An epic clash of cultures in ancient Mesoamerica p. 968

Music is another language pp. 974 & 1043

A primordial body in the Kuiper Belt pp. 980 & 998–1000

Science

$15
28 FEBRUARY 2020
sciencemag.org

AAAS

HUMAN CRISPR

Gene editing meets cancer immunotherapy

pp. 976 & 1001

CONTENTS

28 FEBRUARY 2020 • VOLUME 367 • ISSUE 6481

968

NEWS

IN BRIEF

960 News at a glance

IN DEPTH

962 Strategies shift as coronavirus pandemic looms
The virus seems unstoppable, but mitigating its speed and impact is possible
By J. Cohen and K. Kupferschmidt
EDITORIAL p. 959

963 Preprints bring 'firehose' of outbreak data
COVID-19 has upended the ways researchers share findings and collaborate *By K. Kupferschmidt*
EDITORIAL p. 959

965 Satellite megaconstellations menace giant survey telescope
Darkening the orbiters could limit damaging light trails in wide-field images from the Vera C. Rubin Observatory *By D. Clery*

966 Safety benefits of 'biased' opioids scrutinized
Mouse studies challenge premise of efforts to reduce painkillers' fatal side effect *By K. Servick*

967 Cheap balloon-borne telescopes aim to rival space observatories
Stable pointing system enables SuperBIT to deliver images almost as sharp as those from the Hubble Space Telescope *By A. Cho*

FEATURES

968 The arrival of strangers
New evidence points to a clash between two ancient Mesoamerican cultures, Teotihuacan and the Maya *By L. Wade*
PODCAST

INSIGHTS

PERSPECTIVES

974 Splitting speech and music
Brain asymmetries for words and melodies of songs depend on opposite acoustic cues
By D. Sammler
REPORT p. 1043

976 Knocking out barriers to engineered cell activity
CRISPR-Cas9 gene-edited T cells show safety and long-term engraftment in humans
By J. R. Hamilton and J. A. Doudna
RESEARCH ARTICLE p. 1001

978 Maternal microbial molecules affect offspring health
Intestinal molecules during pregnancy in mice may protect offspring from metabolic disease *By J. Ferguson*
RESEARCH ARTICLE p. 1002

979 Cobalt in lithium-ion batteries
Replacements are sought for cobalt, a costly element used in lithium-ion battery cathodes
By M. Li and J. Lu

980 A deep dive into the abyss
The flyby of the Kuiper Belt object Arrokoth provides quick and tantalizing observations
By D. C. Jewitt
RESEARCH ARTICLES pp. 998, 999, 1000

982 Artificial intelligence in cancer therapy
Artificial intelligence can optimize cancer drug discovery, development, and administration *By D. Ho*

POLICY FORUM

984 Recalibrating global data center energy-use estimates
Growth in energy use has slowed owing to efficiency gains that smart policies can help maintain in the near term
By E. Masanet et al.

BOOKS *ET AL.*

987 Out of sight
A historian investigates the Cold War competition to create an invisible aircraft
By F. R. van der Linden

988 When life meets research
The personal and professional collide in a scientist's story of early human development *By J. Rossant*

LETTERS

989 Giant salamanders: Farmed yet endangered
By C. Lu et al.

PHOTO: MAX SHEN/GETTY IMAGES

Published by AAAS

989 The U.S. military is not sustainable
By O. Belcher et al.

990 Brazilian national parks at risk
By R. Ruaro et al.

RESEARCH

IN BRIEF

995 From *Science* and other journals

RESEARCH ARTICLES

Outer solar system

998 The geology and geophysics of Kuiper Belt object (486958) Arrokoth *J. R. Spencer* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAY3999

999 Color, composition, and thermal environment of Kuiper Belt object (486958) Arrokoth *W. M. Grundy* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAY3705

1000 The solar nebula origin of (486958) Arrokoth, a primordial contact binary in the Kuiper Belt *W. B. McKinnon* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAY6620
PERSPECTIVE p. 980

1001 Clinical trials
CRISPR-engineered T cells in patients with refractory cancer *E. A. Stadtmauer* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.ABA7365
PERSPECTIVE p. 976

1002 Microbiota
Maternal gut microbiota in pregnancy influences offspring metabolic phenotype in mice *I. Kimura* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAW8429
PERSPECTIVE p. 978

1003 Plant science
Pectin homogalacturonan nanofilament expansion drives morphogenesis in plant epidermal cells *K. T. Haas* et al.

1008 Immunology
Structure of the secretory immunoglobulin A core *N. Kumar* et al.

REPORTS

1014 Immunology
Structural insights into immunoglobulin M *Y. Li* et al.

1018 Optics
Ultrafast control of vortex microlasers *C. Huang* et al.

1021 Organic chemistry
Aminoalkyl radicals as halogen-atom transfer agents for activation of alkyl and aryl halides *T. Constantin* et al.

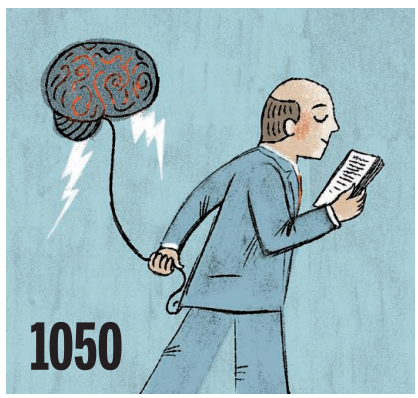
1026 Fluid dynamics
Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations *M. Raissi* et al.

1030 Batteries
Kinetic pathways of ionic transport in fast-charging lithium titanate *W. Zhang* et al.

1035 Ecological speciation
Ecologically diverse clades dominate the oceans via extinction resistance *M. L. Knope* et al.

1039 Cell biology
Polymerization in the actin ATPase clan regulates hexokinase activity in yeast *P. R. Stoddard* et al.

1043 Neuroscience
Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody *P. Albouy* et al.
PERSPECTIVE p. 974



DEPARTMENTS

959 Editorial
The costs of secrecy *By H. Holden Thorp*
NEWS STORIES pp. 962 & 963

1050 Working Life
Adapting to my brain *By Luis Batista*

ON THE COVER

CRISPR-Cas9 technology acts as molecular scissors to edit DNA. Genetically modified T cells are revolutionizing cancer therapy but do not successfully treat all patients. Researchers used CRISPR to cut out sections of DNA on three genes in T cells from patients with advanced cancer. This pioneering therapeutic application of genome editing paves the way for next-generation cell-based therapies for human disease. See pages 976 and 1001.
*Illustration: V. Altounian/*Science

Science Staff 959
AAAS News & Notes 991
Science Careers 1048

An ecologically, taxonomically diverse coral reef community teeming with fish and corals at Miloli'i, Hawaii, USA

CREDITS: (PHOTO) JOHN H.R. BURNS; (ILLUSTRATION) ROBERT NEUBECKER

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2020 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2148; Foreign postage extra: Air assist delivery: $98. 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 available from backissues.sciencemag.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Published by AAAS

Science AAAS

Editor-in-Chief Holden Thorp, hthorp@aaas.org

Executive Editor Monica M. Bradford

Editors, Research Valda Vinson, Jake S. Yeston **Editor, Insights** Lisa D. Chong

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), Brent Grocholski, 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, Keith T. Smith (UK), Jelena Stajic, Peter Stern (UK), Valerie B. Thompson, Brad Wible, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Priscilla N. Kelly, Tage S. Rai, Seth Thomas Scanlon (UK), Yury V. Suleymanov **LETTERS EDITOR** Jennifer Sills **LEAD CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmec **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Chris Filiatreau, Julia Katris, Nida Masiulis, Suzanne M. White **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Ope Martins, Shannon McMahon, Jerry Richardson, Alana Warnke, Alice Whaley (UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Jeremy Dow, Alexander Kief, Ronmel Navas, Hilary Stewart (UK), Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ASI DIRECTOR, OPERATIONS** Janet Clements (UK) **ASI SR. OFFICE ADMINISTRATOR** Jessica Waldock (UK)

News Editor Tim Appenzeller

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** Martin Enserink **DEPUTY NEWS EDITORS** Elizabeth Culotta, Lila Guterman, David Grimm, Eric Hand (Europe), David Malakoff **SR. CORRESPONDENTS** Daniel Clery (UK), Jon Cohen, Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brainard, Catherine Matacic **NEWS REPORTERS** Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Kelly Servick, Robert F. Service, Erik Stokstad, Paul Voosen, Meredith Wadman **INTERNS** Eva Frederick, Rodrigo Perez Ortega **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), Elisabeth Pain (Careers), Charles Piller, Michael Price, Tania Rabesandratana (Barcelona), Emily Underwood, Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **CAREERS** Donisha Adams, Rachel Bernstein (Editor), Katie Langin (Acting Editor) **COPY EDITORS** Julia Cole (Senior Copy Editor), Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

Creative Director Beth Rakouskas

DESIGN MANAGING EDITOR Marcy Atarod **GRAPHICS MANAGING EDITOR** Alberto Cuadra **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **WEB CONTENT STRATEGY MANAGER** Kara Estelle-Powers **SENIOR DESIGNER** Chrystal Smith **DESIGNER** Christina Aycock **GRAPHICS EDITOR** Nirja Desai **INTERACTIVE GRAPHICS EDITOR** Xing Liu **SENIOR SCIENTIFIC ILLUSTRATORS** Valerie Altounian, Chris Bickel **SCIENTIFIC ILLUSTRATOR** Alice Kitterman **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **SENIOR PHOTO EDITOR** Emily Petersen **PHOTO EDITOR** Kaitlyn Dolan

Chief Executive Officer and Executive Publisher Sudip Parikh

Publisher, Science Family of Journals Bill Moran

DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYSIS** Eric Knott **DIRECTOR OF ANALYTICS** Enrique Gonzales **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **SENIOR BUSINESS ANALYST** Cory Lipman, Meron Kebede **FINANCIAL ANALYST** Alexander Lee **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **SENIOR SALES COORDINATOR** Shirley Young **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **QUALITY TECHNICAL MANAGER** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT SPECIALISTS** Steve Forrester, Jacob Hedrick, Antoinette Hodal, Lori Murphy **DIGITAL PRODUCTION MANAGER** Lisa Stanford **CONTENT SPECIALIST** Kimberley Oster **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **DESIGNER, CUSTOM PUBLISHING** Jeremy Huntsinger **SR. TRAFFIC ASSOCIATE** Christine Hall **SPECIAL PROJECTS ASSOCIATE** Sarah Dhere

ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING MANAGER** Aimee Aponte **JOURNALS MARKETING MANAGER** Shawana Arnold **MARKETING ASSOCIATES** Tori Velasquez, Mike Romano, Ashley Hylton **DIGITAL MARKETING SPECIALIST** Asleigh Rojanavongse **SENIOR DESIGNER** Kim Huynh

DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING Sean Sanders **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst

DIRECTOR, PRODUCT & PUBLISHING DEVELOPMENT Chris Reid **DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT** Sarah Whalen **ASSOCIATE DIRECTOR, PRODUCT MANAGMENT** Kris Bishop **SR. PRODUCT ASSOCIATE** Robert Koepke **DIGITAL PRODUCT STRATEGIST** Michael Hardesty **SPJ ASSOCIATE** Samantha Bruno Fuller

DIRECTOR, INSTITUTIONAL LICENSING Iquo Edim **ASSOCIATE DIRECTOR, RESEARCH & DEVELOPMENT** Elisabeth Leonard **MARKETING MANAGER** Kess Knight **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGER** Marco Castellan **MANAGER, AGENT RELATIONS & CUSTOMER SUCCESS** Judy Lillibridge **SENIOR OPERATIONS ANALYST** Lana Guz **FULFILLMENT COORDINATOR** Melody Stringer **SALES COORDINATOR** Josh Haverlock

DIRECTOR, GLOBAL SALES Tracy Holmes **US EAST COAST AND MID WEST SALES** Stephanie O'Connor **US WEST COAST SALES** Lynne Stickrod **US SALES MANAGER, SCIENCE CAREERS** Claudia Paulsen-Young **US SALES REP, SCIENCE CAREERS** Tracy Anderson **ASSOCIATE DIRECTOR, ROW** Roger Goncalves **SALES REP, ROW** Sarah Lelarge **SALES ADMIN ASSISTANT, ROW** Bryony Cousins **DIRECTOR OF GLOBAL COLLABORATION AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu **ASSOCIATE DIRECTOR, INTERNATIONAL COLLABORATION** Grace Yao **SALES MANAGER** Danny Zhao **MARKETING MANAGER** Kilo Lan **ASCA CORPORATION, JAPAN** Kaoru Sasaki (Tokyo), Miyuki Tani (Osaka) **COLLABORATION/CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David **RIGHTS AND LICENSING COORDINATOR** Jessica Adams **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **CONTRACTS AND LICENSING ASSOCIATE** Lili Catlett

---

MAIN HEADQUARTERS
Science/AAAS
1200 New York Ave. NW
Washington, DC 20005

SCIENCE INTERNATIONAL
Clarendon House
Clarendon Road
Cambridge, CB2 8FH, UK

SCIENCE CHINA
Room 1004, Culture Square
No. 59 Zhongguancun St.
Haidian District, Beijing, 100872

SCIENCE JAPAN
ASCA Corporation
Sibaura TY Bldg. 4F, 1-14-5
Shibaura Minato-ku
Tokyo, 108-0073 Japan

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

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

REPRINTS AND PERMISSIONS
sciencemag.org/help/reprints-and-permissions

MEDIA CONTACTS
scipak@aaas.org

MULTIMEDIA CONTACTS
SciencePodcast@aaas.org
ScienceVideo@aaas.org

INSTITUTIONAL SALES AND SITE LICENSES
sciencemag.org/librarian

PRODUCT ADVERTISING & CUSTOM PUBLISHING
advertising.sciencemag.org/products-services
science_advertising@aaas.org

CLASSIFIED ADVERTISING
advertising.sciencemag.org/science-careers
advertise@sciencecareers.org

JOB POSTING CUSTOMER SERVICE
employers.sciencecareers.org
support@sciencecareers.org

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
sciencemag.org/subscriptions

MEMBER BENEFITS
aaas.org/membercentral

AAAS BOARD OF DIRECTORS
CHAIR Steven Chu
PRESIDENT Claire M. Fraser
PRESIDENT-ELECT Susan G. Amara
TREASURER Carolyn N. Ainslie
CHIEF EXECUTIVE OFFICER Sudip Parikh
BOARD Cynthia M. Beall, Rosina M. Bierbaum, Ann Bostrom, Stephen P.A. Fodor, S. James Gates, Jr., Laura H. Greene, Kaye Husbands Fealing, Maria M. Klawe, Robert B. Millard, Alondra Nelson, William D. Provine

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.

