformation change in the catalytic center, so that the phosphate oxygen–rich cleft near the P4 stem together with the scissile phosphate and the O3′ leaving group of pre-tRNA coordinate two catalytic Mg^{2+} ions to form the active state that is primed for catalysis (Fig. 7, A and B). In particular, universally conserved U93 undergoes a dramatic conformational change to mediate an inner-sphere coordination with the catalytically important M1 Mg^{2+} ion, helping to properly position the local pre-tRNA structure at the active site (Fig. 7B). Consistent with our data, previous studies suggested that nucleotide U51 in *B. subtilis* RNase P (equivalent to U93 in yeast Rpr1) contributes to a divalent metal ion–dependent conformational change during the RNase P–catalyzed reaction (*18*, *24*). We proposed that this pre-tRNA–induced Mg^{2+} ion–dependent conformational change in the catalytic center of RNase P corresponds to the conformational change between the initial ES complex and the cleavage step. Consequently, the active site observed in the pre-tRNA–bound complex structure could represent the active ES* state or a transition configuration between ES and ES* states when the M2 ion is still not fully stabilized in the catalytic center (Fig. 8D) (*18*–*20*). We further speculated that the conformation of the initial ES complex should be largely identical to that of the ES complex except for the catalytic center, which instead adopts the conformation of the active site in the apo structure of RNase P (Fig. 8D).

Our EM structure of the pre-tRNA–bound RNase P complex revealed the configuration of the catalytic center of RNase P in the presence of a pre-tRNA substrate (Fig. 7A). On the basis of this structural information, our QM/MM simulations visualized the mechanistic details of phosphodiester bond hydrolysis of pre-tRNA, which is a two-Mg^{2+}-ion–mediated S_N2 reaction (Fig. 8C). Our simulations also reveal a two-step proton transfer pathway from the nucleophilic water to the leaving O3′ group that involves a neighboring bulk water molecule during the cleavage process (Fig. 8, B and C). A similar

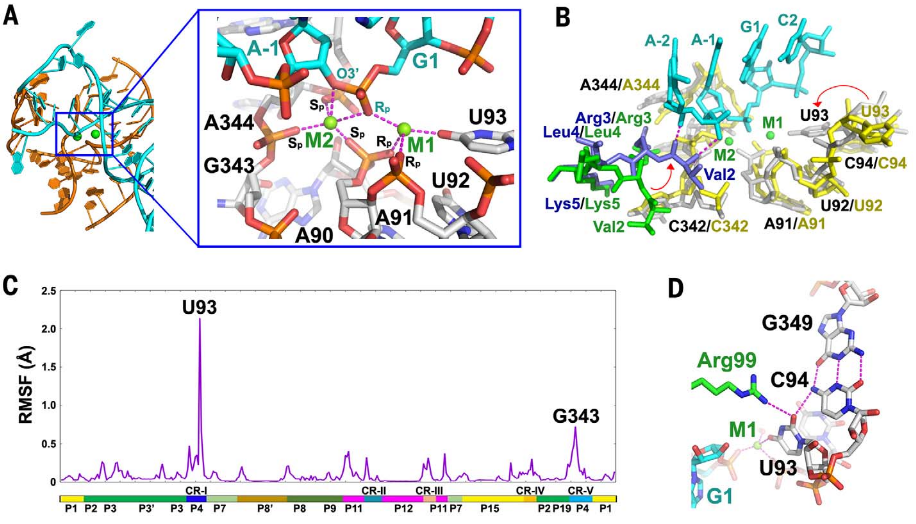


Fig. 7. Configuration of the active site and the conformational change of the catalytic center induced through pre-tRNA binding. (**A**) (Left) An overall illustration of the catalytic center of the yeast RNase P. The Rpr1 RNA and pre-tRNA are colored in orange and cyan, respectively. Mg^{2+} ions (M1 and M2) are shown as green spheres. (Right) A close-up view of the catalytic center. The Rpr1 RNA and pre-tRNA are colored in white and cyan, respectively. The coordination of the two Mg^{2+} ions are highlighted in magenta dashed lines. (**B**) Comparison of the active center between the apo and pre-tRNA–bounded states of yeast RNase P. Pop5, pre-tRNA, and Rpr1 in the pre-tRNA–bound state are colored in slate, cyan, and silver, respectively, and Pop5 and Rpr1 in apo RNase P are colored in green and yellow, respectively. Conformational changes of Rpr1 U93 and the N terminus of Pop5 are denoted. (**C**) The RMSF of Rpr1 nucleotides along the morphed trajectory, which is generated through linearly interpolating the conformations between the apo and pre-tRNA–bound states of Rpr1 based on the cryo-EM structures. (**D**) A close-up view of the detailed interactions among the M1 Mg^{2+} ion, Arg^{99} of $Pop1_{NTM}$, and the C94-G349 base-pair of Rpr1.

bulk water that facilitates the proton transfer process has also been suggested in the cleavage reaction catalyzed by group II introns (*50*).

The M2 site in our structure is different from the second Mg^{2+} (*Tm*M2) site identified in the crystal structure of *T. maritima* RNase P in complex with a mature tRNA (fig. S18) (*23*). The *Tm*M2 site was located in experiments in which the 5′-leader was soaked into the crystals in the presence of Sm_3 (*23*). In contrast to the M1 and M2 sites in our structure, the geometry of the two Mg^{2+} ions in the *T. maritima* RNase P–tRNA complex is not consistent with the octahedral Mg^{2+} coordination geometry (*27*, *51*). In addition, soaking a 5′-leader in a mature tRNA-bound

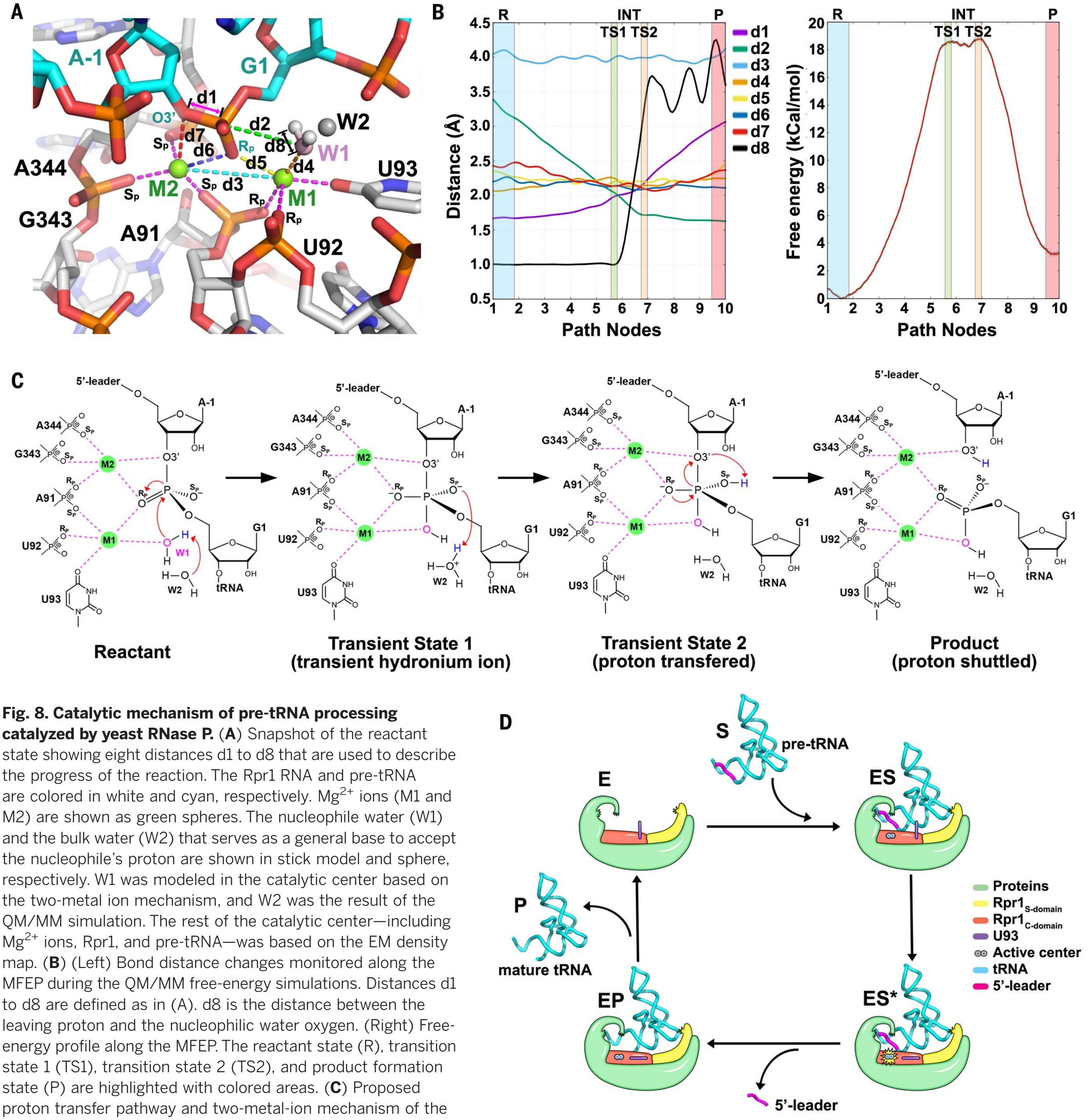


Fig. 8. Catalytic mechanism of pre-tRNA processing catalyzed by yeast RNase P. (**A**) Snapshot of the reactant state showing eight distances d1 to d8 that are used to describe the progress of the reaction. The Rpr1 RNA and pre-tRNA are colored in white and cyan, respectively. Mg^{2+} ions (M1 and M2) are shown as green spheres. The nucleophile water (W1) and the bulk water (W2) that serves as a general base to accept the nucleophile's proton are shown in stick model and sphere, respectively. W1 was modeled in the catalytic center based on the two-metal ion mechanism, and W2 was the result of the QM/MM simulation. The rest of the catalytic center—including Mg^{2+} ions, Rpr1, and pre-tRNA—was based on the EM density map. (**B**) (Left) Bond distance changes monitored along the MFEP during the QM/MM free-energy simulations. Distances d1 to d8 are defined as in (A). d8 is the distance between the leaving proton and the nucleophilic water oxygen. (Right) Free-energy profile along the MFEP. The reactant state (R), transition state 1 (TS1), transition state 2 (TS2), and product formation state (P) are highlighted with colored areas. (**C**) Proposed proton transfer pathway and two-metal-ion mechanism of the phosphodiester bond cleavage reaction of pre-tRNA catalyzed by yeast RNase P. (**D**) Kinetic model of RNase P–mediated pre-tRNA processing. RNase P first recognizes the pre-tRNA substrate through a double-anchor mechanism to form the initial low-affinity ES complex. Binding of pre-tRNA then induces a local conformational change in the catalytic center of RNase P. In particular, nucleotide U93 within stem P4 of the Rpr1 RNA undergoes a conformational change to mediate an inner-sphere coordination of the catalytically important Mg^{2+} ion, so that the ES complex is isomerized to the active ES* state. The activated ES* complex then catalyzes the phosphodiester bond cleavage of pre-tRNA through a two-metal-ion S_N2 mechanism to release the 5′-leader of pre-tRNA. Last, the mature tRNA dissociates from the holoenzyme in a slow, rate-limiting step, and RNase P is ready for the next round of catalysis.

complex does not represent a pre-tRNA substrate–bound state. Superposition analysis showed that the soaked 5′-leader did not occupy the same position as the unprocessed 5′-leader in the yeast RNase P–pre-tRNA complex (fig. S17), arguing that the *Tm*M2 site in the *T. maritima* RNase P–tRNA complex is not a site for a catalytic metal ion.

The cryo-EM structures reported here represent a major step forward for mechanistic understanding of the RNase P function. Our data support that all RNase P ribozymes share an RNA-based, substrate-induced catalytic mechanism of pre-RNA processing. Whereas bacterial RNase P RNA is catalytically active by itself, eukaryotic RNase P is a protein-controlled ribozyme; its protein components not only directly participate in substrate recognition but also stabilize the catalytic RNA in a conformation optimal for pre-tRNA binding and cleavage reaction.

Methods and materials summary

Yeast strain constructions and affinity purification of the RNase P holoenzyme were performed as described previously, with minor modifications (*16*). Briefly, protein subunits Pop4 and Rpr2 were tagged with protein A and 3×Flag, respectively. A two-step affinity purification scheme was used to purify yeast RNase P. For substrate-bound structural studies, the in vitro transcribed yeast pre-tRNA^{Phe} was mixed with the highly purified yeast RNase P at 4°C and immediately subjected to cryo-EM grid preparation. For cryo-EM data collection and processing, the specimens were first prepared by using a lacey carbon film with a continuous ultrathin carbon film. Images were taken on a Gatan K2 summit camera mounted on an FEI Titan Krios electron microscope operated at 300 kV. Image processing includes motion correction, CTF estimation, particles classification, and refinement. De novo atomic model building, rigid docking of known structures, and homologous structure modeling were combined to generate the atomic model for the entire *S. cerevisiae* RNase P holoenzyme. QM/MM free-energy simulation was used to investigate the catalytic reaction of yeast RNase P. Detailed descriptions for all the materials and methods are provided in the supplementary materials.

REFERENCES AND NOTES

1. S. Altman, L. A. Kirsebom, in *Ribonuclease P: In The RNA World* R. Gesteland, T. Cech, J. Atkins, Eds. (Cold Spring Harbor Laboratory Press, ed. 2, 1999), chap. 14, pp. 351–378.
2. O. Esakova, A. S. Krasilnikov, Of proteins and RNA: The RNase P/MRP family. *RNA* **16**, 1725–1747 (2010). doi: 10.1261/rna.2214510; pmid: 20627997
3. D. Evans, S. M. Marquez, N. R. Pace, RNase P: Interface of the RNA and protein worlds. *Trends Biochem. Sci.* **31**, 333–341 (2006). doi: 10.1016/j.tibs.2006.04.007; pmid: 16679018
4. S. C. Walker, D. R. Engelke, Ribonuclease P: The evolution of an ancient RNA enzyme. *Crit. Rev. Biochem. Mol. Biol.* **41**, 77–102 (2006). doi: 10.1080/10409230600602634; pmid: 16595295
5. T. Pan, Higher order folding and domain analysis of the ribozyme from Bacillus subtilis ribonuclease P. *Biochemistry* **34**, 902–909 (1995). doi: 10.1021/bi00003a024; pmid: 7827048
6. A. Loria, T. Pan, Domain structure of the ribozyme from eubacterial ribonuclease P. *RNA* **2**, 551–563 (1996). pmid: 8718684
7. L. Sun, F. E. Campbell, N. H. Zahler, M. E. Harris, Evidence that substrate-specific effects of C5 protein lead to uniformity in binding and catalysis by RNase P. *EMBO J.* **25**, 3998–4007 (2006). doi: 10.1038/sj.emboj.7601290; pmid: 16932744
8. S. Niranjanakumari, T. Stams, S. M. Crary, D. W. Christianson, C. A. Fierke, Protein component of the ribozyme ribonuclease P alters substrate recognition by directly contacting precursor tRNA. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15212–15217 (1998). doi: 10.1073/pnas.95.26.15212; pmid: 9860948
9. L. Sun, M. E. Harris, Evidence that binding of C5 protein to P RNA enhances ribozyme catalysis by influencing active site metal ion affinity. *RNA* **13**, 1505–1515 (2007). doi: 10.1261/rna.571007; pmid: 17652407
10. Y. Kouzuma *et al.*, Reconstitution of archaeal ribonuclease P from RNA and four protein components. *Biochem. Biophys. Res. Commun.* **306**, 666–673 (2003). doi: 10.1016/S0006-291X(03)01034-9; pmid: 12810070
11. H. Y. Tsai, D. K. Pulukkunat, W. K. Woznick, V. Gopalan, Functional reconstitution and characterization of *Pyrococcus furiosus* RNase P. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 16147–16152 (2006). doi: 10.1073/pnas.0608000103; pmid: 17053064
12. A. Terada, T. Honda, H. Fukuhara, K. Hada, M. Kimura, Characterization of the archaeal ribonuclease P proteins from *Pyrococcus horikoshii* OT3. *J. Biochem.* **140**, 293–298 (2006). doi: 10.1093/jb/mvj144; pmid: 16829535
13. W.-Y. Chen, D. K. Pulukkunat, I. M. Cho, H.-Y. Tsai, V. Gopalan, Dissecting functional cooperation among protein subunits in archaeal RNase P, a catalytic ribonucleoprotein complex. *Nucleic Acids Res.* **38**, 8316–8327 (2010). doi: 10.1093/nar/gkq668; pmid: 20705647
14. A. Perederina, O. Esakova, C. Quan, E. Khanova, A. S. Krasilnikov, Eukaryotic ribonucleases P/MRP: The crystal structure of the P3 domain. *EMBO J.* **29**, 761–769 (2010). doi: 10.1038/emboj.2009.396; pmid: 20075859
15. C. W. Chan, B. R. Kiesel, A. Mondragón, Crystal structure of human Rpp20/Rpp25 reveals quaternary level adaptation of the alba scaffold as structural basis for single-stranded RNA binding. *J. Mol. Biol.* **430**, 1403–1416 (2018). doi: 10.1016/j.jmb.2018.03.029; pmid: 29625199
16. K. Hipp, K. Galani, C. Batisse, S. Prinz, B. Böttcher, Modular architecture of eukaryotic RNase P and RNase MRP revealed by electron microscopy. *Nucleic Acids Res.* **40**, 3275–3288 (2012). doi: 10.1093/nar/gkr1217; pmid: 22167472
17. E. Kikovska, N.-E. Mikkelsen, L. A. Kirsebom, The naturally trans-acting ribozyme RNase P RNA has leadzyme properties. *Nucleic Acids Res.* **33**, 6920–6930 (2005). doi: 10.1093/nar/gki993; pmid: 16332695
18. J. Hsieh, C. A. Fierke, Conformational change in the *Bacillus subtilis* RNase P holoenzyme—pre-tRNA complex enhances substrate affinity and limits cleavage rate. *RNA* **15**, 1565–1577 (2009). doi: 10.1261/rna.1639409; pmid: 19549719
19. J. Hsieh *et al.*, A divalent cation stabilizes the active conformation of the B. subtilis RNase P x pre-tRNA complex: A role for an inner-sphere metal ion in RNase P. *J. Mol. Biol.* **400**, 38–51 (2010). doi: 10.1016/j.jmb.2010.04.050; pmid: 20434461
20. B. P. Klemm *et al.*, The diversity of ribonuclease P: Protein and RNA catalysts with analogous biological functions. *Biomolecules* **6**, 27 (2016). doi: 10.3390/biom6020027; pmid: 27187488
21. J. A. Beebe, C. A. Fierke, A kinetic mechanism for cleavage of precursor tRNA(Asp) catalyzed by the RNA component of *Bacillus subtilis* ribonuclease P. *Biochemistry* **33**, 10294–10304 (1994). doi: 10.1021/bi00200a009; pmid: 7520753
22. A. Mondragón, Structural studies of RNase P. *Annu. Rev. Biophys.* **42**, 537–557 (2013). doi: 10.1146/annurev-biophys-083012-130406; pmid: 23654306
23. N. J. Reiter *et al.*, Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468**, 784–789 (2010). doi: 10.1038/nature09516; pmid: 21076397
24. X. Liu, Y. Chen, C. A. Fierke, Inner-sphere coordination of divalent metal ion with nucleobase in catalytic RNA. *J. Am. Chem. Soc.* **139**, 17457–17463 (2017). doi: 10.1021/jacs.7b08755; pmid: 29116782
25. J. Hsieh, S. C. Walker, C. A. Fierke, D. R. Engelke, Pre-tRNA turnover catalyzed by the yeast nuclear RNase P holoenzyme is limited by product release. *RNA* **15**, 224–234 (2009). doi: 10.1261/rna.1309409; pmid: 19095620
26. J. C. Kurz, C. A. Fierke, The affinity of magnesium binding sites in the *Bacillus subtilis* RNase P x pre-tRNA complex is enhanced by the protein subunit. *Biochemistry* **41**, 9545–9558 (2002). doi: 10.1021/bi025553w; pmid: 12135377
27. T. A. Steitz, J. A. Steitz, A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6498–6502 (1993). doi: 10.1073/pnas.90.14.6498; pmid: 8341661
28. M. R. Stahley, S. A. Strobel, Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* **309**, 1587–1590 (2005). doi: 10.1126/science.1114994; pmid: 16141079
29. N. Toor, K. S. Keating, S. D. Taylor, A. M. Pyle, Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82 (2008). doi: 10.1126/science.1153803; pmid: 18388288
30. M. R. Stahley, S. A. Strobel, RNA splicing: Group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. *Curr. Opin. Struct. Biol.* **16**, 319–326 (2006). doi: 10.1016/j.sbi.2006.04.005; pmid: 16697179
31. W. D. Hardt, J. M. Warnecke, V. A. Erdmann, R. K. Hartmann, Rp-phosphorothioate modifications in RNase P RNA that interfere with tRNA binding. *EMBO J.* **14**, 2935–2944 (1995). doi: 10.1002/j.1460-2075.1995.tb07293.x; pmid: 7540978
32. M. E. Harris, N. R. Pace, Identification of phosphates involved in catalysis by the ribozyme RNase P RNA. *RNA* **1**, 210–218 (1995). pmid: 7585250
33. J. M. Warnecke, J. P. Fürste, W. D. Hardt, V. A. Erdmann, R. K. Hartmann, Ribonuclease P (RNase P) RNA is converted to a Cd^{2+}-ribozyme by a single Rp-phosphorothioate modification in the precursor tRNA at the RNase P cleavage site. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8924–8928 (1996). doi: 10.1073/pnas.93.17.8924; pmid: 8799129
34. Y. Chen, X. Li, P. Gegenheimer, Ribonuclease P catalysis requires Mg2+ coordinated to the pro-RP oxygen of the scissile bond. *Biochemistry* **36**, 2425–2438 (1997). doi: 10.1021/bi9620464; pmid: 9054547
35. E. L. Christian, N. M. Kaye, M. E. Harris, Helix P4 is a divalent metal ion binding site in the conserved core of the ribonuclease P ribozyme. *RNA* **6**, 511–519 (2000). doi: 10.1017/S1355838200000042; pmid: 10786842
36. S. M. Crary, J. C. Kurz, C. A. Fierke, Specific phosphorothioate substitutions probe the active site of *Bacillus subtilis* ribonuclease P. *RNA* **8**, 933–947 (2002). doi: 10.1017/S1355838202025025; pmid: 12166648
37. E. L. Christian, K. M. J. Smith, N. Perera, M. E. Harris, The P4 metal binding site in RNase P RNA affects active site metal affinity through substrate positioning. *RNA* **12**, 1463–1467 (2006). doi: 10.1261/rna.158606; pmid: 16822954
38. A. V. Kazantsev, A. A. Krivenko, N. R. Pace, Mapping metal-binding sites in the catalytic domain of bacterial RNase P RNA. *RNA* **15**, 266–276 (2009). doi: 10.1261/rna.1331809; pmid: 19095619
39. E. L. Christian, N. M. Kaye, M. E. Harris, Evidence for a polynuclear metal ion binding site in the catalytic domain of ribonuclease P RNA. *EMBO J.* **21**, 2253–2262 (2002). doi: 10.1093/emboj/21.9.2253; pmid: 11980722
40. A. Torres-Larios, K. K. Swinger, T. Pan, A. Mondragón, Structure of ribonuclease P—a universal ribozyme. *Curr. Opin. Struct. Biol.* **16**, 327–335 (2006). doi: 10.1016/j.sbi.2006.04.002; pmid: 16650980
41. A. V. Kazantsev *et al.*, Crystal structure of a bacterial ribonuclease P RNA. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13392–13397 (2005). doi: 10.1073/pnas.0506662102; pmid: 16157868
42. E. Hartmann, R. K. Hartmann, The enigma of ribonuclease P evolution. *Trends Genet.* **19**, 561–569 (2003). doi: 10.1016/j.tig.2003.08.007; pmid: 14550630
43. S. Xiao *et al.*, Functional characterization of the conserved amino acids in Pop1p, the largest common protein subunit of yeast RNases P and MRP. *RNA* **12**, 1023–1037 (2006). doi: 10.1261/rna.23206; pmid: 16618965
44. M. M. Getz, A. J. Andrews, C. A. Fierke, H. M. Al-Hashimi, Structural plasticity and Mg2+ binding properties of RNase P P4 from combined analysis of NMR residual dipolar couplings and motionally decoupled spin relaxation. *RNA* **13**, 251–266 (2007). doi: 10.1261/rna.264207; pmid: 17194721
45. K. S. Koutmou *et al.*, NMR and XAS reveal an inner-sphere metal binding site in the P4 helix of the metallo-ribozyme ribonuclease P. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2479–2484 (2010). doi: 10.1073/pnas.0906319107; pmid: 20133747
46. S. Xiao, F. Scott, C. A. Fierke, D. R. Engelke, Eukaryotic ribonuclease P: A plurality of ribonucleoprotein enzymes. *Annu. Rev. Biochem.* **71**, 165–189 (2002). doi: 10.1146/annurev.biochem.71.110601.135352; pmid: 12045094
47. S. Wu *et al.*, Transition-state stabilization in *Escherichia coli* ribonuclease P RNA-mediated cleavage of model substrates.

Nucleic Acids Res. **42**, 631–642 (2014). doi: 10.1093/nar/gkt853; pmid: 24097434

48. K. S. Koutmou, J. J. Day-Storms, C. A. Fierke, The RNR motif of B. subtilis RNase P protein interacts with both PRNA and pre-tRNA to stabilize an active conformer. *RNA* **17**, 1225–1235 (2011). doi: 10.1261/rna.2742511; pmid: 21622899
49. R. J. Maraia, T. N. Lamichhane, 3′ processing of eukaryotic precursor tRNAs. *Wiley Interdiscip. Rev. RNA* **2**, 362–375 (2011). doi: 10.1002/wrna.64; pmid: 21572561
50. L. Casalino, G. Palermo, U. Rothlisberger, A. Magistrato, who activates the nucleophile in ribozyme catalysis? An answer from the splicing mechanism of group II introns. *J. Am. Chem. Soc.* **138**, 10374–10377 (2016). doi: 10.1021/jacs.6b01363; pmid: 27309711
51. W. Yang, J. Y. Lee, M. Nowotny, Making and breaking nucleic acids: two-Mg^{2+}-ion catalysis and substrate specificity. *Mol. Cell* **22**, 5–13 (2006). doi: 10.1016/j.molcel.2006.03.013; pmid: 16600865

ACKNOWLEDGMENTS

We thank the staff members of the Electron Microscopy System at the National Facility for Protein Science in Shanghai (NFPS), Zhangjiang Laboratory, China, and the Core Facilities for Protein Science at the Institute of Biophysics (IBP), Chinese Academy of Sciences, for providing technical support and assistance in data collection. We thank M. Cao for his help on data collection and analysis. Mass spectrometry experiments were performed at NFPS. **Funding:** This work was supported by grants from the National Natural Science Foundation of China (31525007 to M.L. and 21625302 to G.L.) and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB08010201 to M.L.). **Author contributions:** M.L., P.L., G.L., J.W., and H.C. conceived the study. M.L. supervised the whole project. P.L., S.N., and J.C. purified the yeast RNase P complex. P.L., M.T., X.W., and S.S. prepared the cryo-EM sample. M.T., P.L., X.W., S.Q., and G.C. collected the EM micrographs and processed the data. J.W. and M.L. built the atomic model. Y.Z., G.L., and X.P. performed the MD and QM/MM simulation analyses. M.L. and P.L. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The accession numbers for the structure reported in this paper are PDB: 6AGB and 6AH3 and EMDB: EMD-9616 and EMB-9622.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/eaat6678/suppl/DC1
Materials and Methods
Figs. S1 to S18
Table S1
References (*52*–*101*)

22 March 2018; accepted 18 September 2018
Published online 27 September 2018
10.1126/science.aat6678

REPORT

BATTERIES

Suppressing corrosion in primary aluminum–air batteries via oil displacement

Brandon J. Hopkins*, Yang Shao-Horn, Douglas P. Hart

Primary aluminum–air batteries boast high theoretical energy densities, but negative electrode corrosion irreversibly limits their shelf life. Most corrosion mitigation methods are insufficient or compromise power and energy density. We suppressed open-circuit corrosion by displacing electrolyte from the electrode surface with a nonconducting oil during battery standby. High power and energy density are enabled by displacing the oil with electrolyte for battery discharge. The underwater-oleophobic wetting properties of the designed cell surfaces allow for reversible oil displacement. We demonstrate this method in an aluminum–air cell that achieves a 420% increase in usable energy density and 99.99% reduction in corrosion, which lowers self-discharge to a rate of 0.02% a month and enables system energy densities of 700 watt-hours per liter and 900 watt-hours per kilogram.

Single-use and mechanically rechargeable aluminum–air (Al–air) batteries are low-cost options for applications that require ultrahigh energy densities that electrically rechargeable batteries have yet to achieve (*1*, *2*). Such applications include long-range drones (*3*), off-grid power supplies (*4*, *5*), and range extenders for electric vehicles (*6*, *7*). A long-standing barrier to their increased adoption, however, has been severe open-circuit corrosion that curtails their shelf life (*8*, *9*). Despite recent advances in corrosion inhibition (*10*), Al–air batteries can irreversibly self-discharge more than 80% a month depending on battery design (*11*), while rechargeable lithium-ion batteries self-discharge only 5% a month (*12*).

Methods to mitigate corrosion in Al–air batteries typically require a substantial trade-off between power and energy density and open-circuit corrosion. Mitigation methods include negative electrode alloying (*2*), electrolyte additives (*2*), gel electrolytes (*13*), nonaqueous electrolytes (*14*), and draining the electrolyte after shutdown (*4*, *15*, *16*). Alloying combined with aqueous electrolyte additives yields high power and energy density but also high corrosion currents near 1 to 10 mA $\mathrm{cm_{geo}}^{-2}$ (geometric area) (*1*, *2*, *10*). Alloys generally use greater than 99.99 weight % (wt %) purity Al combined with a variety of elements such as Zn, In, Mg, Ga, Sn, Bi, B, and Mn (*2*). Studied electrolyte additives include ZnO, $\mathrm{SnO3}^{2-}$, In(OH)3, $\mathrm{BiO3}^{3-}$, $\mathrm{Ga(OH)4}^{-}$, $\mathrm{MnO4}^{2-}$, Cl^{-}, $\mathrm{NO3}^{-}$, $\mathrm{SO4}^{2-}$, and $\mathrm{SnO_3}^{2-}$ (*2*). Conversely, gel and nonaqueous electrolytes using polyacrylic acid and ionic liquids achieve lower corrosion rates but only fractions of the power and energy density achieved with more conductive aqueous electrolytes (*13*, *14*). Ionic electrolytes attain the lowest reported open-circuit corrosion currents, near 0.002 mA $\mathrm{cm_{geo}}^{-2}$ (*14*). Draining the electrolyte after shutdown can result in considerable electrode corrosion and hazardous clogging of the hydraulic system of an Al–air battery pack (*15*, *16*). After draining, remaining electrolyte clings to the hydrophilic Al electrodes, induces corrosion, dries up, and leaves crusts of by-products and electrolyte solutes (*15*, *16*). Rinsing the Al electrode with water or a pH-neutralizing agent after electrolyte draining improves restart performance but can significantly reduce system energy density if the rinsing system is contained in the battery pack (*4*, *5*, *16*). Such rinsing systems require additional tanks for water and electrolyte along with a neutralizing agent that is consumed with every shutdown (*4*, *5*, *16*). If the trade-off between power and energy density and open-circuit corrosion could be sufficiently mitigated, the use of Al–air batteries would extend beyond current niche applications.

We overcame this performance trade-off by redesigning the conventional flowing electrolyte metal–air battery (Fig. 1A) (*5*, *16*, *17*) with a separator that displays underwater oleophobicity (Fig. 1, B and C) (*18*). This separator allows for the reversible displacement of electrolyte from the negative electrode surface with an oil when the battery is on standby. The nonconducting displacing oil dramatically reduces the diffusion rate of electrolyte to the negative electrode surface, enabling ultralow open-circuit corrosion. We achieved high power and energy density by displacing the oil with electrolyte for battery discharge. Such a method could be used to suppress open-circuit corrosion in primary Al–air cells and potentially other batteries (*8*, *9*).

To demonstrate this concept, we constructed a primary flowing-electrolyte Al–air battery (fig. S1). The cathode was a commercially available air electrode (Quantum Sphere). The air electrode consisted of carbon particles and nano-manganese catalyst pressed on a nickel mesh layered with an air-breathing, waterproof expanded polytetrafluoroethylene (ePTFE) membrane that enables a reduction reaction with oxygen from the atmosphere (Eq. 1) (*19*). The anode was a 0.25-mm-thick nonporous Al foil (99.999 wt % purity) (Sigma-Aldrich) that oxidizes during discharge (Eq. 2). The selected aqueous electrolyte was the conventionally used, highly conductive 4 M sodium hydroxide (NaOH) (98 wt % purity) paired with a common corrosion inhibitor, 0.05 M sodium stannate ($\mathrm{Na_2SnO_3}$) (95 wt % purity) (*20*).

$$\begin{gathered}\mathrm{O_2 + 2H_2O + 4e^- \rightarrow 4OH^-}\\ E_\mathrm{c} = 0.3\ \mathrm{V\ versus\ Hg/HgO}\end{gathered} \qquad (1)$$

$$\begin{gathered}\mathrm{Al + 4OH^- \rightarrow Al(OH)_4^- + 3}e^-\\ E_\mathrm{a} = -2.4\mathrm{V\ versus\ Hg/HgO}\end{gathered} \qquad (2)$$

We chose low–surface tension displacing oils and hydrophilic separators for testing in the constructed cell because such systems are likely to display underwater oleophobicity and be resistant to oil fouling. This conclusion can be drawn from a two-phase Young's equation (Eq. 3) relating the wetting properties of fluids in air to the wetting properties of those fluids submerged in a different fluid (*21*).

$$\cos\theta_{o-e} = \frac{\gamma_{o-\mathrm{air}}\cos\theta_{o-\mathrm{air}} - \gamma_{e-\mathrm{air}}\cos\theta_{e-\mathrm{air}}}{\gamma_{o-e}} \qquad (3)$$

For this application, the variable θ_{o-e} is the Young contact angle of an oil droplet surrounded by electrolyte on the negative electrode or separator surface. The variables $\theta_{o-\mathrm{air}}$ and $\theta_{e-\mathrm{air}}$ are the Young contact angles of an oil and electrolyte droplet surrounded by air on either the negative electrode or separator surface, respectively. The surface tension of the oil and electrolyte are respectively $\gamma_{o-\mathrm{air}}$ and $\gamma_{e-\mathrm{air}}$, with an oil-electrolyte interfacial tension expressed as γ_{o-e}. Equation 3 suggests that to achieve underwater oleophobicity, where θ_{o-e} is greater than 90°, the numerator in the expression right of the equals sign must be negative. The more negative the numerator, the larger the underwater contact angle. To achieve a negative numerator, a low–surface tension oil ($\gamma_{o-\mathrm{air}}$) can be selected along with a hydrophilic negative electrode and separator ($\theta_{e-\mathrm{air}} < 90°$). We therefore selected two low–surface tension candidate oils, silicone and perfluoropolyether (PFPE) oil (*22*), that have large working temperatures (–40° to 100°C), act as ionic insulators to inhibit corrosion, and cannot saponify in high-pH solutions because they contain no triglycerides. We also chose two hydrophilic separators, Celgard 5550 and a PTFE separator (Advantec MFS) commercially treated with an additive to be hydrophilic (*11*), that display high working temperatures and chemical compatibility with the electrolyte.

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*Corresponding author. Email: bhopkins@mit.edu

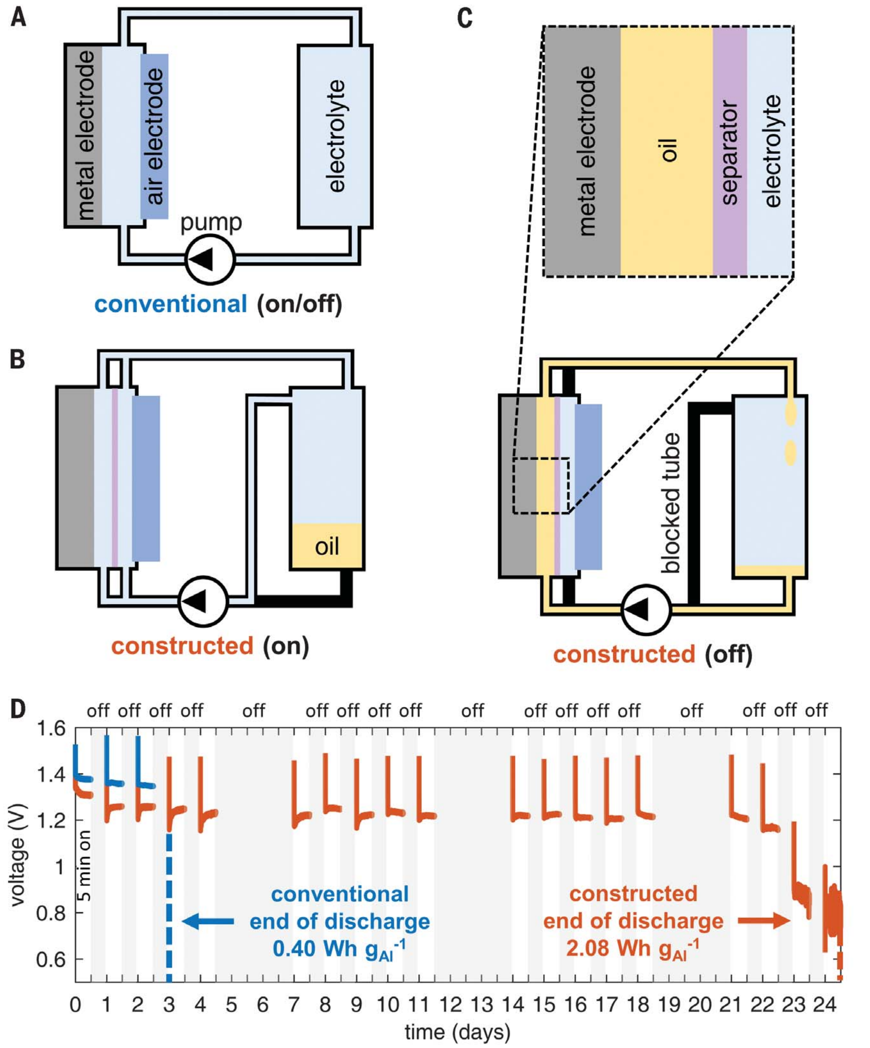


Fig. 1. Oil displacement method. (**A**) Schematic of the conventional flowing electrolyte metal–air battery (*16*, *17*). (**B**) Schematic of the constructed oil displacement system for a flowing electrolyte metal–air battery. Electrolyte is continuously pumped during operation. When not in use, oil is pumped to displace the corrosive electrolyte for a specified duration. (**C**) Schematic of the constructed oil displacement system when not in use, with a magnified view of the interfaces of the metal electrode and separator. (**D**) Voltage versus time for on-off cycling of a primary Al–air battery with (A) the conventional cell design and [(B) and (C)] the constructed cell design. Current densities of 150 mA $\text{cm}_{\text{geo}}^{-2}$ were drawn for 5 min with 24- or 72-hour pauses during which no current was drawn. The conventional cell stopped operating at the start of day 3, yielding an energy density of 0.40 ± 0.07 Wh $\text{g}_{\text{Al}}^{-1}$. By contrast, the constructed cell lasted more than 24 days, yielding an energy density of 2.08 ± 0.07 Wh $\text{g}_{\text{Al}}^{-1}$.

We identified appropriate separator pore sizes using the Young-Laplace equation to ensure that the pores suppressed the passage of oil through the separator at appropriate pumping pressures. The Young-Laplace equation, $\Delta P = 4\gamma_{o-e}/D$, estimates the pressure differential ΔP that is required for oil to leak through a circular pore with a diameter D given an interfacial tension between the oil and electrolyte of γ_{o-e}. For example, the measured interfacial tension between the electrolyte and PFPE oil was 61.6 ± 0.1 mN m^{-1} (*11*). Commercially available pore diameters for hydrophilic PTFE separators range from 0.1 to 1 μm. We selected a 1-μm pore size to minimize ionic resistance across the separator. If we assume severe thermal expansion of 100% owing to numerous heating cycles, 2-μm pore diameters would allow for high pumping pressures of 123 kPa (*23*).

The oil displacement method is enabled by using an oil that displays underwater oleophobicity on both the negative electrode and separator surfaces. To demonstrate this principle, we measured advancing and receding underwater contact angles of silicone oil on the Celgard separator and on the Al electrode and of PFPE oil on the PTFE separator and on the Al electrode (fig. S5). Both the silicone and PFPE oil displayed underwater oleophobicity on the Al electrode (Fig. 2, A and B, top left), but the silicone oil wetted the Celgard, whereas the PFPE oil beaded up on the hydrophilic PTFE separator (Fig. 2, A and B, bottom left). We then performed on-off cycling with Al–air cells using the respective oil-separator combinations (Fig. 2, A and B, right). Current densities of 150 mA $\text{cm}_{\text{geo}}^{-2}$ were drawn for 25 min with 1-hour pauses during which no current was drawn (*11*). At the start of each pause, the displacing oils were pumped into the batteries for 50 s, requiring less than 50 mW of power. Before the next discharge, the electrolyte was pumped into the batteries for 50 s. After the silicone oil was pumped in, subsequent discharge voltages were negative, suggesting that the silicone oil fouled the Celgard separator (Fig. 2A, right). The total energy density achieved with the silicone-Celgard combination was 0.66 ± 0.08 watt-hour (Wh) per gram Al $(\text{g}_{\text{Al}}{}^{-1})$ (*11*). By contrast, the PFPE-PTFE combination showed positive voltages after each pause (Fig. 2B, right). The total energy density achieved with the PFPE-PTFE combination was 3.55 ± 0.08 Wh $\text{g}_{\text{Al}}^{-1}$ (*11*), which is comparable with the energy density achieved by discharging the battery without pauses and without introducing oil, 3.54 ± 0.08 Wh $\text{g}_{\text{Al}}^{-1}$ (fig. S6) (*11*). These energy-density results imply that the Al electrode and hydrophilic PTFE separator resisted fouling by the PFPE oil.

Perturbations and electrolyte pumping trigger the displacement of oil by electrolyte. The Al electrode and separator's underwater oleophobicity (fig. S5) promote the rupturing of thin oil films on their surfaces, as suggested by thin film rupture theory (*24*). The perturbations and electrolyte pumping induce flow instabilities (*24*, *25*) that displace the oil from the electrode and separator surfaces. For displacement of electrolyte by oil, the bulk electrolyte is displaced by the flowing oil that thins the remaining electrolyte films, which are eventually consumed by corrosion on the Al electrode.

With the selected PFPE displacing oil and hydrophilic PTFE separator, we showed that the constructed cell (Fig. 1, B and C) achieved a greater usable energy density than that achieved with a conventional cell (Fig. 1A) by drawing current densities of 150 mA $\text{cm}_{\text{geo}}^{-2}$ for 5 min with 24- or 72-hour pauses in between (Fig. 1D) (*11*). During pauses for the conventional cell, the electrolyte remained in the interelectrode gap, with corrosion currents of 1 to 10 mA $\text{cm}_{\text{geo}}^{-2}$ (figs. S2 and S3) (*11*). For the constructed cell with the oil displacement system, PFPE oil was pumped into the cell for 50 s at the start of each pause to displace the corrosive electrolyte. Once the oil displaced the electrolyte, corrosion currents dropped to 0.0001 to 0.0011 mA $\text{cm}_{\text{geo}}^{-2}$ (figs. S2 and S4) (*11*), which are below those achieved by state-of-the-art ionic electrolytes (*14*). Before each discharge, the electrolyte was pumped back into the cell for 50 s. At the start of day 3, the conventional cell yielded a negative voltage when 150 mA $\text{cm}_{\text{geo}}^{-2}$ were drawn. The Al electrode was corroded through its thickness in the conventional cell, which was monitored through the transparent battery casing. The resulting energy density of the conventional cell was 0.40 ± 0.07 Wh $\mathrm{g_{Al}}^{-1}$ (*11*). By contrast, the constructed cell using the oil displacement system lasted more than 24 days and yielded an energy density of 2.08 ± 0.07 Wh $\mathrm{g_{Al}}^{-1}$ (*11*). Commercial Al–air anodes are relatively thick, 0.5 cm (*26*), and nonporous to mitigate corrosion. They sustain minimal loss in power as electrode thickness increases because no ionic diffusion occurs within the solid Al electrode during discharge (*27*). Consequently, a scaled-up 0.5-cm-thick Al electrode would achieve a self-discharge rate near 0.02 ± 0.02% a month, which could be tuned based on oil film thickness (*11*).

The performance of the constructed cell with the oil displacement system was comparable with the performance of Al–air cells in the literature. The open-circuit potential of the cell with and without the PTFE separator was 1.8 V as found in the literature (*20*). At 150 mA $\mathrm{cm_{geo}}^{-2}$, the coulombic efficiency with and without the separator was 91%, which also matched literature values (*20*). We measured the voltage (Fig. 3A and fig. S7) and capacity (Fig. 3B and fig. S8) as a function of current density. The prototype with the hydrophilic PTFE separator achieved a peak power density of 300 ± 17 mW $\mathrm{cm_{geo}}^{-2}$ (Fig. 3A and fig. S9) and energy density of 3.73 ± 0.08 Wh $\mathrm{g_{Al}}^{-1}$ (fig. S10). Reported peak powers of flowing-electrolyte Al–air batteries with 2- to 3-mm interelectrode gaps range from 350 to 620 mW $\mathrm{cm_{geo}}^{-2}$ and obtain peak energy densities near 4.3 Wh $\mathrm{g_{Al}}^{-1}$ (*1*, *17*). Al–air batteries with gaps near 10 mm, however, achieve peak powers as low as 75 mW $\mathrm{cm_{geo}}^{-2}$ (*28*). We attribute our values to the PTFE separator and to the 7-mm interelectrode gap of the constructed prototype. Without the separator, the prototype reached a peak power of 350 ± 17 mW $\mathrm{cm_{geo}}^{-2}$ (fig. S9). The separator, however, minimally affected power and energy density for energy-efficient current densities below 250 mA $\mathrm{cm_{geo}}^{-2}$ (fig. S10). For example, at 150 mA $\mathrm{cm_{geo}}^{-2}$, the power and energy density achieved with the separator was 192 ± 5 mW $\mathrm{cm_{geo}}^{-2}$ (Fig. 3A) and 3.54 ± 0.08 Wh $\mathrm{g_{Al}}^{-1}$ (fig. S6). Excluding the separator yielded 199 ± 4 mW $\mathrm{cm_{geo}}^{-2}$ (fig. S9) and 3.75 ± 0.08 Wh $\mathrm{g_{Al}}^{-1}$ (fig. S6). The large 7-mm gap allowed for appropriate separator sealing and easy assembly and disassembly of the prototype cell with bolts (fig. S1). Production-level sealing methods such as laser or ultrasonic welding would enable smaller interelectrode gap sizes.

Oil accounted for 13% of the total mass of an Al–air battery pack with an oil displacement system (Fig. 3C). Al–air electric vehicle packs achieve peak powers of 34.8 kW and maximum system energy densities of 936 Wh liter^{-1} and 1300 Wh kg^{-1}, with total volumes and masses of 371 liters and 267 kg (*1*). Calculations include practical pack components such as an electrolyte pump, battery casing, and electrolyte filtration system. The energy density values assume that water is refilled four times before complete discharge and that discharge is continuous. To enable energy-efficient startup/shutdown capability, we added 20 liters of displacing oil to the pack model (*1*). This additional oil resulted in an increase of 36 kg of mass, assuming PFPE oil was used and represented 13% of the total pack mass.

We calculated scaled-up corrosion losses per shutdown cycle for an Al–air battery pack with an oil displacement system using data from the 24-day experiment (Fig. 1D). We attribute the reduction in capacity observed in the 24-day experiment with respect to the capacity achieved by continuous discharge (fig. S6) to corrosion associated with the thin layer of electrolyte that remained on the Al electrode after bulk electrolyte displacement. We observed minimal change in energy density when pauses were 1 hour (Fig. 2B) and minimal changes in voltage between pauses that lasted 24 and 72 hours (Fig. 1D). These observations suggest that little corrosion occurred after 1 hour and that remaining electrolyte layers were consumed by corrosion after ~24 hours. We therefore assumed equal volumes of remaining electrolyte on the negative electrode after each discharge of the 24-day experiment, yielding a loss of 14 $\mathrm{g_{Al}}$ $\mathrm{m_{geo}}^{-2}$ per pause due to corrosion. If 150 mA $\mathrm{cm_{geo}}^{-2}$ is the normal operating current density, an approximate voltage of 1.22 V will be maintained (Fig. 1D) with a capacity of 2.697 Ah $\mathrm{g_{Al}}^{-1}$ (Fig. 3B). Assuming operation at the stated voltage and capacity, system-level corrosion yields a loss of 267 Wh per shutdown with 5.8 $\mathrm{m_{geo}}^{2}$ of Al (*11*). We estimated system-level pumping losses to be 2% of operating power at 150 mA cm^{-2} (*17*), with a displacement time of 10 min, yielding a loss of 40 Wh per startup/shutdown cycle. The large difference in corrosion and pumping losses suggests that there is room for energy-density improvement by pumping optimization (*29*). The corrosion and pumping losses

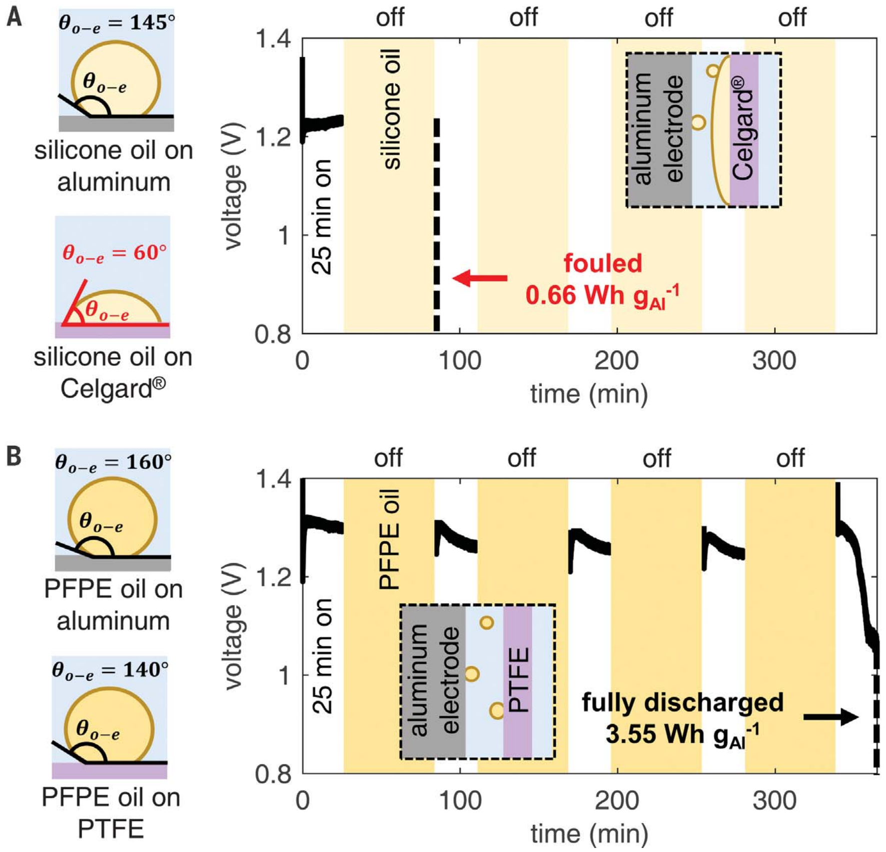


Fig. 2. Selecting a displacing oil and separator. (**A**) Contact angle schematics of silicone oil on the Al electrode (top left) and on a Celgard separator (bottom left). (Right) Voltage versus time for on-off cycling by using the silicone oil and Celgard separator. Current densities of 150 mA $\mathrm{cm_{geo}}^{-2}$ were drawn for 25 min, with 1-hour pauses during which no current was drawn. At the start of each pause, silicone oil was pumped into the battery for 50 s. Before the following discharge, electrolyte was pumped into the battery for 50 s. (Inset) Schematic of the Al electrode and separator surfaces. The silicone oil fouls the Celgard separator but not the Al electrode. (**B**) Corresponding experiments presented in (A), using a PFPE oil and hydrophilic PTFE separator. (Inset) A fouling-resistant hydrophilic PTFE separator and Al electrode. The Al–air battery using the silicone-Celgard combination yielded a negative voltage after the first 1-hour pause, resulting in an energy density of 0.66 ± 0.08 Wh $\mathrm{g_{Al}}^{-1}$, whereas the Al–air battery using the PFPE-PTFE combination reached full discharge capacity, yielding an energy density of 3.55 ± 0.08 Wh $\mathrm{g_{Al}}^{-1}$. Contact-angle photos can be found in fig. S5.

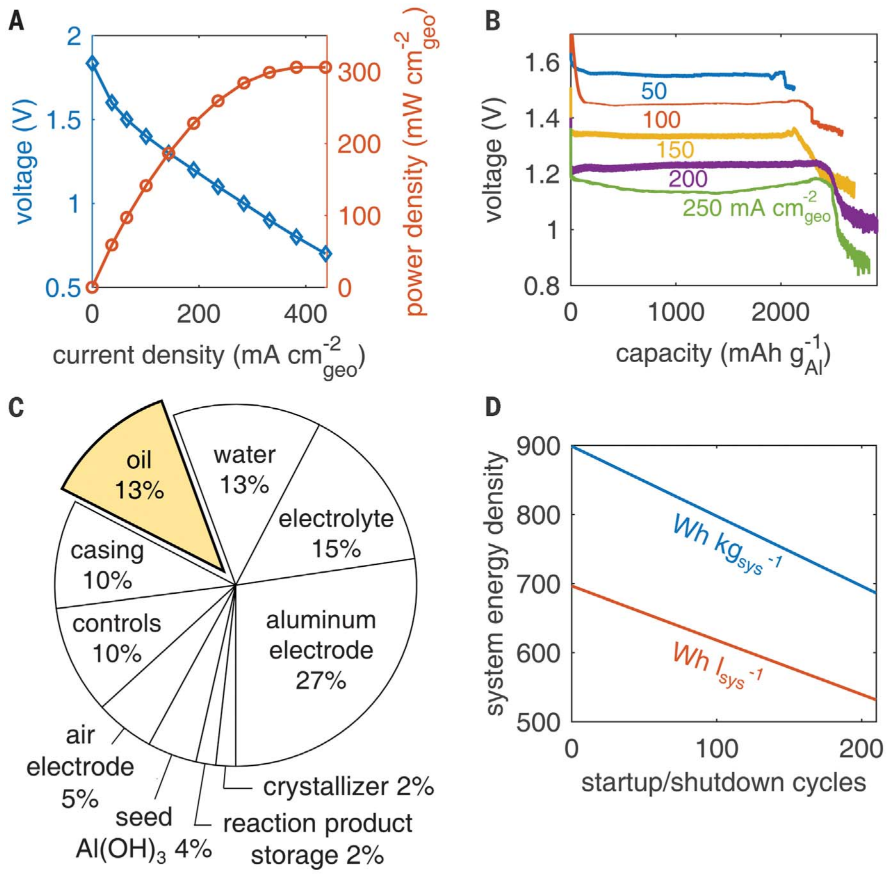


Fig. 3. Constructed-cell performance and pack-level characteristics. (**A**) Voltage and power density versus current density of the constructed primary aqueous Al–air battery using the hydrophilic PTFE separator. (**B**) Voltage versus capacity of the same battery used in (A) for discharging currents of 50, 100, 150, 200, and 250 mA cm$_{geo}^{-2}$. (**C**) Estimated mass distribution of an Al–air battery pack (*1*) with an oil displacement system. The oil accounted for 13% of the total mass of the system. (**D**) Usable system energy density versus the number of startup/shutdown cycles that occur before complete discharge.

per startup/shutdown cycle enable the calculation of usable system energy density as a function of the number of startup/shutdown cycles before complete discharge (Fig. 3D).

We estimated maximum usable pack-level energy densities of 700 Wh liter$_{sys}^{-1}$ (system volume) and 900 Wh kg$_{sys}^{-1}$ (system mass) (Fig. 3D) with a standard operational power of 11 kW. To calculate a lower-bound energy density limit, we assumed that each startup/shutdown cycle yields corrosion damage comparable with a 24-hour shutdown. The usable system energy density (E_{sys}) is a function of the energy stored in the Al (E_{Al}), the number of startup/shutdown cycles (n), the losses due to corrosion and pumping per startup/shutdown cycle (c_{loss} and p_{loss}, respectively), and the system volume or mass that includes the added mass and volume of the oil displacement system (v_{sys} and m_{sys}, respectively), resulting in $E_{sys} = [E_{Al} - n(c_{loss} + p_{loss})]/v_{sys}$ for system volumetric energy density. Using the constructed prototype's power at a current density of 150 mA cm$_{geo}^{-2}$ yields a pack power of 11 kW. Such powers are sufficient for many applications, including industrial tools (*30*), off-grid power units (*15*), and vehicular propulsion (*31*). Applications with few startup/shutdown cycles would have energy densities near 700 Wh liter$_{sys}^{-1}$ and 900 Wh kg$_{sys}^{-1}$. By contrast, applications that require hundreds of startup/shutdown cycles before complete discharge would achieve energy densities near 530 Wh liter$_{sys}^{-1}$ and 680 Wh kg$_{sys}^{-1}$ (*11*). These calculations and data suggest that an Al–air pack with an oil displacement system achieves high system energy densities, power densities that are comparable with aqueous flowing electrolyte Al–air systems, and open-circuit corrosion currents that are lower than those attained with nonaqueous Al–air cells.