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*
James Analytis, *U. of California, Berkeley*
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*
Peter Bearman, *Columbia U.*
Carlo Beenakker, *Leiden U.*
Yasmine Belkaid, *NIAID, NIH*
Philip Benfey, *Duke U.*
Gabriele Bergers, *VIB*
Bradley Bernstein, *Mass. General Hospital*
Alessandra Biffi, *Harvard Med. School*
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*
György Buzsáki, *New York U. School of Med.*
Blanche Capel, *Duke U.*
Annmarie Carlton, *U. of California, Irvine*
Nick Chater, *U. of Warwick*
Zhijian Chen, *UT Southwestern Med. Ctr.*
Ib Chorkendorff, *Denmark TU*
James J. Collins, *MIT*
Robert Cook-Deegan, *Arizona State 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.*
Claude Desplan, *New York U.*
Sandra Díaz, *Universidad Nacional de Córdoba*
Ulrike Diebold, *TU Wien*
Hong Ding, *Inst. of Physics, CAS*
Jennifer Dionne, *Stanford U.*
Dennis Discher, *U. of Penn.*
Gerald Dorn, *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*
Andrea Encalada, *U. San Francisco de Quito*
Nader Engheta, *U. of Penn.*
Karen Ersche, *U. of Cambridge*
Barry Everitt, *U. of Cambridge*
Vanessa Ezenwa, *U. of Georgia*
Michael Feuer, *The George Washington U.*
Toren Finkel, *U. of Pittsburgh Med. Ctr.*
Gwenn Flowers, *Simon Fraser U.*
Peter Fratzl, *Max Planck Inst. Potsdam*
Elaine Fuchs, *Rockefeller U.*
Eileen Furlong, *EMBL*
Jay Gallagher, *U. of Wisconsin*
Daniel Geschwind, *U. of California, Los Angeles*
Karl-Heinz Glassmeier, *TU Braunschweig*
Ramon Gonzalez, *U. of South Florida*
Elizabeth Grove, *U. of Chicago*
Nicolas Gruber, *ETH Zürich*
Hua Guo, *U. of New Mexico*
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*
Jian He, *Clemson U.*
Carl-Philipp Heisenberg, *IST Austria*
Ykä Helariutta, *U. of Cambridge*
Janet G. Hering, *Eawag*
Hans Hilgenkamp, *U. of Twente*
Kai-Uwe Hinrichs, *U. of Bremen*
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*
V. Narry Kim, *Seoul Nat. U.*
Robert Kingston, *Harvard Med. School*
Nancy Knowlton, *Smithsonian Institution*
Etienne Koechlin, *École Normale Supérieure*
Alexander L. Kolodkin, *Johns Hopkins U.*
Julija Krupic, *U. of Cambridge*
Thomas Langer, *Max Planck Inst. Cologne*
Mitchell A. Lazar, *U. of Penn.*
Ottoline Leyser, *U. of Cambridge*
Wendell Lim, *U. of California, San Francisco*
Jianguo Liu, *Michigan State U.*
Luis Liz-Marzán, *CIC biomaGUNE*
Jonathan Losos, *Washington U. in St. Louis*
Ke Lu, *Chinese Acad. of Sciences*
Christian Lüscher, *U. of Geneva*
Jean Lynch-Stieglitz, *Georgia Inst. of Tech.*
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. Nacional Autónoma de México*
Graham Medley, *London School of Hygiene & Tropical Med.*
Jane Memmott, *U. of Bristol*
Baoxia Mi, *U. of California, Berkeley*
Edward Miguel, *U. of California, Berkeley*
Tom Misteli, *NCI, NIH*
Yasushi Miyashita, *U. of Tokyo*
Alison Motsinger-Reif, *NIEHS, NIH* **(S)**
Daniel Nettle, *Newcastle U.*
Daniel Neumark, *U. of California, Berkeley*
Beatriz Noheda, *U. of Groningen*
Helga Nowotny, *Vienna Science, Research & Tech. Fund*
Rachel O'Reilly, *U. of Birmingham*
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.*
Katherine Pollard, *U. of California, San Francisco*
Elvira Poloczanska, *Alfred-Wegener-Inst.*
Julia Pongratz, *Ludwig Maximilians U.*
Philippe Poulin, *CNRS*
Jonathan Pritchard, *Stanford U.*
Félix A. Rey, *Institut Pasteur*
Trevor Robbins, *U. of Cambridge*
Joeri Rogelj, *Imperial College London*
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, *Med. U. of Vienna*
Alexander Schier, *Harvard U.*
Wolfram Schlenker, *Columbia U.*
Susannah Scott, *U. of California, Santa Barbara*
Rebecca Sear, *London School of Hygiene & Tropical Med.*
Vladimir Shalaev, *Purdue U.*
Jie Shan, *Cornell U.*
Beth Shapiro, *U. of California, Santa Cruz*
Jay Shendure, *U. of Washington*
Steve Sherwood, *U. of New South Wales*
Brian Shoichet, *U. of California, San Francisco*
Robert Siliciano, *Johns Hopkins U. School of Med.*
Lucia Sivilotti, *U. College London*
Alison Smith, *John Innes Centre*
Richard Smith, *U. of North Carolina* **(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*
V. S. Subrahmanian, *Dartmouth College*
Ira Tabas, *Columbia U.*
Sarah Teichmann, *Wellcome Sanger Inst.*
Rocio Titiunik, *Princeton U.*
Shubha Tole, *Tata Inst. of Fundamental Research*
Wim van der Putten, *Netherlands Inst. of Ecology*
Reinhilde Veugelers, *KU Leuven*
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*
Xiaowei Zhuang, *Harvard U.*
Maria Zuber, *MIT*

EDITORIAL

The costs of secrecy

"Without freedom of speech there is no modern world, just a barbaric one." These words from China's most famous artist and activist, Ai Weiwei, have never been more important. Ai Weiwei would probably agree that China's actions in the coronavirus crisis require the voice of the scientific community, and he wouldn't be surprised that getting folks to say something has been a challenge.

I didn't want to be the person to write this editorial. I felt that it would best come from someone inside China with a direct connection to the situation. Such a person could help dispel or reinforce the scraps of information coming from the intrepid journalists and the few courageous eyewitnesses. But over the past few weeks, I've been discouraged by the responses of such individuals who declined or didn't respond to an invitation to write a forthcoming editorial about China's secrecy on coronavirus. Some of these scientists and experts even expressed doubt that I'd find such a gutsy author at any organization in China to write such a piece.

Maybe I should have asked Ai Weiwei.

It's no wonder I had trouble. Although we don't know if the lack of information is the result of active suppression or just fear of punishment, Xu Zhangrun, a professor of law at Tsinghua University, was placed under house arrest and cut off from the internet for publishing an essay earlier this month that was critical of China's handling of the public health crisis. His essay called for an independent body to investigate the origin of the coronavirus, which shouldn't be a matter of controversy much less imprisonment.

Although China's actions have been more forthcoming than in the 2002–2003 severe acute respiratory syndrome (SARS) epidemic, it is only within the last few days that experts from the World Health Organization have been allowed into Wuhan, where the outbreak in China was first reported. And the courageous Li Wenliang, the Chinese physician who first sounded the alarm in December, was punished by authorities for his online comments before he then contracted the virus and died.

A public health crisis is not a good time for censorship. Like all other top journals, we are awash in papers on coronavirus disease 2019 (COVID-19) and the virus that causes it, SARS–coronavirus 2 (SARS-CoV-2). We were proud to cosign a letter led by the Wellcome Trust setting out the terms under which the *Science* family of journals would document the virus and the disease. Under these terms, we are strongly encouraging preprints, making all data and the published paper free immediately, and expediting review. Last week, we published—only 9 days after receiving it—the structure of the SARS-CoV-2 spike protein, which could be important in therapeutic design. The contrast between the mobilization of the scientific community and the political actions of China is striking.

We will never be able to better handle future public health crises without learning lessons from previous experiences. And if every experience is shrouded in secrecy, enforced by a repressive government, then we will never solve this problem. Yet, the COVID-19 crisis seems only to add to the legacies of obfuscation and repression left by Chernobyl, Fukushima, and SARS. Will Western governments do better? We don't know. The United States Department of State ignored recommendations of the U.S. Centers for Disease Control and Prevention by bringing home infected individuals from the coutries where they contracted the virus. And with the spread of the virus to Italy and South Korea, we'll see if any lessons have been learned.

> "...the COVID-19 crisis seems only to add to the... obfuscation and repression left by Chernobyl, Fukushima, and SARS."

The solution to all of this is unclear. The scientists working within repressive regimes deserve our support and admiration. And scientific collaboration with Chinese scientists has been crucial in understanding the outbreak and the biology of the virus. Preventing a pandemic is impossible without international collaboration. This week, Harvard University and the Guangzhou Institute announced that they will work together to investigate SARS-CoV-2, which is progress.

The only thing we can do at this point is mobilize the voices that can speak out to ask for greater transparency and collaboration. But it will take more than just an examination of China alone, or even this outbreak alone, to shape the future. To avoid the tragic costs of silence, we must keep pushing for more transparency and truth on all fronts.

–H. Holden Thorp

PHOTO: CAMERON DAVIDSON

H. Holden Thorp
Editor-in-Chief,
Science journals.
hthorp@aaas.org;
@hholdenthorp

10.1126/science.abb4420

Published by AAAS

> “MeTooSTEM is beyond salvaging as an organization.”
>
> **Virologist Angela Rasmussen**, in a 21 February letter resigning from the nonprofit that helps survivors of sexual harassment in science. She and others allege that its founder, BethAnn McLaughlin, sidelined people of color and bullied volunteers. McLaughlin has denied the allegations.

IN BRIEF

Edited by **Jeffrey Brainard**

PUBLISHING

China sours on publication metrics

Ministries in China last week said Chinese universities should de-emphasize publication citations when evaluating scientific research, a striking shift from a long-standing preference in Chinese academe for relying on citations when deciding whether to promote faculty members or award degrees. The joint statement by the ministries of education and science and technology said an excessive and distorted reliance on the Science Citation Index, part of the Web of Science owned by Clarivate Analytics, has hurt the country’s research efforts by emphasizing quantity over quality of scientists’ publications. Although not specifically mentioned in the statement, many institutions used to give researchers monetary rewards for landing publications in journals with high journal impact factors, although that practice has already fallen out of favor. The statement calls for overhauling evaluations of scientists by using qualitative and quantitative measures that consider innovativeness, although it is short on details for implementation. It comes as some academics and policy organizations in Europe and the United States have sought a similar shift in academic culture there.

U.K. science minister demoted

LEADERSHIP | The United Kingdom has a new science minister—the fourth appointment in 1 year—and the position has been downgraded. As part of a Cabinet reshuffle, Amanda Solloway took over the post from Chris Skidmore. She has a background in retail and human resources and is the first science minister without a university degree. The role was changed from senior minister of state to undersecretary, a sign of lesser influence, and responsibility for universities was switched to a separate position. “The role of science minister is too important, and too critical right now, for this to be anything other than an important, visible, decision-making role,” said John Womersley, former head of the United Kingdom’s Science and Technology Facilities Council. Despite the status change, Prime Minister Boris Johnson has said he will prioritize science and innovation, and his chief adviser, Dominic Cummings, is expected to push that agenda.

China bans wildlife trade

PUBLIC HEALTH | A key Chinese legislative body this week called for stricter enforcement of laws governing trade in wild animals, which is believed to be linked to the COVID-19 outbreak, China’s Xinhua News Agency reported. Conservationists caution that the government has not adequately enforced previous such restrictions. China issued a temporary ban during the outbreak of the severe acute respiratory syndrome coronavirus in Hong Kong and mainland China in 2002–03, which was traced to eating wild animals. China is the largest consumer of many threatened species for food and traditional medicines.

UC to subsidize PLOS fees

PUBLISHING | The University of California (UC) last week announced a 2-year agreement with the open-access publisher PLOS to subsidize papers by authors on the university system’s 10 campuses. UC will contribute at least $1000 per paper, and more if authors lack other funds; PLOS journal charges vary, with *PLOS ONE*’s author fee set at $1595. PLOS agreed to hold author fees at 2019 levels during the agreement, the first subsidy arrangement it has inked with any university. Some UC campuses already subsidize author fees charged by other publishers to free journal articles from subscription paywalls, but in some cases, the PLOS subsidy will be more generous. UC wants to sign more deals with publishers of open-access journals to make it as easy for researchers to publish there as in subscription journals, says Jeff MacKie-Mason, co–lead negotiator for UC’s deals with publishers.

Bayer, BASF lose herbicide case

BIOTECHNOLOGY | A U.S. federal jury has awarded a peach farmer in Missouri a $265 million verdict against agrochemical giants Bayer and BASF, finding that a widely used herbicide damaged his trees. Bader Farms owner Bill Bader said the weed-killing chemical, dicamba, drifted into his orchard from neighboring farms that grew soybeans and cotton genetically modified to resist the herbicide. The award, announced 14 February, includes $250 million in punitive damages. The companies will appeal. They face more than 130 other lawsuits, also alleging dicamba drifts from targeted crops. Dicamba is being used more

Plaintiffs in a lawsuit blamed the herbicide dicamba for stunted peaches grown at their farm.

PHOTO: BRYCE GRAY/ST. LOUIS POST-DISPATCH VIA AP

Published by AAAS

A team rescues a man from his flooded home in Monmouth, U.K., after Storm Dennis.

PHOTO: BEN BIRCHALL/PA IMAGES VIA GETTY IMAGES

METEOROLOGY

New computing horsepower to aid forecasts of U.K. storms

The United Kingdom's Met Office will build what would be the world's fastest supercomputer for weather prediction, costing up to £1.2 billion over the next decade, the agency announced last week. The new computer, expected to have a peak speed of nearly 100 petaflops when operational in 2022, will support improved forecasts of severe weather like Storm Dennis, which caused widespread flooding in the United Kingdom this month. The U.S. National Oceanic and Atmospheric Administration, seeking to return its weather model to world-leading status, also announced plans last week to purchase two new supercomputers at a total price of up to $505 million over 10 years, bringing its capacity to 40 petaflops. That upgrade, although short of the Met Office's, will likely put the agency's capacity on par with upgrades planned by the European Centre for Medium-Range Weather Forecasts, widely regarded as producing the best current weather model.

widely as weeds become resistant to Bayer's glyphosate-based herbicide Roundup. Other plaintiffs have sued the company alleging that glyphosate causes cancer; in one such case, a jury last year awarded $2 billion in damages, later reduced to $87 million.