REFERENCES AND NOTES

1. S. Yang, H. Knickle, *J. Power Sources* **112**, 162–173 (2002).
2. D. R. Egan *et al.*, *J. Power Sources* **236**, 293–310 (2013).
3. A. Kindler, L. Matthies, Micro- and nanotechnology sensors, systems, and applications VI. *Proc. SPIE* 10.1117/12.2051820 (2014).
4. B. M. L. Rao, R. Cook, W. Kobasz, G. D. Deuchars, in *IEEE 35th International Power Sources Symposium* (1992), pp. 34–37.
5. W. Wang, Y. Qi, High-power aluminum-air battery system (2017); https://patents.google.com/patent/US9705166B2/en.
6. C. News, Electric car with massive range in demo by Phinergy, Alcoa (2014); www.cbc.ca/news/technology/electric-car-with-massive-range-in-demo-by-phinergy-alcoa-1.2664653.
7. S. B. Sherman, Z. P. Cano, M. Fowler, Z. Chen, *AIMS Energy* **6**, 121–145 (2018).
8. M. A. Rahman, X. Wang, C. Wen, *J. Electrochem. Soc.* **160**, A1759–A1771 (2013).
9. Y. Li, J. Lu, *ACS Energy Lett.* **2**, 1370–1377 (2017).
10. D. Gelman, I. Lasman, S. Elfimchev, D. Starosvetsky, Y. Ein-Eli, *J. Power Sources* **285**, 100–108 (2015).
11. Materials and methods are available as supplementary materials.
12. L. Lu, X. Han, J. Li, J. Hua, M. Ouyang, *J. Power Sources* **226**, 272–288 (2013).
13. Z. Zhang *et al.*, *J. Power Sources* **251**, 470–475 (2014).
14. R. Revel, T. Audichon, S. Gonzalez, *J. Power Sources* **272**, 415–421 (2014).
15. T. A. Dougherty *et al.*, in *Telecommunications Energy Conference, 1995. INTELEC '95., 17th International* (1995), pp. 821–827.
16. D. Tzidon, I. Yakupov, Shutdown system for metal-air batteries and methods of use thereof (2015); https://patents.google.com/patent/US9627726B2/en.
17. E. Behrin, R. L. Wood, J. D. Salisbury, D. J. Whisler, C. L. Hudson, "Design analysis of an aluminum-air battery for vehicle operations, transportation systems research" (Livermore, 1983).
18. B. J. Hopkins, D. P. Hart, Corrosion mitigation in metal-air batteries (2017); https://patents.google.com/patent/WO2017176390.
19. QuantumSphere, "QSI-Nano manganese gas diffusion electrodes" (Santa Ana, CA, USA, 2009); www.tomo-e.co.jp/upload/cProductsJA/25QU073-cProductsJA_content-004.pdf.
20. M. L. Doche, F. Novel-Cattin, R. Durand, J. J. Rameau, *J. Power Sources* **65**, 197–205 (1997).
21. M. Liu, S. Wang, Z. Wei, Y. Song, L. Jiang, *Adv. Mater.* **21**, 665–669 (2009).
22. Chemours, "Krytox^{TM} performance lubricants product overview" (2010); www.chemours.com/Lubricants/en_US/assets/downloads/Krytox_Overview_LowRes_H-58505-5.pdf.
23. A. Tang, J. Bao, M. Skyllas-Kazacos, *J. Power Sources* **248**, 154–162 (2014).
24. A. Sharma, *Langmuir* **9**, 861–869 (1993).
25. X. Chen *et al.*, *Energy Environ. Sci.* **9**, 1760–1770 (2016).
26. A. Yadgar, Y. Miller, D. Tzidon, Protected anode structure suitable for use in metal/air batteries (2016); https://patents.google.com/patent/US20160013529A1.
27. S. H. Yang, H. Knickle, *J. Power Sources* **124**, 572–585 (2003).
28. M. Rota, C. Comninellis, S. Moller, F. Holzer, O. Haas, *J. Appl. Electrochem.* **25**, 114–121 (1995).
29. K. Bode *et al.*, *Heat Transf. Eng.* **28**, 202–209 (2007).
30. E. Abele, M. Weigold, S. Rothenbücher, *Manuf. Technol.* **56**, 387–390 (2007).
31. M. Gebrehiwot, A. Van den Bossche, *Int. J. Automot. Technol.* **16**, 707–713 (2015).

ACKNOWLEDGMENTS

Funding: Research was supported by the Massachusetts Institute of Technology (MIT) Lincoln Laboratory (awards 7000308296, 7000344422, and 7000401832). **Author contributions:** B.J.H. and D.P.H. conceptualized the oil displacement method. B.J.H., Y.S-H., and D.P.H. developed the methodology to test and characterize the performance of the oil displacement method. B.J.H. performed all experimental investigation, formal analysis, experimental validation, data curation, and software programing to process and present collected data. B.J.H. designed, built, and tested all battery systems presented in the work. Y.S-H. provided electrochemical instrumentation. B.J.H. prepared the initial manuscript and figures that D.P.H. and Y.S-H. reviewed and edited. D.P.H. supervised and acquired funding for the research. **Competing interests:** B.J.H. and D.P.H. are inventors on Patent Cooperation Treaty serial no. PCT/US2017/020093 covering the oil displacement method. **Data and material availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/658/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
References (*32*–*36*)
Data File S1

11 June 2018; accepted 21 September 2018
10.1126/science.aat9149

QUANTUM OPTICS

Photon-mediated interactions between quantum emitters in a diamond nanocavity

R. E. Evans[1]*, **M. K. Bhaskar**[1]*, **D. D. Sukachev**[1]*, **C. T. Nguyen**[1], **A. Sipahigil**[1,2], **M. J. Burek**[3], **B. Machielse**[1,3], **G. H. Zhang**[1], **A. S. Zibrov**[1], **E. Bielejec**[4], **H. Park**[1,5], **M. Lončar**[3], **M. D. Lukin**[1]†

Photon-mediated interactions between quantum systems are essential for realizing quantum networks and scalable quantum information processing. We demonstrate such interactions between pairs of silicon-vacancy (SiV) color centers coupled to a diamond nanophotonic cavity. When the optical transitions of the two color centers are tuned into resonance, the coupling to the common cavity mode results in a coherent interaction between them, leading to spectrally resolved superradiant and subradiant states. We use the electronic spin degrees of freedom of the SiV centers to control these optically mediated interactions. Such controlled interactions will be crucial in developing cavity-mediated quantum gates between spin qubits and for realizing scalable quantum network nodes.

Photon-mediated interactions between quantum emitters are an important building block of quantum information systems, enabling entanglement generation and quantum logic operations involving both stationary qubits and photons (*1*, *2*). Progress in cavity quantum electrodynamics (QED) with trapped atoms and ions (*3*), superconducting qubits (*4*), and self-assembled quantum dots (*5*) has created possibilities for engineering such interactions. In particular, coherent multi-qubit interactions mediated via a cavity have been demonstrated in the microwave domain using circuit QED (*6*). Extending such interactions to the optical domain could enable key protocols in long-distance quantum communication (*1*, *7*). This goal is challenging owing to the difficulty of achieving strong cavity coupling and individual control of multiple resonant quantum emitters. Recently, cavity-modified collective scattering has been observed in two-ion (*8*) and two-atom (*9*) systems. Signatures of cavity-mediated interactions between quantum dots have also been reported (*10*, *11*). However, the realization of controlled, coherent optical interactions between solid-state emitters is difficult because of inhomogeneous broadening and decoherence introduced by the solid-state environment (*5*, *11*).

We realize controllable optically mediated interactions between negatively charged silicon-vacancy (SiV) color centers coupled to a diamond photonic crystal cavity (Fig. 1A) (*12*). SiV centers in diamond are atom-like quantum emitters featuring nearly lifetime-limited optical linewidths with low inhomogeneous broadening, both in bulk (*13*) and in nanostructures (*14*). We integrate SiV centers into devices consisting of a one-dimensional diamond waveguide with an array of holes defining a photonic crystal cavity with quality factor $Q \sim 10^4$ and simulated mode volume $V \sim 0.5[\lambda/(n = 2.4)]^3$ (Fig. 1B) (*15*). SiV centers are positioned at a single spot in these devices with 40-nm precision by targeted implantation using a focused beam of ^{29}Si ions, yielding around five SiV centers per device (*12*). The diamond waveguide is tapered and adiabatically coupled to a tapered single-mode fiber, enabling collection efficiencies from the waveguide into the fiber of more than 90% (*15*). These devices are placed in a dilution refrigerator with an integrated confocal microscope (*16*). At 85 mK, SiV centers are completely polarized into the orbital ground state (*17*). Here, we use optical transitions between the lowest-energy orbital states in the electronic ground and excited states. The SiV electronic spin degeneracy is lifted by applying a magnetic field up to 10 kG (*17*, *18*). The cavity resonance frequency ω_c is tuned using gas condensation (*16*).

The coupling between SiV centers and the cavity is characterized by scanning the frequency of a laser incident on one side of the device from free space while monitoring the transmitted intensity in the collection fiber. The resulting transmission spectrum (Fig. 1C) reveals strong modulation of the cavity response resulting from the coupling of spectrally resolved SiV centers to the cavity mode. For instance, two SiV centers near the cavity resonance each result in almost-full extinction of the transmission through the cavity (Fig. 1C, lower spectrum) (*19*). By contrast, when the cavity is detuned from the SiV by several cavity linewidths (κ), the spectrum shows a narrow peak near each SiV frequency (Fig. 1D), corresponding to an atom-like dressed state of the SiV-cavity system with high transmission (*6*). The resonance linewidth (Γ) changes by more than an order of magnitude depending on the SiV-cavity detuning ($\Delta = \omega_c - \omega_{\mathrm{SiV}}$). This can be understood through Purcell enhancement, which predicts $\Gamma(\Delta) \approx \gamma + \frac{4g^2}{\kappa}\frac{1}{1+4\Delta^2/\kappa^2}$ where g is the single-photon Rabi frequency, κ is the cavity energy decay rate, and γ is twice the decoherence rate due to free-space spontaneous emission and spectral diffusion. For the strongest-coupled SiV in the device used in Fig. 1, linewidths range from $\Gamma(0) = 2\pi \times 4.6$ GHz on resonance to $\Gamma(7\kappa) = 2\pi \times 0.19$ GHz $\approx \gamma$ when the cavity is far detuned. The measured $\Gamma(0)$ corresponds to an estimated lifetime of 35 ps compared to the natural SiV lifetime of 1.8 ns (*12*). These measurements give cavity QED parameters $\{g, \kappa, \gamma\} = 2\pi \times \{7.3, 48, 0.19\}$ GHz, corresponding to a cooperativity (the key cavity-QED figure of merit) $C = 4g^2/\kappa\gamma \sim 23$ (*16*). This order-of-magnitude improvement in SiV-cavity cooperativity over previous work (*12*, *20*) primarily results from the decreased cavity mode volume (*15*).

As is evident from Fig. 1C, SiV centers are subject to inhomogeneous broadening, resulting predominantly from strain within the device (*14*, *21*). This broadening is smaller than that of other solid-state emitters compared to their lifetime-limited linewidths (*5*, *10*, *11*). Indeed, the frequencies of some SiV centers within the same devices are nearly identical. We study the cavity-mediated interaction between a pair (SiV 1 and SiV 2 in Fig. 1) of such nearly-resonant SiV centers (SiV-SiV detuning $\delta = 2\pi \times 0.6$ GHz) coupled to the cavity in the dispersive regime, that is, with large SiV-cavity detuning ($\Delta = 2\pi \times 79$ GHz $> \kappa$, Fig. 2A). To identify resonances associated with individual SiV centers, we selectively ionize either SiV into an optically inactive charge state by applying a resonant laser at powers orders of magnitude higher than those used to probe the system (*16*). This allows measurement of each of the SiV centers' spectra individually, with the other parameters (such as Δ) fixed (Fig. 2A, gray data).

When both SiV centers are in the optically active charge state, the splitting between the resonances increases. The new resonances (Fig. 2A, black data) also display different amplitudes compared with the single-SiV resonances and are labeled as bright ($|S\rangle$) and dark ($|D\rangle$) states. The linewidths of $|S\rangle$ ($|D\rangle$) are also enhanced (suppressed) compared to those of the individual SiV centers (Fig. 2B, inset). At a cavity detuning of the opposite sign ($\Delta = 2\pi \times -55$ GHz), the sign of the energy splitting δ_{SD} between $|S\rangle$ and $|D\rangle$ is reversed (Fig. 2B). The observation that Δ affects δ_{SD} indicates that this effect arises from the cavity.

To understand these observations, we describe the system of two SiV centers coupled to a cavity mode using the Hamiltonian (*6*, *22*):

$$\hat{H}/\hbar = \omega_c \hat{a}^\dagger \hat{a} + \omega_1 \hat{\sigma}_1^\dagger \hat{\sigma}_1 + \omega_2 \hat{\sigma}_2^\dagger \hat{\sigma}_2 + \hat{a}^\dagger (g_1 \hat{\sigma}_1 + g_2 \hat{\sigma}_2) + \hat{a}(g_1^* \hat{\sigma}_1^\dagger + g_2^* \hat{\sigma}_2^\dagger)$$

---

[1]Department of Physics, Harvard University, Cambridge, MA 02138, USA. [2]Institute for Quantum Information and Matter and Thomas J. Watson, Sr., Laboratory of Applied Physics, California Institute of Technology, Pasadena, CA 91125, USA. [3]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. [4]Sandia National Laboratories, Albuquerque, NM 87185, USA. [5]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: lukin@physics.harvard.edu

where ω_i is the frequency of the ith SiV center and $\hat{a}$ and $\hat{\sigma}_i$ are the cavity photon annihilation and ith SiV center's electronic state lowering operators. Coherent evolution under $\hat{H}$ is modified by cavity (κ) and SiV (γ) decay and decoherence (*16*). In the dispersive regime, $\hat{H}$ yields an effective Hamiltonian for two resonant ($\delta = 0$) SiV centers (*6*, *22*): $\hat{H}_{\text{eff}}/\hbar = J(\hat{\sigma}_1\hat{\sigma}_2^\dagger + \hat{\sigma}_1^\dagger\hat{\sigma}_2)$ where $J = \frac{g^2}{\Delta}$ (in our system, $g_1 \approx g_2 \equiv g$). Thus, the two SiV centers undergo a flip-flop interaction at rate J mediated by the exchange of cavity photons (Fig. 2C) . This interaction hybridizes the two SiV centers, forming collective eigenstates from the SiV ground $|g\rangle$ and excited $|e\rangle$ states which, for $\delta = 0$, are $|S\rangle = \frac{1}{\sqrt{2}}(|eg\rangle + |ge\rangle)$ and $|D\rangle = 1/\sqrt{2}(|eg\rangle - |ge\rangle)$ and are split by $2J$ (Fig. 2D) (*6*). The symmetric superradiant state $|S\rangle$ has an enhanced coupling to the cavity of $\sqrt{2}g$ (making it "bright" in transmission) and an energy shift of $2J = 2g^2/\Delta$, whereas the antisymmetric combination $|D\rangle$ is completely decoupled from the cavity ("dark" in transmission) and has zero energy shift (*6*, *11*). As δ/J increases, $|D\rangle$ becomes visible and the individual SiV eigenstates are eventually recovered. The energy shift of state $|S\rangle$ is away from the cavity resonance, explaining the reversed energy difference δ_{SD} upon changing the sign of Δ (Fig. 2B) . By comparing the data in Fig. 2 to theory accounting for finite δ (Fig. 2, solid curves), the SiV-SiV interaction strength $J = 2\pi \times 0.6$ GHz is extracted. The splitting δ_{SD} (which is at least $2J$) is larger than the measured linewidths (for a single SiV, $\Gamma(\Delta = 79 \text{ GHz}) = 2\pi \times 0.4$ GHz), allowing these states to be spectrally resolved.

Next, the SiV center's long-lived electronic spin degree of freedom (*18*) is used to control the SiV-cavity transmission and two-SiV interaction. We apply a magnetic field to lift the degeneracy of the spin sublevels in the ground (spinstates $|\uparrow\rangle$ and $|\downarrow\rangle$) and optically excited ($|\uparrow'\rangle$ and $|\downarrow'\rangle$) states. The Zeeman shifts are different for each orbital state and depend on the magnitude and orientation of the field with respect to the SiV center's symmetry axis, yielding spectrally distinguishable spin-selective optical transitions at frequencies $\omega_\uparrow$ and $\omega_\downarrow$ (Fig. 3A). In general, the splitting between these frequencies is maximized for off-axis magnetic fields (*18*). Any off-axis magnetic field also prevents the optical transitions from being perfectly cycling, allowing the SiV center to be initialized into $|\uparrow\rangle$ by pumping at $\omega_\downarrow$ and vice versa (*23*, *24*). Coupling spin-selective transitions to the cavity mode enables high-contrast spin-dependent modulation of the cavity transmission (*25*).

We demonstrate this effect using an SiV center in the dispersive regime ($\Delta \sim 2\kappa$). Here, the optical transition linewidth is narrow, allowing us to resolve these transitions in a 9-kG field well-aligned with the SiV center's symmetry axis where the transitions are highly cycling (branching fraction ~$1 - 10^{-4}$) (*17*). The spin is initialized in $|\downarrow\rangle$ or $|\uparrow\rangle$ via optical pumping, and the system is probed in transmission. When the spin is in $|\downarrow\rangle$, the interaction with the probe at $\omega_\uparrow$ is negligible, and the probe is reflected by the detuned cavity (Fig. 3B, red curve). When the spin is in $|\uparrow\rangle$, single photons at frequency $\omega_\uparrow$ are transmitted via the SiV-like dressed state (blue curve) for a time (12 ms) determined by the cyclicity of the transition (*17*). The distribution of photon counts detected in a 7-ms window when the spin is initialized in $|\downarrow\rangle$ (red) versus $|\uparrow\rangle$ (blue) (Fig. 3C) is well-resolved, allowing the spin state to be determined in a single shot with 97% fidelity (*16*).

We also perform this experiment in the resonant-cavity regime and observe spin-dependent transmission switching with 80% contrast (*16*).

The combination of spin control, high-cooperativity coupling, and a small inhomogeneous distribution of SiVs enables controllable optically mediated interactions between multiple SiV centers. We focus on two SiV centers (SiV 1 and SiV 2 in Fig. 1) in the dispersive

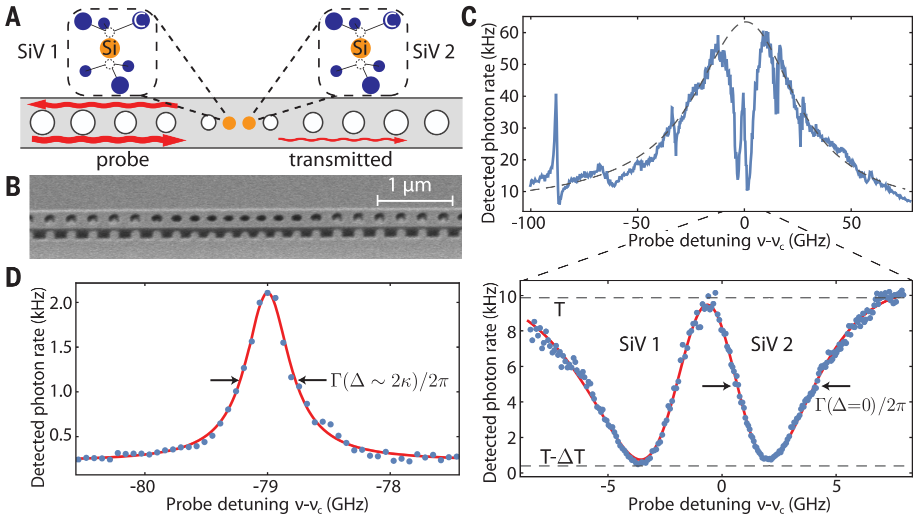


Fig. 1. High cooperativity SiV-photon interface. (**A**) Schematic of a diamond nanocavity containing two SiV centers. (**B**) Scanning electron micrograph of a nanocavity. (**C**) Transmission spectrum of the coupled SiV-cavity system (blue). The broad Lorentzian response of an empty cavity (dashed) is modulated by cavity-coupled SiVs. Near the cavity resonance (lower panel), two SiVs each result in greater than 95% extinction in transmission and are broadened by the Purcell effect [$\Gamma(\Delta = 0) = 2\pi \times 4.6$ GHz]. (**D**) In the dispersive regime ($\Delta = 2\pi \times 79$ GHz ~ 2κ), SiVs appear as narrow peaks in transmission [$\Gamma(\Delta = 2\pi \times 0.5$ GHz]. The solid lines in (D) and the lower panel of (C) are fits to a model (*16*).

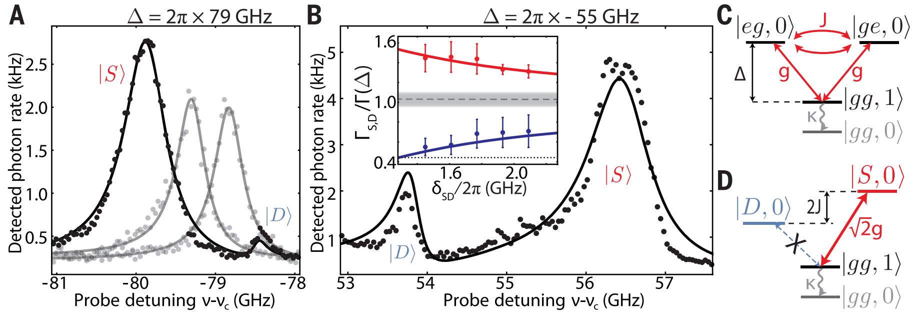


Fig. 2. Cavity mediated SiV-SiV interactions. (**A**) Transmission spectrum of two nearly resonant SiVs ($\delta = 2\pi \times 0.56$ GHz) at cavity detuning ($\Delta = 2\pi \times 79$ GHz). When both SiVs are coupled to the cavity, superradiant (bright) $|S\rangle$ and subradiant (dark) $|D\rangle$ states are formed (black). Spectra of noninteracting SiVs are shown in gray. (**B**) Transmission spectrum of the same SiVs at $\Delta = 2\pi \times -55$ GHz and $\delta = 2\pi \times 2$ GHz. Inset: ratio of $|S\rangle$ (red) and $|D\rangle$ (blue) state linewidths to the single-SiV linewidth at ($\Delta = 2\pi \times 79$ GHz) as a function of δ_{SD}. The resonance frequencies of these SiVs slowly drift as a result of spectral diffusion (*14*), allowing us to measure the linewidths of $|S\rangle$ and $|D\rangle$ at different δ_{SD}. The gray dashed line and shaded region are the average and standard deviation of the single-SiV linewidths. The dotted line corresponds to the SiV linewidth γ without Purcell enhancement, demonstrating that $|D\rangle$ at minimum δ_{SD} is almost completely decoupled from the cavity. Solid lines in (A) and (B) are predictions based on independently-measured SiV parameters (*16*). (**C**) Energy diagram of two SiV centers coupled to a cavity mode. Interaction via exchange of cavity photons at rate $J = g^2/\Delta$ leads to collective $|S\rangle$ and $|D\rangle$ eigenstates (**D**).

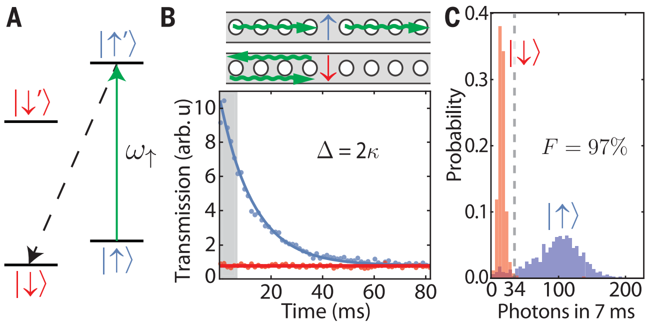


Fig. 3. Cavity-assisted spin initialization and readout. (**A**) Simplified level structure of the SiV in a magnetic field. An optical transition at frequency $\omega_\uparrow$ (green arrow) initializes the SiV spin into $|\downarrow\rangle$ by optical pumping via a spin-flipping transition (dashed line). (**B**) Spin-dependent optical switching in the dispersive regime. State $|\downarrow\rangle$ is not coupled to the probe at frequency $\omega_\uparrow$, which is therefore reflected (red). Initialization into $|\uparrow\rangle$ results in transmission of the probe field (blue). (**C**) Photon number distributions for transmission in the dispersive regime for initialization into $|\uparrow\rangle$ (blue) and $|\downarrow\rangle$ (red) are well-resolved (mean $n_\uparrow = 96$, $n_\downarrow = 16$) in a 7-ms window [gray region in (C)], demonstrating single-shot spin-state readout with 97% fidelity.

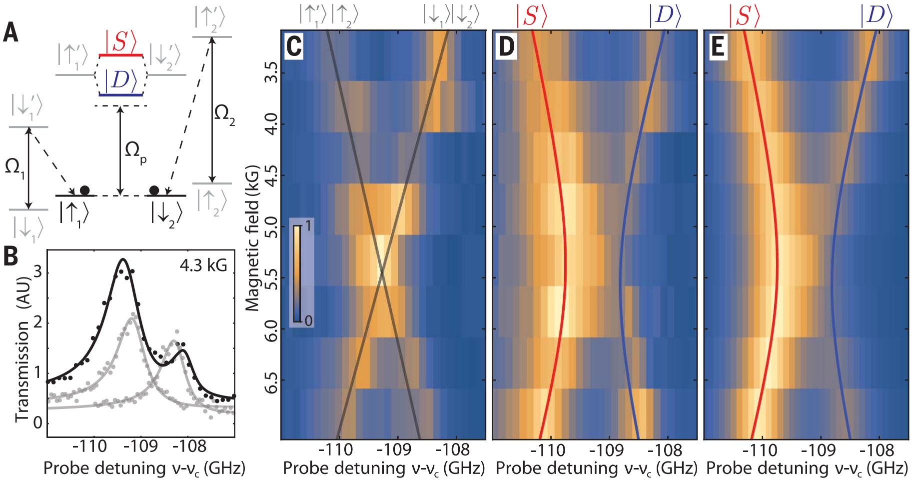


Fig. 4. Deterministic control of photon-mediated interactions via SiV spin states. (**A**) Energy diagram of two SiV centers in a magnetic field. The spins can be initialized in $|\uparrow_1\rangle|\downarrow_2\rangle$ with Ω_1 and Ω_2 and probed by sweeping Ω_p. When Ω_1 (Ω_2) is off, Ω_p is sufficiently strong to ensure initialization in $|\downarrow_1\rangle$ ($|\uparrow_2\rangle$) as it sweeps across the relevant SiV 1 (2) transition. (**B**) Spin-dependent transmission spectra at 4.3 kG. Gray: Spectra of single SiVs in the noninteracting combinations of spin states. Black: Spectrum of two interacting SiVs, demonstrating formation of $|S\rangle$ and $|D\rangle$. (**C**) Composite spectrum of the system at different two-SiV detunings, demonstrating a level crossing characteristic of noninteracting systems. The solid lines are the fitted single-SiV energies of $|\uparrow_1'\rangle|\uparrow_2\rangle$ and $|\downarrow_1\rangle|\downarrow_2'\rangle$ as a function of magnetic field. (**D**) An avoided crossing is visible in cavity transmission when the spins are prepared in the interacting state $|\uparrow_1\rangle|\downarrow_2\rangle$. (**E**) Predicted spectrum. The red and blue lines in (D) and (E) are predicted energies of $|S\rangle$ and $|D\rangle$ (*16*).

regime ($\Delta = 2\pi \times 109$ GHz) with $\{g1 \approx g_2, \kappa, \gamma_1 \approx \gamma_2\} = 2\pi \times \{7.3, 39, 0.5\}$ GHz ($C \approx 11$) and an initial two-SiV detuning $\delta = 2\pi \times 5$ GHz (*16*). We sweep the magnitude of a magnetic field oriented almost orthogonal to the SiV symmetry axis and tune transitions $|\uparrow_1\rangle \rightarrow |\uparrow_1'\rangle$ and $|\downarrow_2\rangle \rightarrow |\downarrow_2'\rangle$ (which have opposite Zeeman shifts) in and out of resonance (Fig. 4A). At each magnetic field, a continuous field Ω_1 or Ω_2 is used to optically pump either SiV 1 or SiV 2 into the spin state resonant with a weak probe field Ω_p measuring the transmission spectrum of the system, thus enabling control measurements where only one spin is addressed by Ω_p at a time (Fig. 4B, gray). The single-spin transmission spectra at each field are summed to form a composite spectrum of the two-SiV system (Fig. 4C), which displays an energy level crossing of the two SiV transitions characteristic of noninteracting systems.

Measurements were then made in the interacting regime by preparing the spins into $|\uparrow_1\rangle|\downarrow_2\rangle$ by simultaneously applying Ω_1 and Ω_2. The two-SiV transmission spectrum demonstrates the formation of superradiant and subradiant states (Fig. 4B, black) that exist only for this combination of spin states. Spectra as a function of applied magnetic field (Fig. 4D) demonstrate an avoided crossing arising from spin-dependent interactions between the two SiV centers (*6*). These experimental observations agree with an analysis based on a complete microscopic model of the system (*16*) including independently measured SiV-cavity parameters (Fig. 4E). For these experiments (Fig. 4), an active preselection sequence is used to stabilize the SiV spectral positions (*16*). Similar observations were reproduced in a separate device on the same chip (*16*).

This optically mediated interaction between quantum emitters could be used to realize key quantum information protocols, including cavity-assisted entanglement generation (*2*, *7*, *22*), efficient Bell-state measurements (*19*, *26*), and robust gates between emitters in distant cavities (*1*). To implement these schemes with high fidelity, qubits should be encoded in long-lived electronic spin states. Recent work has demonstrated that the SiV spin can be used as a long-lived quantum memory (*17*) that can be manipulated with microwave (*17*, *27*) and optical fields (*28*). Although the present work relies on magnetic field tuning of the SiV frequencies, recently developed techniques that include strain (*21*) or Raman tuning (*12*) can be applied to individual devices, allowing our approach to be extended to the chip scale. The infidelity associated with many deterministic cavity-mediated operations scales as $C^{-1/2}$ (*26*) or C^{-1} (*7*) and can therefore be mitigated with improved cooperativity. Although the cooperativity $C \sim 20$ achieved here is among the largest demonstrated in the optical domain, it can be improved by at least an order of magnitude by increasing the cavity Q/V and by reducing sources of spectral diffusion limiting γ (*16*). Alternatively, the cooperativity could be enhanced by using different quantum emitters, such as the GeV (*29*) or SnV (*30*) centers in diamond with higher quantum efficiencies (*29*). Near-unity fidelities can also be achieved with existing cooperativities using heralded schemes where errors can be suppressed via error detection with an auxiliary qubit (*26*). Furthermore, our system could be used to efficiently generate nonclassical states of light (*31*), which are useful in, for example, measurement-based quantum computing. On-chip integration and GHz-level bandwidths make our system well-suited for exploring potential applications in quantum networking, including the implementation of efficient quantum repeaters (*26*) and distributed quantum computing.

REFERENCES AND NOTES

1. J. I. Cirac, P. Zoller, H. J. Kimble, H. Mabuchi, *Phys. Rev. Lett.* **78**, 3221–3224 (1997).
2. A. Imamoğlu *et al.*, *Phys. Rev. Lett.* **83**, 4204–4207 (1999).
3. S. Welte, B. Hacker, S. Daiss, S. Ritter, G. Rempe, *Phys. Rev. X* **8**, 011018 (2018).
4. A. Wallraff *et al.*, *Nature* **431**, 162–167 (2004).
5. P. Lodahl, S. Mahmoodian, S. Stobbe, *Rev. Mod. Phys.* **87**, 347–400 (2015).
6. J. Majer *et al.*, *Nature* **449**, 443–447 (2007).
7. M. J. Kastoryano, F. Reiter, A. S. Sørensen, *Phys. Rev. Lett.* **106**, 090502 (2011).
8. B. Casabone *et al.*, *Phys. Rev. Lett.* **114**, 023602 (2015).
9. R. Reimann *et al.*, *Phys. Rev. Lett.* **114**, 023601 (2015).
10. A. Laucht *et al.*, *Phys. Rev. B* **82**, 075305 (2010).
11. H. Kim, D. Sridharan, T. C. Shen, G. S. Solomon, E. Waks, *Opt. Express* **19**, 2589–2598 (2011).
12. A. Sipahigil *et al.*, *Science* **354**, 847–850 (2016).
13. A. Sipahigil *et al.*, *Phys. Rev. Lett.* **113**, 113602 (2014).
14. R. E. Evans, A. Sipahigil, D. D. Sukachev, A. S. Zibrov, M. D. Lukin, *Phys. Rev. Appl.* **5**, 044010 (2016).

15. M. J. Burek *et al*., *Phys. Rev. Appl.* **8**, 024026 (2017).
16. Materials and methods are available as supplementary materials online.
17. D. D. Sukachev *et al*., *Phys. Rev. Lett.* **119**, 223602 (2017).
18. C. Hepp *et al*., *Phys. Rev. Lett.* **112**, 036405 (2014).
19. E. Waks, J. Vuckovic, *Phys. Rev. Lett.* **96**, 153601 (2006).
20. J. L. Zhang *et al*., *Nano Lett.* **18**, 1360–1365 (2018).
21. S. Meesala *et al*., *Phys. Rev. B* **97**, 205444 (2018).
22. S.-B. Zheng, G.-C. Guo, *Phys. Rev. Lett.* **85**, 2392–2395 (2000).
23. L. J. Rogers *et al*., *Phys. Rev. Lett.* **113**, 263602 (2014).
24. B. Pingault *et al*., *Phys. Rev. Lett.* **113**, 263601 (2014).
25. A. Javadi *et al*., *Nat. Nanotechnol.* **13**, 398–403 (2018).
26. J. Borregaard, P. Kómár, E. M. Kessler, M. D. Lukin, A. S. Sørensen, *Phys. Rev. A* **92**, 012307 (2015).
27. B. Pingault *et al*., *Nat. Commun.* **8**, 15579 (2017).
28. J. N. Becker *et al*., *Phys. Rev. Lett.* **120**, 053603 (2018).
29. M. K. Bhaskar *et al*., *Phys. Rev. Lett.* **118**, 223603 (2017).
30. T. Iwasaki *et al*., *Phys. Rev. Lett.* **119**, 253601 (2017).
31. S. E. Economou, N. Lindner, T. Rudolph, *Phys. Rev. Lett.* **105**, 093601 (2010).

ACKNOWLEDGMENTS

We thank D. Twitchen and M. Markham from Element Six for substrates, J. Borregaard and K. De Greve for discussions, and D. Perry for implantation assistance. **Funding:** Support was provided by the NSF, CUA, DoD/ARO DURIP, AFOSR MURI, ONR MURI, ARL, Vannevar Bush Faculty Fellowship, DoD NDSEG (M. K. B.), and NSF GRFP (B. M. and G. Z.). Devices were fabricated at the Harvard CNS (NSF ECCS-1541959). Implantation was performed at Sandia National Laboratories through the Center for Integrated Nanotechnologies, operated for the DOE-SC (contract DE-NA-0003525) by Sandia Corporation, a Honeywell subsidiary. **Author contributions:** R.E., M.K.B., D.S., C.N., and A.S. performed experiments and analyzed data. M.J.B. and B.M. designed and fabricated nanocavities. A.Z. and G.Z. assisted with experiments. E.B. perfomed implantation. H.P., M.L., and M.D.L. supervised experiments and analysis. All authors contributed to analysis and manuscript preparation. **Competing interests:** The authors declare no competing interests. The views expressed here do not necessarily represent the DOE or U.S. Government. **Data and materials availability:** All data are available in the manuscript or supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/662/suppl/DC1
Materials and Methods
Figs. S1 to S9
References (*32*–*36*)

26 June 2018; accepted 3 September 2018
Published online 20 September 2018
10.1126/science.aau4691

NANOMATERIALS

Controlled crack propagation for atomic precision handling of wafer-scale two-dimensional materials

Jaewoo Shim[1,2]*, Sang-Hoon Bae[1,2]*, Wei Kong[1,2]*, Doyoon Lee[1,2]*, Kuan Qiao[1,2], Daniel Nezich[3], Yong Ju Park[4], Ruike Zhao[1,5], Suresh Sundaram[6], Xin Li[6], Hanwool Yeon[1,2], Chanyeol Choi[1,2], Hyun Kum[1,2], Ruoyu Yue[7], Guanyu Zhou[7], Yunbo Ou[8], Kyusang Lee[1,2,9], Jagadeesh Moodera[8], Xuanhe Zhao[1], Jong-Hyun Ahn[4], Christopher Hinkle[7,10], Abdallah Ougazzaden[6], Jeehwan Kim[1,2,11,12]†

Although flakes of two-dimensional (2D) heterostructures at the micrometer scale can be formed with adhesive-tape exfoliation methods, isolation of 2D flakes into monolayers is extremely time consuming because it is a trial-and-error process. Controlling the number of 2D layers through direct growth also presents difficulty because of the high nucleation barrier on 2D materials. We demonstrate a layer-resolved 2D material splitting technique that permits high-throughput production of multiple monolayers of wafer-scale (5-centimeter diameter) 2D materials by splitting single stacks of thick 2D materials grown on a single wafer. Wafer-scale uniformity of hexagonal boron nitride, tungsten disulfide, tungsten diselenide, molybdenum disulfide, and molybdenum diselenide monolayers was verified by photoluminescence response and by substantial retention of electronic conductivity. We fabricated wafer-scale van der Waals heterostructures, including field-effect transistors, with single-atom thickness resolution.

Because of their in-plane stability and weak out-of-plane interaction, two-dimensional (2D) materials can be stacked together to form a multitude of device types with a broad spectrum of functionalities (*1*–*6*). Construction of 2D material–based heterostructures is often described as stacking Lego blocks (*4*). To tailor the 2D heterostructure characteristics for specific functionalities, it is essential to isolate 2D materials into monolayer films and stack them with monolayer precision. The most common method for assembling these 2D blocks is by using the adhesive-tape method (*7*), with which stacking of micrometer-scale flakes has been demonstrated (*4*). However, this method cannot reliably produce monolayer 2D crystals from bulk materials. The process becomes more complicated if the heterostructure design requires several different types of 2D material monolayers. Multiple monolayer flakes must be initially secured for each 2D material, which becomes extremely time consuming. Moreover, although the isolation of flakes into a nominal monolayer has been demonstrated, the lateral dimensions (hundreds of micrometers) are not sufficient to guarantee the fabrication of large-scale 2D heterostructures (*8*). In parallel, numerous efforts have been made to directly grow 2D heterostructures at the wafer scale (*6*, *9*, *10*). Recently, metal-organic chemical vapor deposition growths of wafer-scale monolayer 2D materials and their heterostructures have been successfully demonstrated for some transition metal dichalcogenides (TMDCs) at a specific growth condition (*11*, *12*).

We introduce a layer-resolved splitting (LRS) technique that can be universally applied to produce 2D material monolayers at the wafer scale. This method requires one short growth of thick 2D materials on the wafer at a relaxed growth condition and subsequently harvests the multilayers to individual monolayers through a wafer-scale splitting process. This method allows for high-throughput production of monolayer 2D materials with single-atom thickness precision for the fabrication of wafer-scale van der Waals (vdW) heterostructures. We demonstrated the wafer-scale LRS for various monolayers of 2D materials, including hexagonal boron nitride (h-BN), tungsten disulfide (WS_2), tungsten diselenide (WSe_2), molybdenum disulfide (MoS_2), and molybdenum diselenide ($MoSe_2$). These films are readily stackable for forming wafer-scale 2D heterostructures. We have revealed the underlying mechanics that allows precise control of crack propagation, enabling LRS of the 2D material multilayers into multiple individual monolayers. The wafer-scale monolayer of TMDCs after LRS exhibits substantial photoluminescence (PL) enhancement uniformly across a 5-cm wafer at a photon energy corresponding to its monolayer. Through this digital control of monolayer 2D materials, we demonstrate wafer-scale vdW heterostructures with single-atom thickness resolution. The heterostructure devices prepared by LRS and a quasi-dry stacking process exhibited substantial improvement in electrical and optical uniformity across the wafer. This finding will open up new opportunities for the 2D material research community, as it provides a reliable pathway to form wafer-scale 2D heterostructures with monolayer resolution.

The schematic of the LRS process (Fig. 1A) shows a thick 2D material with an arbitrary number of layers grown on sapphire. Because of the difficulty in controlling the nucleation of the 2D material, new nucleation sites inevitably appear on top of the initial nucleation layer on the wafer before full substrate coverage. Thus, growths of multilayer 2D materials result in irregular, discontinuous films at the top but leave uniform, continuous films underneath. Once the entire multilayer film is removed from the sapphire wafer, the continuous 2D material films on the bottom can be split into many monolayers with the LRS process. We designed the LRS process on the basis of differences in the interfacial toughness (Γ) of these materials. We used a 600-nm-thick nickel (Ni) film as an atomic-scale adhesive because the reported Γ between 2D materials and Ni ($\Gamma_{\text{2D-Ni}} \sim 1.4\ \text{J m}^{-2}$) (*13*) is three times that of the vdW interface between layers in 2D materials ($\Gamma_{\text{2D-2D}} = 0.45\ \text{J m}^{-2}$) (*14*). The Γ between 2D materials and sapphire ($\Gamma_{\text{2D-Sapphire}}$) has been empirically deduced (see supplementary materials and methods), and its average value ($\Gamma_{\text{2D-Sapphire}} = 0.26\ \text{J m}^{-2}$) is less than $\Gamma_{\text{2D-2D}}$. Application of a bending moment during the liftoff of a Ni/2D material stack on sapphire supplies elastic strain energy to the bottom interfaces (Fig. 1B). The strain energy per unit area is released upon delamination when the desired strain release rate (G) is reached. As shown in Fig. 1C, an external force creates a bending moment, resulting in a torque applied across a small distance in the sample that initiates spalling mode fracture where the cracks propagate downward due to mixed mode I and mode II fracture (*15*). 2D materials grown at the edge of the wafer are typically defective so that the crack propagation can be facilitated. Spalling mode fracture occurs as a result of the external bending moment applied. In addition to the opening mode stress (mode I) acting on the crack tip, a shear field (mode II) is created to guide the cracks into the bulk. Thus, exfoliation of Ni/2D material stacks

---

[1]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA, USA. [4]School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, Republic of Korea. [5]Department of Mechanical and Aerospace Engineering, The Ohio State University, Columbus, OH, USA. [6]School of Electrical and Computer Engineering, Georgia Institute of Technology, UMI 2958 GT-CNRS, GT-Lorraine, Metz, France. [7]Department of Materials Science and Engineering, University of Texas at Dallas, Richardson, TX, USA. [8]Department of Physics, and Plasma Science and Fusion Center, Massachusetts Institute of Technology, Cambridge, MA, USA. [9]Departments of Electrical and Computer Engineering and Materials Science Engineering, University of Virginia, Charlottesville, VA, USA. [10]Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN, USA. [11]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [12]Microsystem Technology Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: jeehwan@mit.edu

can separate the 2D-sapphire interface that has the weakest interfacial toughness and allow clean separation of 2D materials from the wafer.

We first applied the LRS technique to 4-nm-thick WS_2 grown on a sapphire wafer through vapor phase epitaxy, characterizing each step of the process. As shown in the Raman mapping of as-grown WS_2 on the sapphire wafer, the E^1_{2g} peak of the WS_2 indicates that the sapphire wafer is fully covered with WS_2 after the growth (Fig. 1D and fig. S1). Atomic force microscopy (AFM) topology taken from the top of the as-grown WS_2 (Fig. 1E) shows that a rough discontinuous surface morphology developed from uncontrollable nucleation during growth, which is typically observed for the as-grown 2D materials. The LRS process was initiated with an exfoliation step to remove the entire WS_2 stack from the sapphire wafer. A second exfoliation step isolated them into monolayer films. For the first exfoliation, a 600-nm-thick Ni film was deposited on multilayer WS_2, followed by application of thermal release tape on the Ni as a handler. Lifting off the tape/Ni stack successfully separated the weakest WS_2-sapphire interface, resulting in release of the entire WS_2 film from the substrate; no sign of WS_2 was detected from Raman mapping and x-ray photoelectron spectroscopy on the sapphire wafer after exfoliation (Fig. 1F and figs. S2 and S3). Thus, a moment applied by liftoff supplies the Ni/WS_2 stacks with strain energy sufficient to delaminate the weakest 2D-sapphire interface (*16–18*). Successful release of the entire WS_2 film left pristine WS_2 layers at the bottom. The bottom WS_2 layer was continuous and smooth with root mean square (RMS) roughness of 0.5 nm (an average value measured from 15 locations) (Fig. 1G) that resulted from complete merging of the nuclei of the initial layers.

To harvest the continuous WS_2 monolayer, we deposited a Ni layer on the bottom of the WS_2 film while retaining the top tape/Ni/WS_2 stack as exfoliated (Fig. 1A). In a manner similar to the act of peeling the Ni/WS_2 stack off of the sapphire substrate, we applied a moment from the top Ni to initiate spalling mode fracture for guiding the cracks downward. Because $\Gamma_{2D\text{-}Ni}$ is substantially higher than $\Gamma_{2D\text{-}2D}$, the cracks arriving near the bottom Ni propagated through the weaker WS_2-WS_2 interface directly above the bottom Ni layer (Fig. 1C). Thus, the Ni/WS_2 stack separated upon peeling, whereas the bottom Ni strongly adhered to the WS_2 monolayer, which left a monolayer of WS_2 on the bottom Ni layer.

We transferred this monolayer film onto an 20.3-cm Si wafer coated with 90 nm of SiO_2 to investigate the thickness and quality of the exfoliated monolayer WS_2 film (see fig. S4 for photographs of the entire LRS process). As shown in Fig. 2, A and B, the 5-cm wafer-scale WS_2 was transferred intact onto the 20.3-cm SiO_2/Si wafer without prominent wrinkle, folding, or ripple. The average thickness of the transferred WS_2 is 0.7 nm (average value obtained

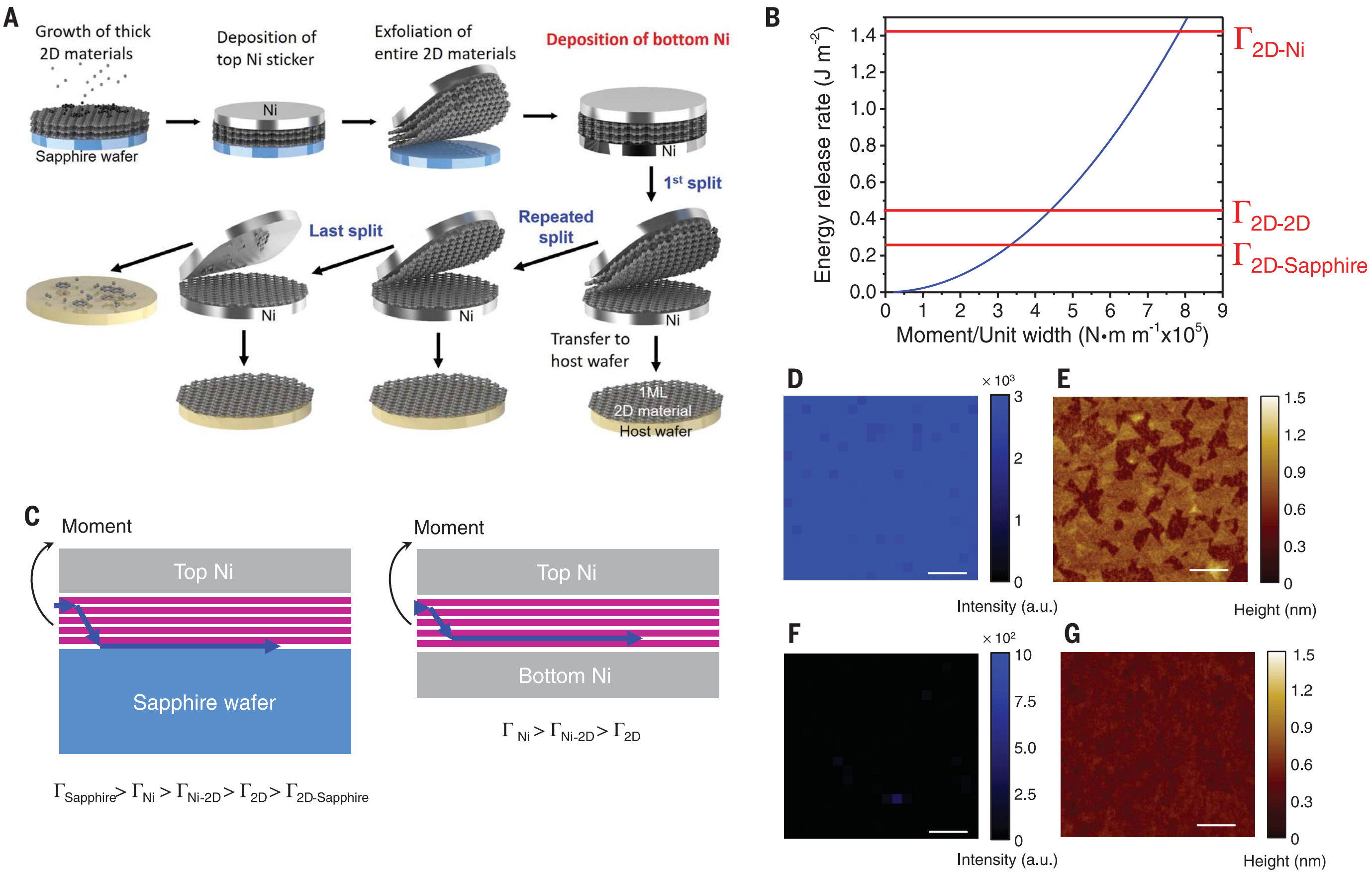


Fig. 1. Layer-resolved splitting (LRS) of 2D materials. (**A**) Schematic illustration explaining the LRS process for 2D materials. ML, monolayer. (**B**) Modeling of energy release rate according to applied moment. (Ni thickness: 600 nm). Γ, interfacial toughness. (**C**) Schematics of crack progression during LRS for initial exfoliation of entire 2D materials from sapphire wafer (left) and exfoliation of the bottom monolayer 2D material (right). (**D**) Raman intensity mapping at the E^1_{2g} peak (353 cm^{-1}) of WS_2 grown on a sapphire substrate, with laser wavelength and power of 532 nm and 2 mW in continuous waveform, respectively, where the spatial resolution is 2 μm. a.u., arbitrary units. (**E**) AFM topology taken from the top of as-grown 4-nm-thick WS_2 on the sapphire wafer. (**F**) Raman mapping image showing the intensity of the E^1_{2g} peak (353 cm^{-1}) on sapphire substrate after exfoliation of the WS_2 layer, with laser wavelength and power of 532 nm and 2 mW, respectively. (**G**) AFM topology taken from the bottom of WS_2 layer after exfoliation. Scale bars for Raman mapping images and AFM topology images are 2 μm and 50 nm, respectively.

from scans on 10 locations) (Fig. 2C). The exact value for the monolayer is 0.616 nm (*19*), and we speculate that the greater observed thickness arose from tip-sample interactions under noncontact measurement mode. Successful isolation of the WS_2 monolayer was confirmed by the substantial enhancement of the peak intensity of the PL spectra (Fig. 2D) at its direct gap of 1.99 eV, as compared with the weak and wide PL characteristic of a thick WS_2 layer at its indirect gap of 1.97 eV (*20*–*25*). Also, the PL intensity was not further degraded (fig. S5),

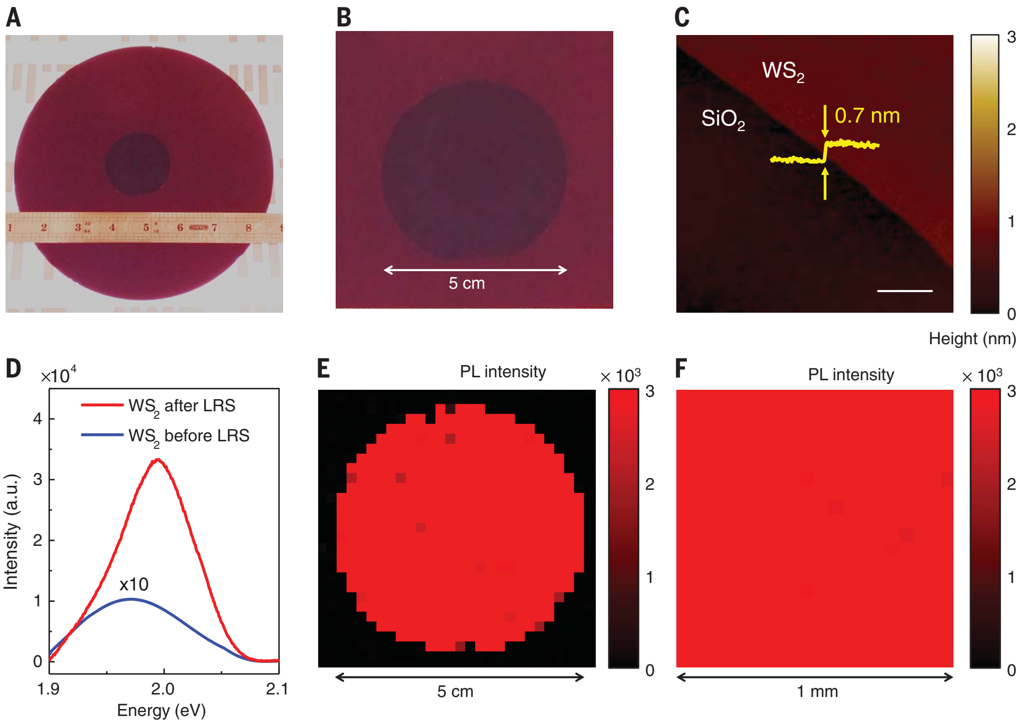


Fig. 2. Wafer-scale monolayer 2D material obtained by LRS process. (**A**) Optical image of 5-cm wafer-scale WS_2 monolayer on 20.3-cm oxidized Si wafer obtained through the LRS process. (**B**) Macrograph of 5-cm wafer-scale WS_2 monolayer. (**C**) AFM image and height profile of the WS_2 monolayer transferred on SiO_2/Si wafer. Scale bar: 1 μm. (**D**) Representative PL spectra of as-exfoliated 4-nm-thick WS_2 (blue line) and monolayer WS_2 obtained by LRS process (red line), in which the PL spectrum for 4-nm-thick WS_2 is multiplied by a factor of 10 to show the clear peak position. (**E**) Wafer-scale PL mapping image at 1.99 eV of WS_2 on the SiO_2/Si wafer, where the spatial resolution is 2 mm. (**F**) Large-scale (1 mm by 1 mm) PL intensity map from 1.99 eV for WS_2 on SiO_2, where the spatial resolution is 50 μm (see fig. S6 for 1-μm resolution map). All PL spectra were taken at the same laser power (2 mW) and wavelength (532 nm).

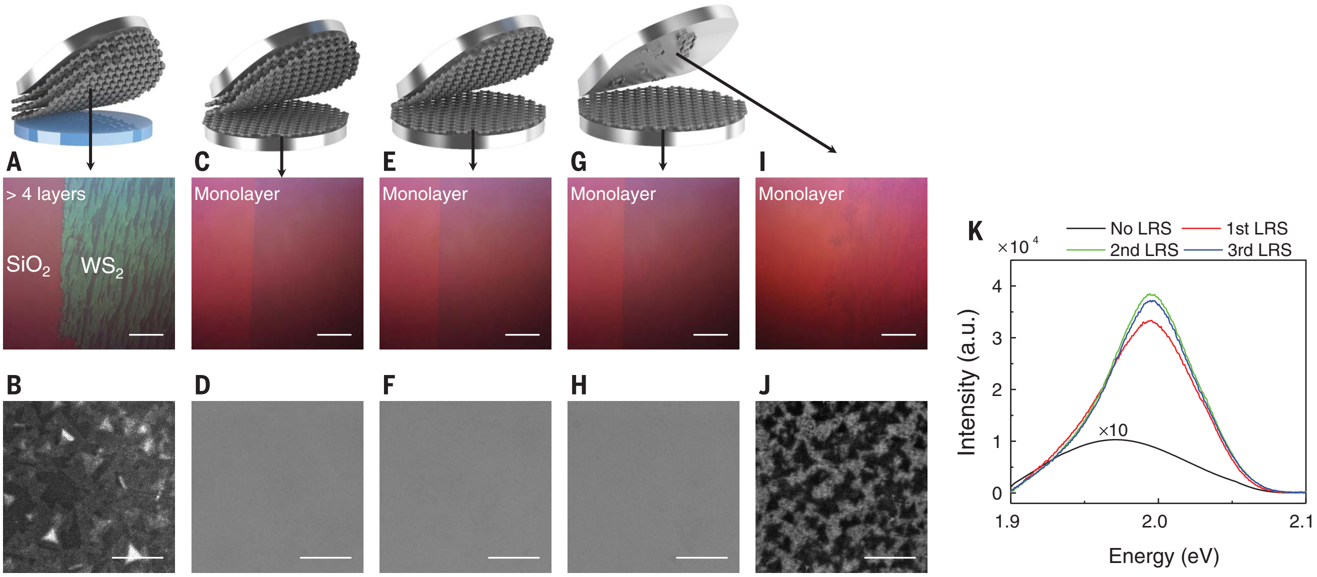


Fig. 3. Split of thick 2D materials into many monolayers via LRS process and their characterization. (**A**, **C**, **E**, **G**, and **I**) Optical micrographs and (**B**, **D**, **F**, **H**, and **J**) plan-view SEM images for as-exfoliated thick WS_2 [(A) and (B)] and monolayers of WS_2 obtained by the first [(C) and (D)], second [(E) and (F)], third [(G) and (H)], and last [(I) and (J)] LRS processes, respectively. The monolayers are all transferred on 90-nm-thick SiO_2/Si substrates. (**K**) Representative PL spectra for as-exfoliated thick WS_2 (black line) multiplied by a factor of 10 to show the clear peak position, as well as monolayers of WS_2 obtained by the first (red line), second (green line), and third (blue line) LRS processes. All PL spectra were taken at the same laser power (2 mW) and wavelength (532 nm). Scale bars for optical microscopy images and plan-view SEM images are 50 μm and 400 nm, respectively.

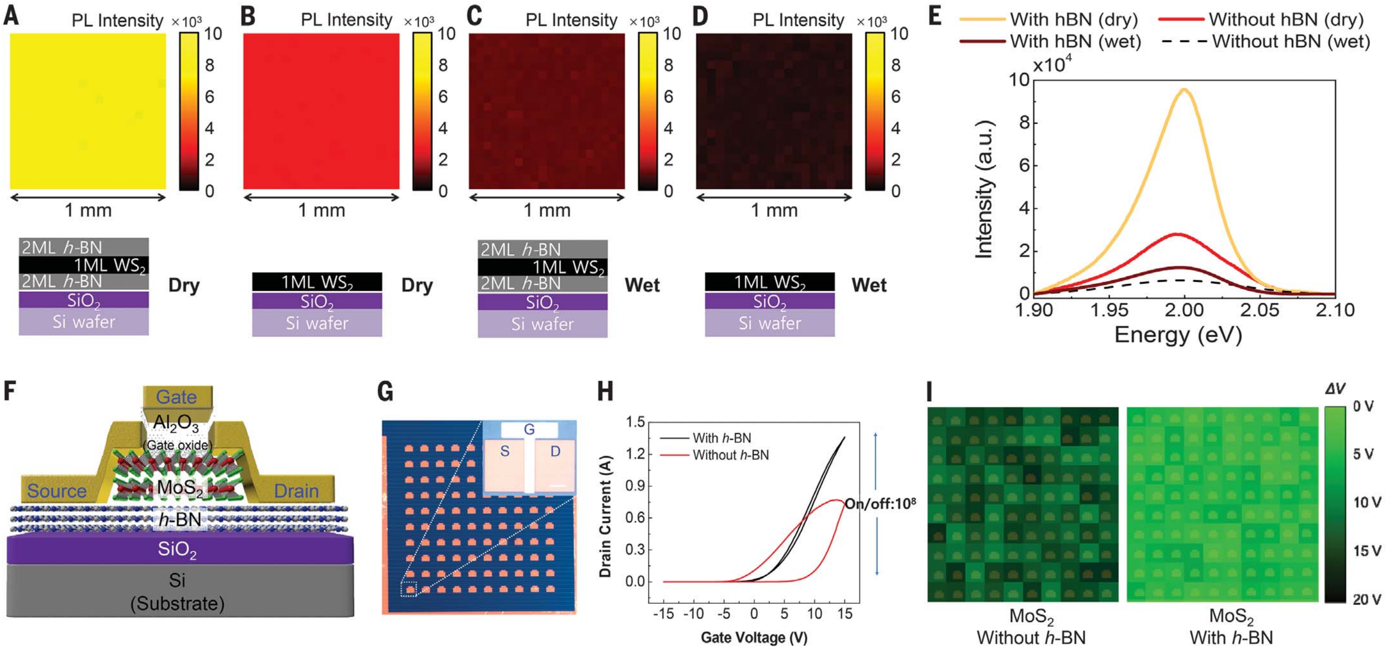


Fig. 4. Wafer-scale 2D heterostructures. PL mapping images at 1.99 eV of (**A**) a double-layer (2 ML) h-BN/monolayer (1 ML) WS_2/2 ML h-BN heterostructure fabricated by quasi-dry stacking, (**B**) 1 ML WS_2 on SiO_2 made by quasi-dry transfer, (**C**) a 2 ML h-BN/1 ML WS_2/2 ML h-BN heterostructure formed by wet stacking, and (**D**) 1 ML WS_2 on SiO_2 made by wet stacking. (**E**) Representative PL spectra of all different structures fabricated by different methods. All PL spectra were taken at the same laser power (2 mW) and wavelength (532 nm). (**F**) Schematic of a MoS_2-based field effect transistor (FFT). (**G**) Macrograph of 10-by-10 FET arrays integrated on a SiO_2/Si wafer with 1-cm–by–1-cm size. Inset shows micrograph of an individual device. Scale bar: 100 μm. The device area is defined by the gap between source (S) and drain (D) electrodes. G, gate. (**H**) Representative drain current–gate voltage (I_D–V_G) characteristics of MoS_2-based FETs at drain voltage (V_{DS}) = 1 V. (**I**) 2D color maps of hysteresis voltage extracted from I_D–V_G curves at V_{DS} = 1 V in transistor arrays made without h-BN (left) and with h-BN (right).

implying that WS_2 was already degraded before the process, given the prolonged time period after growth, which is typically observed in TMDCs after air exposure (*26*). Moreover, the strong PL intensity originating from the monolayer isolation was uniform across the entire 5-cm wafer area (spatial resolution of 2 mm) (Fig. 2E). PL mapping with greater spatial resolution further confirmed uniform monolayer thickness obtained by LRS (see Fig. 2F for a 50-μm resolution map on a 1-mm–by–1-mm area and fig. S6 for 1-μm resolution maps on 20-μm–by–20-μm area). Wafer-scale monolayer thickness was also confirmed by mapping the PL peak position where peaks are all concentrated at its direct gap of 1.99 eV (fig. S7) (*20*–*25*).

All LRS WS_2 films were transferred with a quasi-dry process, in which the WS_2 split from the multilayers was directly dry-bonded to the SiO_2-coated Si wafer, followed by etching to remove the Ni (*16*, *17*, *27*). The average RMS roughness was 0.5 nm after LRS measured from 15 different spots in each material (see fig. S8). We speculate that perfect atomic smoothness (RMS < 0.3 nm) was not obtained because of the polycrystalline nature of the films with a grain size of ~60 nm (fig. S9). Larger-scale morphological inspection of monolayer 2D materials after LRS with Raman microscopy and laser scanning confocal microscopy revealed no prominent surface defects and macroscopic deformation such as wrinkle, folding, or ripple (figs. S10 and S11). More importantly, LRS did not substantially degrade electrical quality of the 2D materials. The average Hall mobility measured from few layers before LRS was 106.8 cm^2 V^{-1} s^{-1} compared with 89.5 cm^2 V^{-1} s^{-1} for the monolayer after LRS (fig. S12) (*28*).

The LRS process was repeated to harvest additional continuous monolayers until the split layer was no longer continuous. In parallel, an entire 4-nm-thick WS_2 film was transferred from another sapphire substrate onto a 90-nm-thick SiO_2-coated Si wafer as a reference. Optical micrographs and scanning electron microscopy (SEM) images show the split layer's nonuniform thickness (Fig. 3, A and B). Repetition of LRS yielded continuous production of uniform continuous monolayer WS_2 (see fig. S13 for Raman spectra showing rightward shift of the E^1_{2g} peak compared with that of as-exfoliated 4-nm-thick WS_2 films) through three cycles (Fig. 3, C to H). After the third cycle, residual WS_2 was observed as discontinuous triangle domains on the top Ni film (Fig. 3, I and J), indicating that the LRS process has reached the final top layer where nucleated islands failed to merge during growth. Furthermore, we measured strong direct-gap emission peaks (at 1.99 eV) only from the WS_2 obtained through the first three cycles (Fig. 3K).

We applied our LRS technique to other 2D materials, including $MoSe_2$, WSe_2, MoS_2, and h-BN, and confirmed successful monolayer splitting for all. As shown in fig. S14, 3-nm-thick h-BN grown on sapphire was split into three h-BN monolayers. In addition, as shown in fig. S15, monolayer isolation of $MoSe_2$, WSe_2, and MoS_2 after LRS was confirmed by PL measurement (see also figs. S16 to S18 for optical microscope images and Raman spectra of $MoSe_2$, WSe_2, and MoS_2 before and after the LRS process). We also characterized these monolayers with AFM, confocal Raman microscopy, and laser scanning confocal microscopy (figs. S8, S10, and S11).

We then fabricated various 2D heterostructures. First, we fabricated WS_2/h-BN heterostructures with the monolayer WS_2 sandwiched between h-BNs to investigate the effect of wafer-scale encapsulation of TMDCs. Reduction of carrier scattering by a surface optical phonon has been observed in the flakes of h-BN/TMDC/h-BN heterostructures (*29*). Wafer-scale h-BN/TMDC/h-BN heterostructures were fabricated with LRS together with the quasi-dry transfer process. Monolayer h-BN prepared by LRS was transferred twice on a SiO_2/Si wafer, followed by monolayer transfer of WS_2 on top, and capped by transferring h-BN twice for encapsulation (see fig. S19 for the schematic description of 2D heterostructure via LRS process). We observed a substantial enhancement of PL intensity uniformly across the wafer (1-mm–by–1-mm PL map shown in Fig. 4A and measured across the wafer with same PL results), as compared with that of a WS_2 monolayer quasi-dry transferred on SiO_2 (see Fig. 4B).

When we fabricated h-BN/WS_2/h-BN heterostructures by using a wet stacking process in which 2D materials were scooped from a solution (see materials and methods for details of the wet stacking process), we observed a substantially weakened PL response compared to that of heterostructures prepared by the quasi-dry stacking process (Fig. 4C), which were attributed to interference coming from poly (methyl methacrylate) residues at the interface (fig. S20). Representative PL spectra for monolayer WS_2 on SiO_2 and double-layer h-BN/monolayer WS_2/double-layer h-BN prepared by quasi-dry and wet stacking are shown in Fig. 4E. Substantial degradation of the PL intensity was observed for the wet-stacked h-BN/WS_2 heterostructure even compared with quasi-dry transferred WS_2 on SiO_2. A 15-fold enhancement of PL intensity was observed for h-BN/WS_2/h-BN quasi-dry stacks compared with that for WS_2 wet stacks on SiO_2 (Fig. 4D).

Our study revealed wafer-scale enhancement in optoelectronic quality of monolayer WS_2 by wafer-scale h-BN encapsulation that should permit fabrication of large-scale optoelectronic devices based on vdW heterostructures. Our quasi-dry stacking method is the key to achieving optoelectronic enhancement in TMDCs by h-BN encapsulation, as the process preserves the quality of the 2D materials. The optical quality of the TMDCs was not severely affected, although the roughness of the surface slightly increases with increasing layer numbers in the heterostructures (fig. S21), possibly because of the presence of nanometer-scale Ni residues after the quasi-dry stacking process (*17*).

We also fabricated arrays of 2D heterostructure devices at the wafer scale. These heterostructure arrays exhibited uniformly enhanced device performance across the wafer. LRS-fabricated MoS_2 and h-BN multilayers were quasi-dry stacked on a SiO_2/Si wafer to form field effect transistors (FETs). Deleterious trap charges from the substrate were avoided by using h-BN (see Fig. 4F for the structure of our 2D heterostructure device) (*30*–*32*). Figure 4G shows an optical image of the 10-by-10 arrays of MoS_2 FETs on a 1-cm^2 wafer. Although FETs both with and without h-BN exhibited high on-off ratios of $>10^7$ (fig. S22), the FETs without h-BN display very large hysteresis in their drain current–gate voltage sweep, which is detrimental to their transistor operation. However, substantial suppression of hysteresis has been observed in FETs with h-BN (Fig. 4H). This reduction of hysteresis was obtained uniformly across the arrays in the wafer. As shown in the map of hysteresis of FETs with and without h-BN, the FET arrays with h-BN showed excellent reduction in hysteresis during a gate voltage sweep applied uniformly across the wafer (Fig. 4I).

The merit of the quasi-dry stacking process was also highlighted by fabricating arrays of WSe_2/graphene vertical transistors (see fig. S23 for an optical image of the 8-by-8 arrays of WSe_2/graphene vertical transistors made on a 1-cm–by–1-cm wafer). As shown in the map of on-off ratios of vertical transistors (fig. S23), our transistor arrays formed by quasi-dry stacking showed excellent device-to-device uniformity with 9.6% variation (see fig. S24 for the statistical distribution). Vertical transistor arrays fabricated by the wet-stacking process had a device-to-device variation of 26%.

We have demonstrated that the LRS technique can precisely isolate monolayers of 2D materials at the wafer scale from multilayers of 2D materials, thus allowing wafer-scale integration of 2D materials. LRS permits high-throughput manufacturing of wafer-scale monolayers of various 2D materials from thick 2D material films grown at relaxed growth conditions. We have demonstrated various 2D heterostructure devices at the wafer by applying LRS together with quasi-dry stacking. These heterostructure devices exhibit uniform device performances across the wafer. This high-throughput manufacturing of 2D heterostructures will become the stepping stone for commercialization of 2D material-based devices.

REFERENCES AND NOTES

1. T. Georgiou *et al.*, *Nat. Nanotechnol.* **8**, 100–103 (2013).
2. L. Britnell *et al.*, *Science* **340**, 1311–1314 (2013).
3. J. Shim *et al.*, *Nat. Commun.* **7**, 13413 (2016).
4. A. K. Geim, I. V. Grigorieva, *Nature* **499**, 419–425 (2013).
5. C. H. Lee *et al.*, *Nat. Nanotechnol.* **9**, 676–681 (2014).
6. Y. Gong *et al.*, *Nat. Mater.* **13**, 1135–1142 (2014).
7. K. S. Novoselov *et al.*, *Science* **306**, 666–669 (2004).
8. S. B. Desai *et al.*, *Adv. Mater.* **28**, 4053–4058 (2016).
9. S. M. Eichfeld *et al.*, *ACS Nano* **9**, 2080–2087 (2015).
10. Y. C. Lin *et al.*, *Nat. Commun.* **6**, 7311 (2015).
11. K. Kang *et al.*, *Nature* **520**, 656–660 (2015).
12. K. Kang *et al.*, *Nature* **550**, 229–233 (2017).
13. Z. Xu, M. J. Buehler, *J. Phys. Condens. Matter* **22**, 485301 (2010).
14. T. Björkman, A. Gulans, A. V. Krasheninnikov, R. M. Nieminen, *Phys. Rev. Lett.* **108**, 235502 (2012).
15. Z. Suo, J. W. Hutchinson, *Int. J. Solids Struct.* **25**, 1337–1353 (1989).
16. J. Kim *et al.*, *Science* **342**, 833–836 (2013).
17. S. H. Bae *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4082–4086 (2017).
18. H. Rydberg *et al.*, *Phys. Rev. Lett.* **91**, 126402 (2003).
19. A. Molina-Sánchez, L. Wirtz, *Phys. Rev. B* **84**, 155413 (2011).
20. H. R. Gutiérrez *et al.*, *Nano Lett.* **13**, 3447–3454 (2013).
21. J. Park *et al.*, *Nanoscale* **7**, 1308–1313 (2015).
22. S. Sasaki *et al.*, *Appl. Phys. Express* **9**, 071201 (2016).
23. N. J. Huo, Y. J. Yang, J. B. Li, *J. Semiconductors* **38**, 031002 (2017).
24. A. L. Elías *et al.*, *ACS Nano* **7**, 5235–5242 (2013).
25. F. Lan *et al.*, *Nanomaterials* **8**, 100 (2018).
26. Z. He *et al.*, *ACS Nano* **10**, 5847–5855 (2016).
27. Y. Kim *et al.*, *Nature* **544**, 340–343 (2017).
28. H. Schmidt, F. Giustiniano, G. Eda, *Chem. Soc. Rev.* **44**, 7715–7736 (2015).
29. S. M. Kim *et al.*, *Nat. Commun.* **6**, 8662 (2015).
30. G. H. Lee *et al.*, *ACS Nano* **9**, 7019–7026 (2015).
31. M. Choi *et al.*, *Sci. Adv.* **4**, s8721 (2018).
32. M. Y. Chan *et al.*, *Nanoscale* **5**, 9572–9576 (2013).

ACKNOWLEDGMENTS

Funding: J.K. acknowledges support from NSF grant no. CMMI-1825731. K.L. acknowledges support from NSF grant no. CMMI-1825256. A.O., S.S., and X.L. acknowledge the French National Research Agency for the cofunding of the h-BN study under the "GANEX" Laboratory of Excellence "LABEX" project. D.N. acknowledges the MIT Lincoln Laboratory Technology Office and Assistant Secretary of Defense for Research and Engineering for funding the growth of wafer-scale WS_2 films. J.M. acknowledge support from NSF grant no. DMR-1700137, ONR grant no. N00014-16-1-2657, and the STC Center for Integrated Quantum Materials under NSF grant no. DMR-1231319. Distribution statement A: Approved for public release. Distribution is unlimited. This material is based on work supported by the Assistant Secretary of Defense for Research and Engineering under Air Force contract no. FA8721-05-C-0002 and/or FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Assistant Secretary of Defense for Research and Engineering. G.Z. and C.H. acknowledge that this work was supported in part by NEWLIMITS, a center in nCORE, a Semiconductor Research Corporation (SRC) program sponsored by NIST through award no. 70NANB17H041. J.K. thanks J.-H. Park of SKKU for fruitful discussion and support. J.K. acknowledges the MIT-MI one-on-one project for partial support. **Author contributions:** J.S., S.-H.B., W.K., and D.L. contributed equally to this work. J.K. conceived the idea, designed experiments, and directed the team. J.S., S.-H.B., W.K., D.L., K.Q., H.Y., C.C., H.K., and K.L. performed fabrication and characterization of all samples and wrote the manuscript. R.Z., Y.O., J.M., and X.Z. contributed to the computational model. D.N. worked on the growth of WS_2 on sapphire. S.S., X.L., and A.O. worked on the growth of h-BN on sapphire. R.Y., G.Z., and C.H. worked on the growth of WSe_2 on sapphire. J.-H.A., and Y.J.P. worked on the growth of MoS_2 on sapphire and transistor devices. All authors contributed to the discussion and analysis of the results. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions of this study are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/665/suppl/DC1
Materials and Methods
Figs. S1 to S24
References (*33*, *34*)

7 April 2018; resubmitted 11 August 2018
Accepted 21 September 2018
Published online 11 October 2018
10.1126/science.aat8126

ORGANIC CHEMISTRY

Enantiodivergent Pd-catalyzed C–C bond formation enabled through ligand parameterization

Shibin Zhao[1,2]*, Tobias Gensch[3]*, Benjamin Murray[1,2], Zachary L. Niemeyer[3], Matthew S. Sigman[3]†, Mark R. Biscoe[1,2]†

Despite the enormous potential for the use of stereospecific cross-coupling reactions to rationally manipulate the three-dimensional structure of organic molecules, the factors that control the transfer of stereochemistry in these reactions remain poorly understood. Here we report a mechanistic and synthetic investigation into the use of enantioenriched alkylboron nucleophiles in stereospecific Pd-catalyzed Suzuki cross-coupling reactions. By developing a suite of molecular descriptors of phosphine ligands, we could apply predictive statistical models to select or design distinct ligands that respectively promoted stereoinvertive and stereoretentive cross-coupling reactions. Stereodefined branched structures were thereby accessed through the predictable manipulation of absolute stereochemistry, and a general model for the mechanism of alkylboron transmetallation was proposed.

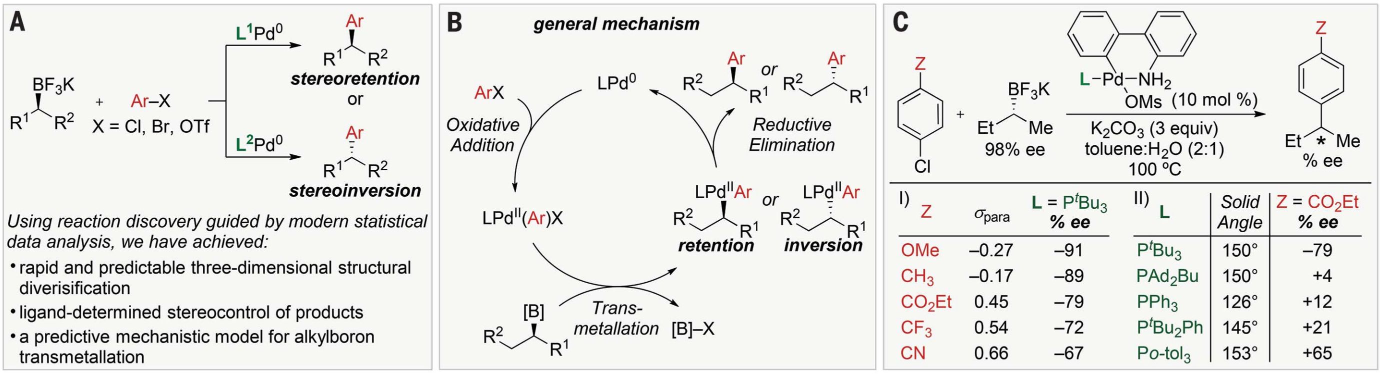
Palladium-catalyzed cross-coupling reactions have revolutionized the construction of $C(sp^2)$–$C(sp^2)$ bonds. Among these cross-coupling processes, the Suzuki-Miyaura reaction has found particularly broad application owing to its extensive reaction scope, as well as the stability, availability, and low toxicity of organoboron reagents (*1*). The 2010 Nobel Prize in chemistry was awarded, in part, to recognize the transformative impact of the Suzuki cross-coupling reaction on chemical synthesis. However, although $C(sp^2)$–$C(sp^2)$ bond construction is now considered routine using the Suzuki reaction, extension of this process to the formation of $C(sp^3)$–$C(sp^2)$ bonds using alkylboron nucleophiles remains a considerable challenge. Of particular interest, a variant using secondary alkylboron nucleophiles with predictable and controllable stereospecificity would establish a powerful synthetic strategy to access molecular geometries with precise three-dimensional control, expanding the exceptional capabilities of the Suzuki reaction (Fig. 1A).

Many efforts have focused on the use of enantioenriched secondary alkylboron nucleophiles in Suzuki cross-coupling reactions (*2*–*4*). Considerable limitations remain because of slow transmetallation of the highly covalent and sterically congested $C(sp^3)$–B bond in these reagents, as well as the propensity of the resulting Pd-alkyl species to undergo β-hydride elimination-reinsertion sequences, which can result in isomerization of the alkyl group and racemization of the stereocenter. To circumvent prohibitively slow transmetallation, as well as competing β-hydride elimination-reinsertion pathways, most stereospecific Suzuki reactions have required the use of secondary alkylboron nucleophiles that are electronically activated via inclusion of a $C(sp^2)$ α-carbon, an α-heteroatom, and/or a strongly coordinating β-carbonyl group (*5*–*17*). In addition, alkylboron nucleophiles can undergo transmetallation via either stereoretentive or stereoinvertive pathways depending on the nature of the substrate, catalyst, and/or reaction conditions. In many cases, the factors controlling the dominant mechanism of transmetallation are not understood (Fig. 1B). Thus, a predictive stereochemical model for transmetallation of alkylboron reagents remains elusive.

Recently, we reported a stereospecific Pd-catalyzed cross-coupling reaction using unactivated secondary alkylboron nucleophiles (*18*). With P^tBu_3 (^tBu, *tert*-butyl) as a supporting ligand, enantioenriched arylation products were obtained with transmetallation proceeding primarily via a stereoinvertive mechanism. Whereas several enlightening mechanistic studies have recently been conducted on the transmetallation of arylboron nucleophiles (*19*–*23*), these studies have not addressed the transmetallation of alkylboron nucleophiles in $C(sp^3)$–$C(sp^2)$ bond-forming processes (*24*, *25*). Thus, unactivated alkylboron nucleophiles constitute an attractive starting point from which to investigate the reaction parameters most influential to the mechanism of alkylboron transmetallation. This mechanistic work should simultaneously facilitate the development of new synthetic methods to rationally incorporate or manipulate stereocenters via cross-coupling strategies. To this end, we report a study using predictive statistical models (*26*, *27*)

---

[1]Department of Chemistry, The City College of New York, 160 Convent Avenue, New York, NY 10031, USA. [2]Ph.D. Program in Chemistry, The Graduate Center of the City University of New York, 365 Fifth Avenue, New York, NY 10016, USA. [3]Department of Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, UT 84112, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: sigman@chem.utah.edu (M.S.S.); mbiscoe@ccny.cuny.edu (M.R.B.)



Using reaction discovery guided by modern statistical data analysis, we have achieved:

- rapid and predictable three-dimensional structural diverisification
- ligand-determined stereocontrol of products
- a predictive mechanistic model for alkylboron transmetallation

I)

Z	σ_{para}	L = P^tBu_3 **% ee**
OMe	–0.27	–91
CH_3	–0.17	–89
CO_2Et	0.45	–79
CF_3	0.54	–72
CN	0.66	–67

II)

L	*Solid Angle*	Z = CO_2Et **% ee**
P^tBu_3	150°	–79
PAd_2Bu	150°	+4
PPh_3	126°	+12
P^tBu_2Ph	145°	+21
P*o*-tol_3	153°	+65

Fig. 1. Reaction development. (**A**) Enantiodivergent Suzuki reactions of secondary alkylboron nucleophiles. (**B**) General mechanism. (**C**) Initial investigation of substrate and ligand influences on stereoselectivity. A positive % ee value indicates net retention; a negative % ee value indicates net inversion. * indicates an enantioenriched stereocenter. L, ligand; Tf, trifluoromethanesulfonyl; Ar, aryl; Et, ethyl; Me, methyl; Ms, methylsulfonyl; *o*-tol, ortho-tolyl.

to relate phosphine ligand properties to stereochemical outcomes obtained from Pd-catalyzed Suzuki reactions of unactivated enantioenriched secondary alkylboron nucleophiles and aryl electrophiles. With statistical models that rely on a next-generation set of molecular descriptors, we achieved a stereoretentive Pd-catalyzed cross-coupling reaction of such nucleophiles. Furthermore, we have identified an improved ligand for the stereoinvertive variant, thus enabling an entirely ligand-controlled enantiodivergent process from a single-enantiomer organoboron nucleophile (*28*). Our statistical models also provide compelling evidence that each transmetallation pathway is intimately tied to specific electronic properties of the supporting ligand, which serves as a predictive guide to the mechanism of alkylboron transmetallation to palladium.

Initial investigations using electronically differentiated aryl chlorides with enantioenriched ^{s}BuBF$_3$K (^{s}Bu, *sec*-butyl) revealed a trend correlating diminished stereofidelity with the use of more electron-deficient coupling partners (section I of Fig. 1C). This observation suggested that subtle electronic effects could influence the mechanism of transmetallation and the resulting stereochemical outcome. Additionally, when the phosphine ligand was varied in an initial screen with a common aryl chloride electrophile, a considerable change in the reaction outcome from stereoinversion to stereoretention was found (section II of Fig. 1C). No obvious correlation was observed between these results and the steric properties (solid angle) of the ligand. Taken together, these outcomes were difficult to interpret and inspired the use of ligand parameterization tools to provide a platform for both predictive ligand performance and mechanistic interrogation.

An expanded inventory of common phosphines with varied properties was evaluated in the Suzuki reaction of enantioenriched ^{s}BuBF$_3$K and ethyl 4-chlorobenzoate. This dataset was then subjected to correlation analysis of phosphine structural features with the stereochemical outcomes as well as the ratios of branched:linear products in these reactions. We devised a workflow and universal parameter set to describe the catalyst properties from the phosphine itself (*29*–*32*). The workflow was initiated by performing a molecular mechanics (MM) conformational search to reveal representative low-energy conformers (Fig. 2A). Next, geometry optimization of the conformers using density

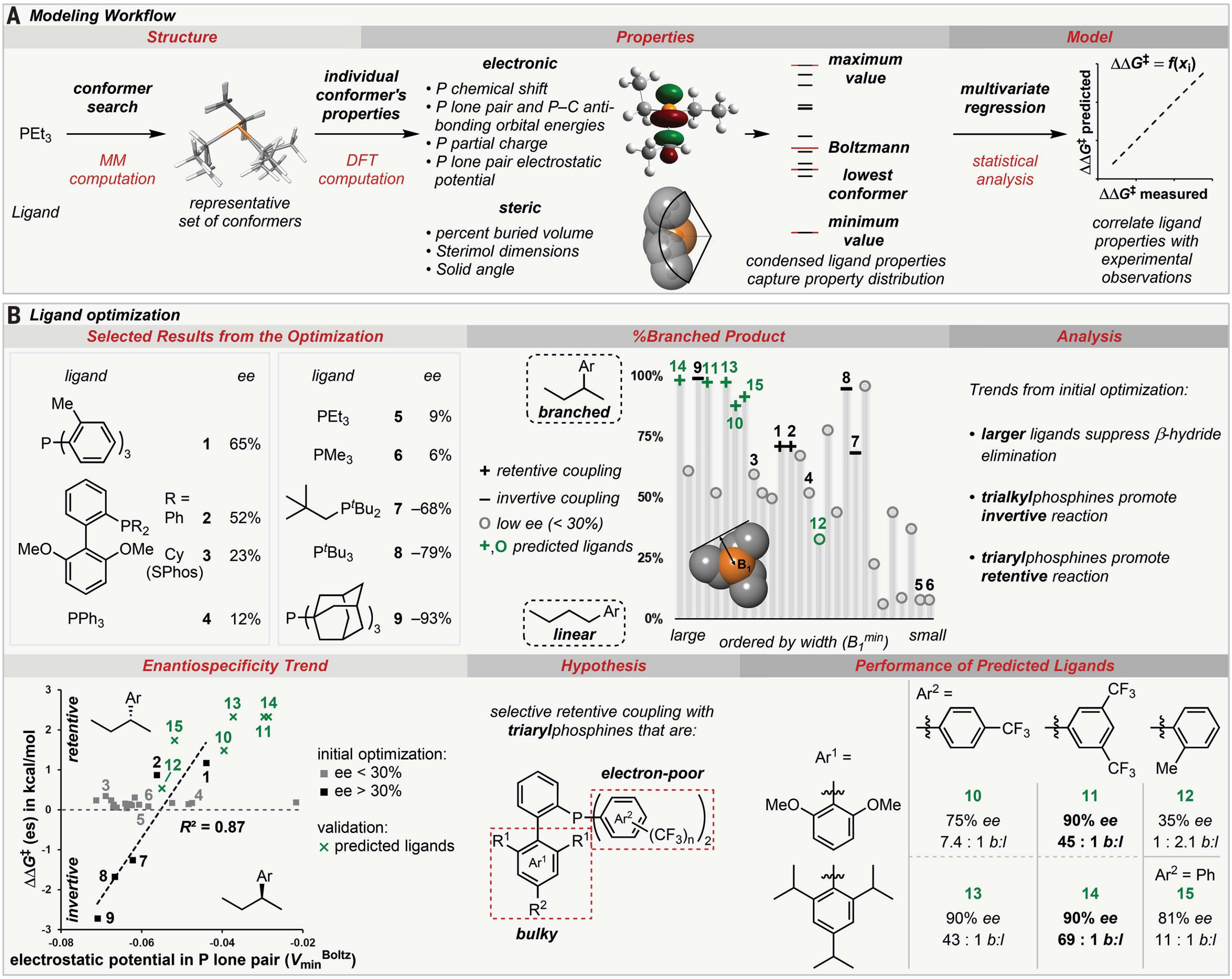


Fig. 2. Phosphine parameterization. (**A**) Workflow of parameter generation and statistical modeling. ΔΔ$G^‡$, relative free energy of activation; $f(x_i)$, function of parameter x_i. (**B**) Application of phosphine parameterization to ligand optimization of the reaction shown in Fig. 1C with Z = CO$_2$Et. *b:l* = branched-to-linear ratio. A positive % ee value indicates net retention; a negative % ee value indicates net inversion. es, enantiospecificity; R^2, coefficient of determination.

functional theory (DFT) was followed by parameter collection. Subsequently, four descriptor subsets were defined to capture the conformational dynamics of the ligands by including the mathematical extreme descriptor values (minimum and maximum), the lowest-energy-conformer values, and the Boltzmann weighted averages. We viewed the specific treatment of representative conformers as a crucial means of describing ensemble properties such as chemical shift while also probing structural flexibility during catalysis.

The final step in the workflow involved the analysis of both the stereofidelity and the branched:linear product ratio. These two readouts

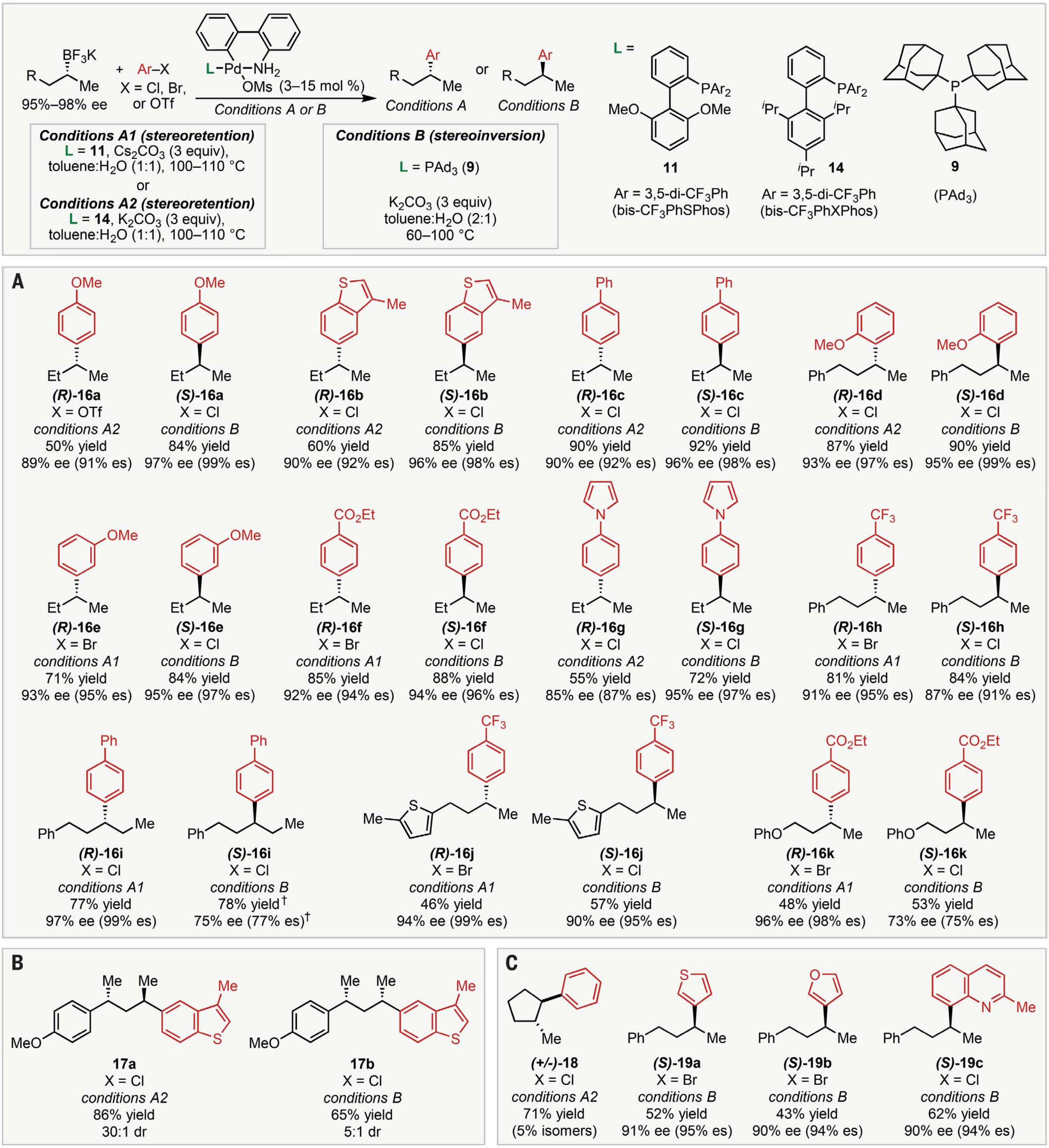


Fig. 3. Stereodivergent Pd-catalyzed cross-coupling reactions using enantioenriched alkylboron nucleophiles. The general reaction scheme is shown at the top. (**A** to **C**) Isolated yields are shown for stereoretentive and stereoinvertive cross-coupling reactions (A), diastereoselective cross-coupling reactions (B), and additional individual reactions (C). ^{i}Pr, isopropyl; dr, diastereomeric ratio. % es = % ee (final product) divided by % ee (starting material). †44% yield and 84% ee (86% es) when run at 60°C.

presumably describe two stages of the reaction mechanism (Fig. 1B): (i) the competing stereoretentive and stereoinvertive transmetallation mechanisms that determine the final stereochemistry of the cross-coupling product and (ii) the competitive β-hydride elimination-isomerization sequences that follow transmetallation. A correlation of the branched:linear ratio with the final enantiopurity of the product reveals that β-hydride elimination is responsible for both racemization and isomerization to the linear side product. Furthermore, a modest trend is observed relating the minimum width B_1 of the phosphine ligand to the branched:linear ratio (Fig. 2B). This is consistent with reports of large ligands facilitating reductive elimination over β-hydride elimination (*33*) and suggests the use of a parameterization approach to take into account the conformational flexibility of ligands.

Because the inherent selectivity of the transmetallation mechanism is masked by deleterious racemization as a consequence of β-hydride elimination, only ligands providing high selectivity were further investigated [>30% enantiomeric excess (ee), Fig. 2B]. The molecular electrostatic potential minimum in the phosphorus lone pair region (V_{min}) has been shown to correlate with the classical Tolman electronic parameter (*34*). Thus, V_{min} serves as an easily computable measure for the overall ligand electronics. A correlation between enantioselectivity and V_{min} was observed within the abridged dataset, indicating that electronic properties of the ligand determine the mechanism of transmetallation. Specifically, electron-rich trialkylphosphines promoted stereoinvertive reactions, whereas the electron-poorer triarylphosphines provided modest selectivity for stereoretention. Use of the bulky, electron-rich ligand PAd_3 (Ad, adamantyl) (**9**), which was recently reported by Carrow and co-workers (*35*), resulted in a particularly large preference for the stereoinvertive outcome. Based on these data, we hypothesized and virtually evaluated ligands for improved stereoretentive outcomes with the following features: (i) large ligand bulk to prevent β-hydride elimination and racemization and (ii) electron-deficient aryl substituents at phosphorus to promote the stereoretentive mechanism and to accelerate reductive elimination (Fig. 2B). Among the proposed ligands was a set of biaryl phosphines (**11** to **15**), as pioneered by Buchwald and colleagues (*36*), featuring various electron-deficient aryl groups at phosphorus. Gratifyingly, ligands **11** and **14** promote the alkyl Suzuki cross-coupling reaction with considerably enhanced selectivity (up to 90% ee) and minimal alkyl isomerization. Thus, parameterization-driven optimization facilitated development of a stereoretentive Suzuki reaction involving unactivated alkylboron nucleophiles. When considered alongside the introduction of **9** to achieve stereoinvertive couplings, predictable control of the absolute sense of enantioselectivity (retention or inversion) can be engendered by simply selecting the appropriate ligand.

Our stereochemical investigations of secondary alkylboron transmetallation in the Suzuki reaction suggested that both enantiomers of a cross-coupling product could be selectively accessed through use of a single enantioenriched alkylboron reagent with the proper selection of the phosphine ligand. The scope of this process is depicted in Fig. 3. Using enantioenriched, unactivated alkyltrifluoroborate nucleophiles, ligand-controlled stereoselectivity was broadly achieved in cross-coupling reactions with aryl electrophiles. Strongly π-accepting ligands bis-CF_3PhSPhos (Ph, phenyl) (**11**) and bis-CF_3PhXPhos (**14**), which emerged from our parameterization-guided optimization, preferentially promote the stereoretentive pathway, whereas strongly σ-donating ligand PAd_3 (**9**) preferentially promotes the stereoinvertive pathway. Because electron-poor palladium catalysts commonly undergo slow oxidative addition with aryl chlorides, we also evaluated aryl bromide and triflate electrophiles in reactions involving **11** and **14**. A particular highlight of this protocol is the uniformity of the conditions used for both the stereoinvertive and stereoretentive reactions: each operates in a toluene and water mixture as solvent, with a carbonate base, and no additional additives. Both reaction variants tolerated the use of electron-rich and electron-deficient aryl electrophiles, as well as an aryl electrophile bearing an ortho substituent. High stereofidelity was achieved for all of these reactions, including those involving alkylboron nucleophiles bearing thiophenyl and phenoxide substituents. Use of an alkylboron nucleophile containing a larger substituent (replacing methyl with ethyl) at the

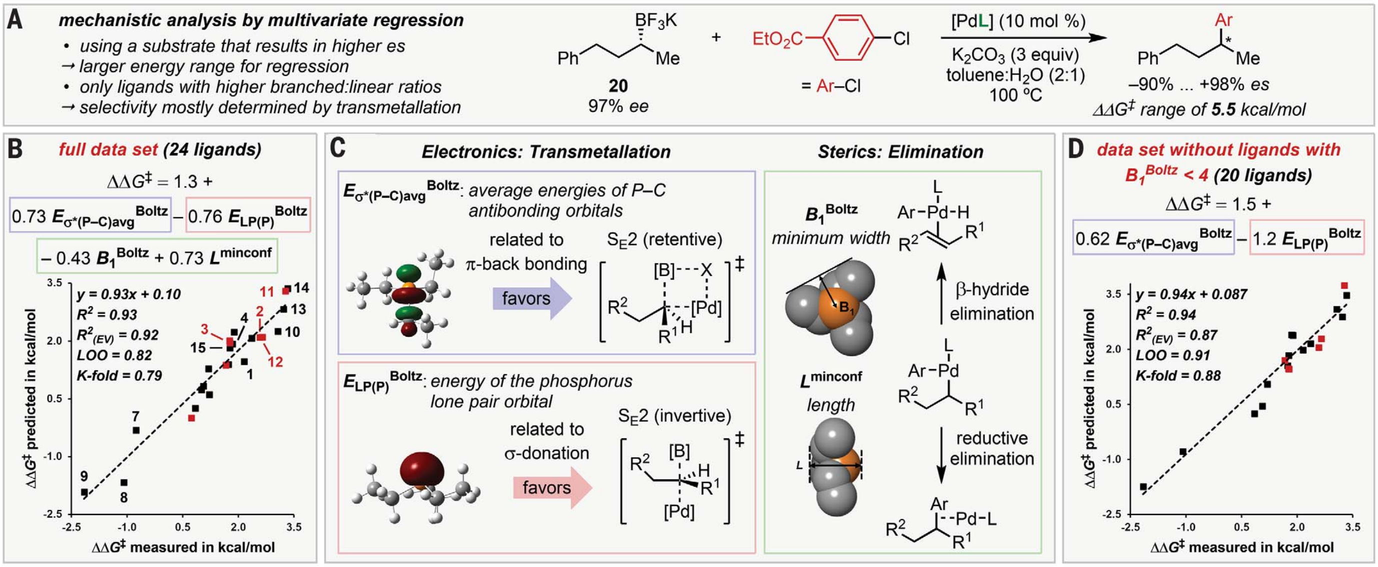


Fig. 4. Mechanistic investigation by multidimensional regression modeling. (**A**) Data from the arylation of **20** was used. [PdL] is the precatalyst as shown in Fig. 3 with varying ligands. * indicates an enantioenriched stereocenter. (**B**) Regression model containing all 24 ligands of this dataset. $E_{\sigma*(P-C)avg}^{Boltz}$ is the Boltzmann-weighted average across the conformers of the average energies of the three P–C σ* antibonding orbitals in each phosphine. $E_{LP(P)}^{Boltz}$ is the Boltzmann-weighted average of the energy of the phosphorus lone-pair orbital. Sterimol B_1^{Boltz} is the least width, and sterimol $L^{minconf}$ is the length of the lowest-energy conformer as seen from opposite the P substituents. Red points in the diagram indicate validation data (EV) not used in the model training. LOO, leave-one-out cross-validation score; K-fold, average threefold cross-validation score. (**C**) Illustration and interpretation of the model terms. (**D**) Regression model after removing the four smallest ligands in this dataset to exclude the influence of competitive β-hydride elimination on the data.

stereogenic center was also well tolerated (**16i**). In the absence of a methyl substituent, the stereoinvertive variant shows modestly reduced selectivity that may be improved by reducing the reaction temperature. This was partially anticipated because of the greater sensitivity of the metal fragment to steric congestion at carbon in the stereoinvertive S_E2 transmetallation mechanism. Diastereomeric products **17a** and **17b** could be generated from a single alkylboron diastereomer (*37*) using **14** and PAd_3, respectively (Fig. 3B). In these reactions, replacement of ligand **14** with PAd_3 resulted in a change in diastereoselectivity from 30:1 to 1:5, a 3.6 kcal/mol free energy of activation difference dependent only on the ligand identity. No erosion of specificity was observed for electron-deficient aryl substrates in stereoinvertive Suzuki reactions using PAd_3, in contrast to analogous reactions using P^tBu_3. Furyl and thiophenyl electrophiles are also compatible with our system (Fig. 3C). As an additional mechanistic probe, *trans*-2-methylcyclopentyltrifluoroborate was subjected to the stereodivergent reaction conditions. Because *trans*-2-methylcyclopentyltrifluoroborate is sterically impeded from undergoing stereoinvertive transmetallation, only the stereoretentive process using **14** should be mechanistically viable. Indeed, we observed that use of ligand **14** smoothly generates **18** with stereoretention, whereas use of PAd_3 results in low alkylboron conversion.

To further probe the origin of the ligand-dependent enantiodivergent process, we interrogated the mechanism of transmetallation using the parameterization strategy described above (Fig. 4). To accomplish this, phenyl-substituted substrate **20** was selected because of enhanced performance and thus a greater output range. A singular aryl chloride electrophile was chosen for this analysis to avoid potential attenuation of the ligand effects by the influence of different counterions (e.g., bromide or triflate). Additionally, 24 ligands were tested, excluding smaller ligands to reduce the complexity associated with β-hydride elimination. Multivariate linear regression revealed that most of the outputs can be expressed in two readily interpretable terms that discriminate the transmetallation pathways: the average energy of the P–C antibonding orbitals $E_{\sigma^*(P-C)}$, representative of π-back bonding, and the energy of the lone pair orbital of phosphorus $E_{LP(P)}$, a measure of the ligand's σ-donation capability (Fig. 4B). This outcome suggests that the stereoinvertive pathway is dependent on strong σ-donation from the ligand, which may stabilize a two-coordinate, cationic palladium complex. Conversely, the stereoretentive pathway is enhanced by π-back bonding, which may stabilize the coordination of a π-donor ligand X (presumably OH^-) to Pd. Including two steric descriptors such as B_1 and L (length of ligand L) improves the model fit by treating the competitive β-hydride elimination that occurs using smaller ligands and decreases the observed specificity. This becomes evident when the four smallest ligands in this dataset are removed from the analysis, which results in an excellent correlation using just the two electronic descriptors with the experimentally observed stereochemical outcomes (Fig. 4D). Multivariate regression analysis thereby provides compelling evidence for the electronic factors favoring each transmetallation mechanism and thus a guideline for future developments in stereospecific cross-coupling reactions.

REFERENCES AND NOTES

1. J. C. H. Lee, D. G. Hall, in *Metal-Catalyzed Cross-Coupling Reactions and More*, A. de Meijere, S. Bräse, M. Oestreich, Eds. (Wiley-VCH, ed. 3, 2014), vol. 1, pp. 65–132.
2. C.-Y. Wang, J. Derosaa, M. R. Biscoe, *Chem. Sci.* **6**, 5105–5113 (2015).
3. C. Sandford, V. K. Aggarwal, *Chem. Commun. (Camb.)* **53**, 5481–5494 (2017).
4. J. P. G. Rygus, C. M. Crudden, *J. Am. Chem. Soc.* **139**, 18124–18137 (2017).
5. D. Imao, B. W. Glasspoole, V. S. Laberge, C. M. Crudden, *J. Am. Chem. Soc.* **131**, 5024–5025 (2009).
6. D. L. Sandrock, L. Jean-Gérard, C. Y. Chen, S. D. Dreher, G. A. Molander, *J. Am. Chem. Soc.* **132**, 17108–17110 (2010).
7. T. Ohmura, T. Awano, M. Suginome, *J. Am. Chem. Soc.* **132**, 13191–13193 (2010).
8. T. Awano, T. Ohmura, M. Suginome, *J. Am. Chem. Soc.* **133**, 20738–20741 (2011).
9. J. C. H. Lee, R. McDonald, D. G. Hall, *Nat. Chem.* **3**, 894–899 (2011).
10. B. M. Partridge, L. Chausset-Boissarie, M. Burns, A. P. Pulis, V. K. Aggarwal, *Angew. Chem. Int. Ed.* **51**, 11795–11799 (2012).
11. G. A. Molander, S. R. Wisniewski, *J. Am. Chem. Soc.* **134**, 16856–16868 (2012).
12. S. C. Matthew, B. W. Glasspoole, P. Eisenberger, C. M. Crudden, *J. Am. Chem. Soc.* **136**, 5828–5831 (2014).
13. C. Sun, B. Potter, J. P. Morken, *J. Am. Chem. Soc.* **136**, 6534–6537 (2014).
14. T. P. Blaisdell, J. P. Morken, *J. Am. Chem. Soc.* **137**, 8712–8715 (2015).
15. S. Miyamura, M. Araki, T. Suzuki, J. Yamaguchi, K. Itami, *Angew. Chem. Int. Ed.* **54**, 846–851 (2015).
16. Y. Lou *et al.*, *Angew. Chem. Int. Ed.* **127**, 12302–12306 (2015).
17. G. L. Hoang, J. M. Takacs, *Chem. Sci.* **8**, 4511–4516 (2017).
18. L. Li, S. Zhao, A. Joshi-Pangu, M. Diane, M. R. Biscoe, *J. Am. Chem. Soc.* **136**, 14027–14030 (2014).
19. C. Amatore, A. Jutand, G. Le Duc, *Chemistry* **17**, 2492–2503 (2011).
20. B. P. Carrow, J. F. Hartwig, *J. Am. Chem. Soc.* **133**, 2116–2119 (2011).
21. A. A. Thomas, S. E. Denmark, *Science* **352**, 329–332 (2016).
22. A. A. Thomas, H. Wang, A. F. Zahrt, S. E. Denmark, *J. Am. Chem. Soc.* **139**, 3805–3821 (2017).
23. A. A. Thomas, A. F. Zahrt, C. P. Delaney, S. E. Denmark, *J. Am. Chem. Soc.* **140**, 4401–4416 (2018).
24. B. H. Ridgway, K. A. Woerpel, *J. Org. Chem.* **63**, 458–460 (1998).
25. K. Matos, J. A. Soderquist, *J. Org. Chem.* **63**, 461–470 (1998).
26. C. B. Santiago, J.-Y. Guo, M. S. Sigman, *Chem. Sci.* **9**, 2398–2412 (2018).
27. M. S. Sigman, K. C. Harper, E. N. Bess, A. Milo, *Acc. Chem. Res.* **49**, 1292–1301 (2016).
28. S. Krautwald, E. M. Carreira, *J. Am. Chem. Soc.* **139**, 5627–5639 (2017).
29. C. A. Tolman, *Chem. Rev.* **77**, 313–348 (1977).
30. J. Jover *et al.*, *Organometallics* **29**, 6245–6258 (2010).
31. Z. L. Niemeyer, A. Milo, D. P. Hickey, M. S. Sigman, *Nat. Chem.* **8**, 610–617 (2016).
32. K. Wu, A. G. Doyle, *Nat. Chem.* **9**, 779–784 (2017).
33. R. Jana, T. P. Pathak, M. S. Sigman, *Chem. Rev.* **111**, 1417–1492 (2011).
34. C. H. Suresh, N. Koga, *Inorg. Chem.* **41**, 1573–1578 (2002).
35. L. Chen, P. Ren, B. P. Carrow, *J. Am. Chem. Soc.* **138**, 6392–6395 (2016).
36. D. S. Surry, S. L. Buchwald, *Chem. Sci.* **2**, 27–50 (2011).
37. S. Balieu *et al.*, *J. Am. Chem. Soc.* **137**, 4398–4403 (2015).

ACKNOWLEDGMENTS

We acknowledge R. Kinthada for contributions to the diastereoselectivity studies. We thank G. Ralph for assistance with chiral high-performance liquid chromatography analysis. We thank L. Cavallo for providing us with the command line tool for computing percent buried volume. **Funding:** We are grateful to the National Institutes of Health (grant SC1GM110010 to M.R.B.), the National Science Foundation (grant CHE-1665189 to M.R.B. and grants CHE-1361296 and CHE-1763436 to M.S.S.), and the Leopoldina Fellowship Programme of the German National Academy of Sciences Leopoldina (LPDS 2017-18 to T.G.). The support and resources from the Center for High Performance Computing (CHPC) at the University of Utah are gratefully acknowledged. Further computational resources were provided by the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the NSF (ACI-1548562) and provided through allocation TG-CHE180003. **Author contributions:** S.Z. and B.M. performed all synthetic experiments and isolated all products. T.G. computed all ligand parameters and performed multivariate analyses. Z.L.N. performed initial multivariate analyses. M.R.B., M.S.S., and T.G. wrote the manuscript. M.R.B. and M.S.S. directed the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All additional data are in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/670/suppl/DC1
Materials and Methods
Supplementary Text
Fig. S1
Tables S1 to S6
NMR Spectra
References (*38*–*80*)

18 May 2018; accepted 6 September 2018
Published online 20 September 2018
10.1126/science.aat2299

NEUROSCIENCE

Nested sequences of hippocampal assemblies during behavior support subsequent sleep replay

Céline Drieu, Ralitsa Todorova, Michaël Zugaro*

Consolidation of spatial and episodic memories is thought to rely on replay of neuronal activity sequences during sleep. However, the network dynamics underlying the initial storage of memories during wakefulness have never been tested. Although slow, behavioral time scale sequences have been claimed to sustain sequential memory formation, fast ("theta") time scale sequences, nested within slow sequences, could be instrumental. We found that in rats traveling passively on a model train, place cells formed behavioral time scale sequences but theta sequences were degraded, resulting in impaired subsequent sleep replay. In contrast, when the rats actively ran on a treadmill while being transported on the train, place cells generated clear theta sequences and accurate trajectory replay during sleep. Our results support the view that nested sequences underlie the initial formation of memory traces subsequently consolidated during sleep.

The sequential activation of neuronal ensembles is a ubiquitous brain coding scheme possibly underlying numerous diverse behaviors (*1*–*4*). Sequential neuronal activity occurs at different time scales, ranging from slow ("behavioral") time scales, where the dynamics are constrained by stimulus or motor time constants, to fast ("endogenous") time scales mostly driven by intrinsic network properties. A prominent example is the hippocampus, where place cells code for an animal's location (*5*). As the animal explores its environment, different place cell ensembles become successively active along the ongoing trajectory, yielding sequences of neuronal activity at the behavioral time scale. During subsequent slow-wave sleep (SWS), the same sequences are endogenously replayed during sharp-wave ripple complexes at a highly accelerated (20×) time scale (*6*), mediating memory consolidation during sleep (*7*–*10*). How is the sequential organization of place cell assemblies maintained across these two time scales, expressed at entirely disjoint moments in time and during different brain states?

Sequential information could be readily stored during exploration in the hippocampal network by the sequential activation of place cells at the behavioral time scale, via a recently discovered form of behavioral time scale plasticity (*11*). An intriguing alternative is that sequential structure is stored via the remarkable ability of the hippocampal network to generate nested sequences of cell assemblies, whereby both slow and fast neural sequences are intermingled in time. This occurs during spatial navigation: Nested within behavioral time scale sequences, the hippocampal network also produces sequences at the theta time scale [one sequence per theta cycle of ~150 ms (*12*–*14*)], allowing cell assemblies to fire within brief delays [~25 ms (*15*)] compatible with classical Hebbian plasticity, such as spike timing–dependent plasticity (*16*). Are these nested sequences of hippocampal cell assemblies required for subsequent sleep replay (*12*, *13*), or are they merely an epiphenomenon deriving from preexisting connectivity within the hippocampal network (*17*–*19*)?

Contrasting these predictions requires a protocol that selectively disrupts fast theta sequences but preserves slow behavioral sequences of place cell assemblies. Further crucial constraints are the ability to trigger or release theta sequence disruption with temporal precision, as well as the necessity to target the entire hippocampal formation, both in terms of extent (to overcome information redundancy along the septotemporal axis of the hippocampus) and in terms of fields (to overcome compensatory network mechanisms such as pattern completion in CA3, which could restore locally induced impairments). During passive transportation in space, place cells remain spatially selective, but their precise timing relative to theta ["phase precession" (*12*)] is altered (*20*) unless the animals actively run on an onboard treadmill (*21*). We thus transported rats on a model train (fig. S1A) and turned the onboard treadmill off (Passive) or on (Active) to respectively perturb nested sequences or leave them intact. The goal was to determine whether intact nested sequences were required for subsequent replay during SWS. The rats were tested in an entirely novel environment, different from the training room. Hence, hippocampal activity was monitored as the animals learned a novel spatial context, which is known to induce the formation of a novel hippocampal map, increase network coordination, boost replay, and enhance plasticity (*22*–*25*). Further, because the rats were tested on a single day, this avoided any confounding network changes that could have resulted from previous experience.

We recorded CA1 pyramidal units and local field potentials (LFPs) in five rats. After a baseline sleep session, the rats underwent three travel sessions (Passive 1, Active, Passive 2) interspersed with sleep sessions (Fig. 1A). The train velocity, number of laps, and travel duration were similar in all conditions for all rats (fig. S1, B to D). We first verified the presence of place cell sequences at the behavioral time scale in all three conditions. Pyramidal cells always coded for the location of the animal in space (Fig. 1B and fig. S2). Their fields remained similar in terms of size (Fig. 1C and fig. S3) and together covered the entire train track (fig. S2C). However, peak firing rates and place cell count were somewhat reduced during passive travels [Fig. 1C and fig. S2B (*20*)]; specific controls for these factors are provided in ensuing analyses (see below). Incidentally, we also confirmed that place fields did not undergo random remapping (Fig. 1, D and E, and fig. S2D). This finding provided further verification that hippocampal dynamics were virtually identical at the slow time scale in all travel sessions.

Clear theta oscillations with similar frequencies were observed in all conditions (Fig. 2A). However, during passive travel, power was slightly reduced at both the fundamental frequency and the first harmonic, the latter resulting in decreased cycle asymmetry, indicative of a change in the internal structure of theta cycles (Fig. 2, A and B). This was expected to alter the precise spike timing of active cell assemblies. Consistently, place cells continued to oscillate slightly faster than theta (i.e., phase precession) during active travel, whereas this was reduced in passive travel (Fig. 2, C and D, and fig. S4, A and B), indicating an overall degradation of phase precession [Fig. 2, E and F; in phase-precessing neurons, phase range was similar in all conditions (*26*) (fig. S4, C to E)]. This degradation was noteworthy, because phase precession is thought to be instrumental for the formation of nested sequences: Place cells normally oscillate slightly faster than theta, so they emit spike bursts earlier and earlier in successive theta cycles; this [possibly combined with additional coordinating mechanisms (*27*, *28*)] results in newly activated cells firing after those that have started firing in earlier cycles, effectively resulting in temporal sequences of activity.

To directly assess how perturbation of phase precession affected theta sequences, we used a Bayesian reconstruction approach (*21*, *29*, *30*). Briefly, theta cycles were subdivided into six phase bins, and the sequential structure of reconstructed positions in these bins ("candidate events") was evaluated using two previously described complementary measures (*30*): (i) trajectory scores [normalized to compare across animals and conditions (*31*)], which assess the quality (linearity) of the reconstructed events (i.e., whether the events represent spatially aligned positions versus mere series of random locations); and (ii) slopes, which estimate the speed at which reconstructed events are played and indicate

---

Center for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, PSL Research University, Paris, France.
*Corresponding author. Email: michael.zugaro@college-de-france.fr

whether the events move through space or remain merely stationary (absence of actual trajectories). Thus, clear theta sequences would be characterized by both high scores and slopes, whereas static representations of current position would result in high scores but low or zero slopes, and random activity would be associated with low scores (but possibly spuriously high slopes). Clear sequential structure was readily visible in individual theta cycles during active, but not passive, travel (Fig. 3A). This was confirmed over all theta cycles for all rats (Fig. 3B and fig. S5). Whereas normalized trajectory scores were significantly better than chance in all conditions (Fig. 3C), slopes were significantly steeper during active travel (Fig. 3D and fig. S5, C and D). Joint analysis of trajectory scores and slopes revealed a much greater proportion of high-value pairs during active travel (Fig. 3E). To confirm and extend these results using an independent measure, we computed the quadrant score (*27*) of each candidate event, which assessed the overall direction of reconstructed trajectories without assuming constant velocity. Quadrant scores were significantly greater for active travel (clear trajectories) and remained very low for passive travel (degraded trajectories) (Fig. 3F and fig. S5B).

Hence, theta sequences were degraded during passive travel. Note that reconstruction quality and quadrant scores were higher in Passive 2 than in Passive 1 (Fig. 3, E and F), implying that sequential structure was slightly less degraded in the second passive travel session (Fig. 3B). This was further supported by higher self-consistency of theta sequences (*31*) in Passive 2 than in Passive 1, whereas Active sequences were the most self-consistent (Fig. 3G and fig. S6A).

We ensured that our results could not be accounted for by differences in the number of simultaneously recorded place cells (fig. S7) and their firing rates (fig. S8). We also controlled for decoding quality (*31*) (fig. S9), ruling out a potential bias due to differences in spatial coding, at both the single-cell and population levels. Finally, because small variations in field locations between individual laps could have altered sequence detection, we also ruled out differences in firing variability between travel conditions (*31*) (fig. S10).