Fetal tissue panel sought

BIOETHICS | The Trump administration is moving forward with a new ethics board that will review proposed research using fetal tissue from elective abortions. In a 20 February notice, the Department of Health and Human Services (HHS) seeks nominations within 30 days for a 15-member panel to include experts in ethics, law, and theology, as well as scientists. The board will advise HHS on whether to fund projects that passed peer review. In June 2019, the administration ended fetal tissue research in government labs and froze awards for new extramural projects. The tissue is used to study HIV, eye diseases, fetal development, and other topics.

India's cow research decried

POLICY | More than 500 scientists have asked the Indian government to withdraw a call for research proposals on indigenous cows and the curative properties of their urine, dung, and milk, including as cancer treatments. In a letter, the scientists say the call, issued 14 February, is "unscientific" and a misdirection of public money as research budgets in India are strapped; young scientists are not receiving their monthly stipends on time, for example. Hinduism considers cows sacred, and some petitioners see the research call as similar to previous efforts by the ruling Hindu nationalist Bharatiya Janata Party to validate faith-based pseudoscience.

Researchers suspect paper mill

MISCONDUCT | Researchers have uncovered what they believe may be a large "paper mill" in China that generated more than 400 journal articles with potentially fabricated images and similar text, suggesting a common source. Elisabeth Bik—a microbiologist and independent consultant who with other researchers uncovered the suspicious activity—described it last week on her blog. The papers contain questionable Western blot and flow cytometry images and similar keywords, titles, and layout. Bik says all authors work at Chinese hospitals; she suspects they may have faked the manuscripts to satisfy requirements to publish in international journals to be eligible for promotions. Some of the journals that published the papers, all based in Western countries, say they are investigating.

Published by AAAS

IN DEPTH

Residents of Casalpusterlengo, an Italian town under lockdown, line up to enter a supermarket.

GLOBAL HEALTH

Strategies shift as coronavirus pandemic looms

The virus seems unstoppable, but mitigating its speed and impact is possible

By **Jon Cohen** *and* **Kai Kupferschmidt**

The global march of COVID-19 is beginning to look unstoppable. In just the past week, a countrywide outbreak surfaced in Iran, spawning additional cases in Iraq, Oman, and Bahrain. Italy put 10 towns in the north on lockdown after the virus rapidly spread there. An Italian physician carried the virus to the Spanish island of Tenerife, a popular holiday spot for northern Europeans, and Austria and Croatia reported their first cases. Meanwhile, South Korea's outbreak kept growing explosively and Japan reported additional cases in the wake of the botched quarantine of a cruise ship.

The virus may be spreading stealthily in many more places. A modeling group at Imperial College London has estimated that about two-thirds of the cases exported from China have yet to be detected.

As *Science* went to press, the World Health Organization (WHO) still avoided using the word "pandemic" to describe the burgeoning crisis, instead talking about "epidemics in different parts of the world." But many scientists say that regardless of what it's called, the window for containment is now almost certainly shut. "It looks to me like this virus really has escaped from China and is being transmitted quite widely," says Christopher Dye, an epidemiologist at the University of Oxford. "I'm now feeling much more pessimistic that it can be controlled." In the United States, "disruption to everyday life might be severe," Nancy Messonnier, who leads the coronavirus response for the U.S. Centers for Disease Control and Prevention, warned on 25 February. "We are asking the American public to work with us to prepare for the expectation that this is going to be bad."

Dye and others say it's time to rethink the public health response. So far, efforts have focused on containment: slowing the spread of the virus within China, keeping it from being exported to other countries, and, when patients do cross borders, aggressively tracing anyone they were in contact with and quarantining those people for 2 weeks. But if the virus, named SARS-CoV-2, has gone global, travel restrictions may become less effective than measures to limit outbreaks and reduce their impact, wherever they are—for instance, by closing schools, preparing hospitals, or even imposing the kind of draconian quarantine imposed on huge cities in China.

"Border measures will not be as effective or even feasible, and the focus will be on community mitigation measures until a vaccine becomes available in sufficient quantities," says Luciana Borio, a former biodefense preparedness expert at the U.S. National Security Council who is now vice president at In-Q-Tel, a not-for-profit venture capital firm. "The fight now is to mitigate, keep the health care system working, and don't panic," adds Alessandro Vespignani, an infectious disease modeler at Northeastern University. "This has a range of outcomes from the equivalent of a very bad flu season to something that is perhaps a little bit worse than that."

Public health experts disagree, however, about how quickly the travel restrictions that have marked the first phase of the epidemic should be loosened. Early this week, the total number of cases stood at more than 80,000 with 2705 deaths—with 97% of the total still in China. Some countries have gone so far as to ban all flights to and from China; the United States quarantines anyone who has been in hard-hit Hubei province and refuses entry to foreign nationals if they have been anywhere in China during the past 2 weeks. Several countries have also added restrictions against South Korea and Iran.

PHOTO: MIGUEL MEDINA/AFP/GETTY IMAGES

Published by AAAS

The restrictions have worked to some degree, scientists say. "If we had not put a travel restriction on, we would have had many, many, many more travel-related cases than we have," says Anthony Fauci, who heads the U.S. National Institute of Allergy and Infectious Diseases.

But many epidemiologists have claimed that travel bans buy little extra time, and WHO doesn't endorse them. The received wisdom is that bans can backfire, for example, by hampering the flow of necessary medical supplies and eroding public trust. And as the list of affected countries grows, the bans will become harder to enforce and will make less sense: There is little point in spending huge amounts of resources to keep out the occasional infected person if you already have thousands in your own country. The restrictions also come at a steep price. China's economy has already taken an enormous hit from COVID-19, as has the airline industry. China also exports many products, from pharmaceuticals to cellphones, and manufacturing disruptions are causing massive supply chain problems.

"It would be very hard politically and probably not even prudent to relax travel restrictions tomorrow," says Harvard University epidemiologist Marc Lipsitch. "But in a week, if the news continues at the pace that it's been the last few days, I think it will become clear that travel restrictions are not the major countermeasure anymore."

Smaller scale containment efforts will remain helpful, says WHO's Bruce Aylward, who led an international mission to China over the past 2 weeks. In a report from the mission that Aylward discussed but did not publicly release, the group concludes that the Chinese epidemic peaked between 23 January and 2 February and that the country's aggressive containment efforts in Hubei, where at least 50 million people have been on lockdown, gave other provinces time to prepare for the virus and ultimately prevent "probably hundreds of thousands" of cases. "It's important that other countries think about this and think about whether they apply something—not necessarily full lockdowns everywhere, but that same rigorous approach."

Yet China's domestic restrictions have come at a huge cost to individuals, says Lawrence Gostin, who specializes in global health policy at Georgetown University Law Center. He calls the policies "astounding, unprecedented, and medieval," and says he is particularly concerned about the physical and mental well-being of people in Hubei who are housebound, under intensive surveillance, and facing shortages of health services. "This would be unthinkable in probably any country in the world but China," he says. (Italy's lockdowns are for relatively small towns, not major cities.)

China is slowly beginning to lift the restrictions in regions at lower risk, which could expose huge numbers of people to the infection, Dye says. "If normal life is restored in China, then we could expect another resurgence," he adds.

Still, delaying illness can have a big payoff, Lipsitch says. It will mean a lower burden on hospitals and a chance to better train vulnerable health care workers on how to protect themselves, more time for citizens to prepare, and more time to test potentially life-saving drugs and, in the longer term, vaccines. "If I had a choice of getting [COVID-19] today or getting it 6 months from now, I would definitely prefer to get it 6 months from now," Lipsitch says. Flattening the peak of an epidemic also means fewer people are infected overall, he says.

Other countries could adopt only certain elements from China's strategy. An updated analysis co-authored by Dye and posted on the preprint server medRxiv concludes that suspending public transport, closing entertainment venues, and banning public gatherings were the most effective mitigation interventions in China. "We don't have direct proof, of course, because we don't have a properly controlled experiment," Dye says. "But those measures were probably working to push down the number of cases." One question is whether closing schools will help. "We just don't know what role kids play" in the epidemic, Lipsitch says. "That's something that anybody who has 100 or more cases could start to study."

Some countries may decide it's better not to impede the free flow of people too much, keep schools and businesses open, and forgo the quarantining of cities. "That's quite a big decision to make with regards to public health," Dye says, "because essentially, it's saying, 'We're going to let this virus go.'"

To prepare for what's coming, hospitals can stockpile respiratory equipment and add beds. More intensive use of the vaccines against influenza and pneumococcal infections could help reduce the burden of those respiratory diseases on the health care system and make it easier to identify COVID-19 cases, which produce similar symptoms. Governments can issue messages about the importance of handwashing and staying home if you're ill.

Whatever the rest of the world does, it's essential that it take action soon, Aylward says, and he hopes other countries will learn from China. "The single biggest lesson is: Speed is everything," he says. "And you know what worries me most? Has the rest of the world learned the lesson of speed?" ■

SCIENTIFIC COMMUNICATION

Preprints bring 'firehose' of outbreak data

COVID-19 has upended the ways researchers share findings and collaborate

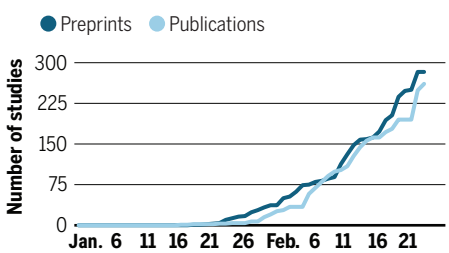
By **Kai Kupferschmidt**

On 22 January, Dave O'Connor and Tom Friedrich invited several dozen colleagues around the United States to join a new workspace on the instant messaging platform Slack. The scientists, both at the Wisconsin National Primate Research Center, had seen news about a new disease emerging in China and realized researchers would need a primate model if they were going to answer some important questions about its biology. "We put out a call to a bunch of investigators and basically said: 'Hey, let's talk,'" O'Connor says. The idea is to coordinate research and make sure results are comparable, Friedrich adds. (They named the Slack workspace the Wu-han Clan, a play on the hip-hop group Wu-Tang Clan.)

The Wu-han Clan is just one example of how the COVID-19 outbreak is transforming how scientists communicate about fast-moving health crises. A torrent of data is being released daily by preprint servers that didn't even exist a decade ago, then dissected on platforms such as Slack and Twitter, and in the media, before formal peer review begins. Journal staffers are working overtime to get manuscripts reviewed, edited, and published at record speeds. The venerable *New England Journal of Medicine* (*NEJM*) posted one COVID-19 paper within

Information revolution

Scientists are sharing more information using preprints than they did during any previous outbreak. The number of published papers is exploding as well.



CREDITS: (GRAPHIC) M. WEILAND/*SCIENCE*; (DATA) PUBMED; MEDRXIV; BIORXIV; CHEMRXIV; ARXIV

Published by AAAS

48 hours of submission. Viral genomes posted on a platform named GISAID, more than 200 so far, are analyzed instantaneously by a phalanx of evolutionary biologists who share their phylogenetic trees in preprints and on social media.

"This is a very different experience from any outbreak that I've been a part of," says epidemiologist Marc Lipsitch of the Harvard T.H. Chan School of Public Health. The intense communication has catalyzed an unusual level of collaboration among scientists that, combined with scientific advances, has enabled research to move faster than during any previous outbreak. "An unprecedented amount of knowledge has been generated in 6 weeks," says Jeremy Farrar, head of the Wellcome Trust.

Sluggish scientific communication has often been a problem during past outbreaks. Researchers sometimes sat on crucial data until a paper was accepted by a high-profile, peer-reviewed journal, because they were worried competitors might run with them. Even if researchers were willing to share their findings early, there wasn't a natural platform to do so.

Lipsitch realized a few years ago that preprint servers, which publish findings prior to peer review, could change that. Scientists could post fresh data rapidly and still get some credit, regardless of where the work was ultimately published. In a 2018 paper, he and others concluded that preprints sped up data dissemination during the Zika epidemic of 2015–16 and the West African Ebola outbreak of 2014–16. Most of the preprints appeared more than 100 days before a journal published the work. But overall, less than 5% of the journal articles about the two epidemics were first posted as a preprint.

The COVID-19 outbreak has broken that mold. Early this week, more than 283 papers had already appeared on preprint repositories (see graphic, p. 963), compared with 261 published in journals. Two of the largest biomedical preprint servers, bioRxiv and medRxiv, "are currently getting around 10 papers each day on some aspect of the novel coronavirus," says John Inglis, head of Cold Spring Harbor Laboratory Press, which runs both servers. The deluge "has been a challenge for our small teams … [they] are working evenings and weekends."

Much of that work, done by staff and outside scientists, involves screening the submissions to weed out pseudoscience and opinion pieces. The manuscripts that make it through vary wildly in quality, says University of Hong Kong epidemiologist Keiji Fukuda. "Some of them are not that helpful and some of them are extremely helpful." Lipsitch calls it "a firehose." Anthony Fauci, head of the U.S. National Institute of Allergy and Infectious Diseases, says he's so busy that he often reads preprints late at night. "Eleven o'clock, 12 o'clock comes and you have 25 of these things to read," Fauci says. "You can't ignore them." But sometimes, "It gets a little confusing what you can really believe."

That's even harder for journalists and the public at large. A 31 January preprint on bioRxiv by scientists in India pointed to "uncanny" similarities between SARS-CoV-2, the virus that causes COVID-19, and HIV, fueling conspiracy theories about genetic engineering. The paper was widely discussed on Twitter and covered by some news outlets—even though some scientists immediately said it was flawed. Inglis points out that the paper received 90 critical comments within 48 hours and was swiftly retracted. (A formal paper debunking the findings was published in *Emerging Microbes and Infections* 2 weeks later.)

Still, such data are becoming part of an "infodemic" of bad information, says virologist Marion Koopmans of Erasmus Medical Center, and the science community needs to debate how to deal with it. "There has been strong advocacy for open science, open data," she says. "OK, this is open science, open data. Now, what do we do?" BioRxiv and medRxiv have both put up prominent notices emphasizing the preliminary nature of the information in preprints. "We urge journalists to include in their reporting the caveats about the use of the information," Inglis says.

Still, Farrar says the benefits of rapid information sharing far outweigh the disadvantages. Moreover, even peer review by a top journal isn't a guarantee that a claim is correct. A 30 January *NEJM* paper that suggested a Chinese woman who showed no symptoms of COVID-19 had transmitted the virus to people in Germany later came under heavy criticism because it turned out the authors had not actually spoken to the woman. A later interview showed that she had some symptoms; the journal has added that information as an appendix.