Taken together, the above results show that place cell sequences at the behavioral time scale were present in all three conditions, but theta sequences were disrupted during passive travel, when stationary network activity continued to reflect the ongoing position at the endogenous time scale. How did this affect subsequent activity during SWS? Candidate replay events were defined as transient surges in aggregate firing rate (*30*) during SWS epochs (fig. S11, A and C), which coincided with ripple events [fig. S11G; results remained unchanged when candidate events were restricted to ripples (fig. S12)]. The average SWS duration, number and rate of candidate events, and ripple occurrence rate were not significantly different across animals (fig. S11, B and D to F). To reconstruct replayed trajectories, we first trained the Bayesian decoder using place cell activity recorded during the preceding travel session, then tested it during SWS on candidate events subdivided into 20-ms nonoverlapping time windows (*31*).

Candidate replay events were evaluated using the same method as theta sequences, whereby genuine sequences are characterized by both

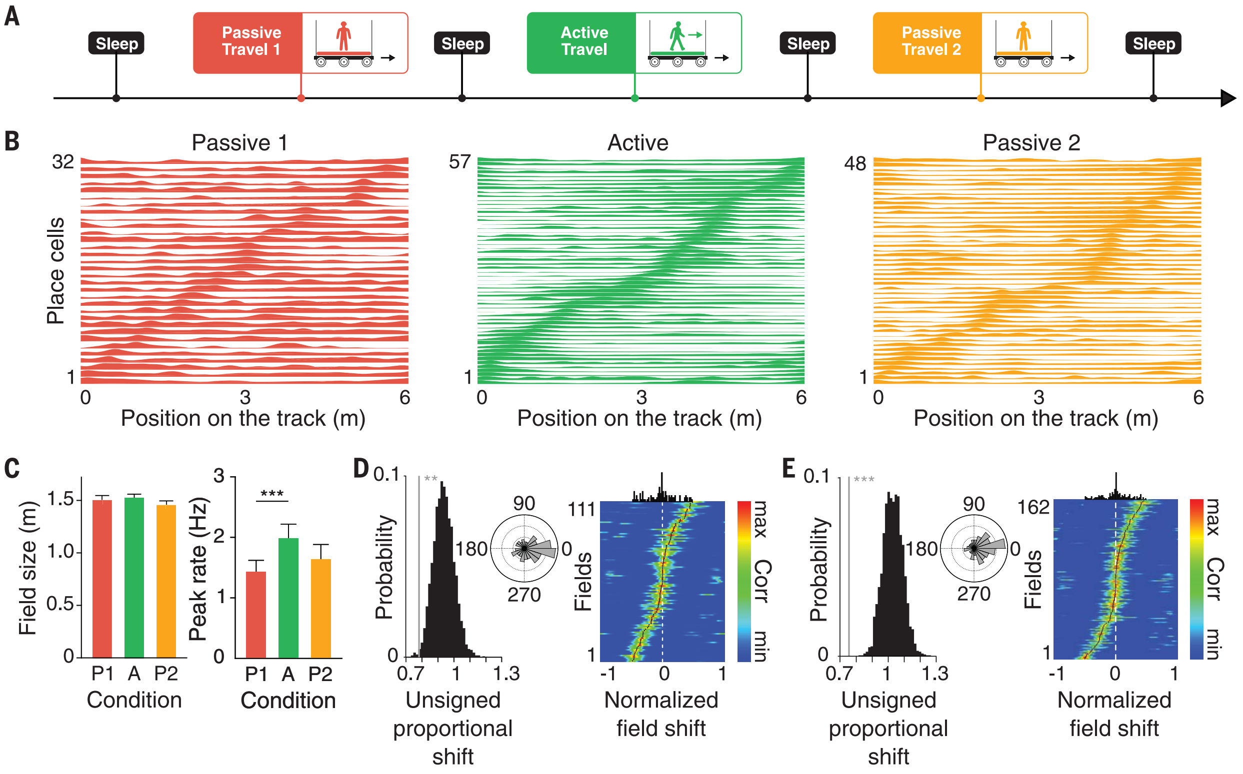


Fig. 1. Maintenance of behavioral time scale properties. (**A**) Behavioral protocol. In a novel environment, rats underwent three successive travel sessions (Passive 1–Active–Passive 2) interspersed with sleep recordings (see also fig. S1). Passive travel was intended to selectively perturb spike timing at the theta time scale but not at the behavioral time scale. (**B**) Linearized normalized firing curves of place cells recorded from one example rat across travel sessions, showing complete track coverage in all conditions. (**C**) Place field size was maintained across travel conditions [Kruskal-Wallis (KW) test, $P > 0.05$]. Peak firing rates were somewhat lower in Passive 1 than in Active (KW test, $^{***}P < 10^{-3}$). (**D**) Firing fields did not remap between Passive 1 and Active. Left: Mean unsigned proportional shift σ (vertical gray line; $^{**}P < 0.01$) and σ distribution for n = 2000 bootstrapped remapping data sets (black histograms). Inset: Circular distributions of angular differences (in degrees) between place field peak locations on the maze (Rayleigh test, $P < 0.001$; V-test against 0, $P < 10^{-4}$). Right: Spatial cross-correlograms of firing fields across successive travel sessions (x axis normalized to field size; black dots represent correlogram modes; black histograms at top, mode distributions). (**E**) Firing fields also did not remap between Active and Passive 2. Data are displayed as in (D). Left: $^{***}P \sim 0$. Inset: Rayleigh test, $P < 10^{-9}$; V-test against 0, $P < 10^{-10}$.

high trajectory scores and slopes. Although candidate events with significant trajectory scores were present in all three sleep sessions, overall the reconstructed trajectories were notably sharper after active travel (Fig. 4A and fig. S13A), and scores were significantly improved relative to baseline (*31*) only in sleep sessions following Active and Passive 2 (Fig. 4B). In addition, slopes were significantly steeper following Active (Fig. 4C and fig. S13B). Joint analysis of trajectory scores and slopes confirmed a much greater proportion of high-value pairs after active travel (Fig. 4D and fig. S13, C and D).

We next addressed the critical question of whether these trajectories did constitute genuine replay of awake behavior, or merely reflected preexisting connectivity patterns independent of experience. We compared the proportion of significant trajectories in each sleep session relative to baseline sleep—that is, relative to the sleep session preceding any exploration of the environment (as noted above, recordings took place in an entirely novel environment). We did not observe replay during sleep following Passive 1 (Fig. 4E), when theta sequences had been selectively disrupted (Fig. 3). Trajectory replay was then boosted in sleep following Active (Fig. 4E), when nested sequences had remained intact (Fig. 3). Finally, an intermediate but significant level of reactivation was observed following Passive 2 (Fig. 4E), when theta sequences had been perturbed to a lesser degree than during Passive 1 (Fig. 3). Although forward and backward trajectories were found in equal proportions during sleep after passive travel, only active travel resulted in a greater proportion of forward replay actually reflecting awake experience (Fig. 4F). Finally, a direct comparison of theta and replay sequences (*31*) highlighted a selective

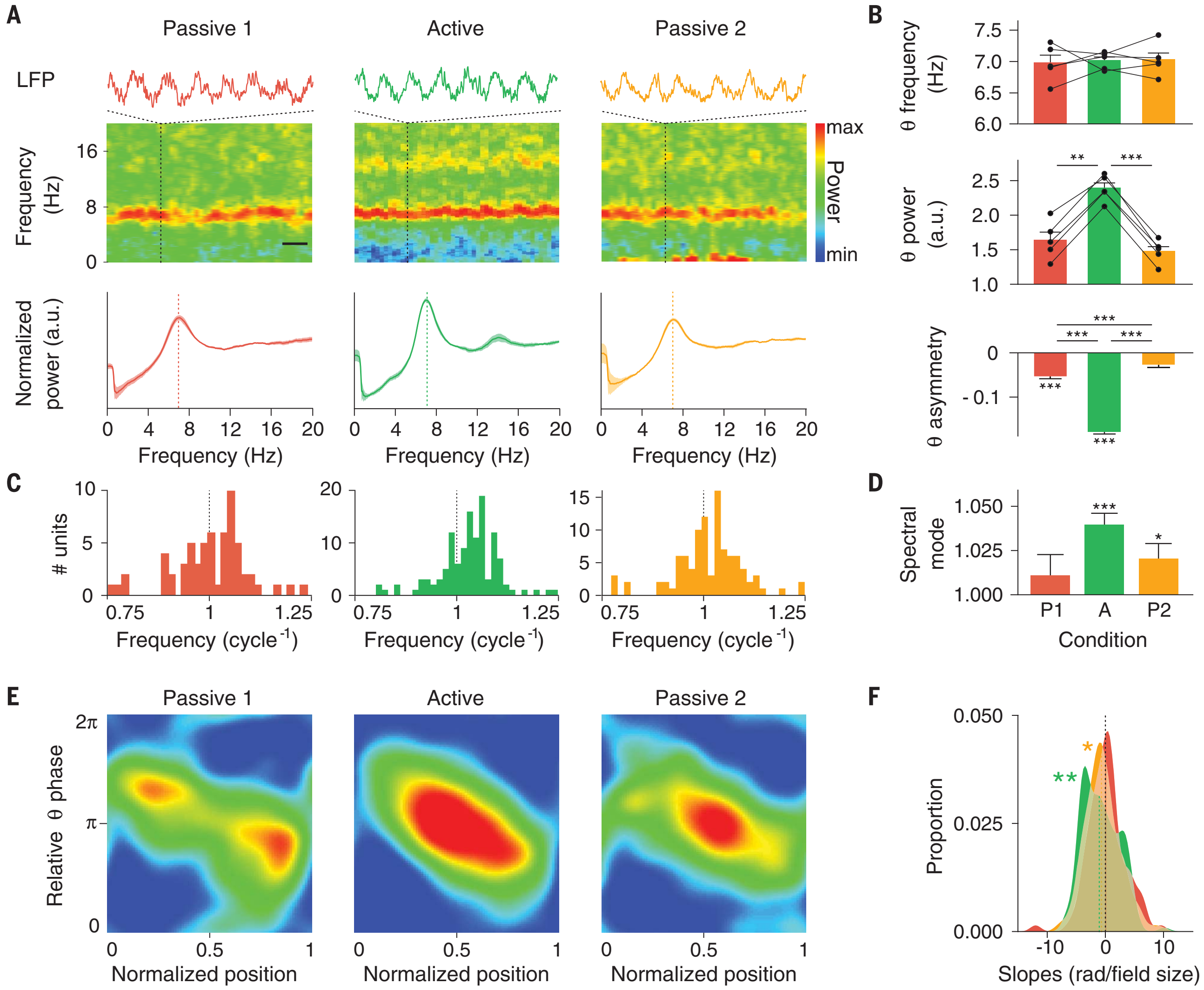


Fig. 2. Perturbation of single-cell theta time scale properties. (**A**) Theta maintenance across travel sessions. Top row: Example raw LFPs (duration, 1 s). Middle row: Power spectrograms (black dashed line, time of LFP traces shown above; black calibration bar, 15 s). Bottom row: Normalized power spectra (mean ± SEM). (**B**) Theta frequency was unchanged across conditions [top; repeated-measures analysis of variance (ANOVA), $P > 0.90$], whereas theta power (middle; repeated-measures ANOVA, $P < 10^{-6}$) and shape (bottom, theta asymmetry; KW test, $P < 10^{-84}$) were altered during passive travel (frequency and power, mean ± SEM). $^{**}P < 0.0033$, $^{***}P < 0.00033$ for post hoc comparisons; $^{**}P < 0.01$, $^{***}P < 0.001$ otherwise. (**C** and **D**) During passive travel, spike bursts recurred at lower rates, closer to baseline theta frequency. (C) Distributions of spectral modes of spike trains (measured relative to theta frequency, x axis trimmed at ±25% around theta frequency). (D) Spectral modes ($^{***}P < 10^{-6}$, $^{*}P < 0.02$). (**E** and **F**) Theta-phase precession was perturbed during passive travel. (E) Average normalized phase precession density plots for significantly phase-precessing cells (blue and red indicate minimum and maximum spike density, respectively). (F) Distribution of phase precession slopes for all place cells (Kolmogorov-Smirnov test, $^{**}P < 0.008$, $^{*}P < 0.04$; colored dashed vertical lines are medians).

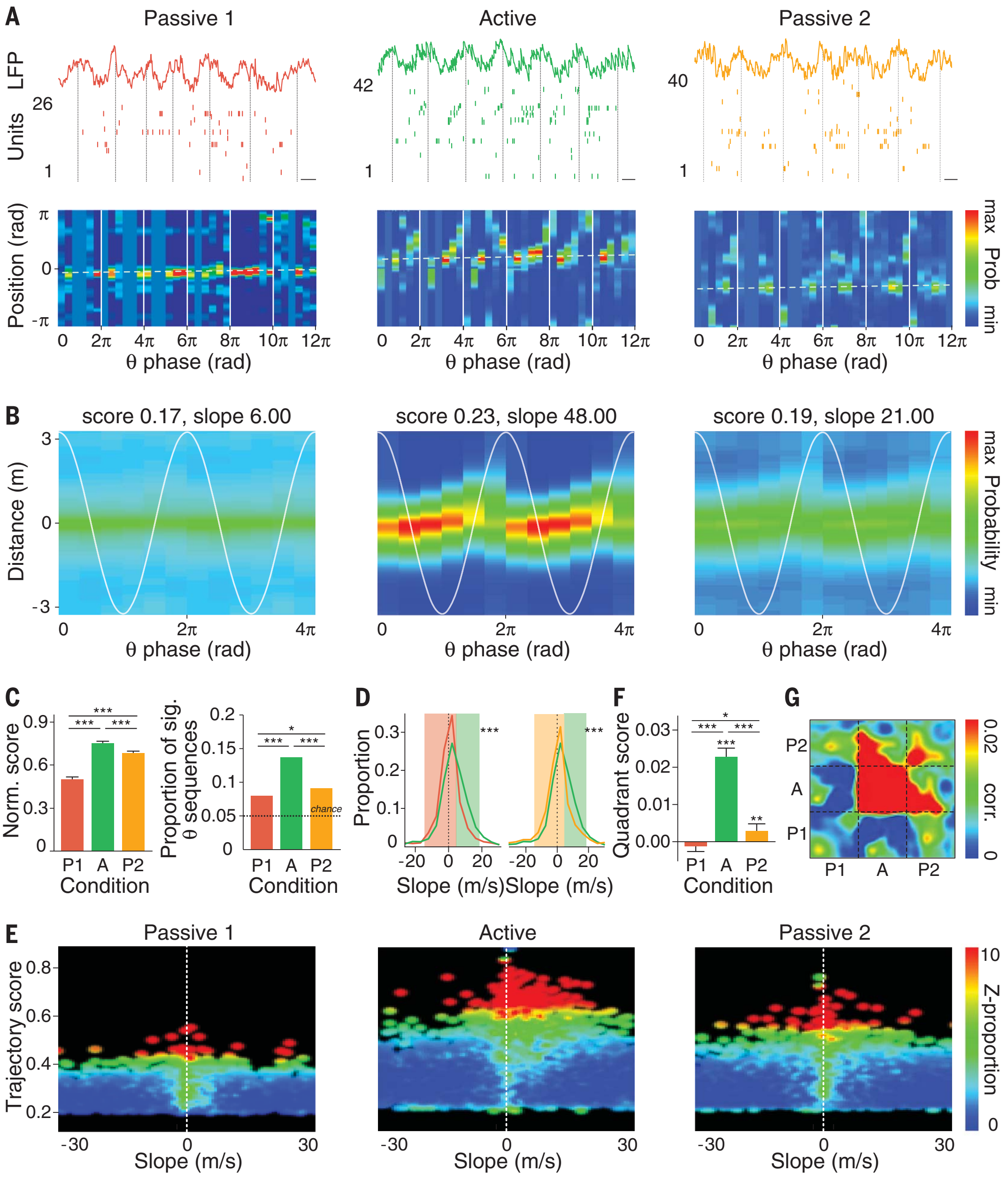


Fig. 3. Perturbation of theta sequences. (**A**) Top: Raw LFPs and place cell spikes in six example theta cycles (black dashed lines, theta peaks; place cells are ordered by their place field location on the track) in each travel condition (black calibration bars, 50 ms). Bottom: Bayesian reconstruction of position encoded in the ongoing activity of the hippocampal network (six phase windows per theta cycle; white vertical lines, theta peaks; white dashed lines, actual position of the animal). (**B**) Average Bayesian reconstruction of position (relative to actual position of the animal) across theta subcycles for all rats. Trajectory score and slope (in cm cycle^{-1}) are indicated above each reconstruction. Two cycles are shown for clarity. (**C**) Left: Normalized score of theta sequences (KW test, $P < 10^{-23}$; $***P < 0.00033$ for post hoc comparisons). Right: Proportion of significant theta sequences (Passive 1: 7.95%, Active: 13.61%, Passive 2: 9.05%; all proportions are significantly greater than shuffled control proportions, Monte Carlo test, $P \sim 0$; binomial proportion tests, $*P < 0.05$, $***P < 0.001$). (**D**) Distributions of significant theta sequence slopes (KS tests; left: Passive 1–Active, $***P < 10^{-16}$; right: Active–Passive 2, $***P < 10^{-18}$). Colored bands indicate significant differences. (**E**) Distribution of theta sequence quality assessed by joint trajectory score and slope (for all animals; color code indicates proportion normalized relative to shuffled control data). (**F**) Quadrant score computed from individual theta cycles (KW test, $***P < 10^{-24}$; $*P < 0.017$, $***P < 0.00033$ for post hoc comparisons; $**P < 0.01$, $***P < 0.001$ otherwise). (**G**) Pairwise bias correlation between awake theta sequences (ordered according to time of occurrence). Sequences were stable only during active travel. Note the absence of correlation during Passive 1 and progressive decay following Active during Passive 2.

correlation between theta sequences in Active and candidate replay events in subsequent sleep (Fig. 4G and fig. S6, B and C).

Our results thus show that during sleep after selective disruption of theta sequences (Passive 1), the proportion of significant trajectories remained at baseline levels observed prior to experience. By contrast, intact nested sequences (Active) resulted in boosted replay during sleep, and trajectories in hippocampal space were preferentially replayed in the same direction as in physical space. In Passive 2, theta sequences were perturbed to a lesser degree than in Passive 1, yielding intermediate levels of trajectory replay during subsequent sleep. This result has three implications. First, repeated experience alone cannot account for the improved replay following Active, because replay was degraded following Passive 2. Second, the improved theta sequences during Passive 2 relative to Passive 1 are consistent with the notion that replay following Active resulted in consolidation (*7*–*10*)—possibly consisting of functional network changes (*12*, *13*)—that carried over to subsequent sessions. Third, during Passive 2, scrambled activation of place cells at the theta time scale appears to have interfered with previously formed and consolidated memory traces (during Active and subsequent sleep), resulting in degraded replay during sleep after Passive 2.

How would nested sequences be altered during passive travel? In the absence of active locomotor signals, spike bursts of pyramidal cells recurred at slightly longer time intervals, consistent with the fact that bursting frequency increases with running speed (*12*, *32*). As predicted by theoretical models (*13*, *14*, *33*, *34*), this decrease in oscillation frequency resulted in impaired theta-phase precession and prevented the formation of theta sequences. Theta sequences have been related to memory performance (*35*, *36*), although the underlying mechanisms

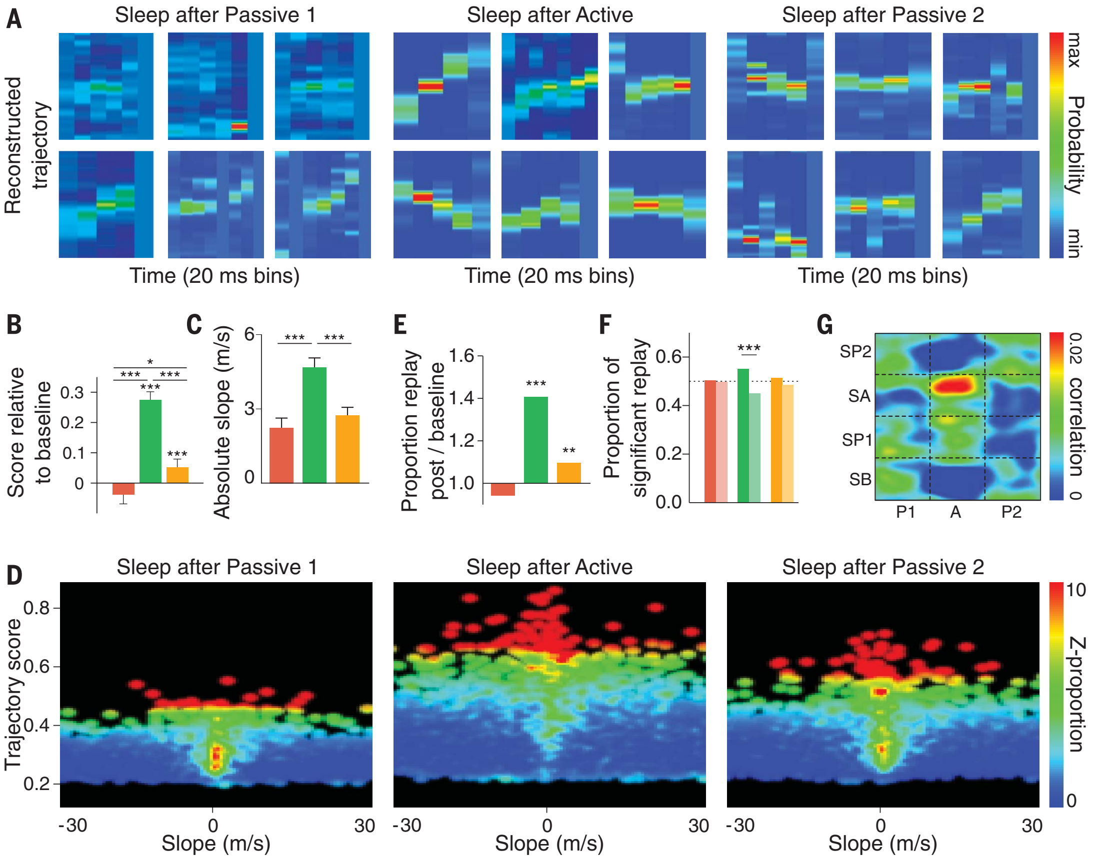


Fig. 4. Replay is degraded following perturbation of theta sequences. (**A**) Examples of significant replay events in sleep sessions (see fig. S16 for more examples). (**B**) Scores of replay events relative to baseline sleep (KW test, $P < 10^{-17}$; $^{*}P < 0.017$, $^{***}P < 0.00033$ for post hoc comparisons; $^{***}P < 0.001$ otherwise). (**C**) Absolute slopes of replay events (KW test, $P < 10^{-27}$; $^{***}P < 0.00033$ for post hoc comparisons). (**D**) Distribution of replay quality assessed by joint trajectory score and slope (for all animals; color denotes proportion normalized relative to shuffled control data). (**E**) Proportion of significant replay events relative to baseline sleep (binomial proportion tests, $^{**}P < 0.01$, $^{***}P < 0.001$). Sleep replay was boosted after active travel relative to after passive travel. (**F**) Proportion of forward (darker colors) versus reverse (lighter colors) replay events (binomial proportion tests, $^{***}P < 0.001$). (**G**) Pairwise bias correlation between awake theta sequences and sleep replay (ordered according to time of occurrence).

have remained unclear. Our results indicate a causal link between theta sequences and sleep replay for memory consolidation, and suggest that behavioral time scale sequences are insufficient to store sequential information for reactivation during subsequent sleep. Nested sequences emerging from independently phase-precessing place cells (*13*, *34*) enabled hippocampal assemblies to fire dozens of milliseconds apart, which is optimal for classical plasticity mechanisms (*16*) and can reinforce synaptic connections (*13*, *14*). This would effectively store sequential organization as network connectivity patterns, which can later be replayed during sleep for long-term consolidation (*7*–*10*). Spatiotemporal spike patterns supporting nested sequences have also been reported in the striatum (*37*) and medial prefrontal cortex (*38*). This may represent a general neural mechanism to encode and store initial memory traces, and plan future actions (*35*, *39*).

REFERENCES AND NOTES

1. A. J. Peters, S. X. Chen, T. Komiyama, *Nature* **510**, 263–267 (2014).
2. T. S. Okubo, E. L. Mackevicius, H. L. Payne, G. F. Lynch, M. S. Fee, *Nature* **528**, 352–357 (2015).
3. K. MacLeod, G. Laurent, *Science* **274**, 976–979 (1996).
4. E. Pastalkova, V. Itskov, A. Amarasingham, G. Buzsáki, *Science* **321**, 1322–1327 (2008).
5. J. O'Keefe, L. Nadel, *The Hippocampus as a Cognitive Map* (Clarendon, 1978).
6. A. K. Lee, M. A. Wilson, *Neuron* **36**, 1183–1194 (2002).
7. G. Buzsáki, *Neuroscience* **31**, 551–570 (1989).
8. G. Girardeau, K. Benchenane, S. I. Wiener, G. Buzsáki, M. B. Zugaro, *Nat. Neurosci.* **12**, 1222–1223 (2009).
9. V. Ego-Stengel, M. A. Wilson, *Hippocampus* **20**, 1–10 (2010).
10. N. Maingret, G. Girardeau, R. Todorova, M. Goutierre, M. Zugaro, *Nat. Neurosci.* **19**, 959–964 (2016).
11. K. C. Bittner, A. D. Milstein, C. Grienberger, S. Romani, J. C. Magee, *Science* **357**, 1033–1036 (2017).
12. J. O'Keefe, M. L. Recce, *Hippocampus* **3**, 317–330 (1993).
13. W. E. Skaggs, B. L. McNaughton, M. A. Wilson, C. A. Barnes, *Hippocampus* **6**, 149–172 (1996).
14. G. Dragoi, G. Buzsáki, *Neuron* **50**, 145–157 (2006).
15. K. D. Harris, J. Csicsvari, H. Hirase, G. Dragoi, G. Buzsáki, *Nature* **424**, 552–556 (2003).
16. J. C. Magee, D. Johnston, *Science* **275**, 209–213 (1997).
17. M. V. Tsodyks, W. E. Skaggs, T. J. Sejnowski, B. L. McNaughton, *Hippocampus* **6**, 271–280 (1996).
18. G. Dragoi, S. Tonegawa, *Nature* **469**, 397–401 (2011).
19. A. D. Grosmark, G. Buzsáki, *Science* **351**, 1440–1443 (2016).
20. A. Terrazas *et al.*, *J. Neurosci.* **25**, 8085–8096 (2005).
21. A. Cei, G. Girardeau, C. Drieu, K. E. Kanbi, M. Zugaro, *Nat. Neurosci.* **17**, 719–724 (2014).
22. S. Leutgeb, J. K. Leutgeb, A. Treves, M. B. Moser, E. I. Moser, *Science* **305**, 1295–1298 (2004).
23. S. Cheng, L. M. Frank, *Neuron* **57**, 303–313 (2008).
24. J. O'Neill, T. J. Senior, K. Allen, J. R. Huxter, J. Csicsvari, *Nat. Neurosci.* **11**, 209–215 (2008).
25. S. Li, W. K. Cullen, R. Anwyl, M. J. Rowan, *Nat. Neurosci.* **6**, 526–531 (2003).
26. A. P. Maurer, S. L. Cowen, S. N. Burke, C. A. Barnes, B. L. McNaughton, *Hippocampus* **16**, 785–794 (2006).
27. T. Feng, D. Silva, D. J. Foster, *J. Neurosci.* **35**, 4890–4902 (2015).
28. S. J. Middleton, T. J. McHugh, *Nat. Neurosci.* **19**, 945–951 (2016).
29. K. Zhang, I. Ginzburg, B. L. McNaughton, T. J. Sejnowski, *J. Neurophysiol.* **79**, 1017–1044 (1998).
30. T. J. Davidson, F. Kloosterman, M. A. Wilson, *Neuron* **63**, 497–507 (2009).
31. See supplementary materials.
32. C. Geisler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 7957–7962 (2010).
33. N. Burgess, M. Recce, J. O'Keefe, *Neural Netw.* **7**, 1065–1081 (1994).
34. A. Chadwick, M. C. W. van Rossum, M. F. Nolan, *eLife* **4**, e03542 (2015).
35. D. Robbe, G. Buzsáki, *J. Neurosci.* **29**, 12597–12605 (2009).
36. Y. Wang, S. Romani, B. Lustig, A. Leonardo, E. Pastalkova, *Nat. Neurosci.* **18**, 282–288 (2015).
37. M. A. A. van der Meer, A. D. Redish, *J. Neurosci.* **31**, 2843–2854 (2011).
38. M. W. Jones, M. A. Wilson, *PLOS Biol.* **3**, e402 (2005).
39. A. M. Wikenheiser, A. D. Redish, *Nat. Neurosci.* **18**, 289–294 (2015).

ACKNOWLEDGMENTS

We thank D. Robbe for comments on an earlier version of the manuscript, Y. Dupraz for technical support, and A. C. Segú for help in animal training. **Funding:** Supported by the Agence Nationale de la Recherche (ANR-15-CE16-0001-02), French Ministry of Research (C.D.), Collège de France (C.D.), and a joint grant from École des Neurosciences de Paris Île-de-France and LabEx MemoLife (ANR-10-LABX-54 MEMO LIFE, ANR-10-IDEX-0001-02 PSL*) (R.T.). **Author contributions:** C.D. and M.Z. designed the study; C.D. performed the experiments; C.D., R.T., and M.Z. designed the analyses; C.D. and R.T. performed the analyses; and C.D. and M.Z. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data of this study are available at http://dx.doi.org/10.6080/K0Z899MM.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/675/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S13
References (*40*–*45*)

22 February 2018; accepted 17 September 2018
10.1126/science.aat2952

PREDATION

Global pattern of nest predation is disrupted by climate change in shorebirds

Vojtěch Kubelka[1,2]*, Miroslav Šálek[3], Pavel Tomkovich[4], Zsolt Végvári[5,6], Robert P. Freckleton[7], Tamás Székely[2,8,9,10]*

Ongoing climate change is thought to disrupt trophic relationships, with consequences for complex interspecific interactions, yet the effects of climate change on species interactions are poorly understood, and such effects have not been documented at a global scale. Using a single database of 38,191 nests from 237 populations, we found that shorebirds have experienced a worldwide increase in nest predation over the past 70 years. Historically, there existed a latitudinal gradient in nest predation, with the highest rates in the tropics; however, this pattern has been recently reversed in the Northern Hemisphere, most notably in the Arctic. This increased nest predation is consistent with climate-induced shifts in predator-prey relationships.

Climate change is affecting organisms at a global scale in several ways (*1*–*4*), including directly altering demographic parameters such as adult survival (*5*) and reproduction (*1*) or through altered trophic interactions (*1*, *6*, *7*). Successful recruitment counters mortality and maintains viable populations; thus, disruption of reproductive performance can have detrimental effects on wild populations (*8*–*10*). Alterations in demographic parameters have been attributed to recent climate change (*1*, *5*, *11*), especially in the Arctic, where the consequences of warming are expected to be more pronounced (*6*, *12*). However, the evidence for impacts of climate change on species interactions is mixed, and to date there is no evidence that such interactions are changing globally (*1*–*3*).

Offspring mortality due to predation has a pivotal influence on the reproductive performance of wild populations (*8*, *13*–*15*), and extreme rates of predation can quickly lead to population declines or even species extinction (*16*). Thus, nest predation is a good indicator of the potential for reproductive recruitment in bird populations (*10*). Disruption to annual productivity through increased nest predation could have a detrimental effect on population dynamics and lead to increased extinction risks (*9*). To explore changes in spatial patterns of reproduction and potential alterations in trophic interactions due to changes in climate, we used nest predation data from shorebirds, a globally distributed group of ground-nesting birds that exhibit high interspecific similarity in nest appearance to potential predators and are exceptionally well-studied in the wild, including ecology, behavior, and demography (*10*, *17*, *18*). We collected data from both published and previously unpublished sources that included 38,191 nests in 237 populations of 111 shorebirds species from 149 locations, encompassing all continents across a 70-year time span (fig. S1 and table S1).

Using our comprehensive dataset in a spatiophylogenetic framework (*19*), we show that rates of nest predation increased over the past 70 years. Daily nest predation, as well as total nest predation (reflecting the full incubation period for a given species), have increased overall worldwide since the 1950s (Figs. 1 and 2, A and B; fig. S2, A and B; and table S2). Thus, total nest predation was historically (until 1999) on average 43 ± 2% (SEM), and this has increased to 57 ± 2% since 2000. However, the extent of change shows considerable geographical variation. In the tropics and south temperate areas, changes in daily and

---

[1]Department of Ecology, Charles University in Prague, Vinicna 7, 128 44, Prague, Czech Republic. [2]Milner Centre for Evolution, University of Bath, Claverton Down, Bath BA2 7AY, UK. [3]Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Kamycka 129, 165 21, Prague, Czech Republic. [4]Zoological Museum, Moscow MV Lomonosov State University, Bolshaya Nikitskaya Str 6, Moscow 125009, Russia. [5]Department of Conservation Zoology, University of Debrecen, Egyetem tér 1, H-4032 Debrecen, Hungary. [6]Hortobágy National Park Directorate, Sumen u. 2, H-4024 Debrecen, Hungary. [7]Department of Animal and Plant Sciences, University of Sheffield, Alfred Denny Building, Western Bank, Sheffield S10 2TN, UK. [8]Department of Evolutionary Zoology and Human Biology, University of Debrecen, Egyetem tér 1, H-4032, Debrecen, Hungary. [9]State Key Laboratory of Biocontrol and College of Ecology and Evolution, Sun Yat-sen University, Guangzhou 510275, China. [10]Ministry of Education Key Laboratory for Biodiversity Sciences and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing 100875, China.
***Corresponding author. E-mail: kubelkav@gmail.com (V.K.); t.szekely@bath.ac.uk (T.S.).**

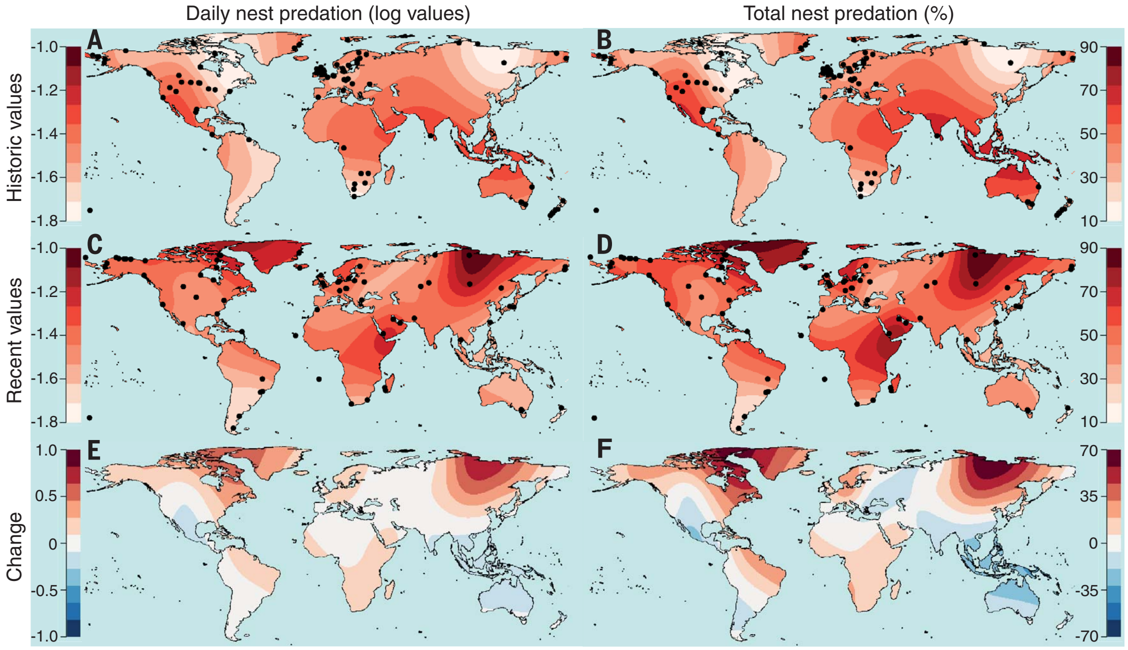


Fig. 1. Nest predation in shorebirds. (**A** and **B**) Historic rates of nest predation (1944–1999, 145 populations). (**C** and **D**) Recent rates of nest predation (2000–2016, 102 populations). (**E** and **F**) Changes between historic and recent nest predation rates. Dots show study locations. [(A), (C), and (E)] Daily nest predation (log transformed) (materials and methods). [(B), (D), and (F)] Total nest predation (percentage) (materials and methods and fig. S1, geographic coverage).

total nest predation were not statistically significant, whereas in the north temperate zone, and especially the Arctic, the increase was pronounced (Figs. 1 and 2, A and B; fig. S2, A and B; and table S2). This pattern holds across major clades of shorebirds (Fig. 2, C and D; fig. S2, C and D; and table S3) and is also observed within local populations, with daily and total nest predation increasing significantly in well-monitored north temperate and Arctic breeding populations (Fig. 2, E and F). Thus, the total nest predation was historically 35 ± 6%, which increased to 64 ± 5% in recent years for these long-term monitored populations (Fig. 2F and tables S4 and S5).

Life-history theory predicts that species that breed close to the equator should exhibit higher rates of nest predation than those of species that breed in temperate and polar latitudes, in part owing to the higher diversity of potential nest predators in the tropics, and there is an empirical support for this prediction (*14*, *15*, *20*, *21*). In line with theoretical expectations, historic rates of nest predation in shorebirds follow the parabolic relationship between both daily and total rates of nest predation and latitude (Fig. 3, fig. S3, and table S6).

However, in recent years, daily nest predation changed only modestly in the tropics and Southern Hemisphere (Fig. 3 and fig. S3), although it increased nearly twofold in the North temperate zone and threefold in the Arctic compared with historic values (Figs. 2A and 3). Thus, 70% of nests are now being depredated in the Arctic (Fig. 2B). As a consequence of latitude-dependent changes in nest predation, predation rates now increase from the equator to the Arctic, in contrast to the historic parabolic latitudinal pattern (Fig. 3, fig. S3, and table S6). Although data from Southern Hemisphere are scanty, they suggest no major changes in nest predation in southern regions (Fig. 1).

It is thought that climate change has influenced trophic interactions (*1*, *6*, *7*, *12*); therefore, to investigate whether altered rates of nest predation are driven by climate, we calculated the changes in ambient temperature in each shorebird population and tested whether the temperature changes predict the shifts in nest predation at a global scale (*19*). We used two proxies of climate change: the slope of annual mean temperature regressed against time and the standard deviation of annual mean temperatures measured over 30 years for each shorebird population. Higher rates of both daily and total nest predation were associated with increased ambient temperatures and temperature variations (Fig. 4). These results are robust to the choice of climatic variables over periods of 20, 30, or 40 years (table S7).

Because predation is the most common cause of breeding failure (*13*, *14*), our results imply declining reproductive success in a widely distributed avian taxon. This decline, unless compensated by higher juvenile or adult survival and/or increased production of clutches, will drive global population declines when recruitment is not sufficient to maintain existing population sizes (*9*, *10*). However, adult survival of long-distance migrants is also decreasing because of recent habitat loss at staging areas (*22*, *23*), and declining chick survival has been reported across Europe (*24*). Therefore, high-latitude breeders are squeezed by both poor breeding performance and reduced adult survival. Although tropical shorebirds may increase the number of breeding

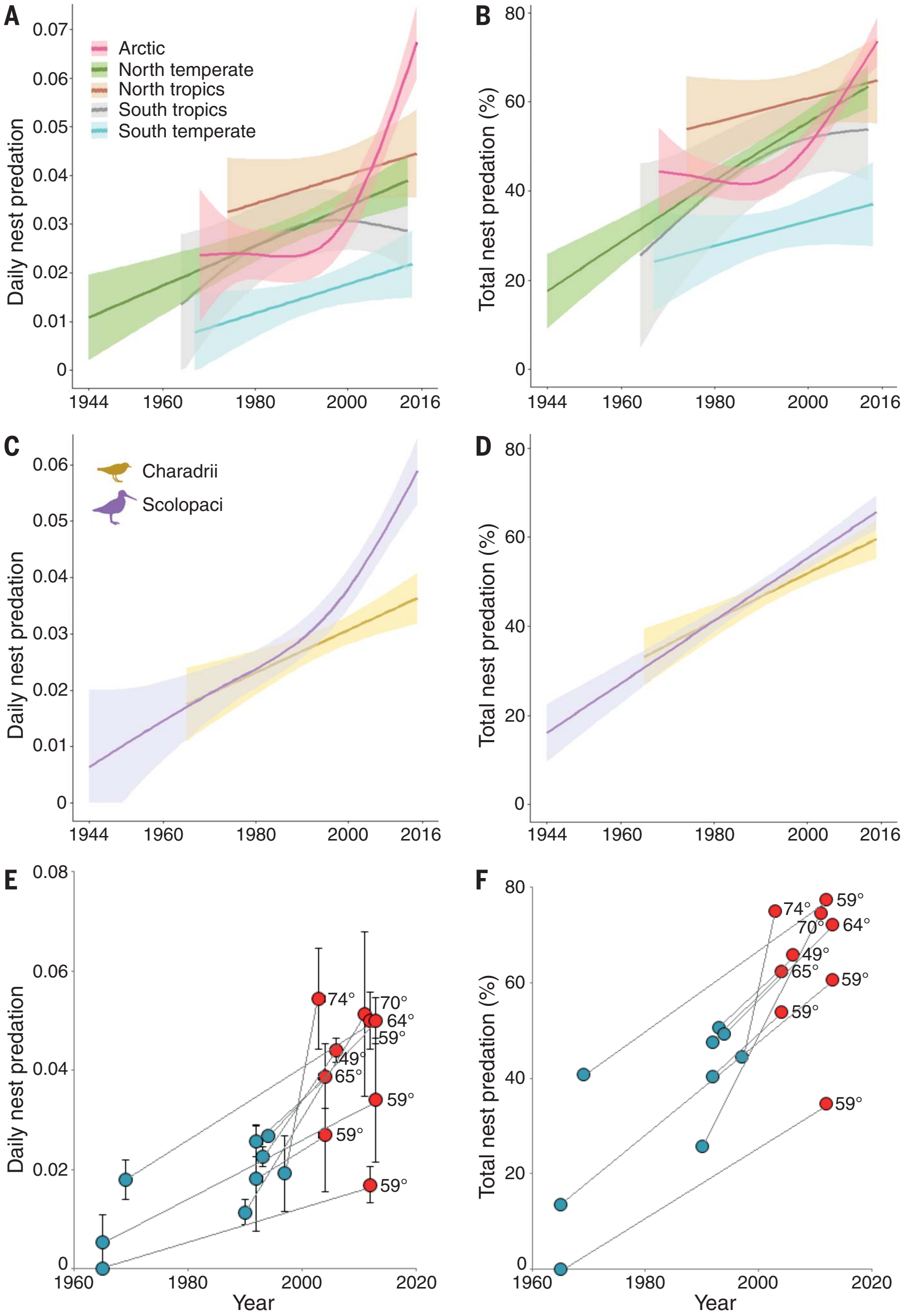


Fig. 2. Temporal changes in nest predation of shorebirds. (**A** and **B**) Nest predation rates for five latitudinal areas. Arctic, *n* = 86 populations; north temperate, *n* = 96 populations; north tropics, *n* = 17 populations; south tropics, *n* = 14 populations; south temperate, *n* = 24 populations. Area definitions are provided in (*19*), and model description is available in table S2. (**C** and **D**) Nest predation rates for plovers and allies (Charadrii, 110 populations) and sandpipers and allies (Scolopaci, 127 populations). Clade definitions are provided in (*19*), and model descriptions are available in table S3. (**E** and **F**) Local changes in nest predation rates for nine populations. Each dot represents mean ± SEM (E) over 2 to 19 breeding seasons for historic data (blue) and recent data (red). Latitude of the population is given next to the recent data (tables S4 and S5, model descriptions). [(A) to (D)] Generalized additive model fits with 95% confidence intervals. [(A), (C), and (E)] Daily nest predation. [(B), (D), and (F)] Total nest predation.

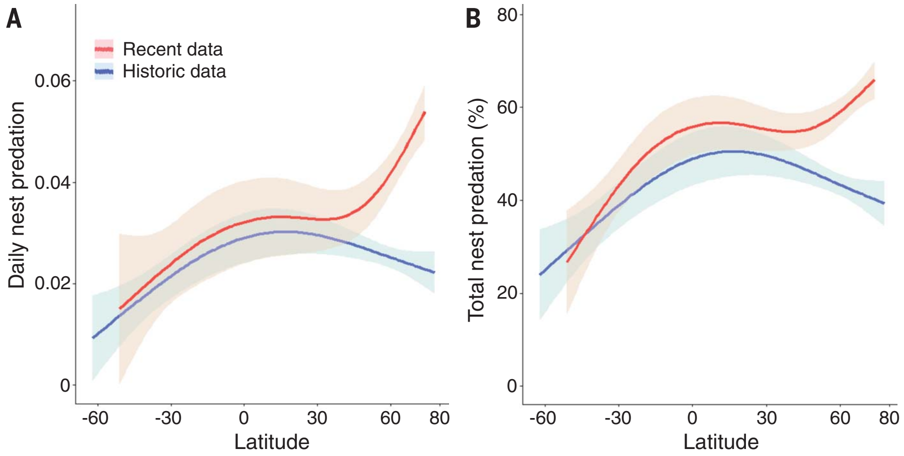


Fig. 3. Latitudinal gradient in historic versus recent nest predation of shorebirds. (**A**) Daily and (**B**) total nest predation rates (historic data 1944–1999, n = 145 populations; recent data 2000–2016, n = 102 populations), generalized additive model fits with 95% confidence intervals. (table S6, model descriptions) (*19*).

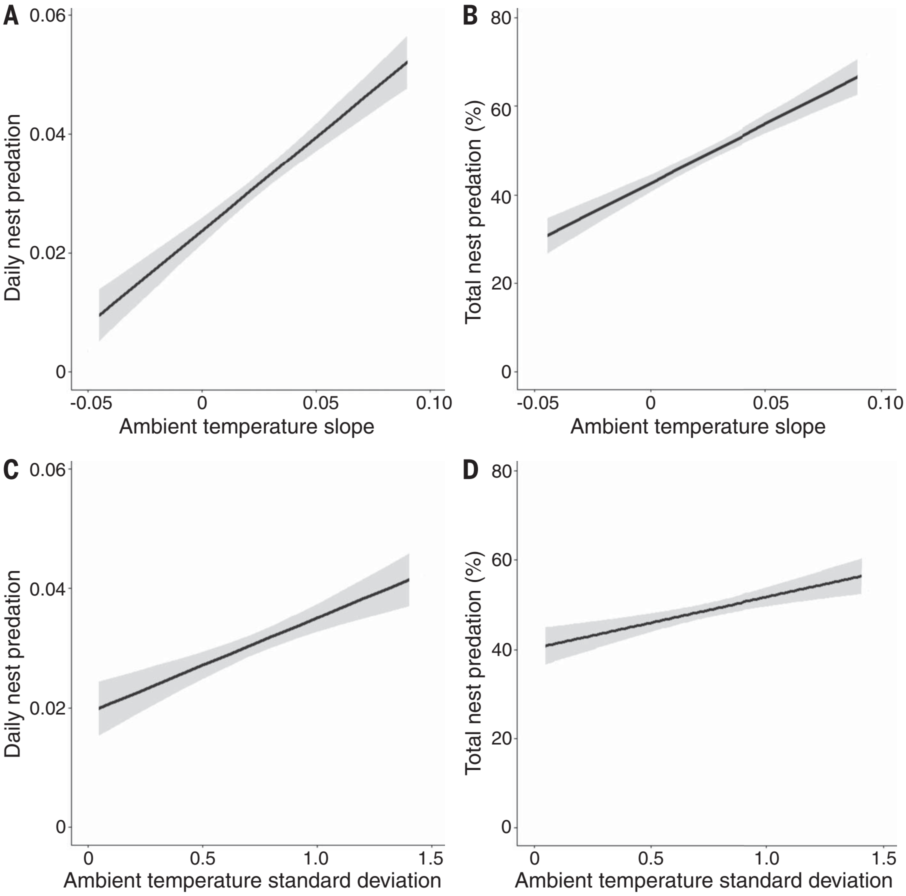


Fig. 4. Climate change predicts nest predation rates in shorebirds. (**A** and **B**) Relationship between (A) daily or (B) total nest predation rates and the slope of mean year temperatures. (**C** and **D**) Relationship between (C) daily or (D) total nest predation rates and the standard deviation of mean year temperatures. [(A) to (D)] Climatic data over 30 years before the last year of data collection; n = 247 values, generalized additive model fits with 95% confidence intervals (table S7, model descriptions) (*19*).

attempts and thus compensate for low breeding success, such compensation is limited at higher latitudes by short polar summers (*6*, *12*). Because most shorebirds are already declining (*18*, *23*, *25*), our results suggest that an important correlate of this decline is the elevated nest predation.

Climate change may influence nest predation rates in several ways (*1*, *6*, *12*). First, lemmings (*Lemmus* spp. and *Dicrostonyx* spp.), small rodents that represent the key component of the Arctic food web, have experienced a crash in their abundances and population cycling due to unsuitable snow cover as a result of ambient temperature increase and fluctuations (*26*–*28*). This change was documented over vast Arctic areas around the year 2000 (*26*–*28*), and the pattern was similar for temperate voles in Europe (*Microtus* spp. and *Myodes* spp.) (*29*, *30*). Changes in rodent abundances may have led to alterations in predator-prey interactions in the Northern Hemisphere, where predators who normally consume mainly rodents increased predation pressure on alternative prey, including shorebird nests (*12*, *28*). Second, the behavior and/or distribution of nest predators may have changed because of climate change; for instance, the distribution or densities of nest predators such as foxes (*Vulpes* spp.) may have increased, or their behavioral activity have changed, making them more successful egg-consumers (*4*, *6*, *12*). Third, vegetation structure may have changed around shorebird nests, leading to increased predation (*6*, *12*, *25*).

The demographic changes we report here have two major implications. First, migrating birds have been presumed to benefit from breeding in the Arctic as a consequence of lower predation pressure (*31*). Currently, however, the productivity of Arctic populations is declining because of high rates of nest predation, which suggests that energy-demanding long-distance migration to northern breeding grounds is no longer advantageous from a nest predation perspective. Thus, the Arctic now represents an extensive ecological trap (*32*) for migrating birds, with a predicted negative impact on their global population dynamics. Second, Arctic birds are likely to decline in the future because of the synergistic effects of the climatically driven increase of predation pressure at their breeding grounds, a trophic mismatch during the chick-rearing period owing to delayed chick hatching relative to the peak of food abundance (*6*, *33*), predicted shrinkage of suitable habitat (*6*, *12*), and reduced adult survival during migration (*22*, *23*). A future scientific challenge with crucial consequences for species conservation lies in disentangling the effects of these drivers on the overall viability of bird species.

We have demonstrated that rapid alterations in species interactions are occurring at a global scale and that these changes are related to altered climate. This underlines the need for understanding the effects of climate change not only for individuals and their populations but also for interactions in complex ecosystems, including prey and predators.

REFERENCES AND NOTES

1. B. R. Scheffers *et al*., *Science* **354**, aaf7671 (2016).
2. J. M. Tylianakis, R. K. Didham, J. Bascompte, D. A. Wardle, *Ecol. Lett.* **11**, 1351–1363 (2008).
3. C. Parmesan, *Annu. Rev. Ecol. Evol. Syst.* **37**, 637–669 (2006).
4. J. L. Blois, P. L. Zarnetske, M. C. Fitzpatrick, S. Finnegan, *Science* **341**, 499–504 (2013).
5. C. Barbraud, H. Weimerskirch, *Nature* **411**, 183–186 (2001).
6. O. Gilg *et al*., *Ann. N. Y. Acad. Sci.* **1249**, 166–190 (2012).
7. W. J. Sydeman, E. Poloczanska, T. E. Reed, S. A. Thompson, *Science* **350**, 772–777 (2015).
8. D. Lack, *Population Studies of Birds* (Oxford Univ. Press, 1966).
9. P. M. Bennett, I. P. F. Owens, *Evolutionary Ecology of Birds: Life Histories, Mating Systems, and Extinction* (Oxford Univ. Press, 2002).
10. M. A. Colwell, *Shorebird Ecology, Conservation, and Management* (Univ. of California Press, 2010).
11. C. D. G. Harley, *Science* **334**, 1124–1127 (2011).
12. H. Meltofte *et al*., *Effects of Climate Variation on the Breeding Ecology of Arctic Shorebirds – Meddelelser om Grønland Bioscience 59* (Danish Polar Center, 2007).
13. T. Caro, *Antipredator Defences in Birds and Mammals* (Univ. of Chicago Press, 2005).
14. R. Ricklefs, *Smithson. Contrib. Zool.* **9**, 1–48 (1969).
15. A. F. Skutch, *Ornithol. Monogr.* **36**, 575–594 (1985).
16. T. M. Blackburn, P. Cassey, R. P. Duncan, K. L. Evans, K. J. Gaston, *Science* **305**, 1955–1958 (2004).
17. J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, E. de Juana, Eds., *Handbook of the Birds of the World Alive* (Lynx Edicions, 2018); www.hbw.com.
18. BirdLife International, IUCN Red List for birds (2018); http://datazone.birdlife.org/species/search.
19. Materials and methods are available as supplementary materials.
20. B. J. M. Stutchbury, E. S. Morton, *Behavioral Ecology of Tropical Birds* (Academic Press, 2001).
21. B. A. DeGregorio, S. J. Chiavacci, T. J. Benson, J. H. Sperry, P. J. Weatherhead, *Bioscience* **66**, 655–665 (2016).
22. T. Piersma *et al*., *J. Appl. Ecol.* **53**, 479–490 (2016).
23. C. E. Studds *et al*., *Nat. Commun.* **8**, 14895 (2017).
24. M. Roodbergen, B. van der Werf, H. Hötker, *J. Ornithol.* **153**, 53–74 (2012).
25. M. Munro, *Nature* **541**, 16–20 (2017).
26. R. A. Ims, J.-A. Henden, S. T. Killengreen, *Trends Ecol. Evol.* **23**, 79–86 (2008).
27. K. L. Kausrud *et al*., *Nature* **456**, 93–97 (2008).
28. Y. Aharon-Rotman *et al*., *Oikos* **124**, 861–870 (2014).
29. T. Cornulier *et al*., *Science* **340**, 63–66 (2013).
30. S. M. Bierman *et al*., *Am. Nat.* **167**, 583–590 (2006).
31. L. McKinnon *et al*., *Science* **327**, 326–327 (2010).
32. M. A. Schlaepfer, M. C. Runge, P. W. Sherman, *Trends Ecol. Evol.* **17**, 474–480 (2002).
33. J. A. van Gils *et al*., *Science* **352**, 819–821 (2016).
34. V. Kubelka *et al*., Data and R codes from: Global pattern of nest predation is disrupted by climate change in shorebirds. Dryad (2018); doi: 10.5061/dryad.45g90h4.

ACKNOWLEDGMENTS

We thank the people who provided us with previously unpublished data or additional information in regards to published articles: A. I. Antonov, C. Carmona, E. Cepáková, R. A. Cerboncini, V. L. D'Amico, P. Donald, L. Eberhart-Phillips, I. Fisher, M. Hancock, G. D. Hevia, V. V. Golovnyuk, A. T. Gonçalves, S. Junker, C. Lishman, B. Katrínardóttir, M. C. López, R. H. Macedo, E. Makrigianni, M. S. Nadeem, E. Santos, K. Seymour, M. Sládeček, M. Soloviev, M. L. Stephens, and V. Štorek. H. J. Hansen from Aarhus University is thanked for data from Zackenberg Northeast Greenland, and Aarhus University is thanked for providing logistical support at Zackenberg Research Station. We thank the people who helped us with gray literature and useful contacts: T. Braga, S. Ferrari, P. Gonzales, N. Karniolova, J. Mlíkovský, R. Sheldon, O. Thorup, T. Sviridova, and N. Warncock. We appreciate the suggestions of S. Ancona, C. Carmona, L. Eberhart-Phillips, D. Frynta, D. Horák, L. D. Hurst, L. Kratochvíl, Vít Kubelka, C. Küpper, P. Linhart, J. Mlíkovský, A. Petrusek, J. Reif, M. Sládecek, D. Storch, and K. Weidinger on various stages of this work. V.K. has special thanks to S. Wilcox, Alexander Library librarian in Oxford; V. Dočekalová, who helped with Shorebird Demographics Network data processing; and A. Tószögyová and L. Eberhart-Phillips for graphical advice and help. J. Reif, K. Weidinger, L. Eberhart-Phillips, D. Hořák, and D. Storch helped us with their critical comments on the previous versions of the manuscript. **Funding:** We appreciate the following funding sources to V.K. (Charles University Grant Agency, GAUK 927516), M.Š. (Ministry of Education, Youth and Sports of the Czech Republic, Kontakt II LH 13278), P.T. (MSU Zoological Museum, AAAA-A16-116021660077-3), and T.S. (Royal Society Wolfson Merit Award, NKFIH-2558-1/2015 and ÉLVONAL-KKP 126949 of Hungarian government). **Author contributions:** V.K., T.S., and M.Š. conceived the study; V.K. collected the data with the help from P.T., M.Š., Z.V., and T.S.; V.K., T.S., M.Š., and R.P.F. developed the methods; V.K. and R.P.F. analyzed the data with input from T.S., Z.V., and M.Š.; and V.K. wrote the paper with inputs from all coauthors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Climatic data are freely available at www.cru.uea.ac.uk/data. Sources of primary nest predation data are presented in table S1. Data and R codes are available at Dryad (*34*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/680/suppl/DC1
Materials and Methods
Figs. S1 to S3
Tables S1 to S8
References (*35–218*)

12 April 2018; accepted 5 September 2018
10.1126/science.aat8695

NEONICOTINOIDS

Neonicotinoid exposure disrupts bumblebee nest behavior, social networks, and thermoregulation

James D. Crall[1,2,3*], Callin M. Switzer[4,5], Robert L. Oppenheimer[6], Ashlee N. Ford Versypt[7,8], Biswadip Dey[9], Andrea Brown[1], Mackay Eyster[10], Claire Guérin[11], Naomi E. Pierce[1], Stacey A. Combes[12], Benjamin L. de Bivort[1,3]

Neonicotinoid pesticides can negatively affect bee colonies, but the behavioral mechanisms by which these compounds impair colony growth remain unclear. Here, we investigate imidacloprid's effects on bumblebee worker behavior within the nest, using an automated, robotic platform for continuous, multicolony monitoring of uniquely identified workers. We find that exposure to field-realistic levels of imidacloprid impairs nursing and alters social and spatial dynamics within nests, but that these effects vary substantially with time of day. In the field, imidacloprid impairs colony thermoregulation, including the construction of an insulating wax canopy. Our results show that neonicotinoids induce widespread disruption of within-nest worker behavior that may contribute to impaired growth, highlighting the potential of automated techniques for characterizing the multifaceted, dynamic impacts of stressors on behavior in bee colonies.