NEJM Editor-in-Chief Eric Rubin concedes there is a tension between rigor and speed. The journal's review process for COVID-19 papers, he notes, is basically the same as always but much faster. "We and authors could do a more careful job if we had more time," he wrote in an email. "But, for now, physicians are dealing with a crisis and the best quality information available quickly is better than perfect information that can't be accessed until it's not helpful."

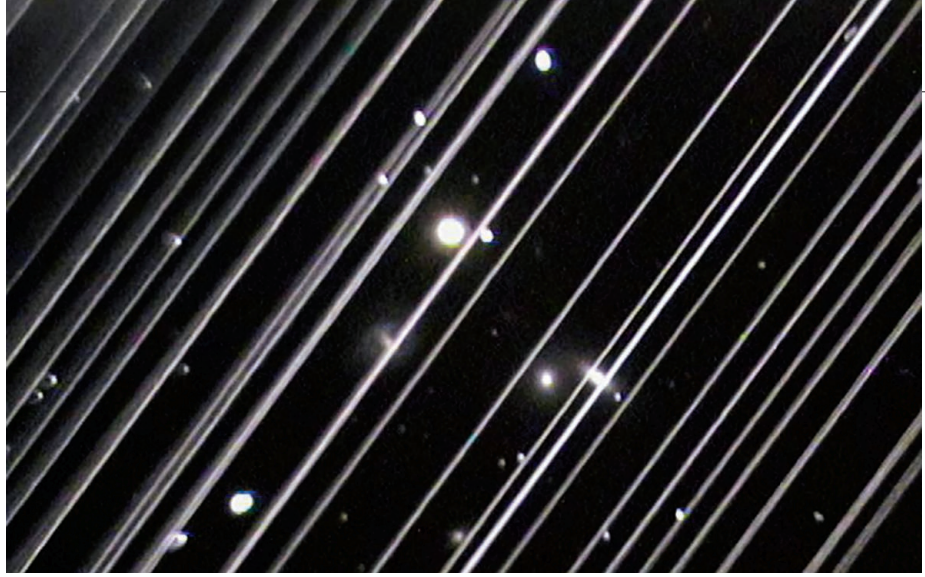
To speed up research, it's also crucial to share things that don't work, O'Connor says—for instance, when experiments show an animal species can't be infected with the novel virus. "That's important information that is not typically shared through traditional channels," he says, which is why groups such as the Wu-han Clan are so handy. Its members also discussed whether to infect animals the traditional way, by putting a liquid virus suspension in their nose, or through an aerosol, a new way of exposure that more closely resembles a sneeze. (They will probably try both.) "By openly sharing plans, we can reduce redundancy," Friedrich says.

It's not clear whether such scientific collaborations will help mitigate the worldwide blow from COVID-19. But many scientists welcome the way the outbreak has already changed the way they communicate. "It feels like things are transitioning to a completely new culture of doing research," says virologist Isabella Eckerle of the Geneva Centre for Emerging Viral Diseases. "It's exciting." ■

Researchers at the Pasteur Institute in Lille, France, at work on the new coronavirus on 20 February.

PHOTO: SYLVAIN LEFEVRE/GETTY IMAGES

Published by AAAS



At twilight, streaks from Starlink satellites could ruin about one-third of Vera C. Rubin Observatory images.

PHOTO: VICTORIA GIRGIS/LOWELL OBSERVATORY

ASTRONOMY

Satellite megaconstellations menace giant survey telescope

Darkening the orbiters could limit damaging light trails in wide-field images from the Vera C. Rubin Observatory

By **Daniel Clery**

Starting in 2022, the Vera C. Rubin Observatory will survey the entire night sky every few nights from a mountaintop in Chile, using a giant mirror to capture faint, fast-changing objects. But something much more mundane is likely to streak into view: thousands of low-flying communications satellites glinting in the Sun. The threat, evident ever since rocket company SpaceX launched its first batch of 60 Starlink satellites in May 2019, has come into sharper focus—as has a potential way to limit the damage.

Unpublished studies, some seen by *Science*, suggest satellite trails could ruin about one-third of the images from the Rubin Observatory during parts of the night. Observatory staff say the problem cannot be avoided with software tweaks or shrewd telescope pointing. The best remedy seems to be to darken the satellites themselves, something SpaceX and the Rubin Observatory are already exploring.

"Darkening is the name of the game," says Tony Tyson, chief scientist of the Rubin Observatory. "I'm cautiously optimistic we will get there with SpaceX." But with other companies starting to build similar megaconstellations, astronomers are racing to keep ahead of the problem.

After the first launch of the Starlinks, which aim to provide internet access to remote parts of the globe, skywatchers could see strings of bright dots with the naked eye as the table-size satellites spread out 290 kilometers up. "We were totally surprised by the brightness," says Patrick Seitzer of the University of Michigan, Ann Arbor, a member of a working group set up by the American Astronomical Society to study the problem.

Just how big the problem will become is hard to predict. Companies keep tight hold of spacecraft design details, says aerospace engineer Hugh Lewis of the University of Southampton, who has adapted a space debris model to help understand the brightness of Starlinks. "Ideally we would want to know everything about the spacecraft, but that's not the world we live in," he says. Astronomers also don't know how big the constellations will be: SpaceX, for example, is aiming for an initial constellation of about 1600 but has applied to the Federal Communications Commission to loft as many as 42,000.

Brightness measurements made by Las Cumbres Observatory, a global network of small telescopes, and others showed that the Starlinks dimmed as they rose to their working orbits 550 kilometers up. They are brightest when they're low on the horizon, around twilight. Most telescopes can avoid the trails, because they observe high in the sky and have a narrow field of view. But long exposures, a large mirror, and a wide field of view all make a telescope vulnerable. "Combine two or three of these and you're in trouble," says Olivier Hainaut, who is studying the problem for the European Southern Observatory.

The Rubin Observatory, with an 8.4-meter mirror that will take pictures of the sky the size of 40 full Moons in 30-second exposures, "is the perfect machine for running into these satellites," Tyson says.

He and his team conducted simulations that suggested the track of a satellite image across their camera would saturate each camera pixel as it passes, and cause leakage into neighboring ones. The resulting artifacts "cannot be removed in software. We have failed in doing that," Tyson says. The team looked at altering schedules to avoid satellite trails, but with such a wide field of view, avoiding thousands of satellites would end up as "a wild goose chase," he says.

So Tyson is pinning his hopes on SpaceX darkening its future satellites. He and his team speak several times a week with engineers at SpaceX, which launched one darkened satellite in January that is just now reaching its final orbit. Tyson's team calculated that if the company can reduce reflections by a factor of 15, the issue will be manageable. Images would still contain trails, but they wouldn't saturate pixels and could be removed digitally. SpaceX and its chief, Elon Musk, are "totally committed to solving this problem," Tyson says, and his team has worked with them to "narrow to a design that may work." Several satellites with this updated dark design will be launched in coming weeks. SpaceX did not respond to requests for comment.

Starlink is, however, just the first constellation to get started. "It's safe to say, the number will rise dramatically," says Connie Walker of the National Optical-Infrared Astronomy Research Laboratory, who chairs an International Astronomical Union commission looking at the threat. The company OneWeb launched 34 satellites on 7 February—a first installment in an initial constellation of 648. The satellites are smaller and orbit higher than Starlinks, so they are fainter, but their altitude means they're illuminated by the Sun most of the night. "They'll need to darken them," says Tyson, who is in contact with OneWeb and other operators, including Amazon, whose Project Kuiper envisions 3200 satellites.

Tyson has his fingers crossed. The constellations won't kill the Rubin Observatory, but they will make its job harder and could jeopardize its chances of making discoveries. "It's there that the existential threat is," he says. ■

Published by AAAS

NEUROSCIENCE

Safety benefits of 'biased' opioids scrutinized

Mouse studies challenge premise of efforts to reduce painkillers' fatal side effect

By **Kelly Servick**

The dark side of opioids' ability to deaden pain is the risk that they might kill their user. The same brain receptors that blunt pain when drugs such as morphine or oxycodone bind to them can also signal breathing to slow down. It's this respiratory suppression that causes most overdose deaths.

So scientists have hoped to design opioids that are "biased" toward activating painkilling signals while leaving respiratory signaling alone. Several companies have cropped up to develop and test biased opioids (*Science*, 17 November 2017, p. 847). But two new studies in mice contest a key hypothesis underlying these efforts—that a signaling protein called beta-arrestin2 is fundamental to opioids' effect on breathing.

"It seems like the premise was wrong," says Gaspard Montandon, a neuroscientist and respiratory physiologist at the University of Toronto. He and others doubt that the good and bad effects of opioids can be disentangled.

Hopes first arose in the late 1990s and early 2000s, as neuroscientist Laura Bohn, biochemist Robert Lefkowitz, and colleagues at Duke University explored the cascades of signals triggered when a drug binds to mu-opioid receptors on a neuron. This binding changes the receptor's structure and its interactions with two types of proteins inside the cell—signaling molecules known as G-proteins, and beta-arrestins, which, among other effects, inhibit G-protein signaling.

It's still not clear how the resulting signal cascades influence cells or brain circuits. But the researchers reported in 1999 that mice engineered to lack the gene for beta-arrestin2 got stronger and longer lasting pain relief from morphine. And in 2005, Bohn and her colleagues found that two morphine-induced side effects, constipation and slowed breathing, were dramatically reduced in these "knockout" mice. The findings suggested that a drug able to nudge the mu-opioid receptors toward G-protein signaling and away from beta-arrestin2 signaling would prompt more pain relief with fewer risks.

A hunt for biased opioids ensued. A commercial front-runner was Trevena, which tested its intravenous drug candidate oliceridine for postsurgical pain. But at high doses, the compound failed to show significantly better respiratory safety than morphine. In 2018, the U.S. Food and Drug Administration rejected oliceridine. Trevena resubmitted its application with additional data this month.

But some researchers question the mouse studies that inspired the hunt. Last year in *Nature Communications*, a group including pharmacologists Andrea Kliewer and Stefan Schulz at Friedrich Schiller University of Jena described experiments with mice carrying mutations in the mu-opioid receptor gene that prevented the receptor from interacting with beta-arrestin2. When given morphine, these animals had constipation and suppressed breathing that was sometimes more severe than in mice without a mutation.

Since then, the Jena team and labs in Bristol, U.K., and Sydney have tried to replicate the 2005 finding directly using beta-arrestin2 knockout mice. All three labs observed opioid-induced constipation and breathing side effects similar to those in mice that could produce beta-arrestin2. The idea that G-protein–biased opioids would be safer was "too easy," Kliewer says.

The results, published 12 February in the *British Journal of Pharmacology*, are "quite convincing," says Charles Chavkin, a pharmacologist at the University of Washington, Seattle. Opioid side effects are unlikely to rely solely on beta-arrestin2, he says.

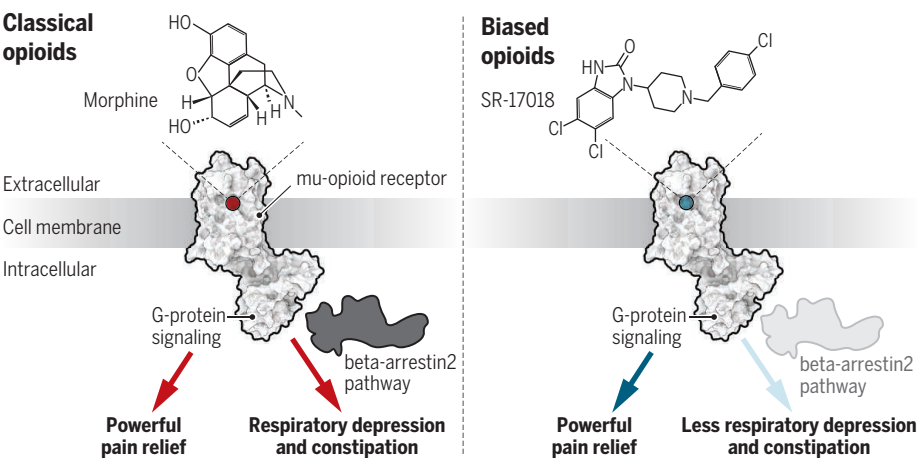
Perhaps the simplest explanation for the conflicting 2005 and 2020 results is that the mice were genetically different. Mice in the initial studies were a mix of multiple strains; further crossing and inbreeding may have changed the way they respond to opioids. "I have every bit of confidence in the early data," says Bohn, now at Scripps Research. "There's no way for us to go back and have that same mix of animal that we had 20 years ago."

Bohn says her original findings were "oversold and oversimplified" by commercial interests. "G doesn't stand for good and beta doesn't stand for bad," she says. Beta-arrestin2 knockout mice were an important starting point, she says, but they're an imperfect model of drug response, in part because they may have somehow compensated to survive without this key protein. Her group is now creating new biased opioids and comparing them with traditional opioids in animal studies that measure breathing suppression and other negative effects such as the development of tolerance and dependence.

"I believe both sets of data, because both groups are careful," says Bryan Roth, a molecular pharmacologist at the University of North Carolina, Chapel Hill, who helped develop a biased opioid known as PZM21 and holds stock in a company that licensed it. Resolving the role of beta-arrestin2 will require more strongly biased opioids, he says. "If the [biased opioid] hypothesis is true, it would be tremendously beneficial for the human race," he says. "It's a hypothesis that should be tested." ■

Accentuate the positive

Researchers have long hoped that opioids that favor G-protein signaling would be safer (below, right), but studies of mice genetically engineered to mimic the effects of that bias have raised doubts.



GRAPHIC: KATHERINE SUTLIFF AND V. ALTOUNIAN/*SCIENCE*

Published by AAAS

PHOTO: SUPERBIT COLLABORATION

Last year, in an 18-hour balloon flight over Canada, SuperBIT held steady and took sharp pictures.

ASTRONOMY

Cheap balloon-borne telescopes aim to rival space observatories

Stable pointing system enables SuperBIT to deliver images almost as sharp as those from the Hubble Space Telescope

By **Adrian Cho**

No instrument has revealed more about the heavens than NASA's Hubble Space Telescope, which cost nearly $5 billion to build and launch. Yet a team of researchers thinks it can match some of Hubble's capabilities with a telescope costing a mere $2 million, hoisted to the edge of space on a balloon. On a September 2019 flight, the Superpressure Balloon-borne Imaging Telescope (SuperBIT) demonstrated the ability to hold steady and image distant stars with exquisite resolution approaching Hubble's, the team is now reporting.

"Like, this actually works," says Barth Netterfield, a SuperBIT member and an astronomer at the University of Toronto. "That's pretty awesome." Richard Ellis, an astronomer at University College London who is not involved in the project, says low-cost, quick-turnaround balloon telescopes open new opportunities. "If somebody comes up with a brilliant idea, a balloon can go for it right away," he says.

Orbiting above Earth's obscuring atmosphere, Hubble sees with a clarity limited only by the size of its 2.4-meter mirror. A balloon-borne telescope floating several tens of kilometers up, above 99% of the atmosphere, would enjoy a similar vantage—if scientists could keep it stably pointed while the balloon drifts and turns. The SuperBIT team, split mainly between the University of Toronto and Princeton University, has conquered that problem, it reports in a paper in press at the *Review of Scientific Instruments*.

To fix on a spot on the sky, the telescope must slowly swivel on all three axes as it dangles from the balloon. SuperBIT employs state-of-the-art bearings and brushless electric motors to make that motion smooth. To orient itself, it refers to gyroscopes, images of the sky, and guide stars. During its 18-hour flight in northern Canada in September 2019, SuperBIT flew so stably that it achieved a resolution of about 260 milliarcseconds, only a factor of five less sharp than Hubble and at the theoretical limit for its 0.5-meter mirror.

To compete with space telescopes, however, a balloon-borne telescope must also stay aloft for many nights. That's where new superpressure balloons come in. To avoid rupturing, an ordinary research balloon must vent helium during the heat of the day, as it expands and rises, before sinking at night. After a few days, the bobbing balloon has too little helium to stay aloft. A sturdier superpressure balloon maintains a steady volume, doesn't vent helium, and cruises at constant altitude for weeks (*Science*, 23 June 2017, p. 1227). In 2009, a NASA superpressure balloon recorded a 54-day flight, says Thomas Hams, program scientist for NASA's balloon program. "We are on the cusp where we will soon see these flights more routinely," he says.

NASA plans to launch one superpressure balloon per year, and SuperBIT researchers hope to ride on one in 2021. That flight would trace the distribution of mysterious dark matter in the universe, says Richard Massey, a SuperBIT co-leader and a cosmologist at Durham University in the United Kingdom. Galaxies are thought to reside within vast clumps and strands of dark matter, whose gravity distorts the images of more distant galaxies, creating correlations among the orientations of the tiny ellipses. "It's a bit like looking through a textured bathroom window at the street lights," Massey says. Known as weak lensing, such distortions can reveal the distribution of dark matter and help trace the universe's evolution.

Massey has done similar work with Hubble. But with its wider-field camera, SuperBIT can cover more sky than Hubble. It could also fly before space telescopes with similar goals, such as the European Space Agency's Euclid, which could launch in 2022.

Balloon-based telescopes present their own challenges. In Antarctica, NASA only launches balloons during the summer, when the weather is milder, but the constant daylight rules out optical astronomy. So SuperBIT's scientific flight would have to launch from midlatitudes, probably from a NASA facility in New Zealand, where it would spend most of its time over the ocean. That makes it difficult to download data to cell towers, so the SuperBIT team has tested a system for dropping hard drives when the payload is over land. A balloon-borne telescope also takes a beating when it lands. "In the flight we just had, it got dragged through a forest and took out 50 trees," Netterfield says.

Balloon-borne telescopes may be limited to niche applications, Ellis says, but the SuperBIT team has grander ambitions. It is working on a 1.5-meter successor to launch after the 2021 flight. For roughly $90 million, it could serve as a partial replacement for Hubble, which launched in 1990 and likely won't last more than another decade, says William Jones, a SuperBIT member and an astrophysicist at Princeton. "A lot of people want the capabilities that Hubble has," he says. "Once you've demonstrated that this works it would make sense to fly it at every opportunity." ■

Published by AAAS

Teotihuacan (left) was once a bustling, cosmopolitan metropolis. Its empire may have included Tikal, an important Maya city 1000 kilometers away (right).

PHOTOS: (LEFT TO RIGHT) MAX SHEN/GETTY IMAGES; W. E. GARRETT/NATIONAL GEOGRAPHIC IMAGE COLLECTION

THE ARRIVAL OF STRANGERS

New evidence points to a clash between two ancient Mesoamerican cultures, Teotihuacan and the Maya

By **Lizzie Wade**, *in San Juan Teotihuacan, Mexico*

On 16 January 378 C.E., a stranger arrived in Tikal, a large Maya city in what is now northern Guatemala. His name was Sihyaj K'ahk' (SEE-yah Kak), or Fire is Born, and he was likely a mighty warrior from a distant land. Many archaeologists think he hailed from Teotihuacan, a metropolis of 100,000 people about 1000 kilometers northwest of Tikal, near today's Mexico City. And he may have come with an army.

The stone Maya monuments that record Sihyaj K'ahk's arrival don't say why he came or how he was received by Chak Tok Ich'aak, or Jaguar Paw, the long-reigning king of Tikal. But the day Sihyaj K'ahk' marched into the city was the day Jaguar Paw died.

The engravings suggest Sihyaj K'ahk' had been sent by a powerful foreign ruler called Spearthrower Owl. Within 2 years, Spearthrower Owl's young son was crowned the new king of Tikal. In portraits carved on stone monuments there, the new king, named Yax Nuun Ayiin, holds an atlatl, a spearthrower used by Teotihuacan warriors, and wears a Teotihuacan-style headdress adorned with tassels. Some images of him and his father on monuments at Tikal are even carved in the flat, geometric style of Teotihuacan art, distinct from the intricate, naturalistic portraits of the Maya. Under the exotic new king and his descendants, Tikal became one of the most powerful cities in the Maya region.

Archaeologists have known the outline of those events for decades, but have long debated their meaning. Now, new evidence from both Teotihuacan and the Maya region has brought the relationship between those two great cultures back into the spotlight—and hints it may have been more contentious than most researchers had thought.

Evidence from Maya writing and art suggests Teotihuacan conquered Tikal outright, adding it to what some archaeologists see as a sweeping empire that may have included several Maya cities. Defaced art in Teotihuacan suggests that about the time Tikal fell under its sway, Teotihuacan may have turned against Maya expatriates who had lived there peacefully for decades.

But doubts about that narrative persist, underlining the challenge of interpreting the archaeological traces of empires that fell short of complete domination. Some researchers say the events of 378 may

Published by AAAS

have been a more limited case of palace intrigue, with the nobles of one powerful region elbowing their way into the politics of another. Archaeologists might even be falling for ancient propaganda: Sihyaj K'ahk' and his army may have been local Maya usurpers who appropriated the symbolism of faraway Teotihuacan. Either way, archaeologists say they are glimpsing a political and cultural collision that helped spark the flourishing of Tikal in the centuries to come.

"It's a thrilling time to be working in this area," says Stephen Houston, an archaeologist at Brown University. "We're getting astounding new finds that amplify what had just been a sketched story before."

MAYA TRAVELERS visiting Teotihuacan during the fourth century would have encountered a city like no other they had ever seen. Three enormous pyramids loomed over the main street, now known as the Avenue of the Dead, their shapes reflecting snow-capped volcanoes visible in the distance. An orderly grid of roads extended from the avenue, and the city's 100,000 residents—far more than in even the largest Maya cities of the time—lived in comfortable, standardized apartment complexes. Economic inequality was strikingly low. Depictions of warriors in Teotihuacan's art, as well as human sacrifices entombed in military regalia, spoke of the city's military might. Merchants from far-flung places such as Oaxaca to the southeast and the Gulf Coast brought goods for Teotihuacan's markets, and pilgrims flocked to the city for religious ceremonies.

Some of those foreigners settled here and set up ethnic enclaves that archaeologists can identify from their foreign household goods and burial practices. "Teotihuacan was a great urban center, almost like Los Angeles or New York City. People from all over Mesoamerica were there," says Karl Taube, an archaeologist at the University of California (UC), Riverside.

Teotihuacanos were likely just as fascinated by the Maya area, about 1000 kilometers away in what is now southern Mexico, Guatemala, Belize, and Honduras. It lay as far to the east as one could get in Mesoamerica, linking it to the mythologically potent rising Sun. Although the cultures shared staples such as maize, the luxury goods prized in Teotihuacan, such as jade, cacao, and brightly colored quetzal feathers, all came from the tropical jungles of the Maya lowlands. "It was a source of wealth and abundance," Taube says. When seen from the chilly, high-altitude plain of Teotihuacan, the lush Maya area would have looked like a paradise replete with elegance and luxury.

Ceramics decorated with naturalistic Maya art (top) were used in a feast at Teotihuacan. In Tikal, a portrait of Spearthrower Owl (bottom), a possible leader of Teotihuacan, was carved on a monument, in Teotihuacan's geometric style.

Diplomacy and trade with the Maya could be tricky, however, because the Maya area was politically fragmented. It was dotted with largely independent city-states knitted together by shared religion and culture, similar to ancient Greece. The most powerful, such as Tikal and its nearby rival Calakmul, commanded the loyalty of smaller cities. But alliances shifted constantly, and no Maya king ever managed to politically unite the entire 390,000-square-kilometer region. Teotihuacan likely had distinct and ever-changing relationships with different Maya cities.

Their interactions left plenty of traces, in exchanges of art, ceramics, and cultural influences. Radiocarbon dating, as well as the exact dates the Maya recorded on their monuments, show definitively that the cultures existed at the same time. Their interactions were most intense in the fourth and fifth centuries C.E. (see timeline, p. 971), the time of the late Roman Empire and part of what archaeologists in Mesoamerica call the Early Classic period. What archaeologists disagree on, often vehemently, is whether that relationship was peaceful and reciprocal or was based on violence and domination.

ON A SUNNY SUMMER morning here, Nawa Sugiyama, an archaeologist at UC Riverside, ducks into a tunnel her team has dug under what was once an impressive pyramid. Just off the Avenue of the Dead and between the imposing Sun and Moon pyramids, the structure sits in what is now called the Plaza of the Columns. (Confusingly, it has no columns and consists of several interconnected plazas and large pyramids.) Crouching under the tunnel's low ceiling, Sugiyama inspects dozens of pieces of broken ceramics painstakingly excavated by her students and the project's workers.

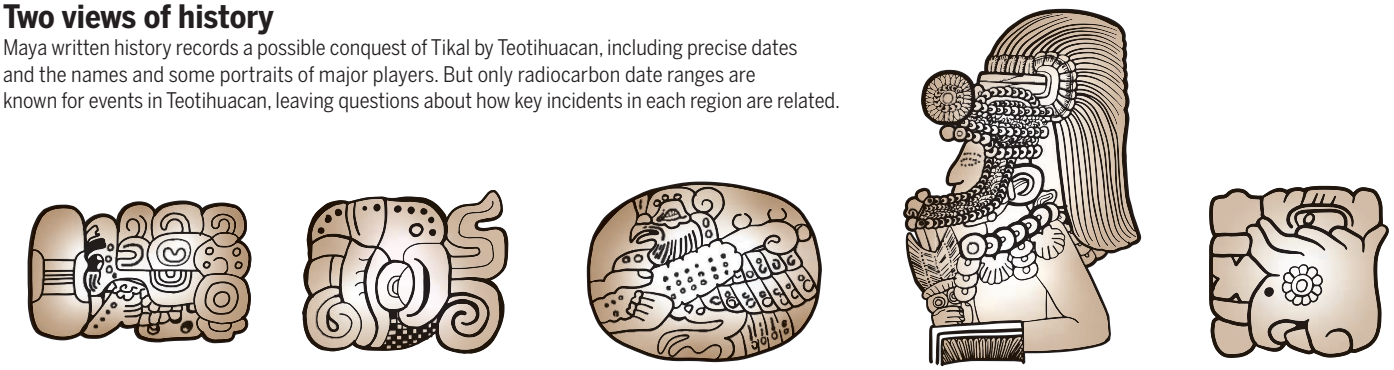
A mix of Maya and Teotihuacan styles, the shards testify not to violence, but to celebration: After the ceramics were broken, they were ceremonially sprinkled into a pit in a type of offering commonly made at the end of a feast in ancient Mesoamerica. The students and workers have excavated more than 10,000 ceramic pieces from this single spot, and this season they uncovered an average of 250 a day. "I've never seen anything like it," Sugiyama says. "We're a little worried that it will never end."

PHOTOS: (TOP TO BOTTOM) CORTESÍA/NOTIMEX/NEWSCOM; KENNETH GARRETT

Published by AAAS

Two views of history

Maya written history records a possible conquest of Tikal by Teotihuacan, including precise dates and the names and some portraits of major players. But only radiocarbon date ranges are known for events in Teotihuacan, leaving questions about how key incidents in each region are related.



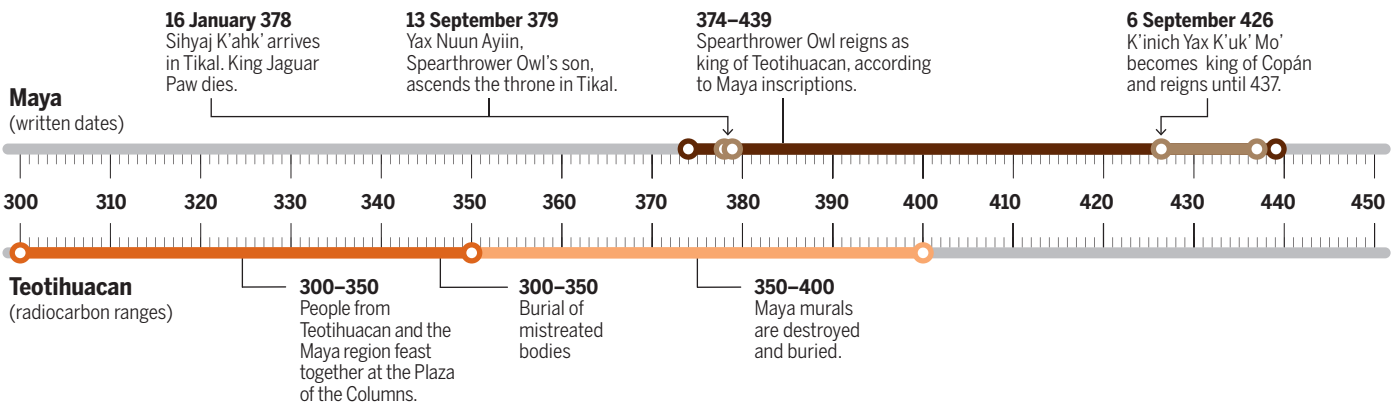
Jaguar Paw (name glyph)

Sihyaj K'ahk' (name glyph)

Spearthrower Owl (name glyph)

Yax Nuun Ayiin (portrait)

K'inich Yax K'uk' Mo' (name glyph)



CREDITS: (GRAPHIC) N. DESAI/*SCIENCE*; (ILLUSTRATIONS, LEFT TO RIGHT) MESOWEB.COM VIA S. MARTIN *ET AL.*, *CHRONICLE OF THE MAYA KINGS AND QUEENS: DECIPHERING THE DYNASTIES OF THE ANCIENT MAYA*. THAMES AND HUDSON, 2000 (2); JUAN PEDRO LAPORTE VIA H. G. CAPISTRÁN, "BÚHOS, LANZADARDOS Y ANTEOJERAS, ELEMENTOS TEOTIHUACANOS EN TIKAL, ¿PRESENCIA O INFLUENCIA?," THESIS, NATIONAL AUTONOMOUS UNIVERSITY OF MEXICO, 2007; JOHN MONTGOMERY VIA THE FOUNDATION FOR THE ADVANCEMENT OF MESOAMERICAN STUDIES, 2000; DAVID STUART

Sugiyama and her team think Teotihuacanos and Maya guests mingled at that ancient feast, perhaps to commemorate completing the pyramid. Most of the ceramic pieces represent fancy servingware, like the fine china people today might bring out for guests. According to radiocarbon dating of burned food scraps, including rabbit bones, maize, and yucca from the tropical Maya region, the feast took place between 300 and 350 C.E.