Animal pollinators support biodiversity and agricultural yields (*1*, *2*), and there is growing concern over the causes and consequences of declining bee populations (*3*, *4*). Mounting evidence indicates that neonicotinoid pesticides can negatively affect both commercial honey bee (*5*, *6*) and wild bee (*7*) populations. Neonicotinoids are agonists of nicotinic acetylcholine receptors and therefore disrupt cholinergic signaling in the insect central nervous system. Neonicotinoids are believed to reduce growth of developing brood (*7*–*9*) by impairing foraging behavior [including navigation (*5*, *10*) and floral learning (*11*, *12*)], leading to reduced colony resource intake (*13*, *14*). Recent work, however, shows that neonicotinoid exposure can impair colony growth without altering foraging (*15*) [and vice versa (*16*)].

In addition to foraging, workers in social insect colonies perform critical tasks within the nest (e.g., larval incubation and feeding, cleaning, and nest construction) that are vital for colony development. Although nest workers are exposed to neonicotinoid residues (*7*, *17*) that may affect behavior [including physiology (*18*, *19*) and locomotion (*20*–*22*)], neonicotinoids' effects on within-nest behaviors are poorly understood.

To investigate imidacloprid's effects on bumblebee (*Bombus impatiens*) nest behavior, we combined a system for automated behavioral tracking of uniquely identified workers [BEEtag (*23*, *24*)] with a robotic observation platform (Fig. 1, A to C), allowing long-term (12-day, Fig. 1 and figs. S1 and S2) tracking of uniquely identified bumblebee workers and queens. Colonies were given ad libitum access to either pure nectar (control, *n* = 9), or nectar containing field-realistic concentrations of imidacloprid, a globally prevalent neonicotinoid [*n* = 9, 6 parts per billion (ppb)].

Chronic imidacloprid exposure impairs a suite of worker behaviors within the nest. Workers in imidacloprid-exposed colonies spent significantly less time active (Fig. 1D and fig. S2). Imidacloprid exposure also reduced rates of nursing among workers (Fig. 1E and fig. S2) and shifted spatial occupancy toward the nest periphery (Fig. 1F and fig. S2).

Behavioral effects of imidacloprid differed markedly between night and day [14:10 light: dark (L:D) cycle] within the colony (Fig. 1); reductions in activity were stronger at night (Fig. 1D and fig. S2) than during the day (but were significant during both night and day, see fig. S2) and effects on daytime activity declined over time (with effects undetectable by the end of the exposure period, Fig. 1D). Reductions in nursing and distance from the nest center were significant at night, but not during the day (Fig. 1, E and F, and fig. S2). Imidacloprid reduced social network density compared to controls, consistent with effects of thiacloprid in honey bees (*25*), although this effect was also only significant at night (Fig. 1G; network density, night: bootstrap $p = 0.0042$; network density, day: bootstrap $p = 0.71$). Imidacloprid increased movement speed in workers, although this effect appears delayed and is only significant during the day (fig. S2). Whereas previous work has shown that imidacloprid exposure can lead to either hyperactivity or immobility depending on dose (*26*), these results show that both effects can occur at the same concentration. Imidacloprid also reduced activity and nursing in queens (fig. S2), consistent with results in honey bees (*27*) and bumblebees (*28*, *29*). Body concentrations of imidacloprid measured after the experiment were independent of foraging activity, confirming that even nonforaging nest workers are exposed to imidacloprid (fig. S3, mean concentration = 2.25 fmol imidacloprid per milligram of body mass).

To confirm that imidacloprid induces direct and rapid changes in nest behavior after exposure, we recorded behavior of workers in four additional *B. impatiens* colonies on the days immediately before and after individually administered, acute consumption of 0 (control), 0.1, or 1.0 ng of imidacloprid (Fig. 2; see supplementary text for justification of doses).

Acute imidacloprid exposure altered nest behavior within 24 hours, with effects qualitatively similar to those of chronic exposure (Fig. 2 and tables S1 to S4). Bees fed 1.0 ng of imidacloprid had reduced activity and nursing, were located further from the nest center, and had reduced social interactions compared to controls (Fig. 2 and table S2). Bees fed 1.0 ng of imidacloprid showed reduced foraging, driven by a reduction in nonforagers initiating foraging after treatment—rather than a decrease in foraging among foragers (tables S2 to S4). Bees fed 0.1 ng of imidacloprid showed no significant differences in behavior compared to controls (fig. 2 and tables S1 to S4).

To gain insight into the mechanisms underlying the multiple behavioral outcomes of imidacloprid exposure, we developed a spatially explicit, agent-based model of worker nest movements (supplementary text and figs. S4 to S6). Bees were modeled as either active (moving) or inactive (not moving) at each time point and Markovian transitions were used to switch between these states (fig. S4). The transition rates were modulated by contact with nestmates and the bee's location on or off the nest and were directly estimated from experiments (fig. S5). When bees were active, they moved with a random walk biased toward the nest center (fig. S4; parameters also fit from experiments).

We then used this model to disentangle the effects of imidacloprid on activity, space use, and social interactions and found evidence that these multiple outcomes of imidacloprid exposure are functionally linked; simulations isolating imidacloprid's direct effects on activity (both spontaneously in isolated workers and when activity is modulated by social contact; fig. S5) resulted in shifts in spatial occupancy and interaction rate within the nest (fig. S6). These effects are

---

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. [2]Planetary Health Alliance, Harvard University, Cambridge, MA, USA. [3]Center for Brain Science, Harvard University, Cambridge, MA, USA. [4]eScience Institute, University of Washington, Seattle, WA, USA. [5]Department of Biology, University of Washington, Seattle, WA, USA. [6]Department of Biological Sciences, University of New Hampshire, Durham, NH, USA. [7]School of Chemical Engineering, Oklahoma State University, Stillwater, OK, USA. [8]Interdisciplinary Toxicology Program, Oklahoma State University, Stillwater, OK, USA. [9]Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ, USA. [10]Biology Department, University of Massachusetts Amherst, Amherst, MA, USA. [11]Department of Ecology and Evolution, Université de Lausanne, Lausanne, Switzerland. [12]Department of Neurobiology, Physiology, and Behavior, University of California, Davis, Davis, CA, USA.
***Corresponding author. Email: jcrall@oeb.harvard.edu**

compounded by imidacloprid's reduction of attraction to the nest center (fig. S6 and supplementary text).

To test neonicotinoids' effects on nest behavior under realistic conditions, we quantified nest thermoregulation performance in free-foraging *B. impatiens* colonies (Fig. 3, A and B). Colonies were given ad libitum access to nectar containing imidacloprid (6 ppb, $n = 9$) or pesticide-free nectar ($n = 9$) within the nest, but foraged outdoors to gather pollen. For each colony, we measured the surface temperature of the brood and air temperature within the nest chamber, in addition to outdoor air temperature (Fig. 3, A and B). Imidacloprid impaired thermoregulation of the developing brood (Fig. 3, C and D, and fig. S7, permutation test, $p = 0.005$, tables S5 and S6) and nest air temperature (Fisher's

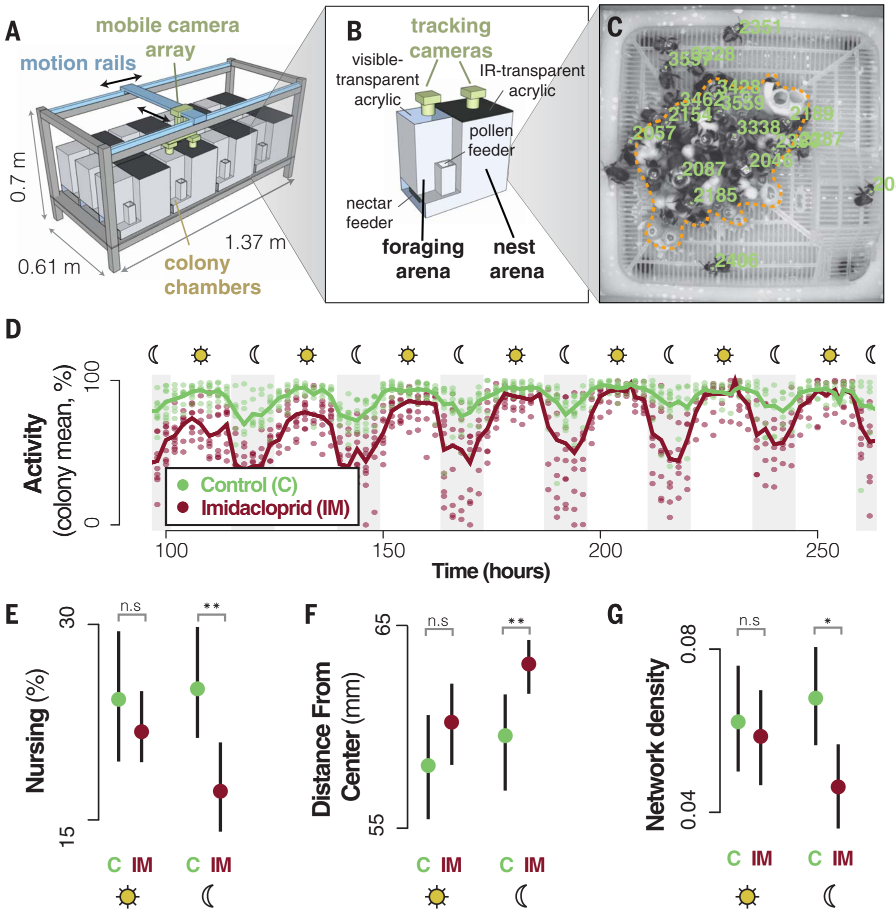


Fig. 1. Chronic exposure to imidacloprid alters nest behavior and social interactions in bumblebee colonies. (**A** and **B**) Schematic diagrams of (A) robotic platform for multicolony (in a 4 by 3 array) behavioral tracking and (B) a single colony chamber. (**C**) Example tracking of nest workers, with unique identification numbers shown in green. Orange dotted line shows the nest structure. (**D**) Colony mean percentage of time active over 7 consecutive days (with time indicating hours after exposure). Filled circles represent mean activity levels for a single colony (averaged across all individual workers) for a single 5-min trial, and solid lines show mean values for treatment groups (control colonies, $n = 9$, in green; imidacloprid-exposed colonies, $n = 9$, in red). Gray blocks and Sun/Moon symbols show the 14:10 hour L:D cycle in the tracking arena. (**E**) Percentage of time engaged in nursing. (**F**) Mean distance to the nest center and (**G**) social network density [proportion of possible pairwise interactions between workers that actually occur, during a single 5-min trial] by treatment group and time of day. $^*p < 0.05$, $^{**}p < 0.005$, based on 10,000 hierarchical bootstrap replicates. Solid markers in (E) to (G) show group means, and black bars indicate 95% bootstrap confidence intervals, with control and imidacloprid-exposed colonies shown in green and red, respectively. n.s., not significant.

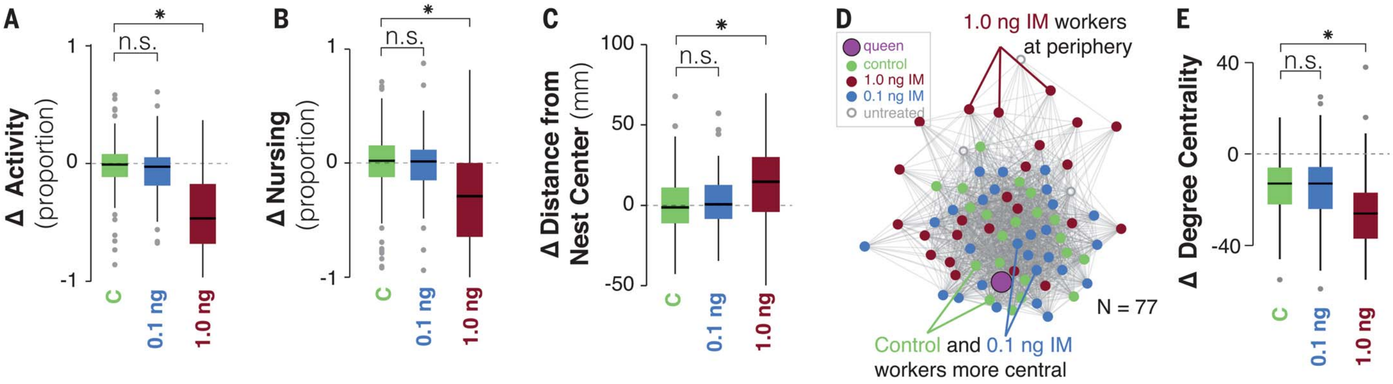


Fig. 2. Effects of imidacloprid on nest behavior occur rapidly after acute exposure. Changes in (**A**) activity, (**B**) nursing, and (**C**) distance from nest center of workers after exposure to different imidacloprid treatments. (**D**) Social network diagram of a representative colony, with nodes positioned by a force-directed algorithm. Circles represent individual bees, with gray lines drawn between bees that interacted during a 1-hour trial. The queen is shown as a purple circle, and untreated workers are shown as open gray circles. (**E**) Change in degree centrality (i.e., number of unique social interactions) after exposure by treatment. Boxplots show median (thick black lines), interquartile range (solid box), and range (thin lines, 75th and 25th percentile ± 1.5 × IQR), with outliers shown in gray. Behavioral changes were calculated as the difference in individual behavior 24 hours after versus 24 hours before exposure. In all panels, workers exposed to 1.0, 0.1, or 0 ng of imidacloprid are shown in red, blue, and green, respectively. $^*p < 0.001$. n.s., not significant.

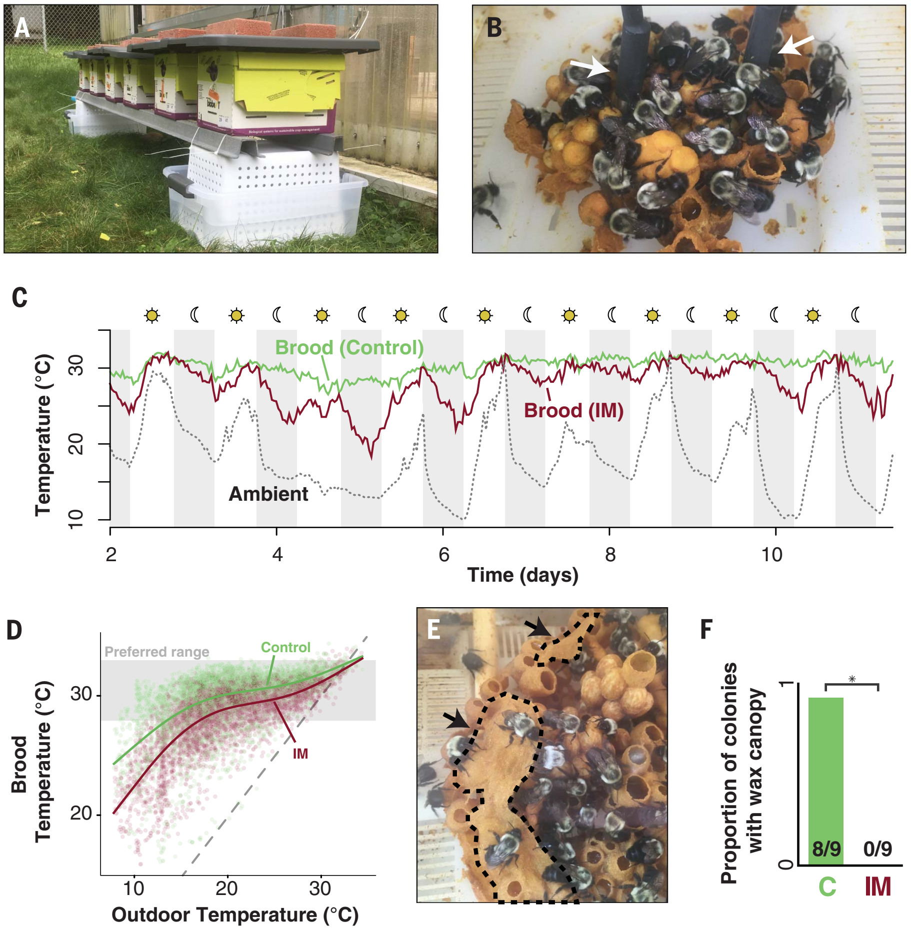


Fig. 3. Chronic imidacloprid exposure disrupts brood thermoregulation in *Bombus impatiens*. (**A**) Outdoor *B. impatiens* colonies with digital temperature sensors (**B**) (white arrows). (**C**) Example brood (solid lines) temperatures from one imidacloprid-exposed colony (red, IM) and one control (green, C) colony. Dotted line shows outdoor air temperature. (**D**) Brood versus outdoor temperatures for control (C, green) and treated (IM, red). Transparent markers show individual measurements across all colonies, and solid lines show LOESS-smoothed trends by treatment. Dashed line: brood temperature = outdoor temperature. (**E**) Example of a partially constructed insulating wax canopy (black arrows and dashed lines) covering brood cells. (**F**) Proportion of colonies that had a partially or completely constructed wax canopy by the end of the experiment, by treatment. Asterisk indicates significant difference between groups (p = 0.0005, permutation test).

exact test, p = 0.009, tables S5 and S6), with stronger effects occurring at lower temperatures (Fig. 3D). This result confirms that neonicotinoids' effects on thermogenesis in individual, isolated honey bees (*18*) and bumblebees (*19*) extend to colony temperature regulation under field conditions. We found a significant interaction between exposure and the direction of temperature change, suggesting that the effect of imidacloprid on thermoregulation may be stronger when air temperature is rising (fig. S7 and tables S5 and S6). Imidacloprid-treated colonies were also less likely to construct an insulating wax canopy around the developing brood, an important behavioral adaptation to cold (*30*) (Fig. 3, E and F, permutation test, p = 0.0005).

Large-scale field studies have revealed that the impacts of neonicotinoids on bee colonies can vary substantially depending on environmental context (*6*, *7*), highlighting the need for improved understanding of the mechanisms by which neonicotinoids affect workers and colonies. Our results suggest that reduced brood growth in neonicotinoid-exposed colonies (*7*) could result from impaired nursing behavior and temperature control by nest workers, in addition to reduced colony resource intake (*13*, *14*). These results support previous findings that neonicotinoids impair worker hygienic behavior in honey bees at higher concentrations [e.g., 50 ppb or higher (*27*)] and over extended exposure periods [e.g., 12 weeks (*31*)].

Our results highlight the multifaceted behavioral impacts of neonicotinoid exposure; imidacloprid's effects on nest behavior vary substantially with time of day (Fig. 1 and fig. S2), exposure affects both mobility and sensory decision-making (Figs. 1 and 2, fig. S5, and supplementary text), and the impacts of imidacloprid on brood thermoregulation are nonlinear (Fig. 3 and tables S5 and S6) and dynamic (fig. S7 and tables S5 and S6). These results illustrate the potential of high-throughput, automated analysis for improving our understanding of the context-specific effects of neonicotinoids, as well as efficiently screening agrochemicals more generally for sublethal impacts on pollinators.

REFERENCES AND NOTES

1. A. M. Klein *et al.*, *Proc. Biol. Sci.* **274**, 303–313 (2007).
2. N. Myers, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2764–2769 (1996).
3. G. K. Meffe, *Conserv. Biol.* **12**, 8–17 (1998).
4. N. Gallai, J.-M. Salles, J. Settele, B. E. Vaissière, *Ecol. Econ.* **68**, 810–821 (2009).
5. J. Fischer *et al.*, *PLOS ONE* **9**, e91364 (2014).
6. B. A. Woodcock *et al.*, *Science* **356**, 1393–1395 (2017).
7. M. Rundlöf *et al.*, *Nature* **521**, 77–80 (2015).
8. P. R. Whitehorn, S. O'Connor, F. L. Wackers, D. Goulson, *Science* **336**, 351–352 (2012).
9. R. J. Gill, O. Ramos-Rodriguez, N. E. Raine, *Nature* **491**, 105–108 (2012).
10. M. Henry *et al.*, *Science* **336**, 348–350 (2012).
11. D. A. Stanley, K. E. Smith, N. E. Raine, *Sci. Rep.* **5**, 16508 (2015).
12. K. Tan *et al.*, *Sci. Rep.* **5**, 10989 (2015).
13. R. J. Gill, N. E. Raine, *Funct. Ecol.* **28**, 1459–1471 (2014).
14. H. Feltham, K. Park, D. Goulson, *Ecotoxicology* **23**, 317–323 (2014).
15. A. N. Arce *et al.*, *J. Appl. Ecol.* **54**, 1199–1208 (2016).
16. D. A. Stanley, A. L. Russell, S. J. Morrison, C. Rogers, N. E. Raine, *J. Appl. Ecol.* **53**, 1440–1449 (2016).
17. E. A. D. Mitchell *et al.*, *Science* **358**, 109–111 (2017).
18. S. Tosi *et al.*, *J. Insect Physiol.* **93-94**, 56–63 (2016).
19. R. Potts *et al.*, *J. Insect Physiol.* **104**, 33–39 (2018).
20. S. M. Williamson, S. J. Willis, G. A. Wright, *Ecotoxicology* **23**, 1409–1418 (2014).
21. P. Medrzycki, R. Montanari, L. Bortolotti, S. Maini, C. Porrini, *Bull. Insectol.* **56**, 59–62 (2017).
22. S. Tosi, J. C. Nieh, *Sci. Rep.* **7**, 15132 (2017).
23. J. D. Crall, N. Gravish, A. M. Mountcastle, S. A. Combes, *PLOS ONE* **10**, e0136487 (2015).
24. J. D. Crall *et al.*, *Nat. Commun.* **9**, 1201 (2018).
25. N. Forfert, R. F. A. Moritz, *J. Apic. Res.* **56**, 467–474 (2017).
26. M. Lambin, C. Armengaud, S. Raymond, M. Gauthier, *Arch. Insect Biochem. Physiol.* **48**, 129–134 (2001).
27. J. Wu-Smart, M. Spivak, *Sci. Rep.* **6**, 32108 (2016).
28. M. Leza, K. M. Watrous, J. Bratu, S. H. Woodard, *Proc. Biol. Sci.* **285**, 20180761–20180769 (2018).
29. G. L. Baron, N. E. Raine, M. J. F. Brown, *Proc. Biol. Sci.* **284**, 20170123 (2017).
30. J. G. Holland, A. F. G. Bourke, *Funct. Ecol.* **29**, 1209–1217 (2015).
31. N. Tsvetkov *et al.*, *Science* **356**, 1395–1397 (2017).
32. J. D. Crall, C. M. Switzer, R. L. Oppenheimer, A. N. Ford Versypt, B. Dey, A. Brown, M. Eyster, C. Guérin, N. E. Pierce,

S. A. Combes, B. L. de Bivort, Neonicotinoid exposure disrupts bumblebee nest behavior, social networks, and thermoregulation. Zenodo (2018).

ACKNOWLEDGMENTS

We thank C. Vidoudez for help in chemical analysis; D. Zucker for help in the design of the robotic testing platform; and S. Donoughe, A. Kao, and members of the Pierce, Combes, and de Bivort labs, as well as three anonymous reviewers, for insightful comments on drafts of this manuscript. **Funding:** This material is based on work supported in part by BioBest, a National Science Foundation Graduate Research Fellowship (J.D.C.), a Winslow Foundation grant (J.D.C), funding from the Rockefeller Foundation (J.D.C.), the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program (C.M.S.), UW Data Science Grant from the Moore and Sloan Foundations (C.M.S.), AFOSR grant FA9550-14-1-0398 (C.M.S.), National Science Foundation grant (CAREER IOS-1253677) to S.A.C., and NSF OIS-1257543 to N.E.P. B.L.dB. was supported by a Sloan Research Fellowship, a Klingenstein-Simons Fellowship Award, and the National Science Foundation under grant no. IOS-1557913. Modeling work by J.D.C., A.N. F.V, and B.D. was supported in part by workshop funding ("Data-Driven Modeling of Collective Behavior and Emergent Phenomena") from the Statistical and Applied Mathematical Sciences Institute (SAMSI). **Author contributions:** J.D.C., C.M.S., N.E.P., S.A.C., and B.L.dB. designed experiments. J.D.C, R.L.O., and C.M.S. performed acute exposure experiments. J.D.C., C.G., and B.L.dD. designed and constructed the robotic tracking arenas and platform. J.D.C., R.L.O., C.M.S., A.B., and M.E. collected data and performed experiments. J.D.C. performed automated tracking, behavioral analysis, and prepared solutions for chronic behavior and field experiments. J.D.C., A.N.F.V, and B.D. designed and implemented the agent-based model. C.M.S. mixed the solutions of imidacloprid for acute experiments and performed statistical analysis for acute experiments and thermoregulation experiments. C.M.S. and J.D.C. performed statistical analysis for chronic imidacloprid experiments. All authors wrote the manuscript. **Competing interests**: The authors declare no competing interests. **Data and materials availability:** Associated data and custom scripts are deposited at Zenodo (*32*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/683/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Tables S1 to S6
References (*33–54*)
Movies S1 to S3

23 February 2018; accepted 26 September 2018
10.1126/science.aat1598

ANTIBIOTIC RESISTANCE

Heterogeneity in efflux pump expression predisposes antibiotic-resistant cells to mutation

Imane El Meouche[1,2] and Mary J. Dunlop[1,2]*

Antibiotic resistance is often the result of mutations that block drug activity; however, bacteria also evade antibiotics by transiently expressing genes such as multidrug efflux pumps. A crucial question is whether transient resistance can promote permanent genetic changes. Previous studies have established that antibiotic treatment can select tolerant cells that then mutate to achieve permanent resistance. Whether these mutations result from antibiotic stress or preexist within the population is unclear. To address this question, we focused on the multidrug pump AcrAB-TolC. Using time-lapse microscopy, we found that cells with higher *acrAB* expression have lower expression of the DNA mismatch repair gene *mutS*, lower growth rates, and higher mutation frequencies. Thus, transient antibiotic resistance from elevated *acrAB* expression can promote spontaneous mutations within single cells.

Antibiotic resistance is a major public health problem and is primarily the result of genetic changes that allow microorganisms to overcome the effects of antimicrobial drugs (*1*). However, genetic changes are not the only way that bacteria can tolerate antibiotics. Transient resistance mechanisms allow cells to temporarily resist drug treatment (*2*), playing a critical role in recalcitrant and recurrent infections (*3*). Examples include bacterial persistence, where cells temporarily enter a dormant state to block drug activity (*4*), and expression of efflux pumps to export antibiotics (*5*–*7*). We asked whether transient antibiotic resistance can lead to permanent antibiotic resistance by providing a window of opportunity in which cells can mutate. A recent study showed that bacterial persistence precedes resistance. In this state, tolerant cells can subsequently acquire mutations conferring resistance (*8*). Additionally, antibiotics often induce stress response mechanisms, which can lead to mutations. For instance, the low-fidelity, mutation-prone polymerases Pol II, Pol IV, and Pol V are induced during the SOS response to DNA damage (*9*). A question remains whether differences in mutation frequency predate antibiotic treatment or whether they are induced by the stress.

We focused on transient resistance arising from heterogeneity in expression of the AcrAB-TolC efflux pump found in many pathogens (*6*, *7*, *10*). The pump recognizes and exports β-lactam, tetracycline, and fluoroquinolone antibiotics, among others (*11*), by using the inner membrane protein AcrB, which works together with the periplasmic linker AcrA and the outer membrane channel TolC (*10*, *12*). *acrA* and *acrB* are commonly arranged together on an operon, whereas *tolC* is expressed elsewhere in the genome.

Recent reports have highlighted the importance of cell-to-cell variability in pump expression (*6*, *7*). For example, in *Escherichia coli*, AcrAB-TolC pumps partition heterogeneously, with pumps accumulating at the old pole, resulting in increased resistance levels in the subset of cells with higher efflux pump expression (*7*). We asked whether cell-to-cell heterogeneity in pump expression results in differences in the spontaneous mutation rate in addition to its known role in producing single-cell differences in transient antibiotic resistance.

We first calculated the spontaneous mutation frequency in *E. coli* strains with and without efflux pumps. To identify mutations not induced by stress, we performed these measurements in the absence of antibiotics. We plated mid-exponential phase cultures on LB agar with and without rifampicin and calculated the mutation

---

[1] Department of Biomedical Engineering and Biological Design Center, Boston University, Boston, MA 02215, USA. [2] School of Engineering, University of Vermont, Burlington, VT 05405, USA.
***Corresponding author. Email: mjdunlop@bu.edu**

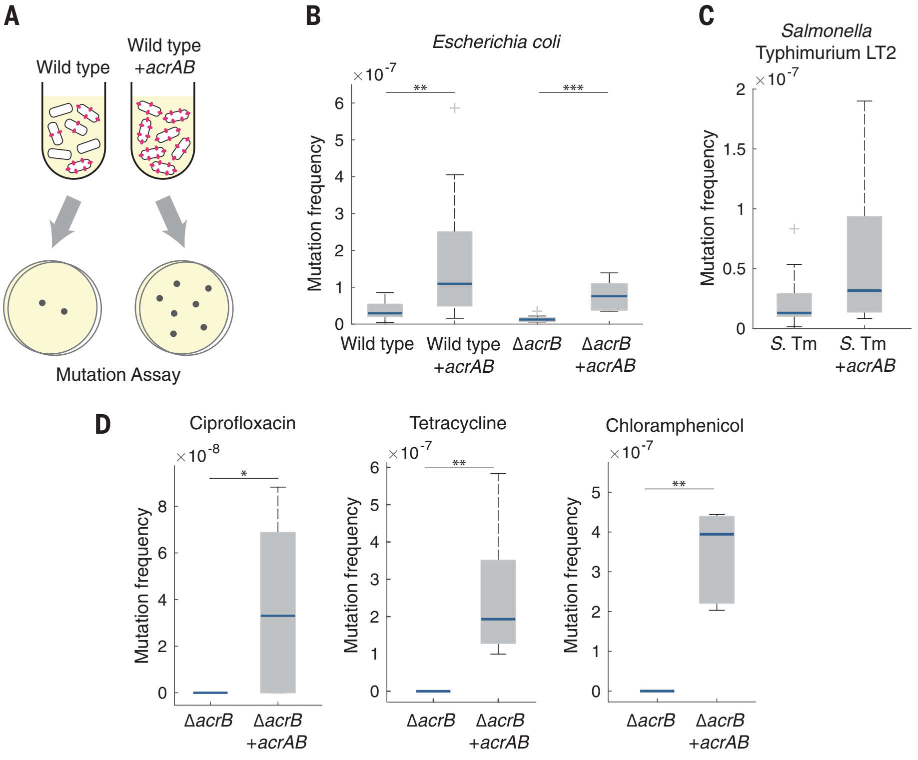


Fig. 1. Overexpression of AcrAB increases the spontaneous mutation frequency. (**A**) Schematic showing an increase in spontaneous mutations in cells with higher efflux pump expression. (**B**) Rifampicin mutation frequency in *E. coli* wild-type and Δ*acrB* strains with and without *acrAB* overexpression. $n \geq 8$ biological replicates. (**C**) Rifampicin mutation frequency in *S.* Typhimurium (*S.* Tm) LT2. $n \geq 12$ biological replicates. (**D**) Ciprofloxacin, tetracycline, and chloramphenicol mutation frequencies in *E. coli* Δ*acrB*. $n \geq 5$ biological replicates. The Δ*acrB* strain in (D) did not produce any mutants in the presence of any of the antibiotics; mutants were observed for all antibiotics in the *acrAB* overexpression strain. For (B) to (D), blue bars show the median values, gray boxes indicate the interquartile range, and whiskers show the maximum and minimum values. Box plot raw data are shown in fig. S9A. Strains without *acrAB* overexpression contained an equivalent plasmid expressing *cfp* in place of *acrAB*. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$, Mann-Whitney rank sum test.

frequency by dividing the number of colony-forming units (CFU) per milliliter on rifampicin plates by the number of CFU per milliliter on LB plates (Fig. 1A) (*13*).

We first compared wild-type *E. coli* strains with and without overexpression of the *acrAB* operon. Although AcrA and AcrB function together with TolC, overexpression of the *acrAB* operon alone is sufficient to increase resistance (*14*, *15*). We found that the wild-type strain had a significantly lower mutation frequency than the strain overexpressing *acrAB* (Fig. 1B). These mutation frequencies correspond to mutation rates of 1.06×10^{-8} and 2.84×10^{-8} mutations per generation in the wild-type and *acrAB* strains, respectively (table S1), providing resistance without incurring a major fitness cost (*16*). Deleting *acrB*, which inactivates the entire efflux pump (*14*), also significantly decreased the mutation frequency relative to that of the wild-type strain (Fig. 1B). Complementing the Δ*acrB* strain with a plasmid containing the *acrAB* operon restored and further increased the mutation frequency over that of the wild type. Using sequencing, we confirmed that the resistance originated from mutations within the *rpoB* gene, which is a known target for rifampicin (*17*) (table S2). Taken together, these results suggest that elevated expression of the AcrAB efflux pump, which plays a critical role in transient resistance, can also increase the frequency of spontaneous mutations.

We asked whether our findings could be generalized to other bacterial species by focusing on the pathogen *Salmonella enterica* serovar Typhimurium LT2 (*10*). We introduced a plasmid overexpressing *acrAB* into *S.* Typhimurium and again observed an increase in the spontaneous mutation frequency (Fig. 1C).

We also measured mutation frequencies in *E. coli* Δ*acrB* with and without *acrAB* overexpression by using ciprofloxacin, tetracycline, and chloramphenicol. In each case, we used antibiotic concentrations that exceeded the minimum inhibitory concentration for the strain with *acrAB* overexpression to ensure that we were measuring the mutation frequency and not differences due to drug efflux (*15*, *18*) (fig. S1). Consistent with our results in rifampicin, overexpression of *acrAB* significantly increased the mutation frequency in all three antibiotics relative to those of the strains lacking the pumps (Fig. 1D).

To test whether the differences in mutation frequency are due to pump activity, we generated a catalytically compromised mutant expressing *acrB* with a Phe^{610}→Ala (F610A) mutation (*19*). In contrast to bacteria with functional AcrB, the Δ*acrB* strain complemented with a plasmid containing *acrAB* F610A had no change in the mutation frequency relative to the Δ*acrB* strain (fig. S2), suggesting that pump activity is critical for the mutation rate differences.

MutS is involved in DNA mismatch repair, which is a crucial step in preventing mutations. MutS deficiency leads to a hypermutable phenotype (*20*). Recent studies have revealed heterogeneity in the expression of DNA repair enzymes between single cells, highlighting the importance of single-cell–level effects in the emergence of resistance (*21*–*23*). To study the link between *acrAB* and *mutS* expression, we constructed a double-color plasmid to report expression simultaneously from the *acrAB* and *mutS* promoters. We fused the *acrAB* promoter to the gene for red fluorescent protein (RFP) (yielding P_{acrAB}-*rfp*) and the *mutS* promoter to the gene for yellow fluorescent protein (YFP) (yielding P_{mutS}-*yfp*). Using fluorescence microscopy, we observed cell-to-cell variation in both reporters, with cells expressing either P_{acrAB}-*rfp* or P_{mutS}-*yfp* but not both (Fig. 2, A and B). This indicates that cells with higher efflux pump expression, which are more antibiotic resistant (*7*), also have less mismatch repair, making them more mutation prone (*21*).

To check whether this effect is due to AcrAB, we introduced the double-color reporter into the Δ*acrB* strain. This decreased the population of cells with high P_{acrAB} and low P_{mutS} expression (Fig. 2, A, C, and D) and potentially indicates positive feedback between AcrAB and its promoter. Complementing the Δ*acrB* strain with a plasmid containing *acrAB* restored cells with higher levels of P_{acrAB} and lower levels of P_{mutS} expression (fig. S3). To rule out spurious plasmid effects as the cause of the inverse relationship, we also tested double-color reporters in which we replaced the P_{acrAB} and P_{mutS} promoters each with a constitutive promoter and no longer observed the reciprocal relationship between the two colors (fig. S4).

Transcriptional fusions between promoters and reporters give an indirect measurement of protein levels; therefore, we next sought to verify our findings by using translational fusions of AcrAB and MutS with fluorescent reporters. We observed a similar relationship between AcrAB-RFP and MutS-YFP, with a subset of cells containing higher levels of AcrAB efflux pumps and lower MutS expression (Fig. 2, E and F).

When we overexpressed *acrAB* in a Δ*mutS* strain, we observed no significant difference in the spontaneous mutation frequency between the strains with and without *acrAB* overexpression (fig. S5). Because MutS is involved in mutation repair, the overall mutation rate is higher in the Δ*mutS* strain than in wild-type cells (*24*). The similar mutation frequencies observed with and without *acrAB* overexpression may be due to the role of MutS as an effector in the AcrAB-dependent mutation increase, or alternatively, the strong mutator phenotype may simply mask any differences.

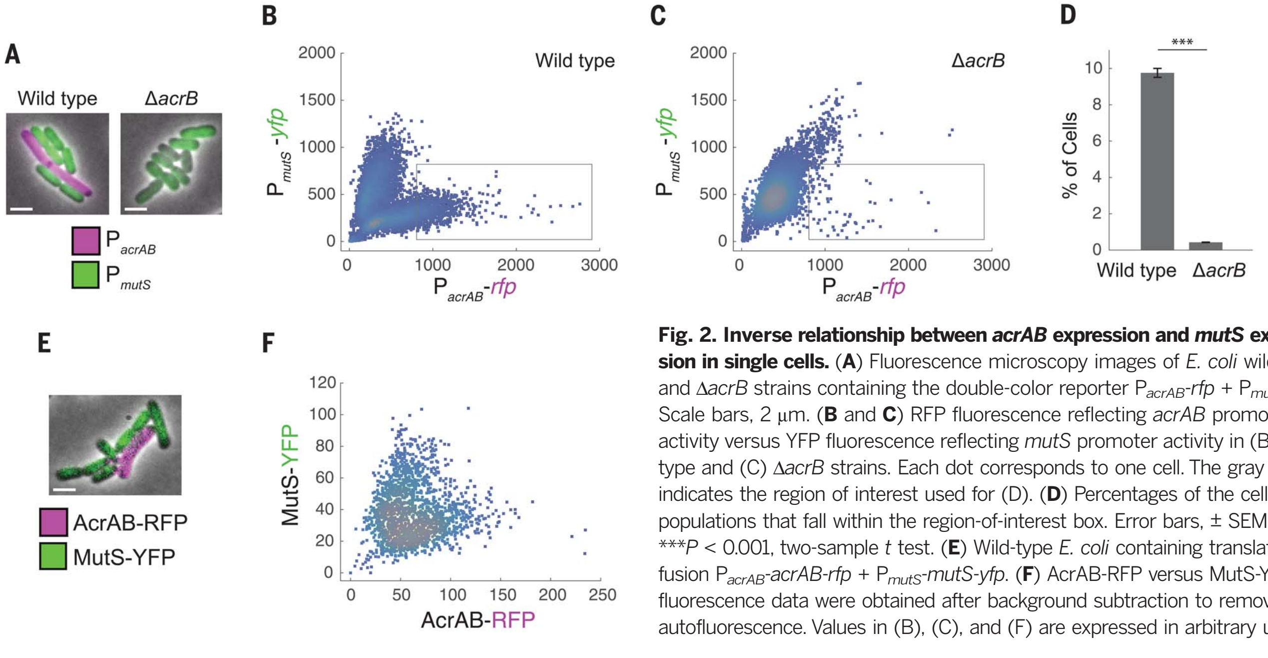


Fig. 2. Inverse relationship between *acrAB* expression and *mutS* expression in single cells. (**A**) Fluorescence microscopy images of *E. coli* wild-type and Δ*acrB* strains containing the double-color reporter P_{acrAB}-*rfp* + P_{mutS}-*yfp*. Scale bars, 2 μm. (**B** and **C**) RFP fluorescence reflecting *acrAB* promoter activity versus YFP fluorescence reflecting *mutS* promoter activity in (B) wild-type and (C) Δ*acrB* strains. Each dot corresponds to one cell. The gray box indicates the region of interest used for (D). (**D**) Percentages of the cell populations that fall within the region-of-interest box. Error bars, ± SEM. ***$P < 0.001$, two-sample *t* test. (**E**) Wild-type *E. coli* containing translational fusion P_{acrAB}-*acrAB-rfp* + P_{mutS}-*mutS-yfp*. (**F**) AcrAB-RFP versus MutS-YFP. All fluorescence data were obtained after background subtraction to remove autofluorescence. Values in (B), (C), and (F) are expressed in arbitrary units.

The AcrAB pump provides transient antibiotic resistance but can be costly to express. Overexpression alters membrane fluidity, slows growth, and can cause cells to pump out essential metabolites (*10*, *25*). Thus, there is a trade-off between pump expression and fitness (*25*). It is well known that mutation rates are dependent on growth rates in *E. coli*, and *mutS* expression is repressed in nutritionally stressed cells (*26*–*28*). We found that the total number of CFU per milliliter decreased when *acrAB* was overexpressed (fig. S6). Using time-lapse microscopy, we grew wild-type cells containing the double-color transcriptional reporter (P_{acrAB}-*rfp* and P_{mutS}-*yfp*) on agarose pads. Single cells with high P_{acrAB}

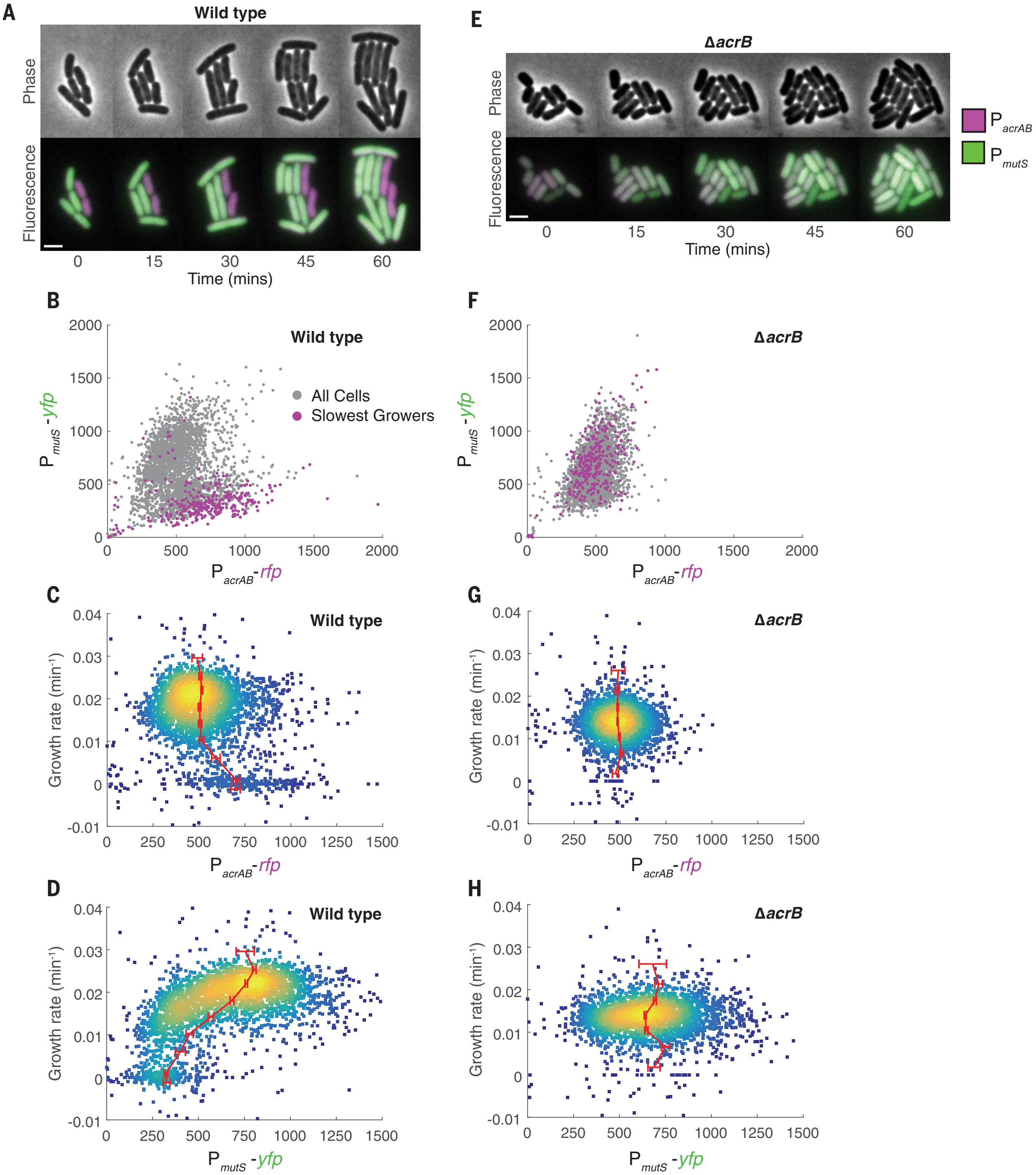


Fig. 3. Reduced growth rate in single cells with high P_{acrAB} and low P_{mutS} expression. (**A**) Time-lapse microscopy images of wild-type cells expressing P_{acrAB}-*rfp* + P_{mutS}-*yfp*. Scale bar, 2 μm. (**B**) P_{acrAB}-*rfp* expression versus P_{mutS}-*yfp* expression in the wild-type strain. The purple dots correspond to cells whose growth rate falls in the bottom 10% of those measured. (**C**) P_{acrAB}-*rfp* expression and (**D**) P_{mutS}-*yfp* expression versus the growth rate in the wild-type strain. (**E**) Δ*acrB* cells expressing P_{acrAB}-*rfp* and P_{mutS}-*yfp*. (**F**) P_{acrAB}-*rfp* expression versus P_{mutS}-*yfp* expression in Δ*acrB* cells. (**G**) P_{acrAB}-*rfp* expression and (**H**) P_{mutS}-*yfp* expression versus the growth rate in the Δ*acrB* strain. Red lines in (C), (D), (G), and (H) plot the mean fluorescence of cells binned across growth rate in increments of 0.004 min^{-1}, where each bin has a minimum of 15 cells. Error bars, ± SEM. Negative growth rates arise when the automated cell identification process identifies a cell in a subsequent frame as having a smaller number of pixels; however, this is an infrequent event (~2% of cells). Values in (B) to (D) and (F) to (H) are expressed in arbitrary units.

and low P_{mutS} expression grew more slowly than those with low P_{acrAB} and high P_{mutS} expression (Fig. 3A and movie S1). We quantified P_{acrAB} and P_{mutS} expression and growth rates across many growing microcolonies (n = 3213 cells) and again observed an inverse relationship between P_{acrAB} expression and P_{mutS} expression (Fig. 3B). Overlaying the growth rate onto these data, we found that the slowest-growing cells were those with high P_{acrAB} and low P_{mutS} expression (Fig. 3, B to D). Measurements with the AcrAB-RFP and MutS-YFP translational fusion strain also showed slower growth in this subpopulation of cells (fig. S7 and movie S2).

Growth rate–dependent effects disappeared in a Δ*acrB* background, with cells growing at similar rates across all levels of P_{acrAB} expression (Fig. 3E and movie S3). Quantification across microcolonies confirmed this finding, and growth rates were roughly constant, regardless of P_{acrAB} or P_{mutS} expression (Fig. 3, F to H). Together, these results demonstrate that *acrAB* expression affects single-cell growth rates and that cell-to-cell differences in pump expression result in a subpopulation of cells with high P_{acrAB} expression, low P_{mutS} expression, and a low growth rate.

The AcrAB efflux pump is regulated by the transcription factor MarA (*29*). Mutations in *marA* and its regulator *marR* frequently arise in clinical isolates and antibiotic resistance studies (*30*, *31*). In addition, MarA expression is heterogeneous and dynamic within isogenic single cells, and its stochastic expression is associated with elevated transient resistance (*5*, *32*). Using fluorescence-activated cell sorting, we found that cells with higher *marA* expression were more mutation prone than those with low *marA* expression and that this effect was due predominantly to the AcrAB pump (fig. S8 and supplementary text).

These results demonstrate a link between transient resistance and heterogeneity in spontaneous mutation frequencies. Our findings indicate that the AcrAB efflux pump, which plays a known role in multidrug resistance, can also affect the initial stages of the evolution of permanent antibiotic resistance. Our results suggest that heterogeneity in AcrAB is correlated with expression of the mismatch repair enzyme MutS in individual cells and that elevated levels of *acrAB* expression decrease the growth rate. In our work, this role for AcrAB was shown in the absence of antibiotic stress, so these differences in mutation frequency are not induced by antibiotic treatment.

Even modest increases in the mutation rate can drive the evolution of resistance under selective pressure. For instance, weak mutator phenotypes have been shown to play a critical role in the evolution of resistance to ciprofloxacin in *E. coli* and *Staphylococcus aureus* (*16*, *33*). Achieving resistance to clinical levels of antibiotics is often a multistep process, requiring several mutational events. As an example, *acrAB*-related genes, including the regulators *acrR*, *marR*, *soxR*, and *marA*, appeared frequently in a microbial evolution and growth arena (MEGA)–plate study in which *E. coli* evolved resistance to trimethoprim (*30*).

Our findings open the door for further studies of the molecular mechanism by which AcrAB affects mutation frequency. Mutation rates have been shown to depend on the cell growth rate and population density (*26*, *27*). Given the link between pump expression and growth, it is likely that other growth-related phenomena are influenced by single-cell–level differences in pump expression. Efflux pumps may contribute to increases in the mutation rate by influencing growth alone or by exporting compounds involved in cell-to-cell interactions and the methyl cycle (*13*). Multidrug efflux pumps and DNA repair enzymes are widespread (*20*, *34*, *35*). Understanding the initial evolutionary trajectory of resistant strains may suggest strategies for treating infections, such as combination therapies involving antibiotics and efflux pump inhibitors (*36*).

REFERENCES AND NOTES

1. M. N. Alekshun, S. B. Levy, *Cell* **128**, 1037–1050 (2007).
2. B. R. Levin, D. E. Rozen, *Nat. Rev. Microbiol.* **4**, 556–562 (2006).
3. L. R. Mulcahy, J. L. Burns, S. Lory, K. Lewis, *J. Bacteriol.* **192**, 6191–6199 (2010).
4. N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, S. Leibler, *Science* **305**, 1622–1625 (2004).
5. I. El Meouche, Y. Siu, M. J. Dunlop, *Sci. Rep.* **6**, 19538 (2016).
6. Y. Pu *et al.*, *Mol. Cell* **62**, 284–294 (2016).
7. T. Bergmiller *et al.*, *Science* **356**, 311–315 (2017).
8. I. Levin-Reisman *et al.*, *Science* **355**, 826–830 (2017).
9. A. A. Al Mamun *et al.*, *Science* **338**, 1344–1348 (2012).
10. X. Z. Li, P. Plésiat, H. Nikaido, *Clin. Microbiol. Rev.* **28**, 337–418 (2015).
11. E. C. Hobbs, X. Yin, B. J. Paul, J. L. Astarita, G. Storz, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16696–16701 (2012).
12. E. B. Tikhonova, H. I. Zgurskaya, *J. Biol. Chem.* **279**, 32116–32124 (2004).
13. R. Krašovec *et al.*, *Nat. Commun.* **5**, 3742 (2014).
14. A. M. Langevin, M. J. Dunlop, *J. Bacteriol.* **200**, e00525-17 (2017).
15. H. Nicoloff, V. Perreten, S. B. Levy, *Antimicrob. Agents Chemother.* **51**, 1293–1303 (2007).
16. H. Orlén, D. Hughes, *Antimicrob. Agents Chemother.* **50**, 3454–3456 (2006).
17. M. G. Reynolds, *Genetics* **156**, 1471–1481 (2000).
18. M. Oethinger, W. V. Kern, A. S. Jellen-Ritter, L. M. McMurry, S. B. Levy, *Antimicrob. Agents Chemother.* **44**, 10–13 (2000).
19. J. A. Bohnert *et al.*, *J. Bacteriol.* **190**, 8225–8229 (2008).
20. E. Denamur, I. Matic, *Mol. Microbiol.* **60**, 820–827 (2006).
21. S. Uphoff *et al.*, *Science* **351**, 1094–1097 (2016).
22. L. Robert *et al.*, *Science* **359**, 1283–1286 (2018).
23. S. Uphoff, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6516–E6525 (2018).
24. T. H. Wu, M. G. Marinus, *J. Bacteriol.* **176**, 5393–5400 (1994).
25. K. B. Wood, P. Cluzel, *BMC Syst. Biol.* **6**, 48 (2012).
26. R. Krašovec *et al.*, *PLOS Biol.* **15**, e2002731 (2017).
27. I. Nishimura, M. Kurokawa, L. Liu, B. W. Ying, *mBio* **8**, e00676-17 (2017).
28. G. Feng, H. C. Tsui, M. E. Winkler, *J. Bacteriol.* **178**, 2388–2396 (1996).
29. T. M. Barbosa, S. B. Levy, *J. Bacteriol.* **182**, 3467–3474 (2000).
30. M. Baym *et al.*, *Science* **353**, 1147–1151 (2016).
31. K. Maneewannakul, S. B. Levy, *Antimicrob. Agents Chemother.* **40**, 1695–1698 (1996).
32. N. A. Rossi, M. J. Dunlop, *PLOS Comput. Biol.* **13**, e1005310 (2017).
33. S. Wang, Y. Wang, J. Shen, Y. Wu, C. Wu, *FEMS Microbiol. Lett.* **341**, 13–17 (2013).
34. P. Hsieh, K. Yamane, *Mech. Ageing Dev.* **129**, 391–407 (2008).
35. M. M. Gottesman, I. H. Pastan, *J. Natl. Cancer Inst.* **107**, djv222 (2015).
36. K. N. Adams *et al.*, *Cell* **145**, 39–53 (2011).
37. I. El Meouche, M. J. Dunlop, Data for: Heterogeneity in efflux pump expression predisposes antibiotic-resistant cells to mutation, Dryad (2018); https://doi.org/10.5061/dryad.n1h9d0d.

ACKNOWLEDGMENTS

We thank R. Del-Rio at the University of Vermont and B. Tilton at the Boston University Flow Cytometry Facilities for assistance. F. Stirling and P. Silver kindly provided the *Salmonella* strain. We thank J. Collins, N. Emery, A. Gutierrez, S. Jain, A. Khalil, J.-B. Lugagne, N. Rossi, T. Wang, and W. Wong for helpful discussions and manuscript critiques. **Funding:** This work was supported by the National Institutes of Health grants R01AI102922 and R21AI137843. **Author contributions:** I.E.M. performed the experiments. I.E.M. and M.J.D. designed the research, analyzed the data, and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data have been archived at Dryad (*37*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/686/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 and S2
References (*38*–*52*)
Movies S1 to S3

20 December 2017; resubmitted 7 June 2018
Accepted 19 September 2018
10.1126/science.aar7981

GENETIC PRIVACY

Identity inference of genomic data using long-range familial searches

Yaniv Erlich[1,2,3,4]*, **Tal Shor**[1], **Itsik Pe'er**[2,3], **Shai Carmi**[5]

Consumer genomics databases have reached the scale of millions of individuals. Recently, law enforcement authorities have exploited some of these databases to identify suspects via distant familial relatives. Using genomic data of 1.28 million individuals tested with consumer genomics, we investigated the power of this technique. We project that about 60% of the searches for individuals of European descent will result in a third-cousin or closer match, which theoretically allows their identification using demographic identifiers. Moreover, the technique could implicate nearly any U.S. individual of European descent in the near future. We demonstrate that the technique can also identify research participants of a public sequencing project. On the basis of these results, we propose a potential mitigation strategy and policy implications for human subject research.

Consumer genomics has gained popularity (*1*). As of April 2018, more than 15 million people have undergone direct-to-consumer (DTC) autosomal genetic tests, with about 7 million kits sold in 2017 alone (*2*). Nearly all major DTC providers use dense genotyping arrays that probe ~700,000 genomic variants and let participants download their raw genotype files in a plain-text format. This has led to the advent of third-party services, such as DNA.Land and GEDmatch, which allow participants to upload their raw genotype files for further analysis (table S1) (*3*). Nearly all of these services offer the option to find genetic relatives by locating identity-by-descent (IBD) segments that can indicate a shared ancestor. Finding genetic relatives can accurately link even distant relatives, such as second or third cousins (*4*–*6*) (fig. S1), and has led to multiple "success stories" within the genetic genealogy community, such as reunions of adoptees with their biological families (*7*).

In the past few months, law enforcement agencies have started exploiting third-party consumer genomics services to trace suspects by finding their distant genetic relatives. This route to identify individuals, dubbed long-range familial search, has been predicted before (*8*). It offers a powerful alternative to familial searches in forensic databases, which can only identify close (first to second degree) relatives (*9*, *10*), and is highly regulated (*11*). In one notable case, law enforcement used a long-range familial search to trace the Golden State Killer (*12*, *13*). Investigators generated a genome-wide profile of the perpetrator from a crime scene sample and uploaded the profile to GEDmatch, a database that contains ~1 million DNA profiles. The GEDmatch search identified a third-degree cousin (*12*). Extensive genealogical data traced the identity of the perpetrator, which was confirmed by a standard DNA test. Between April and August 2018, at least 13 cases were reportedly solved by long-range familial searches (Table 1 and table S2). Most of these investigations focused on cold cases, for which decades of investigation failed to identify the offender. Nonetheless, one case involved a crime from April 2018, suggesting that some law enforcement agencies have incorporated long-range familial DNA searches into active investigations. Parabon NanoLabs, a forensic DNA company, has announced that it set up a division that will use long-range familial searches and has already uploaded 100 cold cases to third-party DTC services (*14*). All of these lines of evidence suggest that long-range familial searches may become a standard investigative tool.

We took an empirical approach to investigate the probability that a long-range familial search will identify an individual. To this end, we analyzed a dataset of 1.28 million individuals who were tested with a DTC provider (*15*). We retained relatives with at least two IBD segments of >6 centimorgans (cM) each to increase the chance of correctly inferring genealogical relationships. Next, we removed pairs with IBD segments greater than 700 cM (i.e., first cousin and closer relationships) to circumvent ascertainment biases owing to the tendency of close relatives to undergo genetic testing together. Finally, considering each individual in turn as our "target," we counted the number of individuals with a total IBD sharing of between 30 and 600 cM with the target (*15*). The low end of our range corresponds to approximately fourth cousins and the high end to second cousins, on the basis of a crowdsourcing project (*16*).

Our results show that nearly 60% of long-range familial searches return a relative with IBD segments with a total length of 100 cM or more (Fig. 1A). This level of IBD sharing usually corresponds to a third cousin or closer relative, similar to the case of the Golden State Killer. Interestingly, these success rates are higher than with surname inference from the Y chromosome, which is another genetic reidentification tactic (*17*). In 15% of the searches with our data, the top match had IBD segments of a total length of at least 300 cM, which corresponds to a second cousin or closer relative.

We validated our results by performing 30 random long-range familial searches in GEDmatch. The results were similar: The top match in GEDmatch shared >100 cM in 76% of the cases [confidence interval (CI) of 59 to 88%] and >300 cM in 10% of the searches (CI of 3 to 25%), similar to the results with our 1.28 million individuals (Fig. 1A).

Long-range familial searches create racial disparity that is the opposite of disparities documented in traditional forensic databases (*11*). About 75% of the 1.28 million individuals were primarily of Northern European genetic background (fig. S2 and table S3), similar to previous reports of DTC genomics data (*18*). Individuals of primarily Northern European background were 30% more likely to have a >100-cM match than individuals whose genetic background was primarily from sub-Saharan Africa (fig. S3).

More broadly, a genetic database needs to cover only 2% of the target population to provide a third-cousin match to nearly any person (Fig. 1B). This assertion relies on a population genetics model that takes into account the probability of sharing at least two IBD segments of length >6 cM and assumes that the population grows at similar rates to the observed growth rates in the Western world during the past 200 years. (*15*) (fig. S4). This model has multiple simplifying assumptions, such as no population structure, no inbreeding, and random sampling of participants, and thus should be interpreted only as a rough guideline. Nevertheless, the model showed consistency between our empirical results and the IBD sharing profile of Northern Europeans in the United States (fig. S5). Using this model, we predict that with a database size of ~3 million U.S. individuals of European descent (2% of the adults of this population), more than 99% of the people of this ethnicity would have at least a single third-cousin match and more than 65% are expected to have at least one second-cousin match. With the exponential growth of consumer genomics (*1*), we posit that such a database scale is foreseeable for some third-party websites in the near future.

Next, we examined the theoretical ability to find the person of interest after finding a relative in a long-range familial search. We focused on reducing the search space using basic demographic information, such as geography, age, and sex. Using genealogical records of population-scale family trees (*19*), we computed the number of relatives of a third-cousin match after filtering them on the basis of place of residence, age, and sex. A study of serial criminals indicates that the place of crime is nearly always within

---

[1]MyHeritage, Or Yehuda 6037606, Israel. [2]Department of Computer Science, Fu Foundation School of Engineering, Columbia University, New York, NY, USA. [3]Center for Computational Biology and Bioinformatics (C2B2), Department of Systems Biology, Columbia University, New York, NY, USA. [4]New York Genome Center, New York, NY, USA. [5]Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel.
*Corresponding author. Email: erlichya@gmail.com

40 km (25 miles) of the criminal's place of residence (*20*). To be conservative, we thus assumed that the location of the target can be estimated within a radius of 160 km (100 miles). We also assumed that the age of the target can be estimated within a ±5-year interval based on eyewitnesses or camera footage, as previously estimated (*21*). Finally, we assumed that the biological sex is known from the DNA sample.

We found that the suspect list can be pruned from basic demographic information. On the basis of counting relevant relatives of the match, the initial list of candidates contains an average of ~850 individuals (Fig. 2A). Our simulations indicate that localizing the target to within 160 km (100 miles) will exclude 57% of the candidates on average (Fig. 2B and table S4). Next, availability of the target's age to within ±5 years will exclude 91% of the remaining candidates (Fig. 2C). Finally, inference of the biological sex of the target will halve the list to just around 16 to 17 individuals, a search space that is small enough for manual inspection. We also considered a scenario of reidentification of anonymized clinical genetic data. The safe-harbor provisions of the Health Insurance Portability and Accountability Act of 1996 (HIPAA) privacy law permit the release of the year of birth. An age specified at a single-year resolution is, as expected, a more powerful identifier compared to a 10-year interval (Fig. 2D). Together with geography (<160 km (100 miles)) and sex, it is expected to reduce the search space to just one to two individuals (Fig. 2E). To conclude, the main barrier is not finding a match or theoretical power to prune the search space. Rather, successfully tracing an individual depends mainly on the accessibility of genealogical data for the matched relative, their accuracy, and the time invested in organizing the genealogical data.

To better understand the risk of reidentifying human subjects, we conducted a long-range familial search on a specific 1000 Genomes Project individual. We selected a female from the CEU (Utah residents with Northern and Western European ancestry) cohort, whose husband has been identified using surname inference (*17*). We extracted her genome from the (publicly available) 1000 Genomes data repository, reformatted her genotype to resemble a file released by DTC

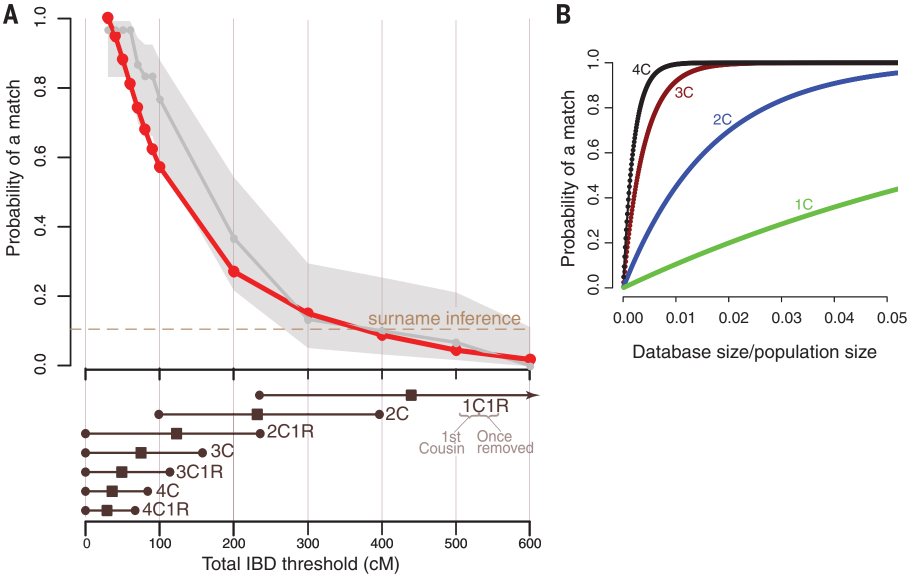


Fig. 1. The performance of long-range familial searches for various database sizes. (**A**) The probability of finding at least one relative for various IBD thresholds (top) with 1.28 million searches of DTC-tested individuals (red) and 30 random GEDmatch searches (gray). Light gray shading indicates the 95% CI for the GEDmatch estimates. The dashed line indicates the probability of a surname inference from Y chromosome data (*17*). The bottom panel shows the 95% CIs (circles) and average total IBD length (squares) for a first cousin once removed (1C1R) to a fourth cousin once removed (4C1R) (*20*). (**B**) A population-genetic theoretical model for the probability of finding relatives up to a certain type of cousinship as a function of the database coverage of the population. 1C to 4C indicate first to fourth cousins.

Table 1. Public cases of long-range familial cases. A "–" indicates data not available.

Case	Announcement	Solved by	Closest match	Comments
Buckskin Girl	9 April 2018	DNA Doe Project	First cousin once removed	
Golden State Killer	24 April 2018	Barbara Rae-Venter	Third cousin	
Lyle Stevik	8 May 2018	DNA Doe Project	Second cousin	Inbreeding complicated the estimation of the match.
William Earl Talbott II	21 May 2018	Parabon	Half–first cousin once removed	Second cousins were identified as well.
Joseph Newton Chandler III	21 June 2018	DNA Doe Project	Second cousin once removed	
Gary Hartman	22 June 2018	Parabon	Half–first cousin	Genealogists were able to overcome a nonpaternity event in the family tree of the suspect.
Raymond "DJ Freez" Rowe	25 June 2018	Parabon	–	
James Otto Earhart	26 June 2018	Parabon	Second cousin	
John D. Miller	15 July 2018	Parabon	–	
Matthew Dusseault and Tyler Grenon	28 July 2018	Parabon	–	
Spencer Glen Monnett	29 July 2018	Parabon	–	This was an active case for a crime that occurred in April 2018.
Darold Wayne Bowden	23 August 2018	Parabon	–	
Michael F. Henslick	29 August 2018	Parabon	–	

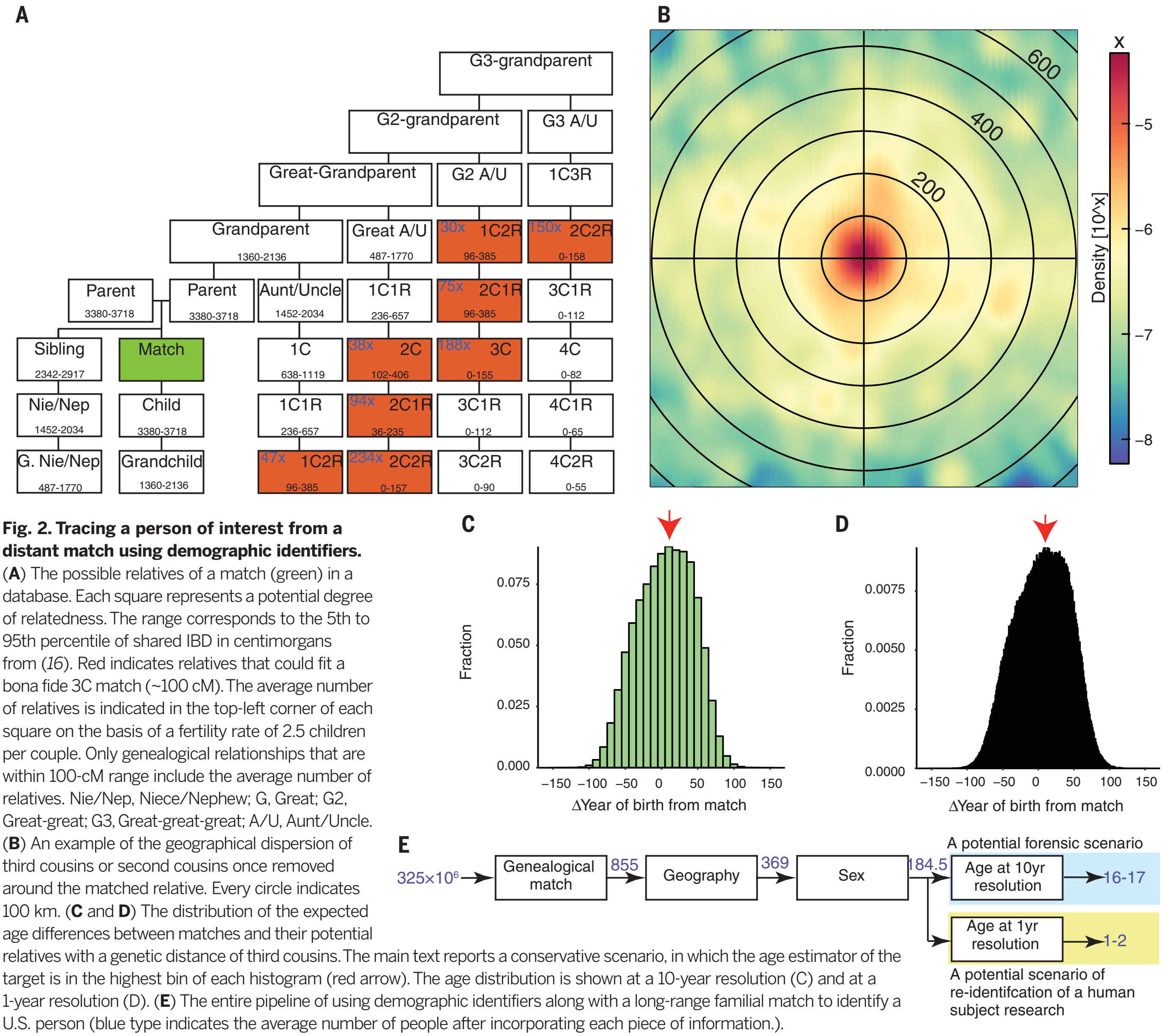


Fig. 2. Tracing a person of interest from a distant match using demographic identifiers. (**A**) The possible relatives of a match (green) in a database. Each square represents a potential degree of relatedness. The range corresponds to the 5th to 95th percentile of shared IBD in centimorgans from (*16*). Red indicates relatives that could fit a bona fide 3C match (~100 cM). The average number of relatives is indicated in the top-left corner of each square on the basis of a fertility rate of 2.5 children per couple. Only genealogical relationships that are within 100-cM range include the average number of relatives. Nie/Nep, Niece/Nephew; G, Great; G2, Great-great; G3, Great-great-great; A/U, Aunt/Uncle. (**B**) An example of the geographical dispersion of third cousins or second cousins once removed around the matched relative. Every circle indicates 100 km. (**C** and **D**) The distribution of the expected age differences between matches and their potential relatives with a genetic distance of third cousins. The main text reports a conservative scenario, in which the age estimator of the target is in the highest bin of each histogram (red arrow). The age distribution is shown at a 10-year resolution (C) and at a 1-year resolution (D). (**E**) The entire pipeline of using demographic identifiers along with a long-range familial match to identify a U.S. person (blue type indicates the average number of people after incorporating each piece of information.).

providers, and uploaded the genotype to GEDmatch. Searching GEDmatch returned two relatives, one from North Dakota and one from Wyoming, with sufficient genetic and genealogical details (Fig. 3). Both relatives shared about 170 to 180 cM with the 1000Genomes sample, which corresponds to six to seven degrees of separation. They also shared 62 cM between each other, indicating that they were distantly related via an ancestral couple who lived four to six generations ago. In about 1 hour of work, we identified the ancestral couple from publicly available genealogical records. Next, we searched for descendants of the ancestral couple that matched the publicly available demographic data of the 1000Genomes sample, such as her expected year of birth and pedigree structure. This step, performed manually, was time consuming and not trivial, because the ancestral couple had more than 10 children and hundreds of descendants. After a full day of work, we eventually excluded all other candidates and traced the identity of our target, which was the same person we had previously reidentified based on surname inference of her husband.

Taken together, we posit that our results warrant a reevaluation of the status quo regarding the identifiability of DNA data, especially of U.S. individuals. Although policy-makers and the general public may be in favor of such enhanced forensic capabilities for solving crimes, it relies on databases and services that are open to everyone. Thus, the same technique could also be exploited for harmful purposes, such as reidentification of research subjects from their genetic data. The Revised Common Rule, which will regulate federally funded human subject research starting in January 2019, does not define genome-wide genetic datasets as identifiable information (*22*). However, the rule permits the U.S. Department of Health and Human Services (HHS) to revise the scope of identifiable private information on the basis of technological developments. In light of our results, we encourage HHS to consider genome-wide information as identifiable.

Finally, we propose a measure to mitigate some of the risks and restore control to data custodians. In our proposal, DTC providers should cryptographically sign the text file containing the raw data available to customers (fig. S6). Third-party services will be able to authenticate that a raw genotyping file was created by a valid DTC provider and not further modified. If adopted, our approach has the potential to prevent the exploitation of long-range familial searches to identify research subjects from genomic data.

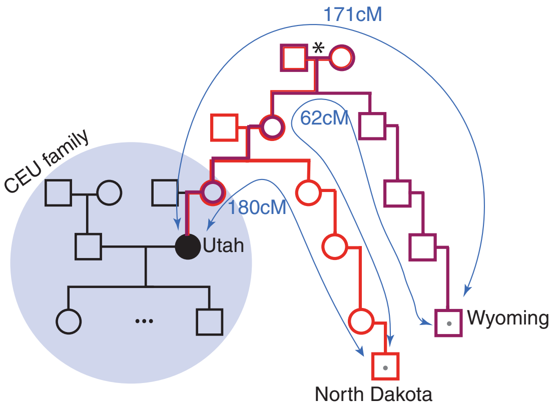


Fig. 3. Tracing a 1000Genomes sample using a long-range familial search. The CEU pedigree is shown in black. To respect the privacy of the family, we omitted the sample identifiers and the exact pedigree structure. A GEDmatch search of the person of interest (black circle) returned two males (squares with gray dots) with a total IBD sharing of 180 and 171 cM to the target, respectively, and 62 cM between themselves. Using public genealogical records, we identified the ancestral couple (asterisk) of the matches and the person of interest.

Moreover, it will complicate the ability to conduct unilaterally long-range familial searches from DNA evidence (*15*). As such, it can complement previous proposals regarding the regulation of long-range familial searches by law enforcement (*23*) and offers better protection in cases in which the law cannot deter misuse. To facilitate consideration of our approach by the community, we provide a demo source code on GitHub that can sign and verify the raw genotype files using a previously published digital signature scheme (*24*). Overall, we believe that technical measures, clear policies for law enforcement in using long-range familial searches, and respecting the autonomy of participants in genetic studies are necessary components for long-term sustainability of the genomics ecosystem.

REFERENCES AND NOTES

1. R. Khan, D. Mittelman, *Genome Biol.* **19**, 120 (2018).
2. L. Larkin, Autosomal DNA testing comparison chart, The DNA Geek; http://thednageek.com/dna-tests/.
3. S. C. Nelson, S. M. Fullerton, *J. Genet. Couns.* **27**, 770–781 (2018).
4. A. Gusev *et al.*, *Genome Res.* **19**, 318–326 (2009).
5. C. D. Huff *et al.*, *Genome Res.* **21**, 768–774 (2011).
6. B. M. Henn *et al.*, *PLOS ONE* **7**, e34267 (2012).
7. International Society of Genetic Genealogy Wiki, Success stories (2018); https://isogg.org/wiki/Success_stories.
8. Y. Erlich, A. Narayanan, *Nat. Rev. Genet.* **15**, 409–421 (2014).
9. J. Ge, R. Chakraborty, A. Eisenberg, B. Budowle, *J. Forensic Sci.* **56**, 1448–1456 (2011).
10. N. A. Garrison, R. V. Rohlfs, S. M. Fullerton, *Nat. Rev. Genet.* **14**, 445 (2013).
11. J. Kim, D. Mammo, M. B. Siegel, S. H. Katsanis, *Investig. Genet.* **2**, 22 (2011).
12. M. Gafni, “Here’s the ‘open-source’ genealogy DNA website that helped crack the Golden State Killer case,” *Mercury News*, 26 April 2018; www.mercurynews.com/2018/04/26/ancestry-23andme-deny-assisting-law-enforcement-in-east-area-rapist-case/.
13. J. Jouvenal, “To find alleged Golden State Killer, investigators first found his great-great-great-grandparents,” *Washington Post*, 30 April 2018; www.washingtonpost.com/local/public-safety/to-find-alleged-golden-state-killer-investigators-first-found-his-great-great-great-grandparents/2018/04/30/3c865fe7-dfcc-4a0e-b6b2-0bec548d501f_story.html?utm_term=.6ff5cff1630e.
14. P. Aldhous, “DNA data from 100 crime scenes has been uploaded to a genealogy website—just like the Golden State Killer,” *BuzzFeed*, 17 May 2018; www.buzzfeed.com/peteraldhous/parabon-genetic-genealogy-cold-cases?utm_term=.tkKXDVOWq#.yyz8oGQWd.
15. See supplementary materials.
16. B. T. Bettinger, The Shared cM Project – Version 3.0 (2017); https://thegeneticgenealogist.com/wp-content/uploads/2017/08/Shared_cM_Project_2017.pdf.
17. M. Gymrek, A. L. McGuire, D. Golan, E. Halperin, Y. Erlich, *Science* **339**, 321–324 (2013).
18. J. Yuan *et al.*, *Nat. Genet.* **50**, 160–165 (2018).
19. J. Kaplanis *et al.*, *Science* **360**, 171–175 (2018).
20. J. Warren *et al.*, *J. Quant. Criminol.* **14**, 35–59 (1998).
21. H. Han, C. Otto, A. K. Jain, “Age estimation from face images: Human vs. machine performance,” paper presented at the 6th International Association for Pattern Recognition (IAPR) International Conference on Biometrics (ICB), Madrid, Spain, 4 to 7 June 2013.
22. Department of Health and Human Services, *Fed. Regist.* **82**, 7149–7274 (2017).
23. N. Ram, C. J. Guerrini, A. L. McGuire, *Science* **360**, 1078–1079 (2018).
24. D. J. Bernstein, N. Duif, T. Lange, P. Schwabe, B.-Y. Yang, *J. Cryptogr. Eng.* **2**, 77–89 (2012).

ACKNOWLEDGMENTS

We thank G. Japhet and A. Gordon for their contributions to the cryptographic signature scheme and Y. Naveh, Y. Ben-David, C. Moore, and the DNA Doe Project for valuable comments. **Funding:** Y.E. holds a Burroughs Wellcome Fund Career Award at the Scientific Interface. S.C. thanks the Israel Science Foundation, grant no. 407/17. **Author contributions:** Y.E. conceived the idea for this study. Y.E. and T.S conducted the analysis of matches using the MyHeritage and Geni.com data. S.C and I.P. developed the theoretical framework to estimate the number of matches. Y.E. and S.C conducted the trace back of the 1000Genomes sample. Y.E., T.S., I.P., and S.C. wrote the manuscript. **Competing interests:** Y.E. and T.S. adapted the code for the cryptographic signatures. Y.E. and T.S are MyHeritage employees. Y.E. is also a consultant of ArcBio. I.P. holds equity in 23andMe. S.C. is a paid consultant of MyHeritage. When multiple companies are mentioned in this manuscript, we listed them in a lexicographic order. **Data and materials availability:** The code for the cryptographic signatures is available on https://github.com/erlichya/signature with an MIT license. The millions of genealogical records for the demographic analysis data are available on http://familinx.org/ under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license. The code for the population genetics simulation and 1000Genomes extraction is available in the supplementary materials under an MIT license. Following the MyHeritage terms, we cannot share the individual-level genomic data. We will share the anonymized IBD network topology on request and subject to the MyHeritage Terms and Conditions and Privacy Policy under the following terms: (i) researchers will need an IRB approval for their study, (ii) the data can only be processed in a MyHeritage facility and cannot be used to reidentify individuals, (iii) the results can only be used for noncommercial purposes, and (iv) MyHeritage does not ask authorship in new publications that use the anonymized IBD network. Researchers who are interested in the data or in pursuing research collaboration opportunities can contact dnaresearch@myheritage.com.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/690/suppl/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S4
References (*25–44*)

14 June 2018; accepted 1 October 2018
Published online 11 October 2018
10.1126/science.aau4832

IMMUNOLOGY

WDFY4 is required for cross-presentation in response to viral and tumor antigens

Derek J. Theisen[1]*, Jesse T. Davidson IV[1,2]*, Carlos G. Briseño[1], Marco Gargaro[3], Elvin J. Lauron[4], Qiuling Wang[5], Pritesh Desai[1,5], Vivek Durai[1], Prachi Bagadia[1], Joshua R. Brickner[1], Wandy L. Beatty[5], Herbert W. Virgin[1,6], William E. Gillanders[2,7], Nima Mosammaparast[1], Michael S. Diamond[1,5,8,9], L. David Sibley[5], Wayne Yokoyama[4], Robert D. Schreiber[1,9], Theresa L. Murphy[1], Kenneth M. Murphy[1,10]†

During the process of cross-presentation, viral or tumor-derived antigens are presented to $CD8^+$ T cells by *Batf3*-dependent $CD8\alpha^+/XCR1^+$ classical dendritic cells (cDC1s). We designed a functional CRISPR screen for previously unknown regulators of cross-presentation, and identified the BEACH domain–containing protein WDFY4 as essential for cross-presentation of cell-associated antigens by cDC1s in mice. However, WDFY4 was not required for major histocompatibility complex class II presentation, nor for cross-presentation by monocyte-derived dendritic cells. In contrast to *Batf3*$^{-/-}$ mice, *Wdfy4*$^{-/-}$ mice displayed normal lymphoid and nonlymphoid cDC1 populations that produce interleukin-12 and protect against *Toxoplasma gondii* infection. However, similar to *Batf3*$^{-/-}$ mice, *Wdfy4*$^{-/-}$ mice failed to prime virus-specific $CD8^+$ T cells in vivo or induce tumor rejection, revealing a critical role for cross-presentation in antiviral and antitumor immunity.

Presentation of antigens as peptides bound to proteins of the major histocompatibility complex (MHC) is the principal mechanism by which innate cells promote antigen-specific T cell immunity (*1*). Classical dendritic cells (cDCs) are particularly efficient antigen-presenting cells and comprise two major functionally distinct subsets, cDC1 and cDC2 (*2*–*4*). The cDC1 lineage (*2*, *5*) is the most efficient at priming cytotoxic $CD8^+$ T cells to exogenously derived antigens, a process termed cross-presentation (*6*–*10*). This specialization was observed in *Batf3*$^{-/-}$ mice that specifically lack cDC1 development and cannot mount cytotoxic T cell responses to viruses and tumors (*10*–*24*). However, because these studies have only analyzed these responses in the context of mice lacking cDC1s, the role of cross-presentation versus other cDC1-specific effector functions, such as interleukin-12 (IL-12)–mediated protection against *Toxoplasma gondii* (*25*), has remained incompletely understood.

Cross-presentation has been studied using different cell types and various forms of antigen, and not all findings have been confirmed in vivo (*26*). DCs generated from monocytes (moDCs) or whole bone marrow cultured in vitro with granulocyte/macrophage colony-stimulating factor with or without IL-4 (*27*–*29*) are heterogeneous, resembling both macrophages and DCs (*30*), and use a cross-presentation program divergent from that of cDC1s in vivo (*26*, *31*, *32*). Studies of moDCs have produced two major models of cross-presentation: one that involves transport of exogenous antigen to the cytosolic proteasome before peptide loading in the endoplasmic reticulum (ER) (*1*, *7*, *33*–*35*), and another where peptide loading occurs directly in phagosomes by fusion with vesicles containing the peptide-loading complex (*36*, *37*). The latter pathway may be regulated by the SNARE family member Sec22b, although two recent studies of *Sec22b*-deficient mice arrived at different conclusions as to the role of Sec22b in T cell priming to cell-associated antigens in vivo (*38*, *39*). These differences highlight the need for systematic investigation into the mechanisms of cross-presentation in vivo (*39*, *40*).

We established a screen for previously unknown cellular components required for cross-presentation, and optimized in vitro conditions to replicate this process in cDC1s. The efficiency and cell type specificity of cross-presentation can vary, depending on whether the antigen is soluble or associated with cells or pathogens (*32*). Bacterial-associated antigen in the form of heat-killed *Listeria monocytogenes* expressing ovalbumin (HKLM-OVA) was efficiently cross-presented by cDC1s to OT-I T cells but was not presented by cDC2s (Fig. 1A). In contrast, soluble ovalbumin (OVA) was cross-presented by both cDC1 and cDC2 lineages, with lower efficiency in cDC2s by a factor of 3 to 10 (Fig. 1B). Presentation of SIINFEKL (Ser-Ile-Ile-Asn-Phe-Glu-Lys-Leu) peptide to OT-I cells was equally efficient in cDC1s and cDC2s, as expected (Fig. 1C). Previous studies have suggested that the majority of antigens undergo translocation to the cytosol during cross-presentation in vivo (*1*, *7*, *35*). We found that cell-associated antigens, which are presented only by cDC1s and not cDC2s, are *Tap1*-dependent, suggesting presentation through the cytosolic pathway (Fig. 1D). In contrast, soluble antigens were presented by both *Tap1*$^{-/-}$ cDC1s and cDC2s, with only slight differences in efficiency relative to wild-type cDCs (Fig. 1E). For these reasons, we concluded that the use of cell-associated antigens in a screen would best emphasize cDC1-specific processing functions.

Screening could be done using either biochemical detection of peptide:MHC complexes (p:MHCs) or a T cell response. The antibody 25-D1.16 can directly measure SIINFEKL:K^b complexes on the cell surface (*41*); 25-D1.16 detected a robust signal from soluble OVA processed by cDC1s (Fig. 1F), but no signal was detected using an immunogenic dose of cell-associated antigen (Fig. 1G). T cells can respond to only a few hundred p:MHCs (*42*, *43*), implying that the detection limit for 25-D1.16 is greater than that for T cells. Thus, we decided to use T cell proliferation as the readout and determined that 10^4 cDCs can produce a reliable and specific signal of OT-I proliferation (Fig. 1H). We considered gene candidates on the basis of expression in cDCs, relative cDC1 specificity, and gene ontology (table S1). We expressed single guide RNAs (sgRNAs) (*44*) for candidates (table S2) by retrovirus under the U6 promoter and infected DC progenitors from Cas9 transgenic mice (*45*) (fig. S1A). Cells were cultured in Flt3L for 7 days, sorted to purify infected cDCs, and tested for cross-presentation (Fig. 1I and fig. S1, B and C).

Cross-presentation was substantially impaired by two independent sgRNAs for *Wdfy4* (WD repeat- and FYVE domain–containing protein 4), a member of the BEACH (Beige and Chediak-Higashi) domain–containing family of proteins (*46*) (Fig. 2A and fig. S1C). *Wdfy4* is highly expressed in mouse and human cDC1s (fig. S2), with 80% species similarity (*47*). WDFY4 is one of nine BEACH domain–containing proteins (BDCPs) (*46*) and has three closely related family members. However, CRISPR targeting using sgRNAs for *Wdfy1*, *Wdfy2*, and *Wdfy3* did not impair cross-presentation, in contrast to *Wdfy4* (Fig. 2B). Thus, *Wdfy4* appears to be unique within this gene family for supporting cross-presentation by cDC1s.

---

[1]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA. [2]Department of Surgery, Washington University School of Medicine, St. Louis, MO 63110, USA. [3]Department of Experimental Medicine, University of Perugia, Perugia, Italy. [4]Division of Rheumatology, Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA. [5]Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110, USA. [6]Vir Biotechnology, San Francisco, CA, USA. [7]Alvin J. Siteman Cancer Center at Barnes-Jewish Hospital and Washington University School of Medicine, St. Louis, MO 63110, USA. [8]Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA. [9]Andrew M. and Jane M. Bursky Center for Human Immunology and Immunotherapy Programs, Washington University School of Medicine, St. Louis, MO 63110, USA. [10]Howard Hughes Medical Institute, Washington University School of Medicine, St. Louis, MO 63110, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: kmurphy@wustl.edu

To evaluate the in vivo function of *Wdfy4*, we obtained mice with exon 4 deleted by CRISPR/Cas9 genome editing, leading to translational termination due to a reading frame shift when exon 3 splices to exon 5 (fig. S3). *Wdfy4*$^{-/-}$ mice were viable, born in normal Mendelian ratios, and displayed normal development of hematopoietic lineages, including cDCs (Fig. 2, C and D, and fig. S4), which expressed *Irf8* and had normal turnover kinetics (fig. S4, H and I), and T cells (fig. S5). In particular, cDC1s developed in *Wdfy4*$^{-/-}$ mice, unlike *Batf3*$^{-/-}$ mice, and expressed CD24, XCR1, and CD103 normally (Fig. 2, C and D, and fig. S4, B and F). However, cDC1s from *Wdfy4*$^{-/-}$ mice showed a striking defect in cross-presentation of both cell-associated and bacterial-associated antigen in vitro (Fig. 2, E and F, and fig. S6A) and showed reduced efficiency for soluble OVA presentation relative to wild-type cDC1s (Fig. 2G). Notably, *Wdfy4*$^{-/-}$ cDC1s cross-presented soluble OVA with the efficiency of cDC2s, which were not influenced by the loss of *Wdfy4* (Fig. 2G). However, *Wdfy4*$^{-/-}$ cDC1s could directly present antigen introduced into the cytoplasm by osmotic shock or virus, a process that is equally efficient in cDC1s and cDC2s (fig. S6, B to D); this finding suggests that *Wdfy4*$^{-/-}$ cDC1s have the capacity to present endogenous antigens on MHC class I complexes.

moDCs can cross-present both soluble and cell-associated antigens in vitro (*27*, *48*, *49*), but their transcriptional program is distinct from that of cDC1 (*31*). We found that moDCs derived from wild-type and *Wdfy4*$^{-/-}$ mice cross-present antigens with the same efficiency, both for cell-associated (fig. S6E) and soluble OVA (fig. S6F), which suggests that moDCs use a *Wdfy4*-independent pathway for cross-presentation. The defect in cross-presentation by *Wdfy4*$^{-/-}$ cDC1s is specific, because MHC class II antigen processing was unchanged in *Wdfy4*$^{-/-}$ mice for both cell-associated and soluble antigens (Fig. 2H and fig. S6G). MHC class II antigen processing by B cells was also normal in *Wdfy4*$^{-/-}$ mice (fig. S7A), which were able to generate germinal center B cells and T follicular helper (T_{FH}) cells in response to immunization with sheep red blood cells (fig. S7, B to E).

cDCs from *Wdfy4*$^{-/-}$ mice expressed normal levels of MHC class I at steady state and after activation (fig. S8, A and B), up-regulated the costimulatory molecules CD80/86, and expressed cytokines normally (fig. S8, C to F). Loss of *Wdfy4* also did not influence gene expression in cDC1s at steady state or after activation in tumor-bearing mice (fig. S8, G and H). Despite their inability to cross-present, *Wdfy4*$^{-/-}$ cDC1s were capable of taking up and degrading soluble antigens normally (fig. S9, A and B) and phagocytosing labeled HKLM-OVA, as seen both microscopically (fig. S9C) and by quantification of this phagocytosis as measured by fluorescence-activated cell sorting (FACS) (fig. S9D).

To explore the mechanism of action of WDFY4, we analyzed various cellular compartments of wild-type and *Wdfy4*$^{-/-}$ cDC1s by confocal microscopy and found minimal differences in distribution of MHC class I stores, ER, early endosomes, lysosomes, late endosomes, or the peptide-loading complex at steady state (fig. S10). In addition, we observed minimal differences in distribution of Rab43 [a molecule previously described to be involved in cross-presentation (*32*)], p62 (autophagic vesicles), Rab7 (late endosomes), or Lamp1 (lysosomes) after antigen phagocytosis (fig. S11). Electron microscopy of WDFY4-deficient ex vivo cDC1s showed the presence of large and numerous lipid bodies throughout the cytoplasm that were not present in wild-type cells (figs. S12 and S13, A and C). However, these lipid bodies were not present in Flt3L-derived cDC1s from *Wdfy4*$^{-/-}$ mice (fig. S13, B and C), which still had a defect in cross-presentation of cell-associated antigen (fig. S13D); this finding suggests that the lipid bodies are not necessary to cause the defect in cross-presentation in *Wdfy4*$^{-/-}$ cDC1s.

To determine the interacting partners of WDFY4, we generated four individually FLAG-tagged subregions of WDFY4 spanning the entire protein (Fig. 3A). We stably transduced these fragments into the murine DC line JAWSII (*50*) and performed affinity purification mass spectrometry (AP-MS) to isolate WDFY4 binding partners. We found 143 candidates enriched by different regions of the WDFY4 protein, with the largest number binding to the FL4 fragment of WDFY4 that contains the pleckstrin homology (PH), BEACH, WD40, and FYVE domains (Fig. 3A and table S3). We performed gene ontology analysis

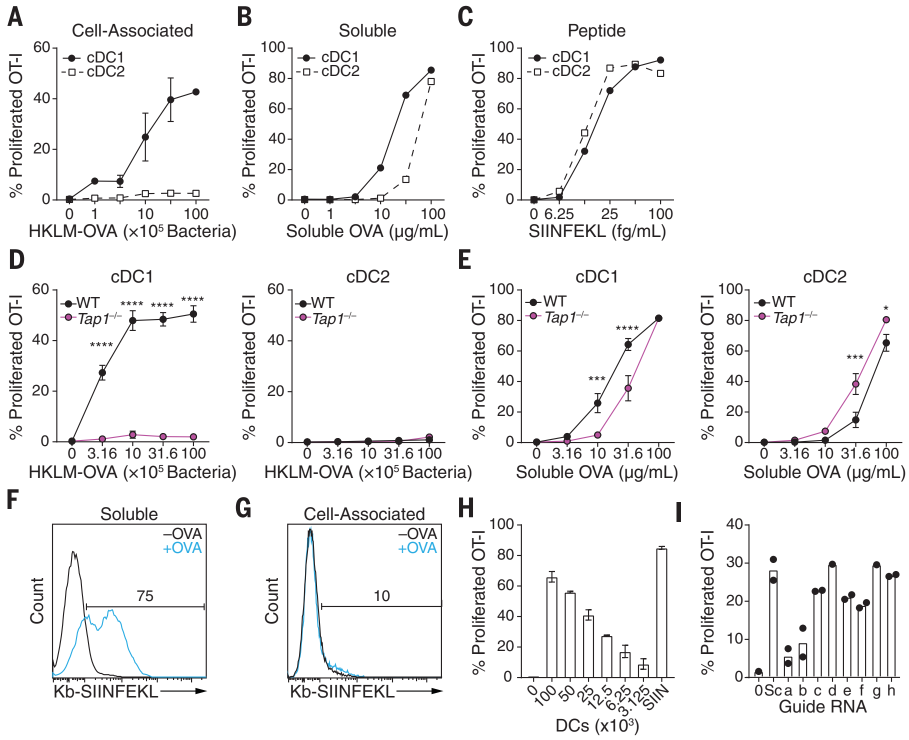


Fig. 1. Establishment of a CRISPR/Cas9 screen for cross-presentation of cell-associated antigens. (**A** to **C**) Sort-purified cDC1s and cDC2s were cultured for 3 days with increasing concentrations of HKLM-OVA (A), soluble OVA (B), or SIINFEKL peptide (C) and CFSE-labeled OT-I T cells and assayed for proliferation ($CFSE^-CD44^+$). (**D** and **E**) Wild-type (WT) or *Tap1*$^{-/-}$ sort-purified cDC1s and cDC2s were cultured for 3 days with varying concentrations of HKLM-OVA (D) or soluble OVA (E) and CFSE-labeled OT-I T cells and assayed for proliferation ($CFSE^-CD44^+$). (**F** and **G**) Sorted cDC1s were cultured with soluble OVA (100 µg/ml) (F) or 10^6 splenocytes osmotically loaded with OVA (G) for 48 hours, stained with 25-D1.16, and analyzed by flow cytometry. (**H**) CFSE-labeled OT-I cells were cultured with the indicated number of whole Flt3L-generated DCs and either 10^7 HKLM-OVA or SIINFEKL peptide (SIIN; 25 fg/ml), and proliferation was measured as in (A). (**I**) c-Kit^{hi} bone marrow progenitors from Cas9 transgenic mice were infected with retroviruses expressing various sgRNAs (table S2) and cultured with Flt3L for 7 days, and infected cDCs were tested for cross-presentation to CFSE-labeled OT-I T cells as in (H). Sc, scramble; activated T cells were gated as $CFSE^-CD44^+$. Data are means ± SEM. $*P < 0.05$, $***P < 0.001$, $****P < 0.0001$ [two-way analysis of variance (ANOVA) with Tukey multiple-comparisons test].

to determine the biological processes most likely influenced by WDFY4 (*51*). The fragments FL1 and FL2 of WDFY4 associated with proteins involved in "protein complex assembly," and therefore they may be involved in forming multimeric protein structures or scaffolding vesicular machinery (Fig. 3B and table S4). FL3 and FL4 associated with proteins involved in "protein localization," "vesicle transport," and "cytoskeletal organization," suggesting a role for WDFY4 in proper subcellular vesicular targeting (Fig. 3C and table S5). Notably, FL4 associated with components critical to the formation, function, and trafficking of endocytic vesicles, including clathrin (*Cltc*, *Clta*) (*52*), subunits of the AP-2 clathrin adaptor complex (*Ap2a1*, *Ap2a2*, *Ap2b1*) (*52*), modulators of cytoskeleton dynamics (*Iqgap1*, *Actn4*) (*53*, *54*), and several members of the vacuolar-type (H$^+$) adenosine triphosphatase complex (*Atp6v0a3*, *Atp6va1*, *Atp6v1f*) (*55*) (Fig. 3D and tables S3 and S6). FL4 also selectively associated with *Hsp90ab1*, a member of the HSP90 chaperone family involved in endosome-to-cytosol translocation of antigen during cross-presentation (*56*–*59*) (Fig. 3, D to F). Although heat shock proteins such as *Hspa8* and *Hsp1a1* can appear as artifacts in AP-MS data because of their function as chaperones (*60*), *Hsp90ab1* is rarely detected in this manner, and therefore its association may represent a functional interaction with WDFY4.

We then sought to determine which vesicles WDFY4 may be acting on by determining its intracellular location. We visualized full-length Twin-Strep-tagged (*61*) WDFY4 in JAWSII cells by confocal microscopy and found that it localized to the periphery of the cytosol near the plasma membrane (Fig. 3G). WDFY4 was poorly colocalized with the cell surface receptor DEC-205, intracellular MHC class I stores, and lysosomes, but demonstrated moderate colocalization with early endosomes and the ER. cDC1s were previously shown to have well-defined and extensive ER structures that may extend throughout the cytosol near vesicular compartments (*62*) and may lead to colocalization with components of the endosomal pathway. WDFY4 showed the highest correlation with the endosomal markers clathrin and Rab11 (Fig. 3, G and H), which suggests that it localizes to an endosomal compartment near the plasma membrane. Taken together, these data suggest that WDFY4 functions in trafficking between the cell surface and endosomes, and thus may regulate multimeric protein assembly required for the proper formation and localization of endocytic vesicles.

We then examined the role of WDFY4 in cross-presentation of cell-associated antigens in vivo. Carboxyfluorescein succinimidyl ester (CFSE)–labeled OT-I T cells showed strong in vivo proliferation induced by immunization with OVA-loaded splenocytes when transferred into *Wdfy4*$^{+/-}$ mice, but not when transferred into *Wdfy4*$^{-/-}$ mice (Fig. 4, A and B), confirming an in vivo defect in cross-presentation. IL-12 produced by cDC1s in response to soluble tachyzoite antigen (STAg) is required for innate immune protection against

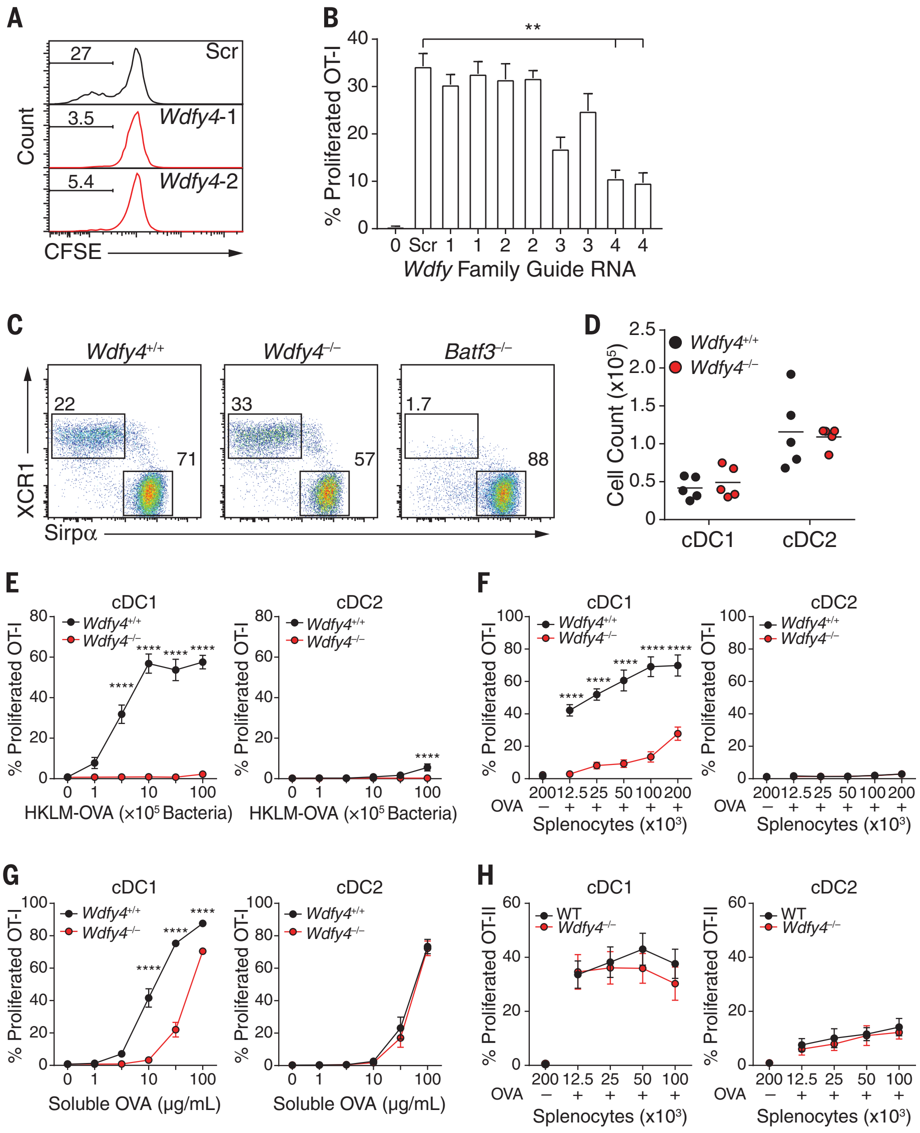


Fig. 2. *Wdfy4* is selectively required for cross-presentation of cell-associated antigens by cDC1s. (**A**) Cross-presentation was measured for Cas9-transgenic cDC1s expressing two sgRNAs (1 and 2; middle and bottom) for *Wdfy4* or a scramble control (Scr; top) that were generated as described in Fig. 1I. T cell proliferation is shown by percentages of CFSE$^-$ OT-I cells. (**B**) Cross-presentation by cDC1s expressing sgRNAs for *Wdfy1*, *Wdfy2*, *Wdfy3*, and *Wdfy4* was measured as described in Fig. 1I. Activated T cells were gated as CFSE$^-$CD44$^+$. Data are means ± SEM of three independent experiments. (**C**) cDC1 and cDC2 development was assessed by flow cytometry in wild-type, *Wdfy4*$^{-/-}$, and *Batf3*$^{-/-}$ mice; plots were pre-gated as B220$^-$CD11c$^+$MHCII$^+$ and then gated as cDC1 (XCR1$^+$Sirpα$^-$) or cDC2 (XCR1$^-$Sirpα$^+$). (**D**) Absolute cell numbers of cDC1s and cDC2s in wild-type and *Wdfy4*$^{-/-}$ mice. Each dot indicates one mouse; bar indicates mean. (**E**) FACS-sorted cDC1s and cDC2s from spleens of wild-type and *Wdfy4*$^{-/-}$ mice were assayed for presentation to OT-I (CFSE$^-$CD44$^+$) in response to the indicated concentrations of HKLM-OVA. (**F**) FACS-sorted cDC1s and cDC2s from spleens of bone marrow chimeric mice with wild-type or *Wdfy4*$^{-/-}$ bone marrow were assayed for presentation to OT-I (CFSE$^-$CD44$^+$) in response to the indicated concentrations of OVA-loaded irradiated splenocytes from MHC class I triple knockout (K$^{b-/-}$, K$^{d-/-}$, β2m$^{-/-}$) mice. (**G**) FACS-sorted cDC1s and cDC2s from spleens of wild-type and *Wdfy4*$^{-/-}$ mice were assayed for presentation to OT-I (CFSE$^-$CD44$^+$) in response to the indicated concentrations of soluble OVA. (**H**) FACS-sorted cDC1s and cDC2s from spleens of wild-type and *Wdfy4*$^{-/-}$ mice were assayed for presentation to OT-II (CFSE$^-$CD44$^+$) in response to the indicated concentrations of OVA-loaded irradiated splenocytes from MHC class II knockout mice. In (F) and (H), OVA– denotes a negative control of splenocytes osmotically pulsed in the absence of OVA. Data are means ± SEM from three independent experiments. *$P < 0.05$, **$P < 0.01$, ****$P < 0.0001$ (two-way ANOVA with Tukey multiple-comparisons test).

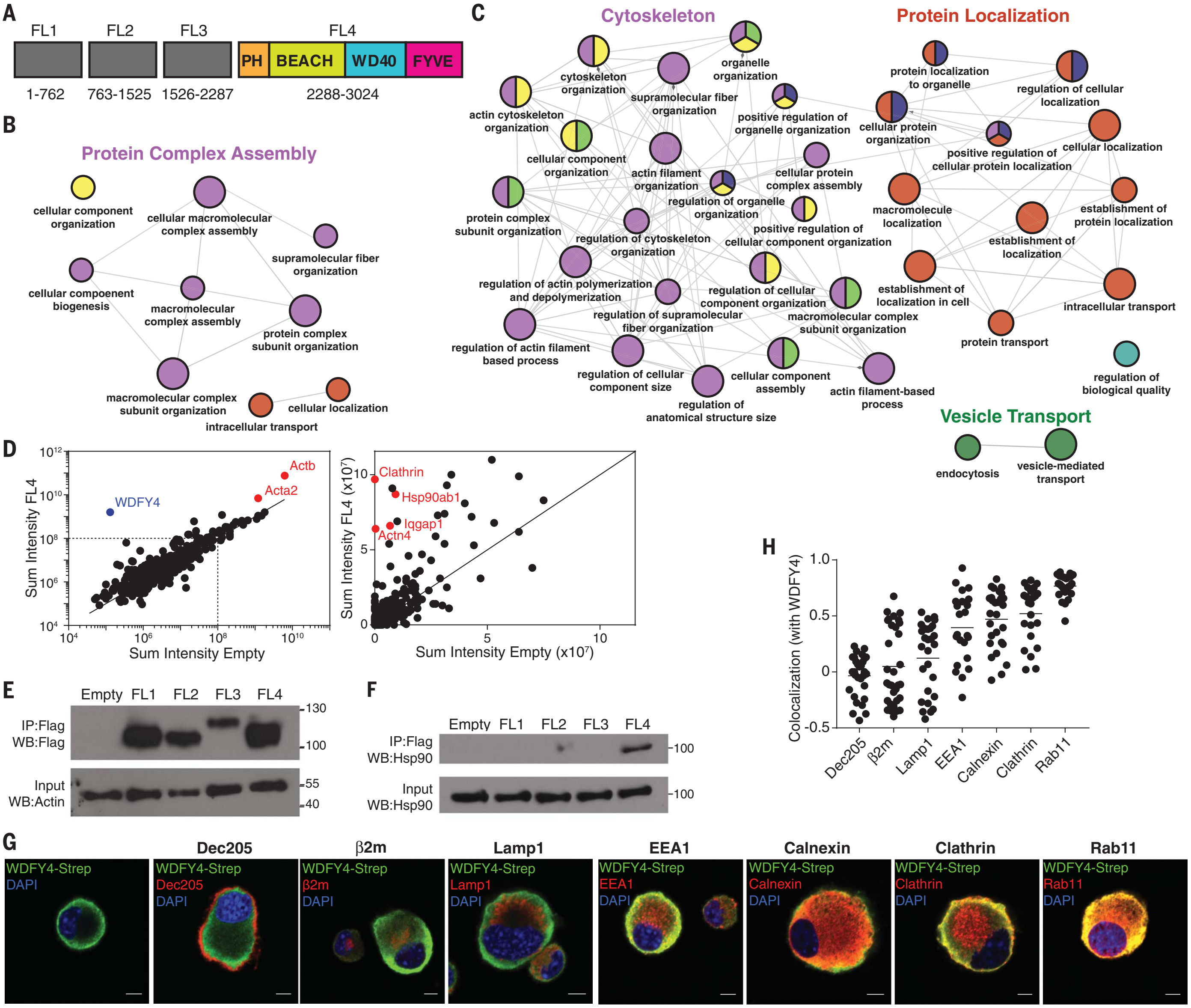


Fig. 3. WDFY4 acts near the plasma membrane and associates with proteins involved in localization and vesicular transport. (**A**) Diagram of truncated fragments of WDFY4 protein, showing predicted domains within the FL4 fragment. Numbers indicate amino acid locations of fragments. (**B**) ClueGO visualization of gene ontology terms enriched after immunoprecipitation of fragments from (A) in the mouse DC line JAWSII, expressing either FL1 or FL2 fragments. Small circles, $P < 0.001$; large circles, $P < 0.0001$. Colors indicate gene ontology (GO) term groups. (**C**) ClueGO visualization of gene ontology terms enriched after immunoprecipitation of fragments from (A) in the mouse DC line JAWSII expressing either FL3 or FL4 fragments. Small circles, $P < 3 \times 10^{-5}$; large circles, $P < 3 \times 10^{-6}$. Colors indicate GO term groups. (**D**) Scatterplot of representative data for sum intensity of proteins found after mass spectrometry between FL4-expressing and empty vector–expressing JAWSII cells. (**E**) Western blot of Flag immunoprecipitates from human embryonic kidney (HEK) 293 cells transfected with empty vector or Flag-tagged WDFY4 fragments 1 to 4 (top) and input control for β-actin (bottom) (**F**) Western blot for endogenous Hsp90 in Flag immunoprecipitates from HEK293 transfected with empty vector or Flag-tagged WDFY4 fragments 1 to 4 (top) and input control for endogenous Hsp90 (bottom). (**G**) Confocal microscopy of JAWSII cells overexpressing full-length Twin-Strep-tagged WDFY4, stained for anti-Strep (green), various cellular markers (red), and 4′,6-diamidino-2-phenylindole (DAPI; blue). Scale bars, 5 μm. (**H**) Quantification of colocalization between WDFY4 and intracellular markers from images in (G). Each dot represents one cell; bar indicates the mean.

T. gondii, as illustrated by the susceptibility of *Batf3*$^{-/-}$ mice to lethal infection by this pathogen (*25*). In contrast, *Wdfy4*$^{-/-}$ mice were resistant to *T. gondii* infection, similar to *Wdfy4*$^{+/-}$ mice (Fig. 4C). These results indicate that cross-presentation is not required for innate protection against *T. gondii* and that *Wdfy4*$^{-/-}$ cDC1s are not globally impaired in function.

We also evaluated CD8^{+} T cell responses of *Wdfy4*$^{-/-}$ mice to cowpox virus infection, a model in which effective CD8^{+} T cell priming is thought to be mediated primarily by *Batf3*-dependent cells through cross-presentation (*13*). *Batf3*$^{-/-}$ mice that lack cDC1s (*10*–*13*) have a defect in priming antigen-specific CD8^{+} T cells to several viruses (*10*–*13*), but these studies only indirectly show that this is due to a lack of cross-presentation, as loss of alternative functions of cDC1s could conceivably be the cause. However, we found that *Wdfy4*$^{-/-}$ mice that retain cDC1 cells that are unable to cross-present also have severely impaired antigen-specific CD8^{+} T cell responses to cowpox virus (Fig. 4, D and E, and fig. S14, A to D). This defect in cross-presentation

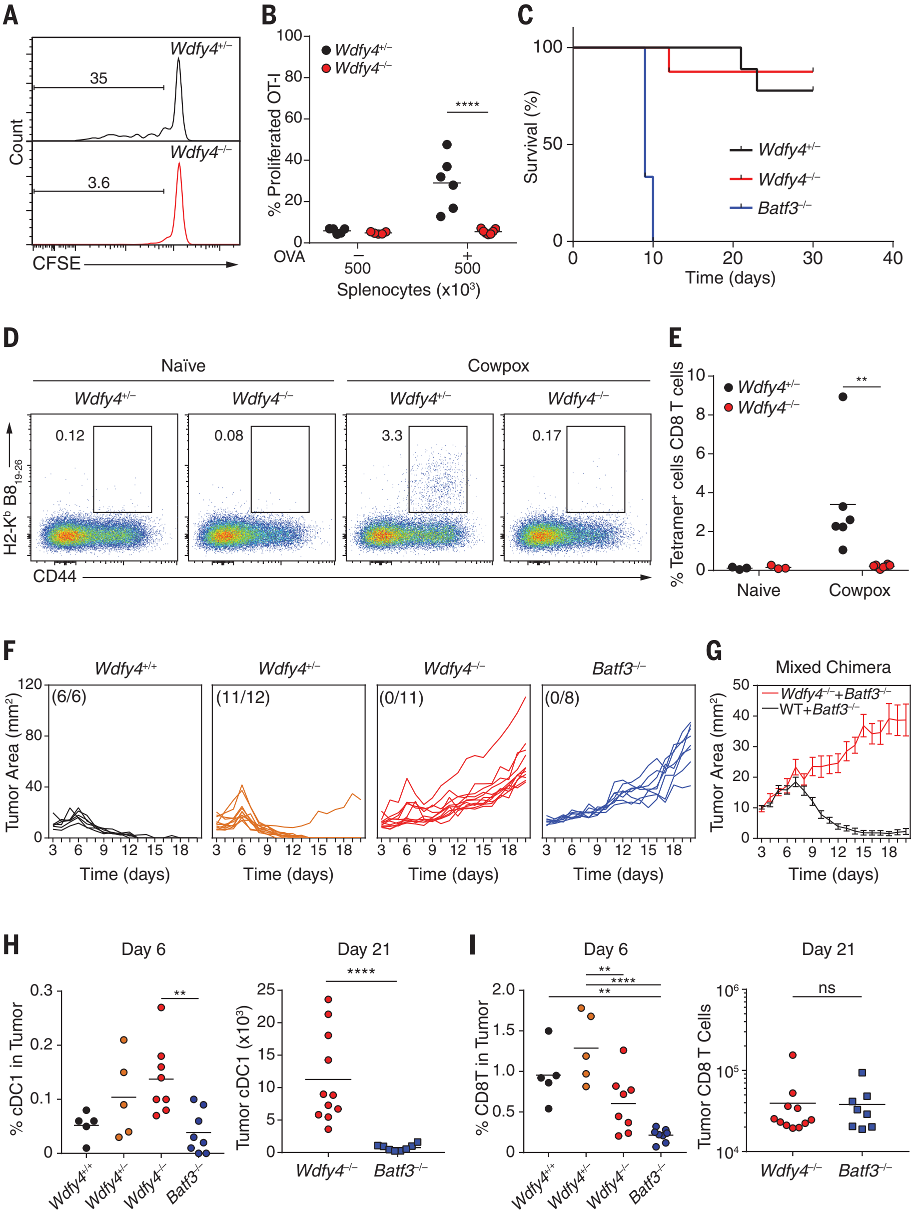


Fig. 4. $Wdfy4^{-/-}$ mice are unable to cross-present in vivo. (**A** and **B**) Representative flow cytometry analysis of in vivo cross-presentation to 500,000 irradiated splenocytes loaded with OVA injected intravenously into mice of the indicated genotypes 1 day after injection of 500,000 CFSE-labeled OT-I cells. Mice were harvested 3 days after antigen injection, quantified in (B). Data are pre-gated on OT-I cells and are shown as percentage of $CFSE^-$ cells (A) or $CFSE^-CD44^+$ cells (B). Data are pooled from three independent experiments; each point represents one mouse. (**C**) Survival of mice of the indicated genotypes to injection of 200 Pru.luc *T. gondii* tachyzoites over 30 days. $Wdfy4^{+/-}$, $n = 9$; $Wdfy4^{-/-}$, $n = 8$; $Batf3^{-/-}$, $n = 3$. (**D** and **E**) Representative flow cytometry plots of CD8 T cells (pre-gated $CD4^-CD3^+CD8^+$) in lungs of naïve or cowpox-infected mice, quantified in (E). Each dot represents one mouse; bar indicates mean. (**F**) Mice of the indicated genotypes were injected with 10^6 fibrosarcoma cells subcutaneously (s.c.) and tumors were measured daily starting at day 3 after injection. (**G**) Mixed bone marrow chimeras with bone marrow of the indicated genotypes were injected into lethally irradiated $CD45.1^+$ wild-type B6 mice. Eight weeks later, mice were injected with 10^6 fibrosarcoma cells s.c. and tumors were measured daily starting at day 3 after injection. Data are means ± SEM of nine mice per group. (**H**) Quantification of cDC1s in tumors at either day 6 or day 21 after injection taken from mice of the indicated genotypes. cDC1s were gated as $B220^-CD11c^+MHCII^+CD103^+CD11b^-$. Each dot indicates one mouse; bar indicates mean. (**I**) Quantification of CD8 T cells in tumors at either day 6 or day 21 after injection taken from mice of the indicated genotypes. $CD8^+$ T cells were gated as $CD45^+TCR\beta^+CD8\alpha^+CD4^-$. Each dot indicates one mouse; bar indicates mean. $^{**}P <$ 0.01, $^{****}P < 0.0001$ (two-way ANOVA with Tukey multiple-comparisons test); ns, not significant.

is not restricted to cowpox virus, because $Wdfy4^{-/-}$ mice also showed a defect in priming $CD8^+$ T cells to infection by West Nile virus (fig. S14E). Furthermore, $Wdfy4^{-/-}$ mice showed normal priming of $CD4^+$ T cells to West Nile, indicating that WDFY4 functions for in vivo cross-presentation to $CD8^+$ T cells, but not for priming of $CD4^+$ T cells (fig. S14F).