Across the plaza from the pyramid, Sugiyama and her collaborators have uncovered a plush compound of buildings once decorated with elaborate Maya murals painted in vivid hues, such as blues and greens, not often seen in Teotihuacan art. Perhaps the Maya people who lived in the Plaza of the Columns were high-status diplomats or members of noble families sent to the capital, like the European nobles who lived or grew up in foreign courts to strengthen alliances and facilitate royal marriages, says David Carballo, an archaeologist at Boston University.

"They're practicing their own customs, which speaks to a peaceful coexistence with the rest of Teotihuacan society," says Verónica Ortega, an archaeologist at Mexico's National Institute of Anthropology and History and a co-director with Sugiyama of the Plaza of the Columns project.

But several decades after the feast, something changed. When the team found the elegant murals, they were no longer attached to the compound's walls, as much of Teotihuacan's art still is. The murals had been smashed to pieces and deeply buried—"absolutely obliterated," Sugiyama says. Faces were cut and scratched off until they were unidentifiable. "It was an act of intentional destruction," Sugiyama says. According to radiocarbon dating of organic matter covering the remains of the murals, the destruction took place between 350 and 400 C.E.

Although Sugiyama and Ortega work together, they interpret the mural destruction differently. Ortega sees it as a ritual that Teotihuacanos and Maya people both participated in—similar to the offering of broken ceramics made at the end of the feast. But Sugiyama points out that scratching out individual faces is an unusual act of targeted erasure that the Maya residents would have been unlikely to embrace.

Sugiyama and Ortega's team also found a nearby pit filled with human skeletons that raises darker questions. The bodies lie in pieces, which is not typical of other burials or sacrifices here. The bone pit could have been simply a workshop for making bone tools—or the remains of a massacre, Sugiyama says. Some skulls have flat backs and slightly pointed tops, and some teeth have holes for jewelry—signs of cranial shaping and adornment styles practiced by the Maya but uncommon in Teotihuacan. Archaeologists will need to study the dietary isotopes and perhaps DNA from the bones to be sure they belonged to Maya people. Researchers would also like to resolve another mystery: Preliminary dating suggests the bones were dumped in the pit about the time of the feast, when relations with the Maya were apparently peaceful.

The radiocarbon dates for the mural destruction tell a clearer story, however. They place it between 350 and 400—within about 25 years of the arrival of Sihyaj K'ahk' in Tikal in 378. "The fact that [the Teotihuacanos] absolutely destroy the murals and then soon after go attack places in the Maya lowlands suggests to me that diplomatic relations had turned sour for some reason," Carballo says. "Turbulent times are coming," Sugiyama agrees.

WHEN SIHYAJ K'AHK' arrived in Tikal, he would have found a smaller, less centralized city than Teotihuacan. Royal palaces and temples perched on hilltops still surrounded by jungle below. Roads cut through the trees, connecting clusters of buildings used by the elite and serving as routes for commoners to follow from their

Published by AAAS

scattered farms to markets and ceremonies in the city center. Tall stone monuments covered in dense writing documented key events in Tikal's history, such as the reigns of kings. Only the city's richest and most powerful people could have read those texts, however: Maya history was written by elites, for elites. (Teotihuacan's script remains undeciphered, partly because researchers don't know its language, whereas ancient Mayan writing is related to a few Mayan languages spoken today.)

In the early 1970s, epigrapher Tatiana Proskouriakoff became the first person in centuries to begin to piece together what happened in Tikal in 378. On the basis of an incomplete reading of the monuments recording the arrival of Sihyaj K'ahk', she spoke of "the arrival of strangers" and proposed they were from central Mexico.

In 2000, David Stuart, an archaeologist and epigrapher at the University of Texas, Austin, offered a more complete understanding of those texts, and he is reanalyzing them now. Thanks to advances in deciphering Mayan script, he can read the glyphs carved into the monuments, including the names and relationships of Sihyaj K'ahk', Jaguar Paw, and Spearthrower Owl. But the historical records raise more questions than they answer.

One especially hot question is whom, exactly, Sihyaj K'ahk' was working for. He apparently was following the orders of Spearthrower Owl, described on the monuments as a foreign king who ruled a faraway land from 374 to 439 C.E. Spearthrower Owl's name is written in a style that echoes Teotihuacan art, and a portrait of him at Tikal was carved in Teotihuacan's unmistakable geometric style, Stuart says. Stuart thinks Spearthrower Owl was the king of Teotihuacan, possibly when the murals at the Plaza of the Columns were destroyed.

But many archaeologists think Teotihuacan had no king. No royal tomb or any depiction of a monarch has ever been found there. Although some researchers argue that only a strong monarch could have ruled such a grand and regimented city, others assert that the city was governed by a council or in some other cooperative way. Carballo points out that most Teotihuacan art focuses on people's clothing and other accoutrements rather than individual features, a sign that the offices they held were more important than their individual identities. Images of birds of prey with atlatls—the key elements of the glyph the Maya used to write Spearthrower Owl's name—show up around Teotihuacan for centuries, far longer than any single person could have ruled. "I think [the glyph] stands for an office, perhaps a military role of some sort in Teotihuacan" that many people could hold over time, Carballo says, rather than an individual monarch.

He and many other archaeologists working in Teotihuacan resist the notion that Maya written history should trump evidence from Teotihuacan itself. Maya cities were ruled by kings, and the Maya expectation of monarchy may have led them to misunderstand Spearthrower Owl's role, Carballo says. Perhaps Sihyaj K'ahk' and the other invaders even promoted that misconception, says Michael Smith, an archaeologist at Arizona State University, Tempe. "Say you're the spin doctor for Sihyaj K'ahk', and you're trying to convince these Maya kings that this guy is really something. What are you going to say? Are you going to say he's from Teotihuacan, where people sort of ruled themselves? Or are you going to say, 'This guy is sent by the king of the biggest city that anyone's ever heard of'?"

Whoever sent Sihyaj K'ahk', Francisco Estrada-Belli, an archaeologist at Tulane University, thinks he didn't stop with Tikal. Estrada-Belli found murals at the city of Holmul, 35 kilometers east of Tikal, showing Teotihuacan warriors accompanying a new king during his ascension to the throne. The building they decorated was constructed to commemorate the first anniversary of Sihyaj K'ahk's arrival in Tikal. Maya records imply that "within a few years, Sihyaj K'ahk' had installed friendly kings at a number of important Maya cities," Estrada-Belli says. "For many of them, this is the beginning of a new dynasty. It's a major turning point."

THE POSSIBILITY REMAINS, however, that Sihyaj K'ahk' and Spearthrower Owl weren't from Teotihuacan at all and were simply invoking that great city to impress a Maya audience. Maya mythology and religion held foreign goods in high esteem, and Teotihuacan was the most prestigious faraway place in Mesoamerica. Little evidence exists of Teotihuacanos living at Tikal, notes Joyce Marcus, an archaeologist at the University of Michigan, Ann Arbor. "The simplest explanation is Tikal's new king"—Spearthrower Owl's son, Yax Nuun Ayiin—"was a Maya usurper who cloaked himself in prestigious foreign attire," she says. "Wearing the trappings of highland Mexican warriors could communicate that the Maya leader had military prowess."

"Conquest is exciting and easy to understand," says Geoffrey Braswell, an archaeologist at UC San Diego. But he, too, thinks the events at Tikal reflect a conflict between Maya groups, one of which adopted the symbols of a foreign power to strengthen their rebellion. Elite groups have practiced similar cultural emulation throughout history, he says. Chinese porcelain became a status symbol in 17th and 18th century Europe, and upper-class Russians spoke French to each other in the 19th century, for example. "It's really quite common," Braswell says.

Isotopic evidence published in 2005 and 2010 seems to be on Marcus and Braswell's side. Archaeologists excavated the tombs of Yax Nuun Ayiin at Tikal and K'inich Yax K'uk' Mo', the founder of a new dy-

Spheres of influence

In the fourth century C.E., Teotihuacan controlled a small empire in central Mexico, perhaps with outposts farther south. It traded with the independent city-states of the Maya region, and may have conquered several.

MAP: X. LIU/SCIENCE

Published by AAAS

nasty at Copán, a Maya city in Honduras. That Copán monarch is depicted wearing Teotihuacan-style dress, including unmistakable "goggles" over his eyes that evoke the rain god of central Mexico. Inscriptions say he was a foreign king and came to Teotihuacan for a ceremony that invested him with the right to rule before assuming Copán's throne. If any Maya kings came from Teotihuacan, it would be those two, Braswell says.

But analysis of the strontium isotopes preserved in his teeth showed that Yax Nuun Ayiin—whom inscriptions clearly name as Spearthrower Owl's son—grew up around Tikal. Researchers couldn't pinpoint the origins of K'inich Yax K'uk' Mo', but in his case, too, the isotopes ruled out central Mexico. "Find us a body [in the Maya lowlands] that's isotopically from Teotihuacan, with a spear in their hand," Braswell says, and he'd be more inclined to think that Teotihuacan conquered Tikal.

Edwin Román Ramírez is looking. An archaeologist at the Foundation for Maya Cultural and Natural Heritage (PACUNAM), he's leading new excavations at Tikal and searching for an ethnic enclave of Teotihuacanos. He expects to announce his first results at a symposium in Guatemala City this summer. He thinks Teotihuacan did conquer Tikal and that soldiers or others from Teotihuacan may have lived there. But, he says, "Their intention, I think, was never to turn [the Maya] into Teotihuacanos." Rather, Tikal likely represented a strategic economic outpost in the Maya region for Teotihuacan.

In fact, any Teotihuacan empire may have relied more on soft power than on overt colonization. The lives of Maya commoners in and around Tikal don't seem to change much after Sihyaj K'ahk's arrival, says Bárbara Arroyo, an archaeologist at Francisco Marroquín University, leading her to question the conquest scenario. Compare that situation with that of the Roman Empire from 27 B.C.E. to 476 C.E. Its imperial footprint is unmistakable, as it posted armies all over Europe, encouraged the adoption of a state religion, remodeled cities, and installed governors who answered directly to Rome.

"Empire is a spectrum," says Sue Alcock, a University of Michigan archaeologist who studies the Greek provinces of the Roman Empire. "It can be very hostile—I come in, I burn your temples, I take your women, I erase you from the earth." Or it can be gentler—"the elites [of both cultures] getting together and reconfiguring the social system."

Teotihuacan definitely controlled a small empire in central Mexico, Smith says. In the Mexican state of Morelos, 85 kilometers south of the city, for example, he found towns full of Teotihuacan-style ceramics and obsidian from the city's heyday. But farther afield, Teotihuacan's empire is a patchwork. It is "strategic about the places it's controlling," says Claudia García-Des Lauriers, an archaeologist at California State Polytechnic University, Pomona. She has mapped and excavated the site of Los Horcones on the coast of the Mexican state of Chiapas, where she says the main pyramid and plaza, used from about 400 to 600 C.E., resemble smaller versions of Teotihuacan's famous Pyramid of the Moon. Located in a narrow mountain pass through which trade passed, Los Horcones would have given Teotihuacan control of the flow of cacao and quetzal feathers from the lush Chiapas coast.

A monument from Tikal known as El Marcador includes the name glyph of Spearthrower Owl in a rosette at the top.

PHOTO: KENNETH GARRETT

Teotihuacan's influence extended as far as the Pacific coast of Guatemala, more than 1000 kilometers from the city. At sites there, archaeologists uncovered Teotihuacan-style household goods, including hundreds of incense burners used for domestic religious rituals, says Oswaldo Chinchilla, an archaeologist at Yale University. He and many other archaeologists think a colony of Teotihuacanos lived in Escuintla, perhaps commanding important land and sea trade routes. With their arrival, "The whole outlook of the sites and their culture changed," he says.

New clues about Teotihuacan's reach might come from PACUNAM's 2016 aerial survey of more than 2000 square kilometers in northern Guatemala, including the area around Tikal (*Science*, 28 September 2018, p. 1355). LIDAR, a laser-based remote sensing technique, revealed tens of thousands of unknown archaeological features, including possible fortifications such as flattened hilltops with watchtowers. "It's this feeling of an intensely guarded landscape," Houston says. Excavations of some sites will begin in May, and he is hoping for clues about whether the fortifications were built by the Maya in response to a Teotihuacan threat, or by Teotihuacanos and their allies once they had taken over Tikal.

One thing is clear: Sihyaj K'ahk's arrival changed the course of Tikal's history. "Following that invasion, Tikal ascends to a new level of greatness," says Thomas Garrison, an archaeologist at Ithaca College. As Tikal's influence spreads, it leads to "the foundation of much of what we know as Classic Maya culture," including a homogenization of written language, Román Ramírez says. "Even though they may have lost to Teotihuacan, Tikal is ultimately the big winner."

In about 550 C.E., Teotihuacan collapsed, its downtown burned in what was perhaps a rebellion by its own citizens. But centuries later, Tikal's kings still celebrated military victories by dressing as Teotihuacan warriors, Stuart says. Whatever happened in 378, its memory lingered far longer than Teotihuacan itself. ■

Published by AAAS

PERSPECTIVES

NEUROSCIENCE

Splitting speech and music

Brain asymmetries for words and melodies of songs depend on opposite acoustic cues

By **Daniela Sammler**

Speech and music are human universals, and people around the world often blend them together into vocal songs (*1*). This entwinement of the speech and music cognitive domains is a challenge for the auditory cognitive system. How do listeners extract words and melodies from a single sound wave? The split is surmised to start in the signal: Speech and musical sounds are thought to differ in details of their acoustic structure and thus activate different receptive preferences of the left and right auditory cortices of the brain (*2*, *3*). On page 1043 of this issue, Albouy *et al.* (*4*) provide evidence for the biophysical basis of the long-debated, yet still unresolved, hemispheric asymmetry of speech and music perception in humans. They show that the left and right auditory regions of the brain contribute differently to the decoding of words and melodies in songs.

Research on the nature of hemispheric

PHOTO: A. LESIK/SHUTTERSTOCK

Published by AAAS

The melody and words in music are processed asymmetrically in the human brain.

asymmetries started in 1861, when French anatomist Pierre Paul Broca astounded his Parisian colleagues with the observation that speech abilities are perturbed after lesions in the left, but not right, brain hemisphere (*5*). This seminal insight not only ended the long-held view of symmetrical brain organization; it also launched the quest for asymmetries in other cognitive domains. One field that lent itself particularly well to comparison was music; it has numerous structural parallels to speech (*6*), and deficits in one but not the other domain were shown to have developed in several neurological patients (although shared deficits in both domains were also reported) (*7*).

What was missing for a long time was a cogent theory that could explain why speech and music should show a divergent hemispheric lateralization. Today, influential neuroacoustic models (*2*, *3*) seek reasons in the specific computational requirements imposed by the structure of speech and musical sounds. For example, proper speech perception hinges strongly (but not solely) on the ability to process short-lived temporal modulations that are decisive for discriminating similar-sounding words, such as "bear" from "pear." By contrast, proper music perception requires, among others, the ability to process the detailed spectral composition of sounds (frequency fluctuations). Biophysical laws constrain the simultaneous extraction of both fast temporal and fine spectral information. The relative specialization of the left and right auditory cortices to better process one or the other cue appears to be an efficient solution to this dilemma. What may be a plausible computational explanation for auditory asymmetries of speech and music has, however, been vigorously contested (*8*), and although there is empirical support, it is often confined to synthetic stimuli (*9*) or acoustically manipulated speech (*10*), not music.

The study of Albouy *et al.* presents a major step forward in this debate by demonstrating direct links among (i) the amount of temporal or spectral detail in the sound structure of songs, (ii) listeners' comprehension of the verbal or melodic content of these songs, and (iii) corresponding left-right asymmetries in the anterior auditory cortex known to represent complex sound categories (*11*). Applying special temporal and spectral filters to recordings of 100 carefully constructed unaccompanied (a cappella) songs, the authors established that the stepwise removal of temporal detail impairs recognition of the words (but not recognition of the melodies), whereas the degradation of spectral detail weakens recognition of the melodies (but not recognition of the words). They used machine learning to identify patterns of brain activity (determined by functional magnetic resonance imaging) that correlated with the words or the melody being heard by participants. This approach was most successful in decoding the words from neural patterns in left (but not right) auditory regions, particularly for songs with high temporal detail. The decoding of the melodies was most accurate from neural patterns in right (but not left) auditory regions and scaled with the spectral detail of the songs. The behavioral (mis)understanding of words or melodies was mirrored in the lateralized neural activity patterns, which suggests that the computations carried out in left or right auditory regions on opposite acoustic features are relevant for participants' split perception of words and melodies in songs.

The conclusions drawn from the findings of Albouy *et al.* make it tempting to think of speech and music as two cognitive domains with sharp borders. However, this dichotomy should be taken with caution. Both speech and music have multiple acoustic and cognitive facets (*8*), many of which are shared by both domains:

Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. Email: sammler@cbs.mpg.de

Published by AAAS

Prosodic nuances in speech melody could never be properly decoded without fine-grained spectral processing, and musicians could never play in sync without perceiving subtle rhythmic deviations. Hence, it seems sensible to assume a weighted contribution of left and right auditory regions to different aspects in both domains, in keeping with recent findings of right-lateralized prosody perception (*12*). Moreover, speech and music perception involves not only the temporal lobe but also large-scale fronto-temporo-parietal networks (*13*, *14*), the concerted activity of which underpins higher-order linguistic and musical processes. The lateralization of these processes, and their dependency from and interaction with auditory encoding, remain pressing questions for future research.

Although previous studies have reported opposite spectrotemporal preferences of left and right auditory regions (*9*, *10*) and have emphasized their optimized tuning to human speech beyond other real-life sounds (*15*), Albouy *et al.* have transferred this approach to vocal musical melodies. Their results emphasize the important role of the right hemisphere for human communication. Following ecological views of brain functioning, these findings also point toward the interesting idea that both speech and music have shaped and optimized—in complementary ways—the general-purpose neural mechanisms through which the human brain represents and processes the sounds of the natural environment today. Whether the observed auditory asymmetries are specific to humans or are shaped by individual experience remains to be ascertained in experiments with nonhuman species, infants, and trained musicians. ■

REFERENCES AND NOTES

1. P. E. Savage, S. Brown, E. Sakai, T. E. Currie, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8987 (2015).
2. R. J. Zatorre, P. Belin, V. B. Penhune, *Trends Cogn. Sci.* **6**, 37 (2002).
3. D. Poeppel, *Speech Commun.* **41**, 245 (2003).
4. P. Albouy, L. Benjamin, B. Morillon, R. J. Zatorre, *Science* **367**, 1043 (2020).
5. P. P. Broca, *Bull. Soc. Anat.* **6**, 330 (1861).
6. W. T. Fitch, *Cognition* **100**, 173 (2006).
7. E. Feuchtwanger, *Amusie: Studien zur Pathologischen Psychologie der Akustischen Wahrnehmung und Vorstellung und Ihrer Strukturgebiete Besonders in Musik und Sprache* (Springer, 1930).
8. C. McGettigan, S. K. Scott, *Trends Cogn. Sci.* **16**, 269 (2012).
9. M. Schönwiesner, R. J. Zatorre, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14611 (2009).
10. A. Flinker *et al.*, *Nat. Hum. Behav.* **3**, 393 (2019).
11. S. Norman-Haignere, N. G. Kanwisher, J. H. McDermott, *Neuron* **88**, 1281 (2015).
12. D. Sammler *et al.*, *Curr. Biol.* **25**, 3079 (2015).
13. A. D. Friederici, *Physiol. Rev.* **91**, 1357 (2011).
14. S. Koelsch, W. A. Siebel, *Trends Cogn. Sci.* **9**, 578 (2005).
15. R. Santoro *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4799 (2017).

10.1126/science.aba7913

MEDICINE

Knocking out barriers to engineered cell activity

CRISPR-Cas9 gene-edited T cells show safety and long-term engraftment in humans

By **Jennifer R. Hamilton**[1,2] *and* **Jennifer A. Doudna**[1,2,3,4,5,6,7]

Engineered T cell therapies are revolutionizing cancer treatment by achieving long-lasting remission in blood-related cancers, such as leukemia and lymphoma. These therapies involve removal of patient T cells, "reprogramming" them to attack cancer cells, and then transferring them back into the patient. Targeted gene inactivation (knockout) using CRISPR-Cas9 can enhance T cell activity (*1*, *2*) and has the potential to expand cell therapy applications. Until now, it has been unknown whether CRISPR-Cas9–edited T cells would be tolerated and thrive once reinfused into a human. On page 1001 of this issue, Stadtmauer *et al.* (*3*) present data from a phase 1 clinical trial (designed to test safety and feasibility) on the first cancer patients treated with CRISPR-Cas9–modified T cells. The findings represent an important advance in the therapeutic application of gene editing and highlight the potential to accelerate development of cell-based therapies.

The production of engineered cell therapies involves transduction of isolated patient T cells with a disarmed virus to express a receptor recognizing an antigen present on the outside of cancer cells (through chimeric antigen receptors, CARs) or on the inside of cancer cells [through T cell receptors (TCRs) specific for cancer-associated peptides]. Once transduced, the engineered T cell population is expanded and then reinfused back into the patient. Although highly effective at treating some types of cancer, the specificity and longevity of engineered T cell activity can be improved. For example, T cell activity is naturally down-regulated through the programmed cell death protein 1 (PD-1) receptor. Systemic inhibition of PD-1 in patients can enhance T cell activity but often triggers adverse autoimmune reactions. Additionally, endogenous TCR expression competes with the transgenic receptor in engineered T cells, interfering with signaling or cell trafficking. Genetic knockout of the TCR and *PDCD1* (the gene encoding PD-1) can enhance engineered human T cell activity in preclinical human tumor xenograft models in mice (*4*, *5*). Stadtmauer *et al.* tested whether patient-derived T cells containing these gene knockouts generated by CRISPR-Cas9 are safe and persistent upon reinfusion into humans.

Six patients with either myeloma or sarcoma were enrolled in the trial, and three met the study's criteria for T cell reinfusion. A two-step process was used to achieve both native TCR and *PDCD1* gene knockout and transgenic TCR expression in the engineered cells (see the figure). In the first step, isolated patient T cells were electroporated with preformed ribonucleoproteins (RNPs) of Cas9 protein and guide RNA that targets the endogenous TCR—TCRα (*TRAC*) and TCRβ (*TRBC*)—and *PDCD1* for genetic disruption. In the second step, cells were transduced with a viral vector to express the transgenic TCR, which recognizes cancer-testis antigen 1 (NY-ESO-1), and expanded in culture to create NY-ESO-1 transduced CRISPR 3X edited

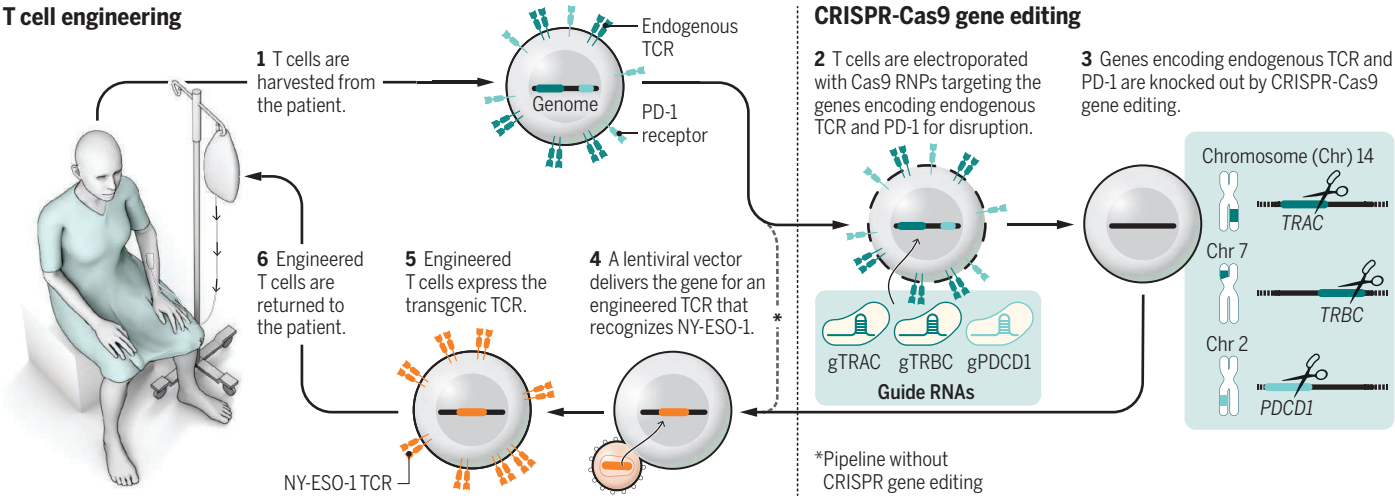
[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. [2]Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA. [3]Department of Chemistry, University of California, Berkeley, CA 94720, USA. [4]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720, USA. [5]Innovative Genomics Institute, University of California, Berkeley, CA 94720, USA. [6]MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [7]Gladstone Institutes, University of California, San Francisco, CA 94114, USA. Email: doudna@berkeley.edu

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

Published by AAAS

Modifying engineered T cells with CRISPR-Cas9 gene editing

Engineered T cells with improved anticancer activity can be generated through the targeted disruption of immunomodulatory genes, such as programmed cell death protein 1 (*PDCD1*, which encodes PD-1), and T cell receptor (TCR) genes (*TRAC* and *TRBC*), using CRISPR-Cas9 delivered as preformed ribonucleoproteins (RNPs). These cells are then modified to express an engineered TCR that recognizes cancer-testis antigen 1 (NY-ESO-1) expressed by cancer cells.



(NYCE) cells. Notably, NYCE cells eliminated NY-ESO-1–expressing cells more effectively than T cells expressing the NY-ESO-1 TCR alone, as would be expected from the successful knockout of the endogenous TCR.

NYCE cells successfully engrafted in all patients and were detected up to 9 months after reinfusion. According to the authors, this persistence compared favorably to the ~1-week half-life of infused, unedited T cells expressing NY-ESO-1 TCR in previous trials. NYCE cells reisolated from a study participant also had gene expression profiles consistent with central memory cells, a mark of stable engraftment. This result contrasts with past studies in which unedited cells expressing NY-ESO-1 TCRs displayed markers of T cell exhaustion (*6*). Together, the CRISPR-Cas9 disruption of endogenous TCR and PD-1 improved the cell-killing ability of the engineered T cells and promoted long-term persistence.

But are CRISPR-Cas9–edited cells safe in humans? It has been unclear whether Cas9-edited cells will be immunogenic and whether residual Cas9—a bacterial protein—will trigger an immune response. Stadtmauer *et al.* report no editing-associated toxicity of the NYCE cells in the three patients. Furthermore, although study participants had preexisting T cells and antibodies specific for Cas9 protein [as observed previously (*7*)], antibody titers did not increase from baseline over the course of the study. The lack of a Cas9 immune response could be attributed either to immunosuppression of the patients receiving NYCE cells, or to the delivery of Cas9 as a nonviral, preformed RNP, which has a limited half-life in cells compared to viral delivery where Cas9 protein is continuously expressed in treated cells. Stadtmauer *et al.* also report minimal off-target editing by CRISPR-Cas9, and the ≤1% of NYCE cells containing chromosomal translocations decreased in patients after reinfusion. Together, these findings provide a guide for the safe production and nonimmunogenic administration of gene-edited somatic cells.