Studies with $Batf3^{-/-}$ mice suggested that cDC1s were required for tumor rejection (*10*). To examine the role of cross-presentation directly, we evaluated the growth of the highly immunogenic 1969 regressor fibrosarcoma (*15*) in wild-type, $Wdfy4^{+/-}$, $Wdfy4^{-/-}$, and $Batf3^{-/-}$ mice (Fig. 4F). Tumors were readily rejected by wild-type mice but not by $Batf3^{-/-}$ mice (Fig. 4F), as expected (*15*). However, tumors were also rejected by heterozygous $Wdfy4^{+/-}$ mice but grew uncontrolled in $Wdfy4^{-/-}$ mice, similar to $Batf3^{-/-}$ mice (Fig. 4F and fig. S15A). These results with germline-deficient $Wdfy4^{-/-}$ mice indicate an in vivo requirement for WDFY4 in tumor rejection but do not pinpoint its function to cDC1s. To test whether the in vivo defect in $Wdfy4^{-/-}$ mice is cDC1-intrinsic, we generated mixed bone marrow chimeras using mixtures of either wild-type:$Batf3^{-/-}$ or $Wdfy4^{-/-}$:$Batf3^{-/-}$ bone marrow (Fig. 4G). Wild-type:$Batf3^{-/-}$ chimeras rejected tumors normally, but $Wdfy4^{-/-}$:$Batf3^{-/-}$ chimeras, in which cDC1s develop only from the $Wdfy4^{-/-}$ bone marrow, failed to control tumor growth (Fig. 4G). These results indicate that the defect in tumor rejection results from loss of Wdfy4 expression in cDC1s. Notably, in $Wdfy4^{-/-}$ mice, cDC1s did infiltrate into tumors as they expanded (Fig. 4H and fig. S15B), yet they induced less recruitment of $CD8^+$ T cells to tumors, similar to the lack of $CD8^+$ T cells in tumors in $Batf3^{-/-}$ mice (Fig. 4I).

WDFY4 is one of nine mammalian BDCPs that typically also contain a PH-like domain and WD repeats (*46*). BDCPs function as protein scaffolds that regulate intracellular vesicle fission and fusion events, and several are associated with human diseases (*46*). For example, mutations in *Lyst* cause Chédiak-Higashi syndrome, a primary immunodeficiency disorder characterized by defective neutrophil phagolysosome formation and cytotoxic T cell degranulation (*63*, *64*). Mutations in *Lrba* result in immune dysregulation in regulatory T cells due to improper trafficking of CTLA4 from endosomes to lysosomes by the clathrin adaptor AP-1 (*65*, *66*). WDFY3, the closest WDFY4 homolog, regulates recruitment of polyubiquitinated protein aggregates to autophagosomes by interactions with p62, Atg5, Atg12, Atg16L, LC3, and TRAF6 (*67*–*70*). Although cross-presentation of cell-associated antigens does not involve autophagy (*71*), WDFY4 conceivably may regulate vesicular trafficking pathways, a concept supported by its localization to submembrane endosomes and its interaction with endocytic and cytoskeletal machinery. These WDFY4-dependent trafficking pathways may be required for translocation of dead-cell antigen ligated by the cDC1-specific receptor CLEC9A (*72*) to specific compartments to promote cross-presentation (*73*). Further investigation of the mechanisms of WDFY4 may elucidate previously unknown components of the cross-presentation pathway and thus offer therapeutic targets for human disease.

REFERENCES AND NOTES

1. J. S. Blum, P. A. Wearsch, P. Cresswell, *Annu. Rev. Immunol.* **31**, 443–473 (2013).
2. T. L. Murphy *et al.*, *Annu. Rev. Immunol.* **34**, 93–119 (2016).
3. V. Durai, K. M. Murphy, *Immunity* **45**, 719–736 (2016).
4. M. Guilliams *et al.*, *Nat. Rev. Immunol.* **14**, 571–578 (2014).
5. K. Crozat *et al.*, *J. Immunol.* **187**, 4411–4415 (2011).
6. M. J. Bevan, *J. Exp. Med.* **143**, 1283–1288 (1976).
7. O. P. Joffre, E. Segura, A. Savina, S. Amigorena, *Nat. Rev. Immunol.* **12**, 557–569 (2012).
8. J. M. den Haan, S. M. Lehar, M. J. Bevan, *J. Exp. Med.* **192**, 1685–1696 (2000).
9. D. Dudziak *et al.*, *Science* **315**, 107–111 (2007).
10. K. Hildner *et al.*, *Science* **322**, 1097–1100 (2008).
11. N. Torti, S. M. Walton, K. M. Murphy, A. Oxenius, *Eur. J. Immunol.* **41**, 2612–2618 (2011).
12. K. Nopora *et al.*, *Front. Immunol.* **3**, 348 (2012).
13. M. D. Gainey, J. G. Rivenbark, H. Cho, L. Yang, W. M. Yokoyama, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E3260–E3267 (2012).
14. M. B. Fuertes *et al.*, *J. Exp. Med.* **208**, 2005–2016 (2011).
15. M. S. Diamond *et al.*, *J. Exp. Med.* **208**, 1989–2003 (2011).
16. T. Toubai *et al.*, *Blood* **121**, 4231–4241 (2013).
17. X. Yu *et al.*, *Cancer Res.* **73**, 2093–2103 (2013).
18. O. A. Ali *et al.*, *Cancer Res.* **74**, 1670–1681 (2014).
19. Y. Zhang *et al.*, *J. Immunol.* **194**, 5937–5947 (2015).
20. A. Tzeng *et al.*, *Cell Rep.* **17**, 2503–2511 (2016).
21. A. R. Sánchez-Paulete *et al.*, *Cancer Discov.* **6**, 71–79 (2016).
22. K. D. Moynihan *et al.*, *Nat. Med.* **22**, 1402–1410 (2016).
23. B. H. Kwan *et al.*, *J. Exp. Med.* **214**, 1679–1690 (2017).
24. S. Spranger, D. Dai, B. Horton, T. F. Gajewski, *Cancer Cell* **31**, 711–723.e4 (2017).
25. M. Mashayekhi *et al.*, *Immunity* **35**, 249–259 (2011).
26. D. Theisen, K. Murphy, *F1000 Res.* **6**, 98 (2017).
27. F. Sallusto, A. Lanzavecchia, *J. Exp. Med.* **179**, 1109–1118 (1994).
28. C. Caux, C. Dezutter-Dambuyant, D. Schmitt, J. Banchereau, *Nature* **360**, 258–261 (1992).
29. K. Inaba *et al.*, *J. Exp. Med.* **176**, 1693–1702 (1992).
30. J. Helft *et al.*, *Immunity* **42**, 1197–1211 (2015).
31. C. G. Briseño *et al.*, *Cell Rep.* **15**, 2462–2474 (2016).
32. N. M. Kretzer *et al.*, *J. Exp. Med.* **213**, 2871–2883 (2016).
33. M. Kovacsovics-Bankowski, K. L. Rock, *Science* **267**, 243–246 (1995).
34. V. G. Morón, P. Rueda, C. Sedlik, C. Leclerc, *J. Immunol.* **171**, 2242–2250 (2003).
35. F. M. Cruz, J. D. Colbert, E. Merino, B. A. Kriegsman, K. L. Rock, *Annu. Rev. Immunol.* **35**, 149–176 (2017).
36. J. D. Pfeifer *et al.*, *Nature* **361**, 359–362 (1993).
37. L. Shen, L. J. Sigal, M. Boes, K. L. Rock, *Immunity* **21**, 155–165 (2004).
38. I. Cebrian *et al.*, *Cell* **147**, 1355–1368 (2011).
39. A. Alloatti *et al.*, *J. Exp. Med.* **214**, 2231–2241 (2017).
40. S. J. Wu *et al.*, *Cell Rep.* **19**, 2645–2656 (2017).
41. A. Porgador, J. W. Yewdell, Y. Deng, J. R. Bennink, R. N. Germain, *Immunity* **6**, 715–726 (1997).
42. C. V. Harding, E. R. Unanue, *Nature* **346**, 574–576 (1990).
43. E. R. Christinck, M. A. Luscher, B. H. Barber, D. B. Williams, *Nature* **352**, 67–70 (1991).
44. L. Cong *et al.*, *Science* **339**, 819–823 (2013).
45. R. J. Platt *et al.*, *Cell* **159**, 440–455 (2014).
46. A. R. Cullinane, A. A. Schäffer, M. Huizing, *Traffic* **14**, 749–766 (2013).
47. UniProt Consortium, *Nucleic Acids Res.* **45**, D158–D169 (2017).
48. P. Schnorrer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10729–10734 (2006).
49. E. Segura, A. L. Albiston, I. P. Wicks, S. Y. Chai, J. A. Villadangos, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20377–20381 (2009).
50. V. McKay, E. E. Moore, U.S. Patent 5,648,219 (1997).
51. G. Bindea *et al.*, *Bioinformatics* **25**, 1091–1093 (2009).
52. M. Kaksonen, A. Roux, *Nat. Rev. Mol. Cell Biol.* **19**, 313–326 (2018).
53. A. C. Hedman, J. M. Smith, D. B. Sacks, *EMBO Rep.* **16**, 427–446 (2015).
54. D. G. Thomas, D. N. Robinson, *Semin. Cell Dev. Biol.* **71**, 68–74 (2017).
55. K. C. Jefferies, D. J. Cipriano, M. Forgac, *Arch. Biochem. Biophys.* **476**, 33–42 (2008).
56. H. Zheng, Z. Li, *J. Immunol.* **173**, 5929–5933 (2004).
57. T. Ichiyanagi *et al.*, *J. Immunol.* **185**, 2693–2700 (2010).
58. T. Imai *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16363–16368 (2011).
59. Y. Kato *et al.*, *Autoimmune Dis.* **2012**, 745962 (2012).
60. D. Mellacheruvu *et al.*, *Nat. Methods* **10**, 730–736 (2013).
61. M. R. Junttila, S. Saarinen, T. Schmidt, J. Kast, J. Westermarck, *Proteomics* **5**, 1199–1203 (2005).
62. F. Osorio *et al.*, *Nat. Immunol.* **15**, 248–257 (2014).
63. C. M. Perou *et al.*, *Nat. Genet.* **13**, 303–308 (1996).
64. D. L. Nagle *et al.*, *Nat. Genet.* **14**, 307–311 (1996).
65. B. Lo *et al.*, *Science* **349**, 436–440 (2015).
66. B. Lo *et al.*, *Blood* **128**, 1037–1042 (2016).
67. A. Simonsen *et al.*, *J. Cell Sci.* **117**, 4239–4251 (2004).
68. T. H. Clausen *et al.*, *Autophagy* **6**, 330–344 (2010).
69. M. Filimonenko *et al.*, *Mol. Cell* **38**, 265–279 (2010).
70. P. Isakson *et al.*, *Autophagy* **9**, 1955–1964 (2013).
71. J. D. Mintern *et al.*, *Autophagy* **11**, 906–917 (2015).
72. D. Sancho *et al.*, *Nature* **458**, 899–903 (2009).
73. S. Zelenay *et al.*, *J. Clin. Invest.* **122**, 1615–1627 (2012).

ACKNOWLEDGMENTS

We thank the Alvin J. Siteman Cancer Center at Washington University School of Medicine for use of the Center for Biomedical Informatics and Multiplex Gene Analysis Genechip Core Facility; E. Tonc for blinded microscopy analysis; D. Oakley and J. Fitzpatrick for assistance with confocal microscopy, which was performed at Washington University Center for Cellular Imaging (WUCCI), supported by Washington University School of Medicine, the Children's Discovery Institute of Washington University and St. Louis Children's Hospital (CDI-CORE-2015-505), the Foundation for Barnes-Jewish Hospital (3770), and the National Institute for Diabetes and Digestive and Kidney Diseases (P30 DK020579); and R. Tomaino and the Taplin mass spectrometry core at Harvard University for performing mass spectrometry experiments. **Funding:** Supported by NIH grant T32 AI007163-40 (D.J.T.); NCI grants T32 CA 009621 (J.T.D.), 5R01 CA 190700 (R.D.S.), P30 CA 091842 (W.E.G.), and 5R01 CA 193318 (N.M.); NIAID grant U19-AI109948 (W.Y.); NIH grant F30DK108498 (V.D.); NSF grant DGE-1143954 (P.B.); NIH grant U19-AI109725 (H.W.V.); and the Howard Hughes Medical Institute (K.M.M.). WDFY4-deficient mice used in this study were generated by the JAX KOMP program (U42OD011185). **Author contributions:** D.J.T. and J.T.D. designed, performed, and analyzed most experiments; C.G.B. and P.B. performed and analyzed experiments; V.D., M.G., E.J.L., W.Y., Q.W., L.D.S., W.L.B., J.R.B., N.M., P.D., and M.S.D. designed, performed, and analyzed experiments; R.D.S. and H.W.V. generated models and designed experiments; and W.E.G., T.L.M., and K.M.M. designed experiments and wrote the manuscript. **Competing interests:** H.W.V. is a founder of Casma Therapeutics, Cambridge, MA. **Data and materials availability:** Microarray data have been deposited in the NCBI gene expression omnibus at accession number GSE118652; the mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data identifier PXD011210. All other data are available in manuscript or supplementary materials. WDFY4 knockout mice were obtained through the Knockout Mouse Project (KOMP) at the Jackson Laboratory (catalog number: 029334).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/694/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S6
References (*74*–*85*)

5 March 2018; resubmitted 17 August 2018
Accepted 27 September 2018
10.1126/science.aat5030

IMMUNOLOGY

An autoimmune disease variant of IgG1 modulates B cell activation and differentiation

Xiangjun Chen[1], Xiaolin Sun[2]*, Wei Yang[3]*, Bing Yang[1]*, Xiaozhen Zhao[2], Shuting Chen[1], Lili He[1], Hui Chen[4], Changmei Yang[1], Le Xiao[1], Zai Chang[3], Jianping Guo[2], Jing He[2], Fuping Zhang[5], Fang Zheng[6], Zhibin Hu[7], Zhiyong Yang[8], Jizhong Lou[4], Wenjie Zheng[9], Hai Qi[10], Chenqi Xu[11], Hong Zhang[12], Hongying Shan[13], Xu-jie Zhou[12], Qingwen Wang[13], Yi Shi[14,15], Luhua Lai[16], Zhanguo Li[2]†, Wanli Liu[1,17]†

The maintenance of autoreactive B cells in a quiescent state is crucial for preventing autoimmunity. Here we identify a variant of human immunoglobulin G1 (IgG1) with a Gly^{396}→Arg substitution (hIgG1-G396R), which positively correlates with systemic lupus erythematosus. In induced lupus models, murine homolog Gly^{390}→Arg (G390R) knockin mice generate excessive numbers of plasma cells, leading to a burst of broad-spectrum autoantibodies. This enhanced production of antibodies is also observed in hapten-immunized G390R mice, as well as in influenza-vaccinated human G396R homozygous carriers. This variant potentiates the phosphorylation of the IgG1 immunoglobulin tail tyrosine (ITT) motif. This, in turn, alters the availability of phospho-ITT to trigger longer adaptor protein Grb2 dwell times in immunological synapses, leading to hyper–Grb2–Bruton's tyrosine kinase (Btk) signaling upon antigen binding. Thus, the hIgG1-G396R variant is important for both lupus pathogenesis and antibody responses after vaccination.

Autoimmune diseases such as systemic lupus erythematosus (SLE) are characterized by the presence of large numbers of self-reactive antibodies that induce deposition of immune complexes (ICs), leading to inflammation and tissue damage (*1*). Autoreactivity is pervasive in the antibody repertoire of human B cells across different developmental stages (*2*). It is especially enriched in the peripheral immunoglobulin G–positive (IgG^{+}) memory B cell pool but is efficiently diminished in the plasma cell compartment in healthy individuals (*2*–*4*). However, these checkpoints fail in patients with autoimmune diseases. It remains unclear how autoreactive IgG^{+} B cells are maintained in a quiescent state under physiological immune homeostasis and how these checkpoints are broken in pathological conditions. IgG–B cell receptor (IgG-BCR) potently enhances memory IgG antibody responses via the evolutionarily conserved cytoplasmic tail of membrane-bound IgG (mIgG-tail) (*5*–*10*). The mIgG-tail amplifies BCR signaling via its phospho–immunoglobulin tail tyrosine (ITT) motif, which recruits the adaptor protein Grb2 to enhance Ca^{2+} mobilization, synergistically with Bruton's tyrosine kinase (Btk) and phospholipase C–γ2 (PLC-γ2) (*5*, *11*).

Here we identified a single-nucleotide polymorphism (SNP) rs117518546, which results in a glycine-to-arginine substitution at codon *396* in human IgG1 (hIgG1-G396R) (fig. S1A and table S1). This SNP was common in East Asian populations (fig. S1B) and was significantly correlated with susceptibility to SLE (tables S2 and S3). The G396R variant frequency was substantially enriched in SLE patients compared with that of criteria-matched controls in three independent cohorts from multiclinical centers in China (1786 healthy controls versus 1838 SLE patients in total, Pearson's chi-square test and binary logistic regression analysis, $P = 6.0 \times 10^{-5}$). Furthermore, the G396R variant was associated with a more severe disease phenotype, including earlier onset, multiple organ involvement, and higher SLE disease activity; specifically aggravated autoantibody production; and inflammation (Fig. 1A and table S4). The variant drove an autoantibody subclass profile shift toward IgG1-isotype predominance in G396R patients (Fig. 1B). Thus, hIgG1-G396R is a risk locus for SLE.

We generated knockin mice harboring the murine homolog mIgG1-G390R (denoted as *Ighg1*$^{T/T}$ or G390R mice) (figs. S1C and S2, A and B). Under normal conditions, there were no significant differences in terms of natural antibodies or other evident phenotypes between wild-type (WT) and G390R mice (fig. S2C). In the bm12 splenocyte–inducible lupus model (*12*), the levels of IgG1 subclass anti–double-stranded DNA (dsDNA), anti–Smith D (SmD), and anti-nuclear antibodies (ANA) were notably increased in G390R mice (Fig. 1, C and D, and fig. S3A). Autoantigen microarrays further confirmed a clear increase in IgG1 but not IgM and IgG2b autoantibody production in G390R mice (Fig. 1E and fig. S3B). Moreover, G390R mice showed enlarged glomeruli and substantial IgG1 deposits containing ICs along with moderate IgG2b deposition (Fig. 1F). Similar results were also observed in a second lupus model by using apoptotic thymocytes (fig. S3, C to E). In aged mice, the G390R variant facilitated autonomous IgG1 autoantibody production (fig. S3F). Thus, the G390R variant promotes autoantibody production during autoimmune disease progression.

Six weeks after the induction of autoimmunity with bm12 splenocytes, there was a fourfold increase in the number of IgG1^{+} germinal center (GC) B cells in G390R mice compared with that of WT mice (Fig. 2A and fig. S4A). IgG1^{+} memory B cells only mildly increased in G390R mice (Fig. 2B). Notably, there was an almost sixfold increase in IgG1^{+} plasma cells in both the spleens and bone marrow of G390R mice (Fig. 2, C to E, and fig. S4A). This phenomenon was also seen in models of apoptotic thymocyte-induced lupus and aging (fig. S4, B and C). In a competitive model, in which WT or G390R B cells were adoptively transferred into B cell–deficient hosts (fig. S4D), more IgG1^{+} plasma cells were differentiated in the G390R group than in the WT group relative to the internal control. The competition index for the G390R group was 50% higher than that for the

---

[1] Ministry of Education Key Laboratory of Protein Sciences, Center for Life Sciences, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Institute for Immunology, School of Life Sciences, Tsinghua University, Beijing 100084, China. [2] Department of Rheumatology and Immunology, Peking University People's Hospital, Beijing Key Laboratory for Rheumatism and Immune Diagnosis (BZ0135), Peking-Tsinghua Center for Life Sciences, State Key Laboratory of Natural and Biomimetic Drugs, Peking University, Beijing 100044, China. [3] School of Life Sciences, Tsinghua University, Beijing 100084, China. [4] Laboratory of RNA Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. [5] Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China. [6] Department of Immunology, School of Basic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China. [7] Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing 211166, China. [8] Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA 94143, USA. [9] Department of Rheumatology and Clinical Immunology, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China. [10] Tsinghua-Peking Center for Life Sciences, Laboratory of Dynamic Immunobiology, School of Medicine, Tsinghua University, Beijing 100084, China. [11] State Key Laboratory of Molecular Biology, Shanghai Science Research Center, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai 200031, China. [12] Renal Division, Peking University First Hospital, Peking University Institute of Nephrology, Key Laboratory of Renal Disease, Ministry of Health of China, Beijing 100034, China. [13] Department of Rheumatism and Immunology, Peking University Shenzhen Hospital, Shenzhen 518036, China. [14] CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100101, China. [15] Research Network of Immunity and Health (RNIH), Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China. [16] BNLMS, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, and Peking-Tsinghua Center for Life Sciences at College of Chemistry and Molecular Engineering, Center for Quantitative Biology, Peking University, Beijing 100871, China. [17] Beijing Key Lab for Immunological Research on Chronic Diseases, Beijing 100084, China.

*These authors contributed equally to this work.

†Corresponding author. Email: liulab@tsinghua.edu.cn (W.L.); li99@bjmu.edu.cn (Z.L.)

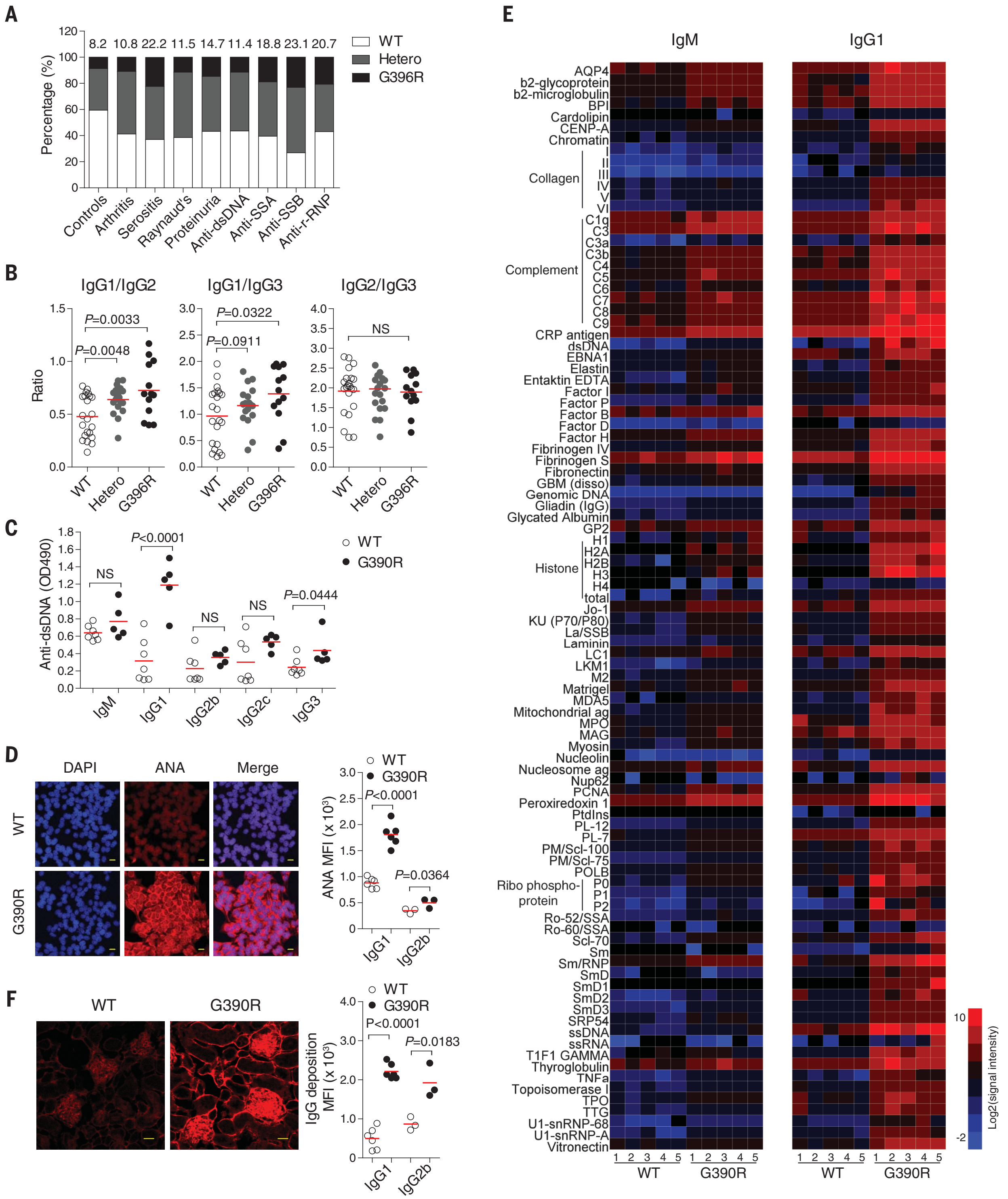


Fig. 1. An SLE-correlated SNP variant increases autoantibody production. (**A**) Genotypic correlation analysis of the human IgG1-G396R variant with clinical manifestations in SLE patients (279 controls and 251 SLE patients, as detailed in table S4). The percentages indicate the portion of G396R homozygotes. Hetero, heterozygous; SSA, Sjögren's syndrome–related antigen A; SSB, Sjögren's syndrome–related antigen B; r-RNP, ribonucleoproteins. (**B**) Ratio of IgG subclass for ANA in WT ($n = 22$), Hetero ($n = 19$), and G396R ($n = 13$) SLE patients. One-way analysis of variance (ANOVA), IgG1/IgG2 ($P = 0.0017$), IgG1/IgG3 ($P = 0.0552$), and IgG2/IgG3 ($P = 0.9268$). (**C**) Anti-dsDNA antibodies detected in serum from bm12-induced WT ($n = 7$) and G390R ($n = 5$) mice 3 weeks after induction. OD490, optical density at a wavelength of 490 nm. (**D**) Immunofluorescence of IgG1$^+$ ANA with HEp-2 reactivity in week 3 serum from bm12-induced mice (WT, $n = 6$; G390R, $n = 6$). Mean fluorescence intensity (MFI) of IgG1 and IgG2b subclass ANA was quantified. DAPI, 4′,6-diamidino-2-phenylindole. (**E**) Autoantigen microarray was used to detect autoantibodies in serum as in (C). (**F**) Immunofluorescence of deposited IgG1 ICs in mice glomeruli at week 6 after induction. Statistical analysis of MFI of IgG1 and IgG2b ICs in glomeruli. The scale bar represents 20 μm in (D) and (F). Unpaired two-tailed Student's *t* tests, (B) to (D) and (F). NS, not significant. Red bars indicate means. At least three sections were analyzed for each sample in (F). Data are representative of at least two independent experiments in (B) to (D) and (F).

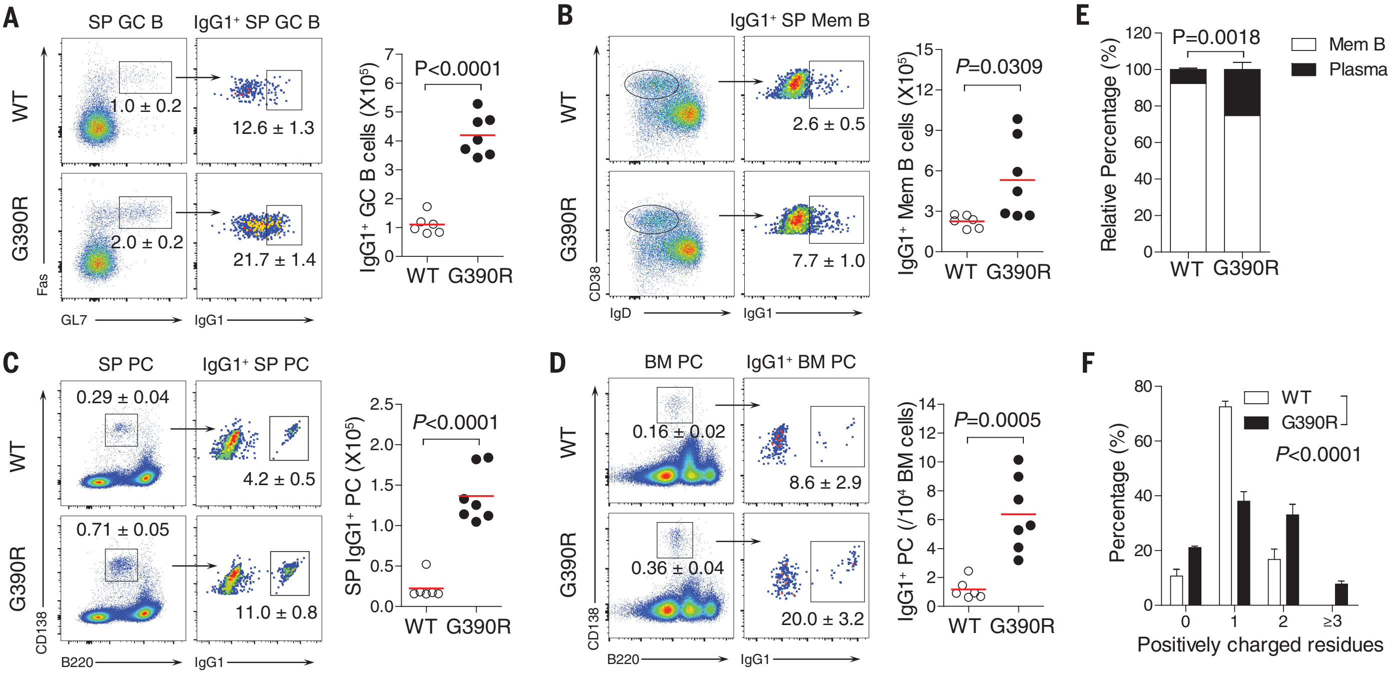


Fig. 2. The IgG1-G390R variant promotes plasma cell accumulation in induced autoimmune models. (**A** to **D**) Flow cytometric analyses of IgG1$^+$ GC B cells, memory B (Mem B) cells, spleen (SP), and bone marrow (BM) plasma cells (PC) in WT (n = 6) and G390R (n = 7) mice at week 6 after bm12 splenocyte induction. GC B cells (GL7$^+$, Fas$^+$) were pregated on B220$^+$ cells. Percentage (means ± SEM) of B cell subsets and corresponding comparison of cell numbers (right) are indicated. (**E**) The relative percentage (means ± SEM) of IgG1$^+$ plasma cells and memory B cells in fate-selected cells of WT in comparison to those of G390R mice. (**F**) Frequency (means ± SEM) of Igγ1 CDR3 with different numbers of positively charged amino acids in IgG1$^+$ bone marrow plasma cells (WT CDR3, n = 64; G390R CDR3, n = 66) from bm12-induced WT (n = 3) and G390R mice (n = 3). Unpaired two-tailed Student's t tests, (A) to (E). Two-way ANOVA, (F). Bars indicate means. Representative data are from at least two independent experiments in (A) to (F).

WT group (fig. S4E). Thus, increased B cell differentiation upon antigen stimulation, especially plasma cell generation, may be responsible for excessive autoantibody production in G390R mice.

Positively charged amino acids in IgH complementarity-determining region 3 (CDR3) are associated with antibody autoreactivity (*13*), and indeed, these types of amino acids were enriched in IgG1$^+$ plasma cells from G390R mice compared with those from WT mice (Fig. 2F). Moreover, the length of the Igγ1 CDR3, which potentially predicts autoreactivity enrichments (*2*), was prominently increased in the G390R population (fig. S4F). We also observed reduced Ig variable (V) gene VH1 and joining (J) gene JH1 usage, but increased VH5, VH14, and JH4 usage in G390R IgG1$^+$ plasma cells (fig. S4G).

The enhanced antibody production by the hIgG1-G396R variant may also have effects on physiological humoral responses to non-self antigens. Thus, we conducted influenza vaccination experiments in healthy human WT controls and G396R homozygous carriers (Fig. 3A). The G396R variant significantly potentiated the generation of influenza virus–specific IgG1 (Fig. 3B) and only showed a relative minor effect on IgG2 production. We also confirmed these effects in 4-hydroxy-3-nitrophenylacetyl (NP) eight-keyhole limpet hemocyanin (KLH)–immunized mice (Fig. 3C). NP-specific IgG1 antibody responses were almost twofold higher in G390R mice and threefold higher during antigen recall responses (Fig. 3D). Affinity maturation did not appear to be significantly affected (fig. S4H). The increase in antibody production was in line with the enhanced numbers of NP-specific IgG1$^+$ GC B and plasma cells in G390R mice (Fig. 3, E and F). Thus, this variant enhances antibody production in both autoimmune disorders and in physiological humoral responses upon vaccination.

Because IgG1$^+$ light-zone GC B cells in bm12-induced G390R mice up-regulated plasma cell fate favoring transcription profiles (Fig. 4A), we then investigated the membrane-proximal signaling upon B cell activation (*8*, *14*). BCR microcluster formation, the synaptic recruitment of Grb2 and Btk, calcium mobilization, and the phosphorylation of downstream signaling molecules [Erk, S6, and nuclear factor–κB (NF-κB)] were significantly enhanced, indicating heightened signaling mediated by the variant (Fig. 4, B and C, and fig. S5, A to H). This effect was also observed in IgG1$^+$ primary B cells from the peripheral blood of G396R homozygotes compared with those from WT SLE patients (Fig. 4D and fig. S5, I and J).

In a fluorescence resonance energy transfer (FRET)–based phospho-ITT activation reporter (fig. S6, A to C), G390R variants showed significantly higher FRET ratio changes compared with WT upon antigen stimulation (Fig. 4E). Thus, the G390R variant promotes the activation of the phospho-ITT–Grb2 signaling module. Additional single-molecule imaging experiments revealed a more confined Grb2 motion in the immunological synapse of the G390R variant compared with that of WT primary B cells (Fig. 4F). Thus, two potential models were proposed to explain the excessive activation of phospho-ITT–Grb2 signaling module: Increased Grb2 binding to the phospho-ITT motif and/or increased ITT-motif phosphorylation. Isothermal titration calorimetry assays demonstrated similar binding between the Grb2 protein and either the WT or G390R phospho-ITT motif peptide (fig. S7A). In the second model, both Syk and Lyn tyrosine kinases are involved in the phosphorylation of the ITT motif (*11*). G390R ITT phosphorylation by Syk was only mildly higher (fig. S7, B and C). Substantially enhanced phosphorylation of the G390R ITT motif by Lyn was observed in vitro (Fig. 4G and fig. S7D) and in vivo (fig. S7E).

Finally, to identify the structural basis for excessive G390R ITT motif phosphorylation by Lyn, we carried out molecular dynamics simulations of the Lyn kinase domain bound to either the WT or the G390R ITT motif. We constructed complex models of substrate peptide and kinase domain based on the structure of the active human Lyn kinase domain (Protein Data Bank 3A4O) (*15*). The antiparallel β sheet formed by the substrate peptide and the activation loop of kinase domain persisted for the length of each

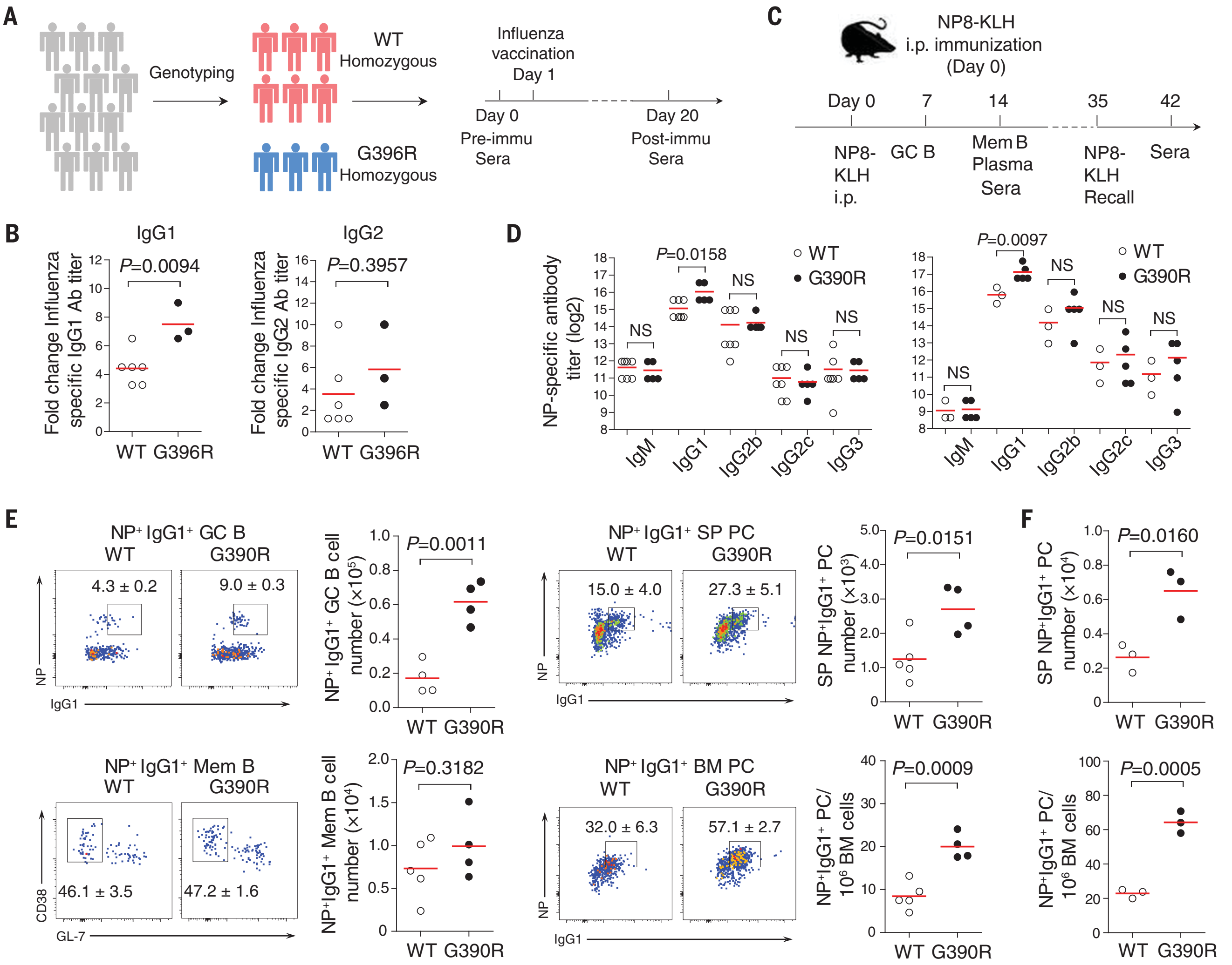


Fig. 3. The IgG1-G390R variant facilitates IgG1 antibody production in physiological humoral responses. (**A**) Schematic diagram of the human influenza vaccination study. WT and G396R homozygous healthy volunteers were influenza vaccinated on day 1. Pre- and postimmune sera were collected on day 0 and day 20. (**B**) Influenza-specific IgG1 and IgG2 antibody responses in vaccinated WT (*n* = 6) and G396R (*n* = 3) healthy volunteers. Fold change of antibody titers on day 20 to day 0 was quantified. (**C**) Diagrammatic representation of the NP8-KLH immunization in mice. GC B, memory B, and plasma cells were analyzed on days 7 (GC B) and 14 (memory B and plasma cells). (**D**) NP-specific antibody titers in WT (*n* = 7 or 3) and G390R (*n* = 5) mice on days 14 (left) and 42 (right) after NP8-KLH immunization. (**E**) Flow cytometric analyses of NP-specific IgG1$^+$ GC (pregated on B220$^+$, GL7$^+$, CD38$^-$) on day 7, memory B cells (pregated on B220$^+$, IgG1$^+$, NP$^+$), spleen and bone marrow plasma cells (pregated on B220^{lo}, CD138$^+$) on day 14 in NP8-KLH immunized WT (*n* = 4 or 5) and G390R (*n* = 4) mice. Percentage (means ± SEM) and B cell–subset numbers were compared. (**F**) Spleen and bone marrow plasma cell numbers in NP recall responses (WT, *n* = 3; G390R, *n* = 3). Unpaired two-tailed Student's *t* tests, (C) to (F). Red bars indicate means. NS, not significant. Data in (C) to (F) are representative of at least two independent experiments.

trajectory. Additionally, the Asp-Phe-Gly (DFG) motif and αC helix of the Lyn kinase domain were maintained in an active conformation (Fig. 4H and fig. S7, F and G). Notably, the Arg^{390} residue in the G390R ITT motif formed an additional hydrogen bond with the backbone carbonyl of Asn^{290} in Lyn kinase domain, which was not available for the Gly^{390} residue in the WT ITT motif (Fig. 4H). Furthermore, the distance between the hydroxyl group of catalytic residue Tyr^{385} and γ-phosphate group of adenosine triphosphate (ATP) was significantly shorter for the G390R variant than for WT (Fig. 4I and fig. S7H), suggesting more stable contacts and potentially more effective phosphotransfer for the G390R variant. Indeed, the binding free energy of the G390R variant was consistently lower than that of the WT ITT motif (Fig. 4H). Together, these computational results suggest the higher binding ability of the G390R variant to active Lyn kinase.

The G390R variant potentiates ITT-motif phosphorylation and alters the availability of the phospho-ITT motif, thereby triggering a significantly longer dwell time of Grb2 in the immunological synapse. This, in turn, changes the synaptic Grb2 recruitment from a "recruit-and-escape" model to a "recruit-and-confine" model, in which the initially recruited but subsequently escaped Grb2 can be recaptured by the proximal phospho-ITT motifs within the synapse (fig. S8). This results in excessive G390R variant IgG1$^+$ B cell activation and differentiation to plasma cells producing increased IgG1 antibodies in the context of both autoimmune disorders and physiological humoral responses upon infection or vaccination. This enhancement of G390R variant IgG1$^+$ B cells may have some bona fide effects on the production of other IgG subclass antibodies by potential mechanisms: (i) Enhanced IgG1$^+$ B cell activation may promote the function of T follicular helper cells (*16*) to potentiate the activation and differentiation of other IgG subclass B cells; (ii) increased IgG1$^+$ antibodies may drive the activation and proliferation of

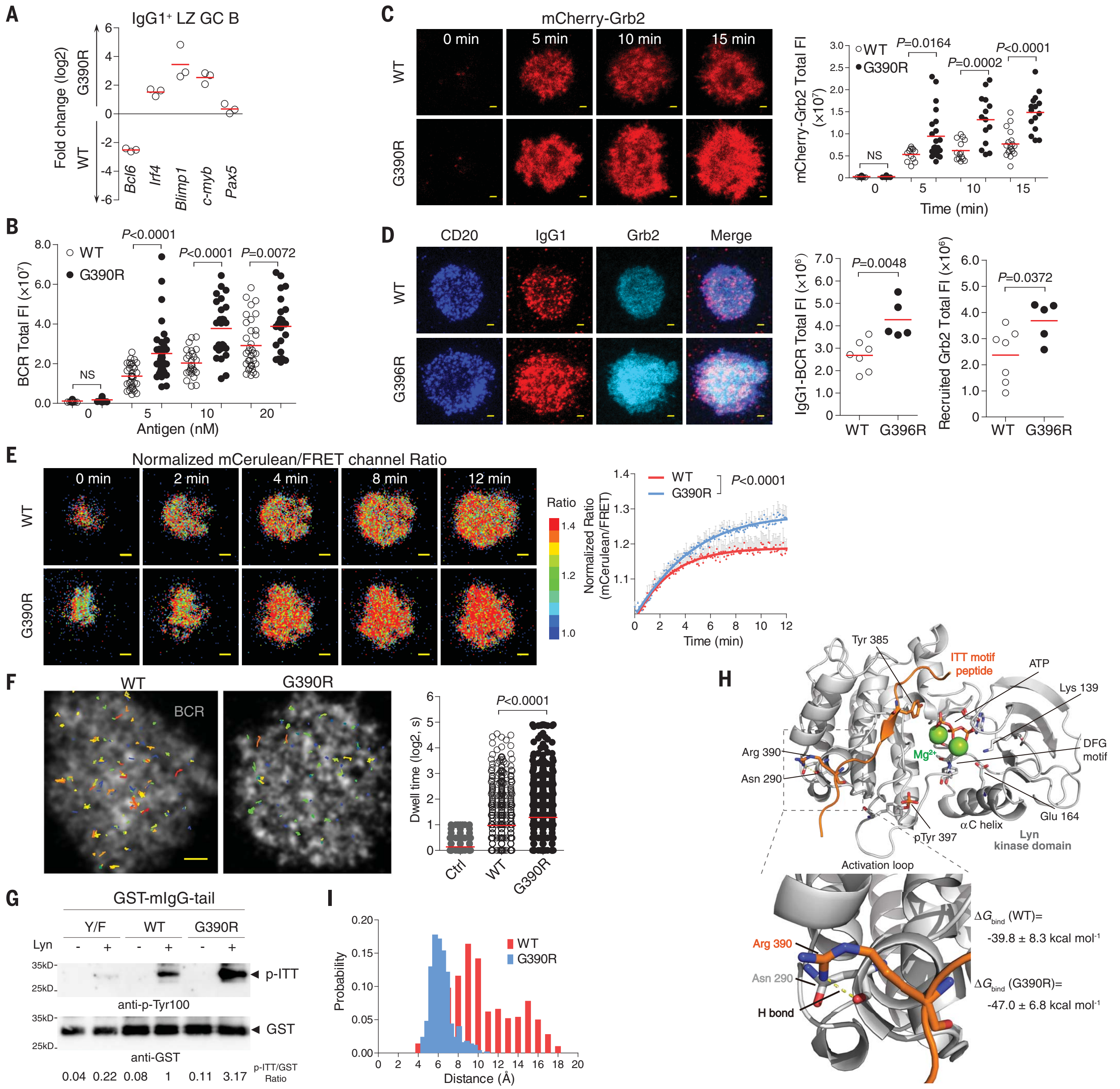


Fig. 4. The IgG1-G390R variant induces excessive IgG-BCR signaling by potentiating phospho-ITT–Grb2 signaling module. (**A**) Transcription profile of *Bcl6*, *Irf4*, *Blimp1*, *c-myb*, and *Pax5* in IgG1$^+$ light-zone GC B cells (B220$^+$, GL-7$^+$, Fas$^+$, CD86^{hi}, CXCR4^{lo}) from WT ($n = 3$) and G390R ($n = 3$) mice upon immunization. (**B**) Total fluorescence intensity (FI) of synaptic accumulated BCRs were compared in IgG1$^+$ class-switched WT and G390R primary B cells with different concentrations of antigen stimulation. (**C**) Dynamics of Grb2 recruitment to the immunological synapse in WT and G390R class-switched IgG1$^+$ primary B cells ($n > 13$). (**D**) Synaptic IgG1-BCR accumulation and Grb2 recruitment in IgG1$^+$ B cells from the peripheral blood of WT ($n = 7$) and G396R homozygous ($n = 5$) SLE patients after surrogate antigen stimulation. (**E**) Heat map indicates ratio dynamics of mCerulean to cpV (FRET channel) fluorescence intensity in cells expressing FRET-based phospho-ITT activation biosensor with WT or G390R ITT motifs after antigen stimulation. Ratio dynamics were quantified, and ratios were normalized to the initial frame upon activation. (**F**) Trajectories of mEos3.1-Grb2 (colored) in immunological synapse (gray). Dwell time of mEos3.1-Grb2 in activated IgG1$^+$ class-switched WT or G390R primary B cells ($n > 10$) are calculated and compared, with resting WT cells as control. (**G**) In vitro phosphorylation assay with purified glutathione *S*-transferase (GST)–fused mIgG-tail protein to compare the phosphorylation of WT or G390R ITT motif by active Lyn kinase. p, phosphorylated. (**H**) Predicted ITT motif peptide binding modes with the active Lyn kinase domain. A snapshot of G390R ITT motif:Lyn kinase complex structure from molecular dynamics simulations and molecular mechanics–generalized Born surface area (MM-GBSA) binding free energy (ΔG_{bind}) are shown. (**I**) Probability distributions of the distance between the hydroxyl group of catalytic residue Tyr^{385} and the γ-phosphate group of ATP in 100-ns simulations with WT or G390R ITT motif peptide docking to the active Lyn kinase domain in 50 to 100 ns of the simulations. The scale bars represent 1.6 μm in (C) to (F). Unpaired two-tailed Student's *t* tests in (B) to (D) and (F). Two-way ANOVA in (E). Red bars indicate means. NS, not significant. Data are representative of three independent experiments in (A) to (G).

dendritic cells, which then enhance general B cell activation and differentiation (*17*); (iii) IgG1$^+$ B cells may undergo class switching again to other IgG subclasses (*18*); (iv) IgG1$^+$ B cells may produce inflammatory cytokines such as interleukin-12 (IL-12) and tumor necrosis factor–α (TNF-α) to regulate the differentiation of other IgG-subclass plasma cells (*19*); and (v) excessive IgG1 antibodies may promote antigen presentation to activate B cells producing other IgG-subclass antibodies (*20*). Thus, we contend that IgG-BCR not only promotes enhanced memory responses but also leads to the development and relapse of autoimmune diseases when dysregulated.

REFERENCES AND NOTES

1. F. Martin, A. C. Chan, *Immunity* **20**, 517–527 (2004).
2. H. Wardemann *et al.*, *Science* **301**, 1374–1377 (2003).
3. J. F. Scheid *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18044–18048 (2011).
4. T. Tiller *et al.*, *Immunity* **26**, 205–213 (2007).
5. N. Engels *et al.*, *Nat. Immunol.* **10**, 1018–1025 (2009).
6. T. Kaisho, F. Schwenk, K. Rajewsky, *Science* **276**, 412–415 (1997).
7. S. W. Martin, C. C. Goodnow, *Nat. Immunol.* **3**, 182–188 (2002).
8. W. Liu, T. Meckel, P. Tolar, H. W. Sohn, S. K. Pierce, *Immunity* **32**, 778–789 (2010).
9. X. Chen *et al.*, *Nat. Commun.* **6**, 8552 (2015).
10. J. Lutz *et al.*, *Nat. Commun.* **6**, 8575 (2015).
11. N. Engels *et al.*, *Nat. Commun.* **5**, 5456 (2014).
12. J. Klarquist, E. M. Janssen, *J. Vis. Exp.* **105**, e53319 (2015).
13. T. Tiller *et al.*, *J. Exp. Med.* **207**, 2767–2778 (2010).
14. S. K. Pierce, W. Liu, *Nat. Rev. Immunol.* **10**, 767–777 (2010).
15. N. Miyano *et al.*, *Bioorg. Med. Chem. Lett.* **19**, 6557–6560 (2009).
16. H. Qi, *Nat. Rev. Immunol.* **16**, 612–625 (2016).
17. M. Guilliams, P. Bruhns, Y. Saeys, H. Hammad, B. N. Lambrecht, *Nat. Rev. Immunol.* **14**, 94–108 (2014).
18. A. Duhlin *et al.*, *J. Immunol.* **197**, 2618–2626 (2016).
19. F. E. Lund, *Curr. Opin. Immunol.* **20**, 332–338 (2008).
20. A. Bergtold, D. D. Desai, A. Gavhane, R. Clynes, *Immunity* **23**, 503–514 (2005).

ACKNOWLEDGMENTS

We thank S. K. Pierce, J. Lukszo, and K. Rajewsky for providing experimental materials. We thank Q. Z. Li and I. Raman in Microarray Core Facility, University of Texas Southwestern Medical Center, for support with the autoantigen microarray. We thank X. Wang, J. Wu, and Z. Wang for discussions. We thank P. Tolar, C. Wu, and D. Long for critical reading of this manuscript. **Funding:** This work is supported by funds from the National Natural Science Foundation of China (8182500030, 81730043, and 81621002 to W.L. and 31530020 to Z.L.), Ministry of Science and Technology of China (2014CB542500-03 to W.L. and 2014CB541901 to J.G.), Beijing Sci-Tech Program (Z171100000417007 to Z.L.), Beijing Nova Program (Z171100001117025 to X.S.), and Sanming Project of Medicine in Shenzhen (SZSM201612009 to Q.W.). **Author contributions:** Conceptualization, X.C. and W.L.; funding acquisition, W.L., Z.L., X.S., J.G., and Q.W.; investigation, X.C., B.Y., L.H., S.C., L.X., C.Y., W.Y., X.Zha., J.G., J.H., X.Zho., H.Z., Q.W., and H.S.; methodology, X.C., X.S., and W.Y.; resources, W.Z., Y.S., Z.C., Z.S., Z.H., F.Zha., F.Zhe., H.C., and J.L.; supervision, W.L., Z.L., and L.L.; writing–original draft, X.C. and W.Y.; and writing–review and editing, W.L., Z.L., X.C., B.Y., X.S., L.L., Z.Y., H.Q., and C.X. All authors contributed to revising the manuscript. **Competing interests:** The authors declare no financial or commercial conflicts of interest. **Data and materials availability:** All data described in this paper are present either in the main text or in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/700/suppl/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S4
References (*21–27*)

11 September 2017; resubmitted 10 April 2018
Accepted 19 September 2018
Published online 4 October 2018
10.1126/science.aap9310

PROTEIN DESIGN

De novo design of self-assembling helical protein filaments

Hao Shen[1,2,3]*, Jorge A. Fallas[1,2]*†, Eric Lynch[2]*, William Sheffler[1,2], Bradley Parry[4,5], Nicholas Jannetty[5,6], Justin Decarreau[7], Michael Wagenbach[7], Juan Jesus Vicente[7], Jiajun Chen[8,9], Lei Wang[9,10], Quinton Dowling[2,11], Gustav Oberdorfer[1,2]‡, Lance Stewart[1], Linda Wordeman[7], James De Yoreo[8,9], Christine Jacobs-Wagner[5,6,12], Justin Kollman[2], David Baker[1,2,13]†

We describe a general computational approach to designing self-assembling helical filaments from monomeric proteins and use this approach to design proteins that assemble into micrometer-scale filaments with a wide range of geometries in vivo and in vitro. Cryo–electron microscopy structures of six designs are close to the computational design models. The filament building blocks are idealized repeat proteins, and thus the diameter of the filaments can be systematically tuned by varying the number of repeat units. The assembly and disassembly of the filaments can be controlled by engineered anchor and capping units built from monomers lacking one of the interaction surfaces. The ability to generate dynamic, highly ordered structures that span micrometers from protein monomers opens up possibilities for the fabrication of new multiscale metamaterials.

Natural protein filaments differ considerably in their dynamic properties. Some, such as collagen, are relatively static, with turnover rates of several weeks (*1*–*4*), whereas others, such as cytoskeletal polymers, are dynamic, growing or disassembling in response to changing physiological conditions (*3*, *5*–*7*). The fraction of the total residue-residue interactions in the filament that are within (rather than between) the monomeric building blocks is generally higher for dynamic polymers; the monomers are usually independently folded structures rather than relatively extended polypeptides (Fig. 1A). Although peptide filaments in the first class have been successfully designed by staggering extended interaction motifs and generating end-to-end interactions between peptide coiled coils (*8*–*14*), the accurate computational design of reversibly assembling filaments built from folded protein monomers has remained an unsolved challenge. Much of the progress in recent years in the computational design of self-assembling nanomaterials has relied on building blocks with internal symmetry, which allows the generation of architectures with tetrahedral, octahedral, and icosahedral point group (*15*–*17*) and two-dimensional (2D) crystal space group (*18*) symmetry through the design of a single new protein-protein interface. By contrast, the building blocks in most reversibly assembling filaments have no internal symmetry, and hence multiple designed interfaces are required to drive the formation of the desired structure. The reduced symmetry also makes the sampling problem more challenging, as the space of possible filament geometries is very large.

To tackle the challenge of de novo designing dynamic protein filaments, we devised a computational approach that exploits the requirement for multiple intermonomer interfaces to reduce the size of the search space (Fig. 1C). Simple helical symmetry results from repeated application of a single rigid-body transform; we also consider architectures in which multiple such simple helical filaments are arrayed with cyclic symmetry. The search is therefore over the six rigid-body degrees of freedom and the discrete degrees of freedom associated with the different cyclic symmetries. The approach starts from an arbitrary asymmetric protein monomer structure and generates a second randomly oriented copy in physical contact by applying a random rotation (three degrees of freedom), choosing a random direction (two degrees of freedom), and sliding the second copy toward the first until they come into contact (Fig. 1C, left) (the sliding into contact effectively reduces the number of degrees of freedom from six for an arbitrary rigid-body transform to five). Successive monomers related by the filament-defining rigid-body transform need not themselves be in contact, and such arrangements are rare in biology. To go beyond this restriction, we consider not only filaments generated by the rigid-body transform relating the two contacting monomers, but also those generated by the nth root of this transform, where n ranges from two to five—with a choice of $n = 4$, for example, the first monomer will be in contact with the fourth monomer (Fig. 1C, bottom, and fig. S1). We also consider filaments with cyclic symmetry generated by the application of n-fold cyclic (C_n) symmetry operations around the superhelical axis, where n is between two and five (Fig. 1C, middle). In all cases, we then generate several repeating turns of the full filament by repeated application of the rigid-body transformation and cyclic symmetry operations, eliminate geometries with clashing subunits, and require the existence of at least one additional interface beyond that generated in the initial sliding-into-contact step. Filament architectures with multiple interacting surfaces predicted to have low energy after design (*19*) are selected, and Rosetta combinatorial sequence optimization is carried out on a central monomer, propagating the sequence to all other monomers. The resulting designs, which span the range of helical parameters (diameter, rise, and rotation) (table S1) of native filaments (Fig. 1B, blue dots), are filtered for high shape complementarity, low monomer-monomer interaction energy, and few or no buried unsatisfied hydrogen bonds.

We chose as the monomeric building blocks a set of 15 de novo–designed helical repeat proteins (*20*) (DHRs), which span a wide range of geometries and hence can give rise to a wide range of filament architectures. In addition to shape diversity, the DHRs have the advantages of very high stability and solubility and are likely to tolerate the substitutions needed to design the multiple interfaces required to drive filament formation. They can also be extended or shortened simply by the addition or removal of one or more of the 30- to 60-residue repeat units, potentially allowing tuning of the diameter of designed filaments. Starting from both the computational design models and the x-ray crystal structures of the DHRs, we generated 230,000 helical filament backbones as described above and selected 124 designs for experimental testing [we refer to these as de novo–designed helical filaments (DHFs); for comparison with filaments generated from native backbones, see fig. S2].

The designs were expressed in *Escherichia coli* under the control of a T7 promoter and purified by immobilized metal affinity chromatography (IMAC). Eighty-five of the designs were recovered in the IMAC eluate, whereas 22 were in the insoluble fraction (17 designs were not found in either fraction). IMAC eluates were concentrated, and filament formation was monitored by negative stain electron microscopy (EM); insoluble designs were characterized by EM either directly in the initial insoluble fraction or after solubilization in guanidine hydrochloride, IMAC, and subsequent removal of denaturant. A total of 34 designs

---

[1] Institute for Protein Design, University of Washington, Seattle, WA 98195, USA. [2] Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. [3] Molecular Engineering Ph.D. Program, University of Washington, Seattle, WA 98195, USA. [4] Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06511, USA. [5] Microbial Sciences Institute, Yale University, West Haven, CT 06516, USA. [6] Howard Hughes Medical Institute, Yale University, West Haven, CT 06516, USA. [7] Department of Physiology and Biophysics, University of Washington, Seattle, WA 98195, USA. [8] Department of Materials Science and Engineering, University of Washington, Seattle, WA 98195, USA. [9] Physical Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA. [10] State Key Laboratory of Crystal Materials, Shandong University, Jinan 250100, China. [11] Department of Bioengineering, University of Washington, Seattle, WA 98195, USA. [12] Department of Microbial Pathogenesis, Yale School of Medicine, New Haven, CT 06510, USA. [13] Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: jaf18@uw.edu (J.A.F.); dabaker@uw.edu (D.B.)
‡Present address: Institute of Biochemistry, Graz University of Technology, Graz, Austria.

(15 soluble and 19 insoluble) were found to form 1D nanostructures (figs. S3 and S4). A subset of the designs were synthesized as SUMO (small ubiquitin-like modifier) fusions to prevent premature filament formation; the SUMO tag was removed by using SUMO protease, and the samples were characterized by negative stain EM (fig. S5).

We chose six designs with a range of model architectures and longer persistence lengths for higher-resolution structure determination by cryo-EM. We determined the filament structures and refined helical symmetry parameters by using iterative helical real-space reconstruction in SPIDER (*21*, *22*), followed by further 3D refinement in Relion (*23*) and Frealign

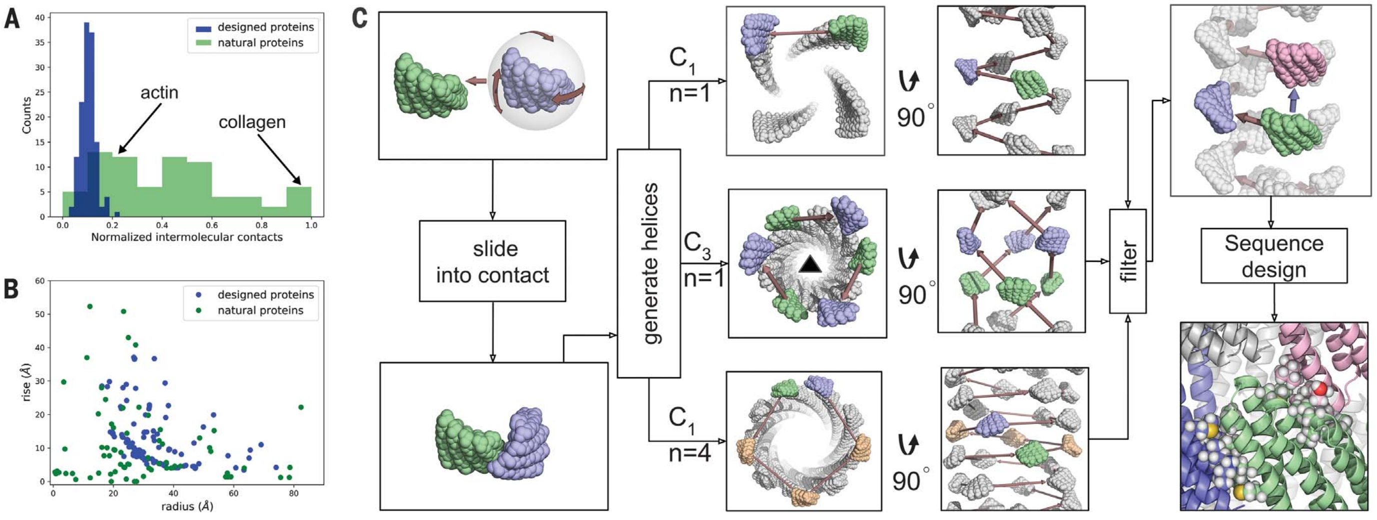
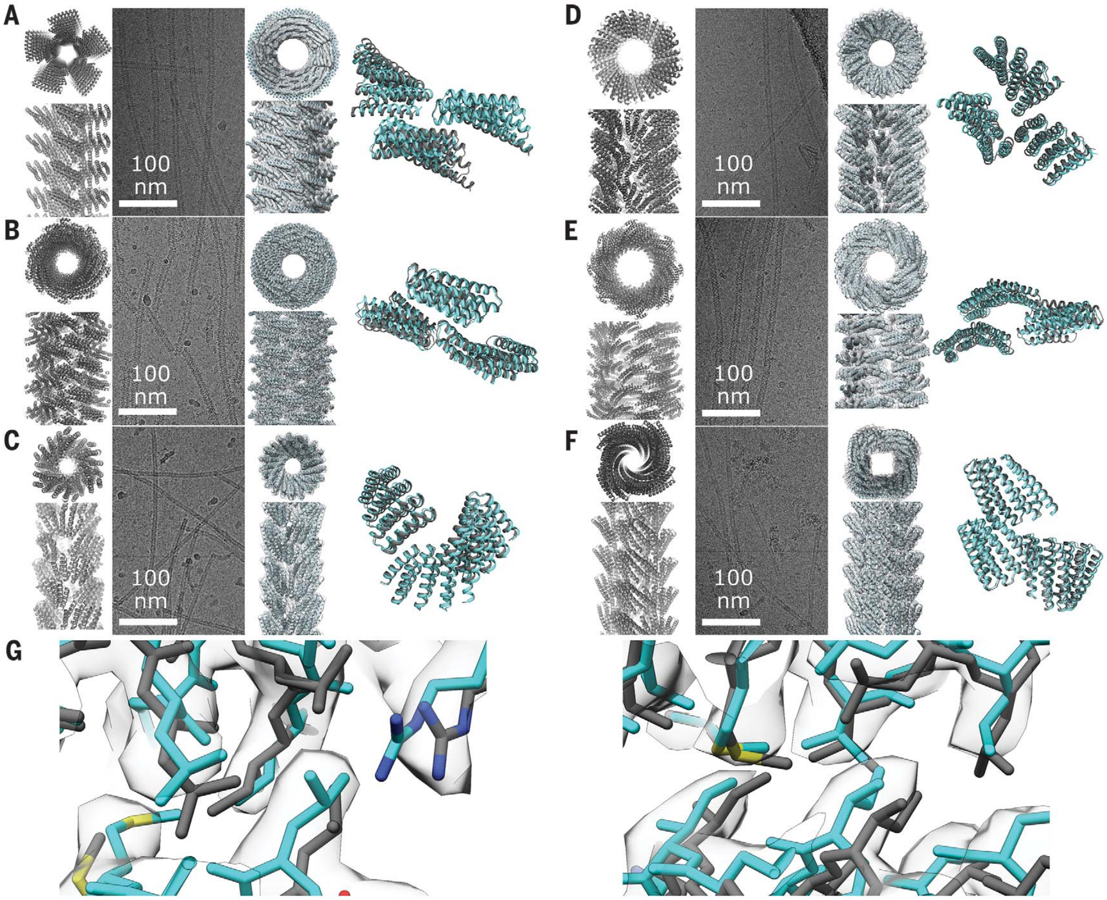


Fig. 1. Filament architectures and computational design protocol. (**A** and **B**) Comparison of properties of our designed filaments (blue) with those of native filaments (green). (A) The fraction of total residue-residue interactions between (rather than within) monomers. (B) Superhelical parameters. (**C**) Computational design protocol.

Fig. 2. Cryo-EM structure determination. (**A** to **F**) (Left to right) Computational model, representative filaments in cryo-EM micrographs, cryo-EM structure, and overlay between the model and structure for (A) DHF58 (RMSD, 3.3 Å), (B) DHF119 (RMSD, 2.3 Å), (C) DHF91 (RMSD, 1.2 Å), (D) DHF46 (RMSD, 2.3 Å), (E) DHF79 (RMSD, 4 Å), and (F) DHF38 (RMSD, 0.9 Å). (**G**) Close-up views of the two main intermonomer interfaces in the filament for DHF119, with the computational model (gray) and cryo-EM structure (cyan) in sticks in the helical reconstruction density (3.4-Å resolution). The high-resolution structure of design DHF119 is very close to the design model.



(*24*). In all six cases, the overall orientation and packing of the monomers in the filament were similar in the experimentally determined structures and design models, but the accuracy with which the details of the interacting interfaces were modeled varied considerably (Fig. 2 and fig. S6). Subtle shifts in the interaction interfaces in several cases altered the designed symmetry; DHF119, for example, was designed to be C_1, but the cryo-EM structure has C_3 symmetry (helical lattice plot comparisons are in fig. S7). Four of the six designed filaments matched the computational models at near-atomic resolution: For DHF38 and DHF91, the experimentally observed rigid-body orientation was nearly identical to that of the design models [0.9- and 1.2-Å root mean square deviation (RMSD) over three chains containing all unique interfaces]; for DHF46 and DHF119, the RMSD over three chains was 2.3 Å, and for DHF91 and DHF58, 3.6 and 4 Å. The structure of DHF119 was solved to 3.4-Å resolution; the backbone and side-chain conformations at the subunit interfaces are very similar to those in the design model (Fig. 2G).

To determine whether the filament diameter could be modulated by changing the number of repeat units in the monomer, we generated a series of DHF58 variants that retain the fiber interaction interfaces but have three, four, five, or six repeats in the protomer. The designs were expressed, purified, and characterized by negative stain EM. Consistent with the computational models (Fig. 3A), the diameter of the filaments changes linearly with the number of repeat units (Fig. 3, B and C).

We monitored assembly dynamics in vitro by solution scattering and in living cells by using fluorescence microscopy with monomers fused to green fluorescent protein (GFP). The extent and kinetics of DHF119 filament formation in vitro were strongly concentration dependent. Filament nucleation was too fast to observe by manual mixing; the rate of the observed elongation phase was linear with respect to the monomer concentration, and extrapolation of the plateau values from progress curves back to zero yielded a critical concentration of 3 μM (fig. S8). Upon dilution below the critical concentration, filaments disassembled in several hours (fig. S9). In *E. coli* after the induction of expression of DHF58-GFP, discrete puncta were first observed, which over time resolved into filaments up to micrometers in length (Fig. 4A and movie S1) (the puncta may simply be filaments below the resolution limit of the microscope, ~250 nm). The filaments formed in vivo have high stability: After bacteria expressing the monomer were lysed by lysogenic phages, the filaments largely retained their shape (movie S2).

Natural systems achieve complexity and diversity of filament-based structures through modulating the nucleation, growth, and cellular location of the polymers. In some natural systems, nucleation and location are controlled by complexes that act as templates that initiate new growth and anchor filaments to specific locations, such as the γ-tubulin ring complex for microtubules and the Arp2/3 complex for actin. We sought to replicate this mechanism of control by designing multimeric anchor constructs, with multiple monomeric subunits held close to the relative orientations in the corresponding filaments by a fusion to designed homo-oligomers with the appropriate geometry (fig. S10) (one of the interaction interfaces is eliminated to restrict fiber growth to one direction). For example, anchor DHF119_C6 (Fig. 4B) is a hexamer in which each monomer consists of a designed oligomerization domain fused to the fiber monomer; the orientations of the monomers in the hexamer are close to those in the filament structure to promote both nucleation and fiber attachment. To study the kinetics of filament formation in vitro in more detail, we attached the anchors to glass slides, added monomers fused to yellow fluorescent protein (YFP), and monitored fiber formation by total internal reflection fluorescence microscopy. The anchors seeded the rapid growth of multiple-micrometer-length fibers over 30 min (Fig. 4C and movie S3; for analysis of growth kinetics of a second fiber, see fig. S11). Few or no fibers were observed to grow from the glass slide surface when it was coated with an anchor designed for a different fiber (movie S4) or with no anchor at all (movie S5). The attachment of a biotinylated anchor to streptavidin-coated beads, followed by incubation with a filament monomer, resulted in an extensive network of filaments emanating from the beads (Fig. 4D, left); by contrast, very few filaments were observed around control beads that lacked the anchor protein (Fig. 4D, right).

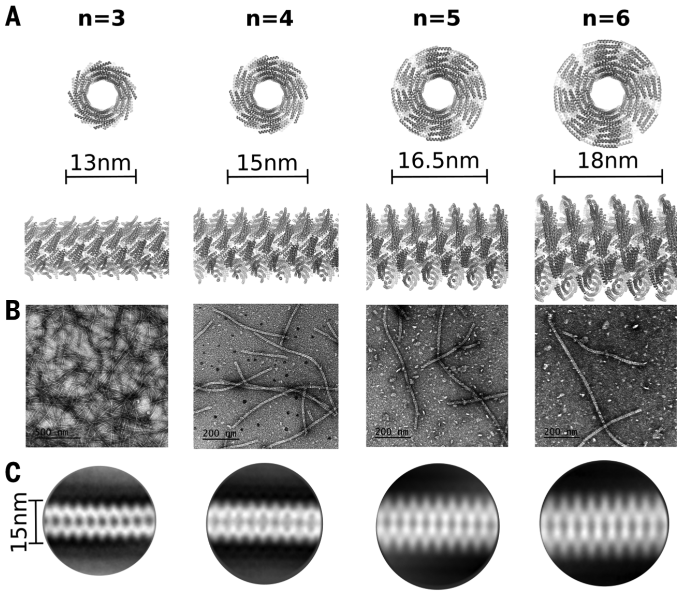


Fig. 3. Modular tuning of fiber diameter. DHF58 filament variants with different numbers of repeats were characterized by EM. (**A**) Cross sections and side views of computational models based on the four-repeat cryo-EM structure. The number of repeats (*n*) is shown at the top. (**B**) Negative stain electron micrographs. (**C**) 2D-class averages.

To determine whether filament dissolution could also be modulated by designed accessory proteins, we produced monomeric capping units lacking one of the two designed interfaces in the DHF119 filament. These caps are expected to add to one end of the filament but not the other, preventing further elongation (because the two ends of the filaments are distinct, there are two types of caps). The addition of increasing concentrations of the caps to already formed filaments resulted in shrinking and ultimate disappearance of the filaments (fig. S12), suggesting that filaments are dynamically exchanging protomers at equilibrium. Monitoring of cap-induced disassembly by atomic force microscopy (AFM) showed that fibers incubated with equal concentrations of the protomers and single-end caps disassemble primarily from one (presumably the uncapped) end whereas, in the presence of both caps, disassembly occurs from both ends (Fig. 4E and fig. S12). In the absence of caps, increasing the monomer concentration led to growth from both ends of the fibers at a rate (~15 nm/min per end at 18 μM monomer) (fig. S12) similar to that observed by fluorescence for anchored fiber growth (8.4 nm/min at 18 μM monomer) (fig. S11). The observed behavior can be understood as follows. At the critical monomer concentration where fibers neither grow nor shrink, the (concentration-dependent) rate of monomer addition to the ends is balanced by the (concentration-independent) disassociation rate. Caps perturb this balance by complexing with monomers in solution (fig. S13, bottom), effectively reducing the free monomer concentration; thus, when both end caps are present, disassembly wins out over growth, leading to a net shrinking of the filaments. When one cap is present, the net rate of subunit addition is greater at the end where both free monomers

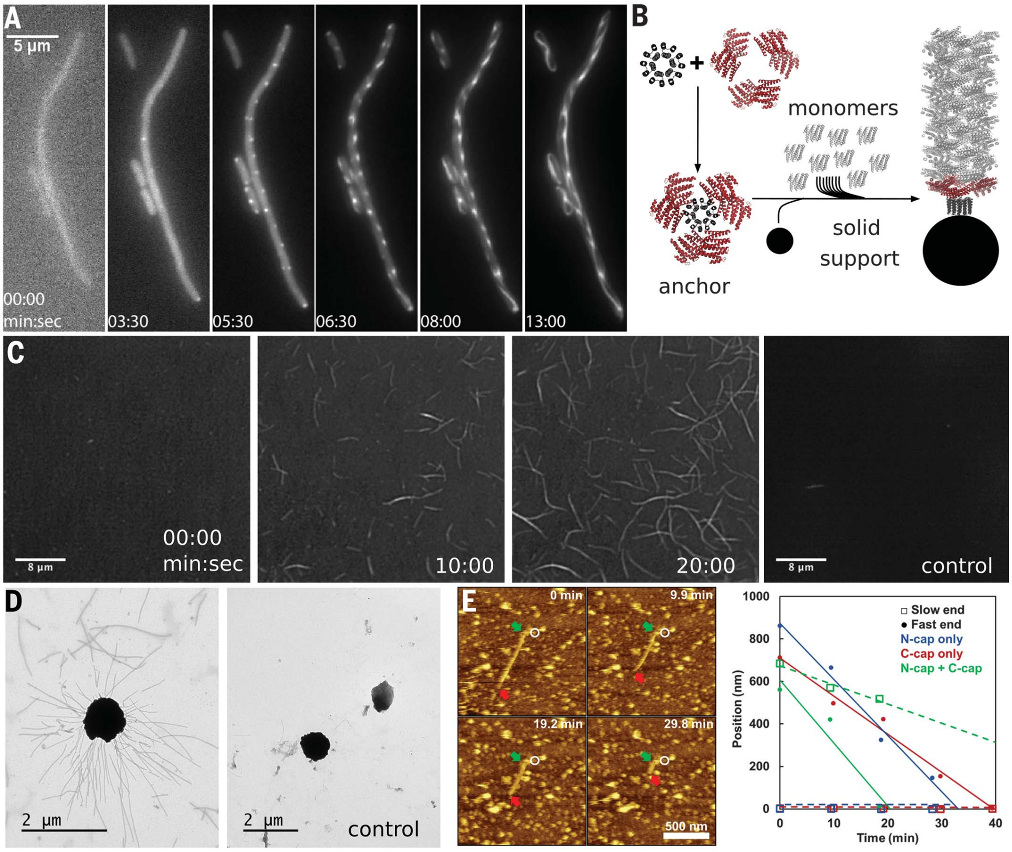


Fig. 4. Characterization of fiber growth and disassembly. (**A**) Kinetics of the assembly of DHF58-GFP filaments in *E. coli*. (**B**) Construction of fiber anchors holding monomers in the rigid-body arrangement found in the filament. (**C**) Kinetics of DHF119-YFP filament assembly in vitro on a glass surface coated with the DHF119_C6 anchor. In the control panel, the glass surface was coated with the noncognate DHF91 anchor. (**D**) DHF119 filaments emanating from DHF119_C6 anchor–coated magnetic beads incubated with the monomer. Beads on the right lack the anchor. (**E**) Disassembly of DHF119 fibers in the presence of capping units monitored by in situ AFM. (Left) Image sequence showing disassembly in the presence of N caps. The white circles mark a fixed position in all images. (Right) Positions of fiber ends versus time in solutions with N caps, C caps, and N caps plus C caps. In all cases, the DHF119 monomer concentration and the total cap concentration are each 3.8 μM (at this concentration of monomer, fibers neither dissolve nor grow in the absence of caps). Because they lack one of the filament interfaces, caps can bind only to one end; disassembly from this end will be slower, as the combined on rate of caps and monomers is greater than the on rate of monomers alone at the other end.

and free caps can add (fig. S13, top right) than at the other end, where only monomers can add (fig. S13, top left). Because the rate of monomer dissociation is the same at both ends, the fibers shrink primarily from one end, as observed.

The ability to program micrometer-scale order from angstrom-scale designed interactions between asymmetric monomers is an advance for computational protein design. In contrast to previous nanomaterial design efforts relying on an already existing interface within symmetric building blocks, proper assembly requires the design of two independent interfaces. The introduction of a small number of hydrophobic substitutions near the periphery of dihedral complexes can promote stacking into extended filaments because each sequence change is replicated multiple times at the stacking interface (*25*); the filaments described here are instead built from monomeric building blocks and have a much wider range of geometries because only a small fraction of possible helical assemblies contain dihedral point group symmetry. Both designed interfaces were accurately recapitulated in four of the six structures solved by cryo-EM; despite the deviations in the interfaces in the other two, the overall filament architecture was reasonably well recapitulated. The ability to program filament dynamics provides a baseline for understanding the much more complex regulation of the dynamic behavior of naturally occurring filaments. The repeat protein building blocks are hyperstable proteins robust to genetic fusion, and therefore, the designed filaments provide readily modifiable scaffolds to which binding sites for other proteins or metal nanoclusters can be added for applications ranging from cryo-EM structure determination to nanoelectronics.

REFERENCES AND NOTES

1. S. Ricard-Blum, F. Ruggiero, M. van der Rest, in *Topics in Current Chemistry*, vol. 247, *Collagen*, J. Brinckmann, H. Notbohm, P. K. Müller, Eds. (Springer, 2005), pp. 35–84.
2. L. C. Serpell, *Biochim. Biophys. Acta* **1502**, 16–30 (2000).
3. H. Herrmann, U. Aebi, *Annu. Rev. Biochem.* **73**, 749–789 (2004).
4. G. J. Rucklidge, G. Milne, B. A. McGaw, E. Milne, S. P. Robins, *Biochim. Biophys. Acta* **1156**, 57–61 (1992).
5. K. C. Holmes, D. Popp, W. Gebhard, W. Kabsch, *Nature* **347**, 44–49 (1990).
6. E. Nogales, M. Whittaker, R. A. Milligan, K. H. Downing, *Cell* **96**, 79–88 (1999).
7. B. Bhyravbhatla, S. J. Watowich, D. L. D. Caspar, *Biophys. J.* **74**, 604–615 (1998).
8. A. M. Smith *et al.*, *Angew. Chem. Int. Ed. Engl.* **44**, 325–328 (2004).
9. L. E. R. O'Leary, J. A. Fallas, E. L. Bakota, M. K. Kang, J. D. Hartgerink, *Nat. Chem.* **3**, 821–828 (2011).
10. C. J. Bowerman, B. L. Nilsson, *Biopolymers* **98**, 169–184 (2012).
11. J. D. Hartgerink, J. R. Granja, R. A. Milligan, M. R. Ghadiri, *J. Am. Chem. Soc.* **118**, 43–50 (1996).
12. E. H. Egelman *et al.*, *Structure* **23**, 280–289 (2015).
13. N. C. Burgess *et al.*, *J. Am. Chem. Soc.* **137**, 10554–10562 (2015).
14. C. Xu *et al.*, *J. Am. Chem. Soc.* **135**, 15565–15578 (2013).
15. F. A. Tezcan, in *Coordination Chemistry in Protein Cages: Principles, Design and Applications*, T. Ueno, Y. Watanabe, Eds. (Wiley, 2013), pp. 149–174.
16. Y. Hsia *et al.*, *Nature* **540**, 150 (2016).
17. N. P. King *et al.*, *Nature* **510**, 103–108 (2014).
18. S. Gonen, F. DiMaio, T. Gonen, D. Baker, *Science* **348**, 1365–1368 (2015).
19. J. A. Fallas *et al.*, *Nat. Chem.* **9**, 353–360 (2017).

20. T. J. Brunette *et al*., *Nature* **528**, 580–584 (2015).
21. E. H. Egelman, *J. Struct. Biol.* **157**, 83–94 (2007).
22. E. H. Egelman, *Ultramicroscopy* **85**, 225–234 (2000).
23. S. H. W. Scheres, *J. Struct. Biol.* **180**, 519–530 (2012).
24. N. Grigorieff, *J. Struct. Biol.* **157**, 117–125 (2007).
25. H. Garcia-Seisdedos, C. Empereur-Mot, N. Elad, E. D. Levy, *Nature* **548**, 244–247 (2017).

ACKNOWLEDGMENTS

We thank C. Bahl for providing vector pCDB24, G. Ueda for assistance with protein expression and purification techniques, C. Li for help with inspecting design models, and Y. Li for help with screening insoluble fractions. We thank the Arnold and Mabel Beckman Cryo-EM Center at the University of Washington for access to electron microscopes. **Funding:** This work was supported by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program (D.B.); the Howard Hughes Medical Institute (D.B., H.S., W.S., N.J., C.J.-W.); Defense Advanced Research Projects Agency (DARPA) grant W911NF-17-1-0318 (J.A.F.); Marie Curie postdoctoral research fellowship grant PIOF-GA-2012-332094 (G.O.); the U.S. Department of Energy, Basic Energy Sciences, Division of Materials Science and Engineering, Biomolecular Materials Program (J.D.Y); GM069429 from the National Institutes of Health (L.Wo.); and the Biomolecular Materials Program and Washington State (L.S.). **Author contributions:** H.S. carried out design calculations. J.A.F., W.S., H.S., L.S., and D.B. developed the protein design method. H.S. and J.A.F. expressed and purified the designed proteins. H.S. carried out the screening of designs and tuning of fiber diameter and performed anchoring experiments by negative stain EM. E.L., H.S., and J.K. carried out the cryo-EM data acquisition and structure determination. B.P., N.J., and C.J.-W. carried out in vivo fluorescence microscopy. J.D., M.W., H.S., J.J.V., and L.Wo. carried out in vitro fluorescence measurements. J.C., L.Wa., and J.D.Y. performed AFM measurements. H.S. and Q.D. assessed assembly kinetics by using ultraviolet scattering. G.O. contributed the oligomeric scaffold for anchor design. All authors discussed results and commented on the manuscript. **Competing interests:** H.S., J.A.F., and D.B. are inventors on U.S. provisional patent application 62/750,435, submitted by the University of Washington, which covers the compositions and uses of the self-assembling helical protein filaments. **Data and materials availability:** The cryo-EM structures and atomic models were deposited in the Electron Microscopy Data Bank and the Protein Data Bank (PDB) with the following accession codes (table S2): DHF119, EMD-9021, and PDB 6E9Z; DHF38, EMD-9020, and PDB 6E9Y; DHF58, EMD-9017, and PDB 6E9T; DHF46, EMD-9016, and PDB 6E9R; DHF79, EMD-9018, and PDB 6E9V; and DHF91, EMD-9019, and PDB 6E9X. All other data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6415/705/suppl/DC1
Materials and Methods
Figs. S1 to S13
Tables S1 and S2
References (*26*–*37*)
Movies S1 to S5
Data S1

5 June 2018; accepted 29 September 2018
10.1126/science.aau3775

MWIR Camera for Real-Time Thermal Analysis

The A6750sc MWIR (midwave-infrared) camera from FLIR Systems is suitable for recording fast thermal events due to its short exposure times and high frame rates. With its cooled InSb (indium antimonide) camera, accurate temperature measurements can be carried out on fast-moving objects without motion blur, and a variety of nondestructive tests can also be performed. The FLIR A6750sc operates in the 3.0 µm–5.0 µm wavelength range, or 1.0 µm–5.0 µm when the broadband option is chosen, enabling precision measurements down to the short-wave infrared (SWIR) range. Its camera produces razor-sharp, detailed thermal images that are ideal for electronics inspections or material tests. Offering short integration times (<1 ms) and fast frame rates (to 125 Hz), the A6750sc can routinely image dynamic thermal processes on surfaces at room temperature. The maximum frame rate in the smallest frame mode is 4.1 kHz. The A6750sc offers the versatility of plug-and-play with third-party software due to its GigEVision and GenICam compatibility.

FLIR Systems
For info: 800-254-0630
www.flir.co.uk/products/a6750sc-mwir

3D/4D Stage

Applied Scientific Instrumentation's (ASI's) new compact 3D/4D stage is a precise, motorized motion-control system designed to move samples around fixed optics. It incorporates three ASI linear stages and an optional motorized rotating stage employed for a theta axis. The linear stages comprising the XYZ elements offer travel options of 25, 50, 100, or 200 mm, and each axis can be chosen separately. The linear stages derive smooth, accurate motion from closed-loop direct current servomotors, crossed-roller bearings, high-precision lead screws, and high-resolution encoders for positioning feedback. Like other ASI stages, they can move uniformly at extremely slow speeds for in-motion acquisition. All the stages offer various speed/accuracy options for a more customized 3D/4D system. The stage elements are rigidly attached together and usually mounted to a breadboard via an adapter plate.

Applied Scientific Instrumentation
For info: 800-706-2284
www.asiimaging.com

Laser Confocal Scanning Microscope

The Olympus LEXT OLS5000 3D laser confocal scanning microscope delivers precise imaging in a fast, easy-to-use platform for R&D and quality control inspection in the automotive, electronic component, and semiconductor industries. 4K scanning technology and optics designed specifically for the OLS5000 enable detection of near-perpendicular features and small steps at near-nanoscale levels. The instrument acquires data quickly and improves user experience with intuitive software designed to automate many common settings. An expansion frame and a dedicated, long working-distance lens perform precise measurements on samples up to 210 mm in height and concavities up to 25 mm deep—even those with uneven surface cracks. The result is simple, accurate, noncontact 3D measurement of a wide variety of samples.

Olympus
For info: 800-225-8330
www.olympus-ims.com

Field Emission Scanning Electron Microscope

With the ZEISS GeminiSEM 450 field emission scanning electron microscope (SEM), users benefit from high resolution, surface-sensitive imaging, and an optical system that obtains the best analytical results—especially when working with low voltages. High-throughput electron backscatter diffraction (EBSD) analysis and low-voltage energy-dispersive X-ray spectroscopy (EDS) deliver superior performance due to the instrument's ability to precisely and independently control spot size and beam current. With the Gemini 2 design, the user can always work under optimized conditions, switching seamlessly between imaging and analytical modes at the touch of a button. In addition, the 450 can handle a wide range of sample types, from classical conductive metals to beam-sensitive polymers. Its variable pressure technology reduces charging on nonconductive samples without compromising Inlens detection capabilities, simultaneously enabling high-resolution EDS analysis by minimizing the skirt effect. The ZEISS GeminiSEM 450 is a flexible instrument suited to a variety of applications in materials science, industrial labs, and life sciences.

ZEISS
For info: 800-233-2343
www.zeiss.com

Inverted Confocal Raman Microscope

The alpha300 R*i* inverted Raman microscope from WITec combines the advantages of data acquisition from below with the established merits of 3D confocal Raman imaging—a powerful, versatile technique that can chemically characterize samples nondestructively and without labeling or other specialized preparation. Specimens in aqueous environments, such as cell cultures, can be examined more effectively. Standardized liquid-sample holder formats can be quickly mounted and measured, accelerating experimental workflow and helping ensure consistency. Materials science investigations will be aided by the very large working area that can accommodate bulky samples and the set focal plane. The motorized sample stage also facilitates the mounting of environmental enclosures and other accessories. Many modular components and upgrade possibilities developed for the alpha300 series are compatible with the R*i* version. Other microscopy techniques associated with inverted microscopes, such as fluorescence, differential interference contrast, and phase contrast, can also be easily integrated.

WITec
For info: 865-984-4445
www.witec.de

Autofluorescence Quenching Kit

The TrueVIEW Autofluorescence Quenching Kit removes unwanted fluorescence in formalin-fixed paraffin embedded (FFPE) tissue samples, blood cells, and structural elements such as collagen and elastin. Autofluorescence often impairs or prevents the use of immunofluorescence assays when the specific stained antigen cannot be distinguished from the interfering background signal. This is particularly problematic with tissue samples preserved using the FFPE method. TrueVIEW specifically targets and eliminates background autofluorescence, retaining the intended specific fluorescent staining and enabling researchers to identify specific markers that cannot be seen otherwise. It is effective across the spectral range from blue to far-red, and is compatible with commonly used fluorophores. Applying TrueVIEW to sections is fast, requiring just 5 min for incubation. Ethanol steps and dilution are not required. The Vector TrueVIEW kit includes VECTASHIELD HardSet Antifade Mounting Medium and contains enough reagent for between 100 and 150 tissue sections.

Vector Laboratories
For info: 800-227-6666
vectorlabs.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

"SciLifeLab is like a Mecca for Swedish research"

SciLifeLab, Science for Life Laboratory, is a Swedish national center for molecular biosciences. To further strengthen the research environment, the institution regularly recruits young, talented research leaders to become SciLifeLab fellows.

Each fellow is recruited by one of the center's host universities and also receives funding from them. One of the 22 fellows currently enrolled in the program is Claudia Kutter, who was recruited from CRUK, Cambridge, by Karolinska Institutet/SciLifeLab in 2016.


Photo: John Senett

Claudia Kutter

"I was actively looking for research institutions where I would be embedded in a vivid and international scientific setting with an active young faculty", says Claudia Kutter.

"Each of us brings a unique line of research and expertise to the center. I appreciate the continuous exchange of experience, the interdisciplinary research approaches that can come together in a synergistic manner, and the formation of interest groups on special research topics."

Potential for future cancer treatment

Her research focuses on genome, chromatin and RNA biology in the context of human disease.

"I am intrigued by the fact that we have one genome that gives rise to different transcriptomes, which then generate functionally diverse cell types", she says.

In particular, her group wants to understand the cellular and molecular defects underlying noncoding RNAs by using a combination of computational and experimental methods. Many of these noncoding RNA molecules have been disregarded as transcriptional noise or by-products but, in fact, harbour tremendous regulatory potential since they can interact with DNA, other RNA molecules and proteins.

"In a sense, a noncoding RNA is like a molecular Swiss army knife with various tools all stowed inside one unit", she says. "Besides contributing to novel insights, our long-term goal is to incorporate noncoding RNAs in biomedical research and therapeutics."

Using genome-engineering methods, Claudia Kutter and her colleauges remove regulatory sequences from DNA entirely, decrease or enhance the expression levels and assess the phenotypical consequences. For example, they enhanced the expression of a noncoding RNA, which stopped cancer cells from proliferating whereas normal cells remained unchanged.

"Noncoding RNAs might become essential for future therapeutic development since we would only target cancer cells and not healthy cells. This could lead to more refined treatments and prevent many of the systemic effects of classical treatment strategies", she explains.

For the future, she wants to ensure that the conditions for her group remain optimal and constructive to allow innovations and embark on interdisciplinary research projects.

"SciLifeLab truly offers an attractive research environment, has state-of-the-art instrumentation and core facilities, and is continuously increasing the research activities - it is like a Swedish Mecca for research. It is beneficial to be in a place without institutional boundaries."

Advancing life sciences

SciLifeLab is both a research institution and a national resource for Swedish researchers with the mission to develop, use and provide advanced scientific technologies and expertise. The center is a joint effort by Karolinska Institutet, KTH Royal Institute of Technology, Stockholm University and Uppsala University. Founded in 2010, SciLifeLab today encompasses more than 60 research groups and 40 infrastructure units across the country.

SciLifeLab

www.scilifelab.se

Creating partnerships to strengthen ocean science


Professor Lixin Wu

The world's oceans hold essential and important reserves of natural resources, critical for the continued and sustainable survival of humanity. These oceans are now confronted with serious environmental challenges due to human activity and global climate change, potentially leading to significant socioeconomic consequences. The international community continues to call for the creation of innovation partnerships to explore effective solutions for the conservation and sustainable use of the oceans and their resources. The field of marine science offers a vitally important foundation for the implementation of these goals, but suffers from an imbalance in global capacity building. In an effort to bring together marine research institutes from around the world to tackle the challenges confronting marine science, the Pilot National Laboratory for Marine Science and Technology (Qingdao) (QNLM) launched the Global Ocean Summit (GOS).

The first GOS was held in Qingdao in September 2016, with the primary purpose of providing a forum for leaders from marine research institutes, universities, and other international organizations to share their ideas and experiences related to ocean science and technology strategies, and to promote collaborative innovation. The director of QNLM, Lixin Wu, as chair of the local organizing committee, put forward a declaration that called for enhancing marine science and technology innovation and creating a shared community of concerned scientists who can assess and address the impact of ocean changes. The GOS 2018 participants believe that these discussions are critical to ensure a secure and sustainable future for the peoples of the world.

In the spirit of the GOS 2016 Declaration, QNLM worked with the Department of Science and Technology of Shandong Province and *Science*/AAAS to organize the second GOS, which took place in July 2018, also in Qingdao. Participating in the GOS 2018 conference were 152 leaders and representatives from 101 marine research institutes and universities in 24 countries, as well as five international organizations involved in ocean research. The theme of the conference was "Enhancing Partnerships on Ocean Observation and Research," and it included in-depth conversations and vigorous discussions covering many challenging issues related to ocean research. The summit drew much interest and attention from global institutions involved in marine-related research and policy, particularly the discussion around topics such as enhancing global collaboration in ocean observation and prediction, polar-seas research, deep-sea research, and ocean sustainability. The summit released the GOS 2018 Recommendations (see facing page), which advocated strengthening cooperation in capacity building; promoting activities such as ocean monitoring, exploration, sustainable utilization, protection, event forecasting, and management; and tackling the challenges presented by global climate change.

The next GOS is scheduled for 2020. In the interim, it is expected that the previous two GOS gatherings will encourage long-term, stable dialogue among the leaders of global marine-related institutions and organizations, and promote the development of new international programs in ocean research.

PHOTOS: QNLM

GOS

GLOBAL OCEAN SUMMIT 2018

Enhancing Partnerships on Ocean Observation and Reseach

Qingdao • China 3-5 July, 2018

GOS 2018 participants

Recommendations from the Global Ocean Summit 2018

Global environmental changes resulting from human activity pose unprecedented challenges to the Earth's ecological systems and, most importantly, to the sustainable development of the world's oceans. In response, we, as GOS 2018 participants, firmly believe that it is the responsibility of the world's marine institutions and universities to deepen our systematic understanding of the oceans in order to tackle these global challenges.

Jointly, we make the following recommendations:

1. We propose to adhere to the spirit of the "Global Ocean Summit 2016 Declaration" and strive for "Enhancing Partnerships on Ocean Observation and Research," to implement United Nations (UN) Sustainable Development Goals, and to support initiatives announced at the UN Ocean Conference 2017. We will do this by establishing a joint global network for sharing knowledge and experience. The network will encourage society's rights and equal access to marine knowledge, develop marine research capacity through sharing of observational and educational resources, and promote the efficient use of megafacilities and infrastructure. The network will actively encourage participation by intergovernmental and nongovernmental organizations.

2. We will collaborate to advance ocean research for the benefit of humanity.

2.1. We are committed to supporting international ocean observation through the sharing of open-access data from international ocean observation programs; boosting research and development of new technologies for ocean observation; ensuring data quality control and assurance; and facilitating the calibration and standardization of key ocean parameter measurements while also forging collaborations to create a sustainable, vertically integrated observation system, encompassing the ocean floor, ocean, and space.

2.2. We will encourage deeper integration of the observation system with high-resolution models to improve ocean environment forecasting and the development of early warning systems to combat extreme weather events and marine ecological disasters.

2.3. We will strengthen research collaboration in polar research, focusing on sea-ice changes and permafrost melting, global sea-level change, and ocean acidification, to improve our understanding of polar climate change and its impact on the oceans.

2.4. We will promote interdisciplinary research of the deep ocean to improve our understanding of ocean temperature and carbon uptake as well as the deep-ocean ecological environment, through the development of deep-ocean, onsite, remote-controlled monitoring technologies.

3. We will encourage our governments to support international ocean observation. We will promote engagement with policy-makers, shareholders, students, and the general public to involve every individual in the development of ocean science for the well-being of humanity.

4. In future Global Ocean Summits, we will explore means to enhance international cooperation in other areas of marine research and monitoring of importance to humanity, including but not limited to coastal development, environmental research on regional seas, and the sustainable utilization of ocean resources (also known as the "blue economy").

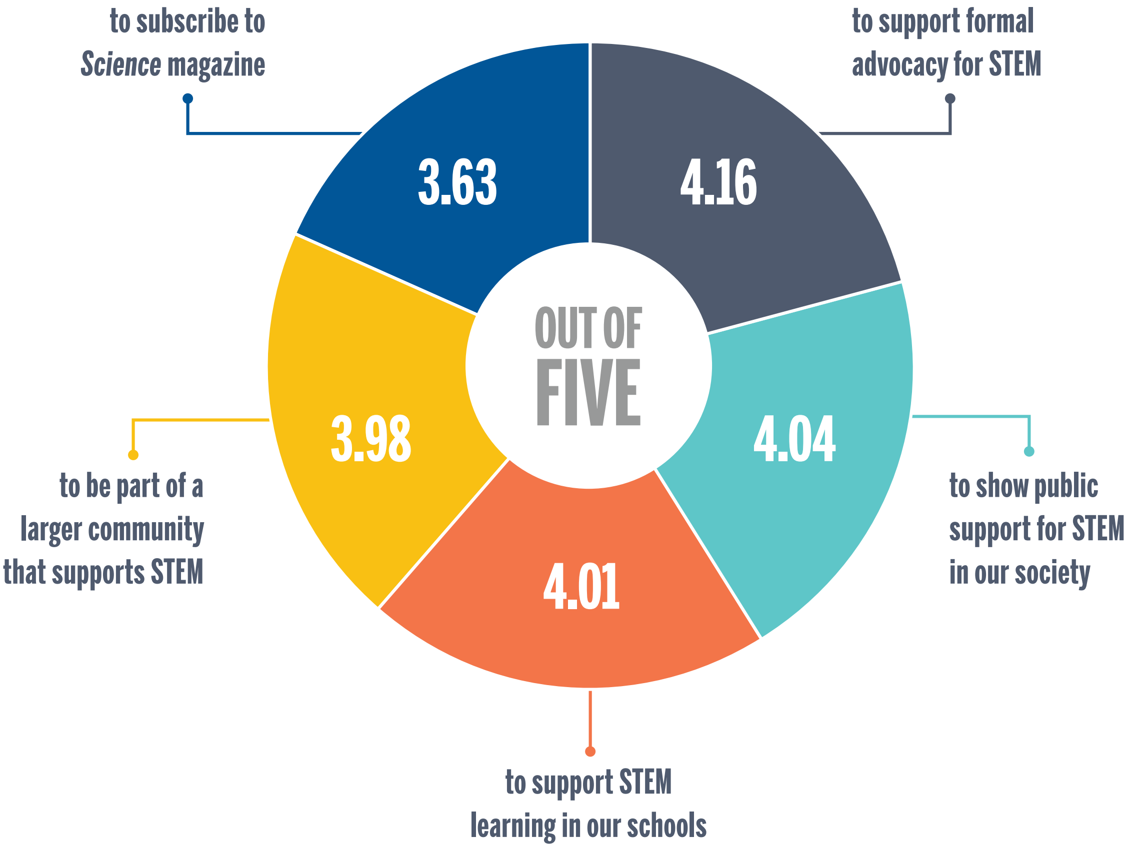
We stand by the mantra: "One earth, one ocean." We commit to contributing our knowledge and wisdom for the continued vitality and abundance of the world's oceans and the maintenance of a sustainable marine environment for all of humanity.

青岛海洋科学与技术试点国家实验室
Pilot National Laboratory for Marine Science and Technology (Qingdao)

PHOTO: QNLM

AAAS IS THE FORCE FOR SCIENCE

According to the 2017 Member Survey, you joined AAAS ...



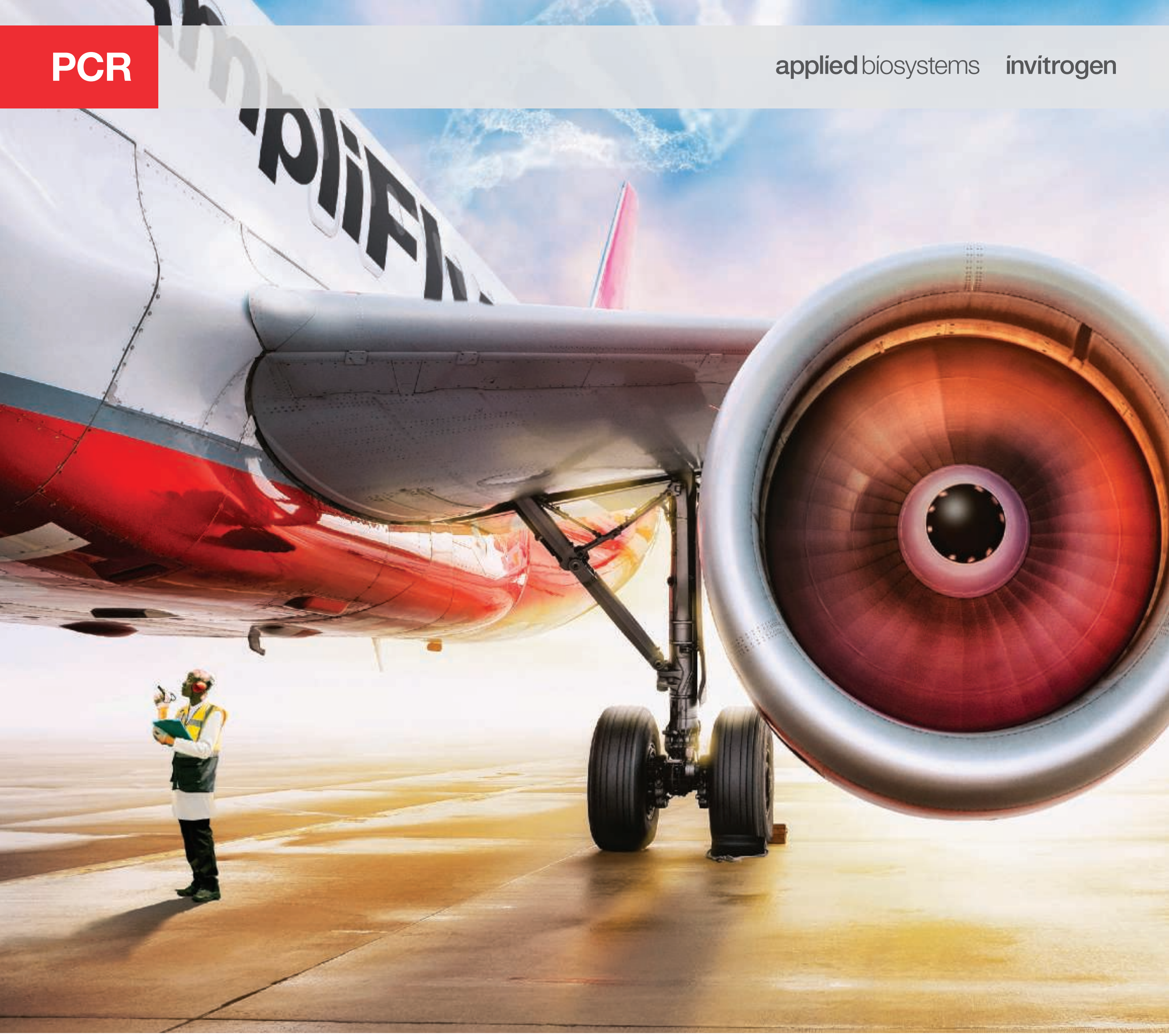
TELL US WHAT'S IMPORTANT TO YOU!

The 2018 Member Survey is launching in September. Look in your inbox for a link.

Your responses help us to better serve science, scientists, and the global community. Don't miss your chance to tell us what's most important to you!

AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE



Science
Careers

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 AAAS

ScienceCareers.org

CASE WESTERN RESERVE
UNIVERSITY EST. 1826

TENURED FACULTY POSITION IN CANCER RESEARCH

Case Western Reserve University School of Medicine
Case Comprehensive Cancer Center

The Case Comprehensive Cancer Center (**http://cancer.cwru.edu/**), an NCI-designated Comprehensive Cancer Center, at CWRU, with affiliates University Hospitals Case Medical Center and Cleveland Clinic, invites applications for tenured faculty positions at the level of Associate or Full Professor. Applicants are being sought that have basic/translational research expertise in one of the following areas of cancer focus: Tumor Immunology, Drug Discovery, Genomics, Brain Tumors, Adolescent and Young Adult, and Women's Cancers. Qualified individuals should have a PhD with established expertise in the specific cancer to be studied. Candidates must also have a track record of a sustained and ongoing nationally funded research program, an outstanding record of collaborative cancer research activities and commitment to mentoring and teaching. Candidates for Full Professor should also demonstrate national and international recognition of their research program as well as success in leadership activities.

Successful candidates will be expected to interact closely with members of established basic and translational cancer center research programs. Where appropriate, identify interests in using human tissue and PDX models of cancer in the research effort. Primary appointment will be in the CWRU SOM department of expertise or in the Cancer Center.

Please send curriculum vitae and a cover letter outlining your research interests, and experience in mentoring electronically to: Stanton L. Gerson, MD, Director, Case Comprehensive Cancer Center, c/o **cancersearch@case.edu**. After initial review, you will be asked for a list of three or more references. Please include "Cancer Faculty Search" in the subject line.

In employment, as in education, Case Western Reserve University is committed to Equal Opportunity and Diversity. Women, veterans, members of underrepresented minority groups, and individuals with disabilities are encouraged to apply.

Case Western Reserve University provides reasonable accommodations to applicants with disabilities. Applicants requiring a reasonable accommodation for any part of the application and hiring process should contact the Office of Inclusion, Diversity and Equal Opportunity at 216-368-8877 to request a reasonable accommodation. Determinations as to granting reasonable accommodations for any applicant will be made on a case-by-case basis.

Science Careers online @sciencecareers.org

Career Feature:

Artificial Intelligence

Issue date: November 30

Book ad by November 15

Ads accepted until November 21 if space allows

129,562
subscribers in print every week

503,472
monthly unique browsers on ScienceCareers.org

56 %
of our weekly readers are Ph.D.s

Artificial Intelligence (AI) is impacting science in new and exciting ways as scientists are using it to better understand society to find solutions to problems across diverse disciplines. This feature will give an overview of AI, and explore the hotspots/centers of excellence and applications for AI. Typical career paths for those working in AI will be explored as well as the opportunities that exist for careers in AI.

Your organization can brand itself as a leader in AI by raising your visibility alongside relevant content while attracting potential candidates. Contact us for further details.

What makes *Science* the best choice for recruiting?

- Read and respected by 400,000 readers around the globe
- Your ad dollars support AAAS and its programs, which strengthens the global scientific community.

Why choose this AI Feature for your advertisement?

- Relevant ads lead off the career section with a special "AI" banner.

Expand your exposure by posting your print ad online:

- Link on the job board homepage directly to AI jobs
- Dedicated landing page for AI positions.

To book your ad:
advertise@sciencecareers.org

The Americas
+ 202 326 6577

Europe
+44 (0) 1223 326527

Japan
+81 3 6459 4174

China/Korea/Singapore/Taiwan
+86 131 4114 0012

Produced by the *Science*/AAAS Custom Publishing Office.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

FOR RECRUITMENT IN SCIENCE, THERE'S ONLY ONE *SCIENCE*.

FIU FLORIDA INTERNATIONAL UNIVERSITY

Department of Biological Sciences

WALTER AND ROSALIE GOLDBERG PROFESSORSHIP

in Tropical Ecology

The Department of Biological Sciences at Florida International University (FIU, http://biology.fiu.edu) is pleased to announce the **Walter and Rosalie Goldberg Professorship in Tropical Ecology**. We seek a top scholar in any area of tropical ecology. The successful candidate will be expected to have an exceptional publication record, bring and maintain an active externally-funded research program, mentor graduate students, and teach courses in their area of expertise. The minimum requirements are a doctoral degree from an accredited institution, full professor or equivalent status, and a demonstrated record of achievement in academic research, teaching and service.

Qualified candidates are encouraged to apply to Job Opening ID 516044 at **http://facultycareers.fiu.edu** and attach a cover letter, curriculum vitae, and statements of research, teaching, and service as a single pdf file. Please also provide names and contact information for at least three references who will be contacted upon submission of application.

To receive full consideration, applications and required materials should be received by **November 30, 2018**. Review will continue until position is filled. Please direct inquiries about this search to the chair of the search committee, Dr. Todd Crowl, tcrowl@fiu.edu.

PennState
Eberly College of Science

Eberly Research Fellows

at Penn State University

The Eberly College of Science at Penn State University invites nominees for the Eberly Research Fellowship program. Eberly Fellowships are designed to attract exceptional early career scientists to Penn State to enhance their career goals in the vibrant, highly collaborative environment of the Eberly College of Science and the broader STEM community of Penn State University. The Eberly College of Science which includes the Departments of Astronomy & Astrophysics, Biology, Biochemistry & Molecular Biology, Chemistry, Mathematics, Physics, and Statistics, ranks in the top 10 universities in the U.S. and has annual research expenditures exceeding $110M. Each of the seven departments is expecting to appoint one or more Eberly Fellows. Nominations for early career scientists with exceptional promise in basic research in physics, chemistry, biology, molecular biology, astronomy, mathematics, and statistics and/or applied research in health, energy, materials, or the environment are encouraged. Interdisciplinary as well as traditional disciplinary research is encouraged. Fellows who wish to also gain training and experience in teaching may elect to receive mentored teaching experience. Eberly Fellow advisors must hold their primary appointment in one of the seven departments of the Eberly College of Science. Co-advisors and cross-disciplinary research are also supported.

Eligibility and appointment

Applicants must be a current doctoral student or have received a doctoral degree in science, statistics, or mathematics within the past three years. Current doctoral students must have their doctoral degree prior to the start of their fellowship. Current doctoral students and postdoctoral fellows at Penn State are not eligible. Eberly Research Fellowships may be held from 1-2 years with annual appointments conditional on satisfactory performance. Fellows will receive a stipend of $65,000 and $5,000 per year in discretionary funds, which can be used for travel and other research expenses.

Nomination and applications

Nominations will be accepted from faculty advisors, graduate program chairs, department chairs, or others who can attest to the nominee's potential as a scientist. Nominations should include the nominees CV. Nominations of women and under-represented minorities are strongly encouraged. The nomination deadline is December 15, 2018 for appointments beginning 6-12 months later. Nominations should be sent to research-fellows@psu.edu

The Eberly Research Fellowship Selection Committee will select the nominees who will then be invited to submit their applications by January 15, 2019. Applications will include (1) a biographical sketch – including publications, accepted, and submitted manuscripts, (2) three letters of reference including one from the doctoral advisor,(3) research statement summarizing research accomplishments and research that you intend to pursue at Penn State, (4) names of one or more potential faculty advisors among the faculty in the seven departments of the Eberly College of Science, Penn State University.

CAMPUS SECURITY CRIME STATISTICS: For more about safety at Penn State, and to review the Annual Security Report which contains information about crime statistics and other safety and security matters, please go to http://www.police.psu.edu/clery/, which will also provide you with detail on how to request a hard copy of the Annual Security Report.

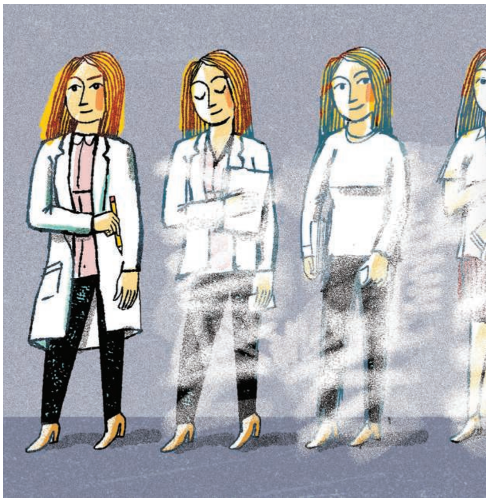
WORKING LIFE

By **Brittany L. Forte**

Finding peace with pencil

I reached for my favorite black pen, eager to begin my annual career development plan—three pages of boxes and lines detailing my vision of what lies ahead after I graduate with a Ph.D. Each of the previous 3 years, I had confidently completed the form in pen. But as I prepared to fill out the form for the fourth time, I hesitated. I had changed during graduate school, and I had to ask myself, “Can I be sure I will stick to the plan?” Having a clear path had offered such reassurance as I worked through the highs and lows of grad school. But this time, I decided, I would set aside the pen and complete the plan in pencil.

My fixation on planning began as a high school freshman. My principal gave an inspirational speech about the importance of a 5-year plan, and I was sold. Once I created my plan, the rest would be easy—I just had to stick to the straight and narrow path I had set for myself. In permanent ink, my 14-year-old self wrote that I would channel my passion for science into a career as a pharmacist. I signed the document and confidently submitted it to my homeroom teacher, who promised to return it to me on the final day of senior year.



Four winding years later, I was horrified to open the envelope. I still loved science, but I had discovered a passion for doing research and now envisioned a future in the lab. Embarrassed that my plans had changed and unsettled by learning that the path to satisfaction was not as clear as I had hoped, I ripped up my 5-year plan and buried the pieces in the trash. I promised myself I would not make such a mistake again. The next time I made a plan, I would execute it flawlessly.

“I was hesitant to change my life’s blueprint yet again.”

Each of my first 2 years as an undergraduate, I recorded my progress against a new plan: After finishing my bachelor’s degree, I would enroll as a Ph.D. student in the same pharmacology lab where I was already conducting research. I completed my annual progress report—in pen. Then, I discussed it with my faculty mentor, and we both signed it in permanent ink. Having a detailed blueprint, verified by a mentor I looked up to, made me feel secure in my future.

I was hesitant to change my life’s blueprint yet again. But as I prepared to fill out my annual progress report with my favorite black pen for the third and final time as an undergrad, just 9 months before graduation, I realized that I wanted to broaden my horizons. I still hoped to get a Ph.D.—but in basic biology instead of pharmacology.

To my surprise, my faculty mentor praised me for realizing that it was time to make a change, as opposed to blindly sticking to my earlier path. But he insisted that I be precise about my new direction: What would I study, and in which lab?

I eagerly devised a new plan. I was comfortable giving up plan A—but if I was going to switch to plan B, it needed to be as clear and detailed as before.

Today, though, in the fourth year of my Ph.D. program, I don’t have a plan A or even a firm plan B. I filled out my most recent career development plan in pencil, smudged it by erasing and changing entries, and left a number of boxes blank. I may yet pursue an academic career, as I had written with such confidence in previous years. Perhaps I will become a scientific editor. Or maybe I will end up going some other direction I haven’t yet considered.

To be entirely honest, not knowing what my future holds is a little uncomfortable. It is hard to give up knowing—or thinking I know—exactly what I want to do and how I will get there. But I’m grateful to have room to grow and develop in unexpected directions. So my new mantra is “be flexible, embrace change and uncertainty—and always write in pencil.” ■

Brittany L. Forte is a Ph.D. candidate in cancer biology at the University of Arizona in Tucson. Do you have an interesting career story? Send it to SciCareerEditor@aaas.org. For more on life and careers, visit sciencecareers.org.

ILLUSTRATION: ROBERT NEUBECKER

Published by AAAS