The big question that remains unanswered by this study is whether CRISPR-edited, engineered T cells are effective against advanced cancer. Phase 1 trials assess safety, so the efficacy of the NYCE cells for treating patients was not evaluated. At the end of the study, one participant had died because of cancer progression, and the other two were receiving other therapies. Although the efficacy of the Cas9-engineered cells is thus ambiguous, the authors point out that their study was restricted to editing protocols available in 2016, when the U.S. Food and Drug Administration reviewed the clinical trial application. Gene disruption efficiencies in this study were modest (15 to 45%), whereas protocols now exist for reliably achieving >90% gene disruptions in human T cells using Cas9 RNPs (*8*, *9*). Moreover, recent efforts have demonstrated CAR transgene insertion at the *TRAC* gene in human T cells, resulting in simultaneous knockout of the endogenous TCR while driving CAR expression by the native promoter (*8*, *10*). Advances in generating precise genetic modifications, as well as other choices of cancer-associated targets, could enhance the efficacy of engineered T cells for the treatment of additional cancers, including solid tumors, which have largely been resistant to the activity of engineered cell therapeutics.

The clinically validated long-term safety of CRISPR-Cas9 gene-edited cells reported by Stadtmauer *et al.* paves the way for next-generation cell-based therapies. Although the safety of other types of gene-edited somatic cells, such as stem cells, remains to be determined, encouraging results show healthy blood production in the first β-thalassemia and sickle cell anemia patients infused with cells modified by CRISPR-Cas9 (*11*). As more gene-based therapies are demonstrated to be safe and effective, the barrier to clinical translation will become cell manufacturing and administration. A restructuring of production processes for engineered cells and new CRISPR-Cas9 delivery strategies for the modification of targeted cells in the body are now imperative to reduce cost and make these revolutionary therapies accessible to all who can benefit. ■

GRAPHIC: V. ALTOUNIAN/*SCIENCE* ADAPTED FROM TAMI TOLPA

REFERENCES AND NOTES

1. C. H. June *et al.*, *Science* **359**, 1361 (2018).
2. X. Liu *et al.*, *Cell Res.* **27**, 154 (2017).
3. E. Stadtmauer *et al.*, *Science* **367**, eaba7365 (2020).
4. L. J. Rupp *et al.*, *Sci. Rep.* **7**, 737 (2017).
5. J. Ren *et al.*, *Clin. Cancer Res.* **23**, 2255 (2017).
6. T. S. Nowicki *et al.*, *Clin. Cancer Res.* **25**, 2096 (2019).
7. C. T. Charlesworth *et al.*, *Nat. Med.* **25**, 249 (2019).
8. T. L. Roth *et al.*, *Nature* **559**, 405 (2018).
9. A. Seki *et al.*, *J. Exp. Med.* **215**, 985 (2018).
10. J. Eyquem *et al.*, *Nature* **543**, 113 (2017).
11. C. R. I. S. P. R. Therapeutics, https://crisprtx.gcs-web.com/static-files/f1e96190-16d1-447c-a4f6-87cc28cc978f (2019).

ACKNOWLEDGMENTS

J.R.H. is supported by the Jane Coffin Childs Fund for Medical Research. J.A.D. is a cofounder of Caribou Biosciences, Editas Medicine, Scribe Therapeutics, and Mammoth Biosciences; scientific advisory board member of Caribou Biosciences, Intellia Therapeutics, eFFECTOR Therapeutics, Scribe Therapeutics, Mammoth Biosciences, Synthego, and Inari; director at Johnson & Johnson; and has research sponsored by Biogen and Pfizer. The Regents of the University of California have patents issued and pending for CRISPR technologies on which J.R.H. and J.A.D. are inventors. We thank B. Shy, D. Nguyen, and C. Tsuchida for their thoughtful review of this Perspective.

Published online 6 February 2020
10.1126/science.aba9844

Published by AAAS

MICROBIOLOGY

Maternal microbial molecules affect offspring health

Intestinal molecules during pregnancy in mice may protect offspring from metabolic disease

By **Jane Ferguson**

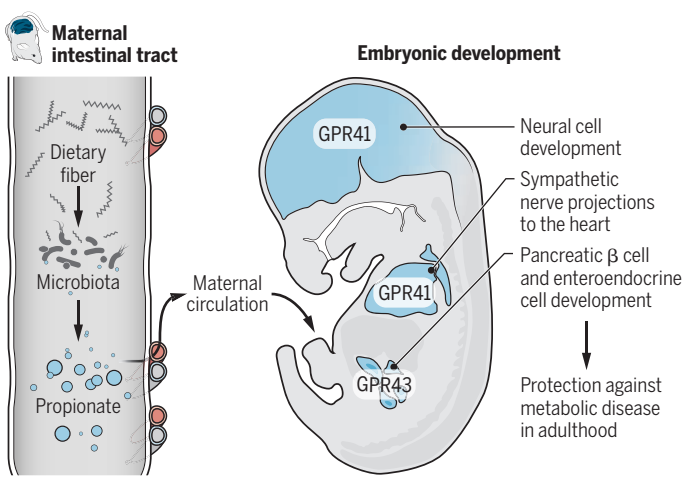
The early-life period is a critical time: Events that affect fetal development can have lifelong implications. Subtle disturbances during human fetal development affect not only major developmental outcomes, but also phenotypes that may not manifest for decades, such as risk of cardiometabolic disease (*1*). Despite this, pregnancy remains a poorly understood physiologic state, and there is relatively little mechanistic knowledge of how the maternal environment affects future disease risk. Studies suggest that maternal microbiota influence cardiometabolic disease risk in offspring; however, the mechanisms underlying this relationship are elusive. On page 1002 of this issue, Kimura *et al.* (*2*) find that, in mice, maternal diet and consequent gut microbiota–derived propionate protects against future obesity and metabolic dysregulation in offspring.

Microbiota, the commensal organisms residing in a particular host location, are underexplored modulators of health and disease. Microbiota in the gut, at the interface of human dietary and systemic metabolism, may have particular relevance for obesity and cardiometabolic disease. However, despite intense interest, there remains a knowledge gap between the myriad of hypothesized functions attributable to gut microbiota and those that have actually been demonstrated through mechanistic interrogation. Key functions of gut microbiota include digestion of diet-derived nutrients, which contain components that are not digestible by the host such as complex polysaccharides and fibers. This microbial action generates short-chain fatty acids (SCFAs), which translocate into the bloodstream and enter host metabolism. SCFAs, including propionate, acetate, and butyrate, can be used as a source of fuel by the host, but also function as active signaling molecules (*3*). SCFAs have thus been implicated as key metabolites linking the microbiota to host metabolic outcomes, including obesity and insulin sensitivity (*4*), and are generally considered to be protective from metabolic diseases, although data are conflicting (*5*).

Division of Cardiovascular Medicine, Vanderbilt University Medical Center, Nashville, TN, USA.
Email: jane.f.ferguson@vumc.org

Microbial protection from metabolic disease

Digestion of fiber by microbiota generates short-chain fatty acids (SCFAs). In pregnant mice, SCFAs, particularly propionate, regulate embryonic development through G protein–coupled receptor 41 (GPR41) and GPR43. This protects offspring from metabolic disease in adulthood.



The mechanisms linking maternal metabolic status during pregnancy to cardiometabolic disease risk in offspring remain elusive. Maternal microbiota are an emerging target of investigation, potentially modulating both maternal health and fetal development. Despite speculation that transfer of microbes from mother to infant may occur in utero, it is generally accepted that microbial colonization does not occur in a meaningful way until birth, when, depending on the method of birth, maternal vaginal or skin microbes colonize the infant gut (*6*). Thus, the maternal microbiota likely influence fetal development in utero through intermediates.

Kimura *et al.* probed the role of microbial SCFA production on fetal development and outcomes in mouse models. Mice whose mothers were housed in a germ-free facility, and consequently did not have gut microbiota during pregnancy, were highly susceptible to high-fat diet–induced obesity later in life. Further, these offspring displayed evidence of glucose intolerance and insulin resistance, suggesting cardiometabolic disease. The absence of maternal microbiota during pregnancy affected offspring irrespective of vaginal or cesarean birth, and there were no significant differences in the gut microbiota of the adult offspring of germ-free versus conventionally housed mice, suggesting that the effects were not mediated through offspring microbiota. However, there were significant differences in the amounts of circulating acetate, propionate, and butyrate in both the mothers and embryos when comparing germ-free and conventionally housed mice.

Kimura *et al.* examined expression of the SCFA receptors, G protein–coupled receptor 41 (*Gpr41*) and *Gpr43*, in the brain, intestine, and pancreas of embryos and adult mice. High expression in embryonic tissues suggested that embryos were sensing maternal-derived SCFAs through *Gpr41* and *Gpr43*. However, they observed lower *Gpr41* expression in the germ-free mice. Loss of *Gpr41* expression in mice resulted in defects in sympathetic nerve projections to the heart, which were also apparent in germ-free offspring. The ability of the SCFAs to activate sympathetic neuronal differentiation was confirmed in vitro, with propionate having the greatest effect. In mice lacking *Gpr43*, the authors found SCFA- and microbiota-dependent defects in embryonic enteroendocrine and pancreatic β-cell development. These data strongly suggest that altered SCFA signaling through GPR41 and GPR43 in embryonic tissues could affect development. Indeed, Kimura *et al.* showed that the offspring of mice fed a low-fiber diet during pregnancy had increased risk of obesity and insulin resistance. The deleterious effects of a low-fiber diet could be rescued with propionate supplementation. When pregnant mice were given antibiotics, there was no difference in the metabolic parameters of the offspring of high-fiber versus low-fiber diet–fed mothers, confirming the importance of maternal microbiota in mediating protection.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

Published by AAAS

Although several studies have found differences in the gut microbiota of lean or obese humans (*7*), the study of Kimura *et al.* suggests that early exposure to microbial products may be causal in later obesity even if there are no differences in the subsequent gut microbiota of the offspring. The offspring of the high-fiber–fed animals were heavier at birth, but protected against obesity later in life. This is consistent with humans, where low birth weight is associated with future obesity (*8*). Supplementation of pregnant mice with propionate protected offspring against future disease. However, whether propionate supplementation had any effects on the mother was not assessed. Propionate supplementation in humans is associated with weight loss and improved metabolic function (*9*), although studies are conflicting, with propionate also reported to increase insulin resistance (*10*).

Given the goal of reducing human metabolic disease, it is crucial to determine whether similar mechanisms control human development. There is an urgent need to better understand how, why, and under what circumstances should attempts be made to modulate gut microbiota or SCFAs in pregnancy. Diets high in fiber are consistent with existing nutritional recommendations, including during pregnancy, but whether these would be effective in increasing SCFAs and protecting offspring from future metabolic disease remains unclear. There is large interindividual variability in the microbial composition of humans, with different microbiota having differing capacities for SCFA production (*11*). Thus, differences in the SCFA-producing capacity of the microbiota may affect risk of offspring obesity despite high consumption of fiber during pregnancy. Although supplementation with propionate may be a convenient option, the safety and efficacy during pregnancy remain to be determined. Further studies in humans are warranted to understand whether modulation of this pathway could be an avenue to improving the metabolic health of the next generation. ■

REFERENCES AND NOTES

1. H. L. Blackmore, S. E. Ozanne, *J. Mol. Cell. Cardiol.* **83**, 122 (2015).
2. I. Kimura *et al.*, *Science* **367**, eaaw8429 (2020).
3. A. Koh *et al.*, *Cell* **165**, 1332 (2016).
4. E. E. Canfora, J. W. Jocken, E. E. Blaak, *Nat. Rev. Endocrinol.* **11**, 577 (2015).
5. D. J. Morrison, T. Preston, *Gut Microbes* **7**, 189 (2016).
6. B. McDonald, K. D. McCoy, *Science* **365**, 984 (2019).
7. V. K. Ridaura *et al.*, *Science* **341**, 1241214 (2013).
8. P. H. Whincup *et al.*, *JAMA* **300**, 2886 (2008).
9. E. S. Chambers *et al.*, *Gut* **68**, 1430 (2019).
10. A. Tirosh *et al.*, *Sci. Transl. Med.* **11**, eaav0120 (2019).
11. P. Louis, H. J. Flint, *Environ. Microbiol.* **19**, 29 (2017).

10.1126/science.aba7673

GRAPHIC: N. CARY/*SCIENCE*

BATTERIES

Cobalt in lithium-ion batteries

Replacements are sought for cobalt, a costly element used in lithium-ion battery cathodes

By **Matthew Li,**[1,2] *and* **Jun Lu**[1]

The use of cobalt in lithium-ion batteries (LIBs) traces back to the well-known $LiCoO_2$ (LCO) cathode, which offers high conductivity and stable structural stability throughout charge cycling. Compared to the other transition metals, cobalt is less abundant and more expensive and also presents political and ethical issues because of the way it is mined in Africa (*1*). Cheaper cathodes have been developed that substitute some of the cobalt with nickel and manganese, and $LiNi_{0.80}Co_{0.15}Al_{0.05}O_2$ (NCA) and $LiNi_{1-x-y}Co_xMn_yO_2$ (NMC, where x and y <1) are used in the majority of the LIBs in electric vehicles. Nonetheless, in NCA and NMC, cobalt enables high-rate performance and to some extent, enhances cycle stability. We outline research efforts that could further decrease or even eliminate cobalt content in LIBs to lower their cost while maintaining high performance.

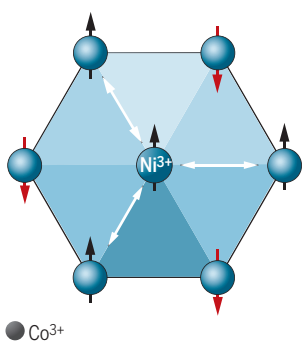
Efforts to replace cobalt have to start with an understanding of what makes cobalt so crucial within the NMC and NCA compositions. Originally, cobalt and manganese were introduced into $LiNiO_2$ (LNO) to stabilize the material itself. Although LNO has a high theoretical energy density, it also has very poor cycling stability and presents potential safety hazards because of lattice instability. For these reasons, cobalt was added as a stabilizer. In comparison to LCO, it is difficult to synthesize pure layered LNO, which facilitates Li^+ ion transport, and more often the undesired rock salt structure forms.

Also, nickel is inherently unstable by itself in the transition-metal layer of the oxide as it has a relatively strong magnetic moment. Three triangularly placed Ni^{2+} cations will always have two opposing magnetic moments, creating "magnetic frustration" (*2*). Because Li^+ ions do not have a magnetic moment, they preferentially exchange with some of the nickel ions. The loss of a spin at one site relieves magnetic frustration (see the figure). The strong interlayer antiferromagnetic coupling between nickel in the transition-metal layer and the migrated nickel in the lithium layer creates a superexchange interaction that further stabilizes the Li^+ ion (*2*). Overall, this lithium-nickel mixing deteriorates performance because the lithium-deficient LiO_2 interslab layer decreases in thickness. This thinner layer severely hinders transport of lithium ions and ultimately results in very rapid degradation of the LNO composition (*3*).

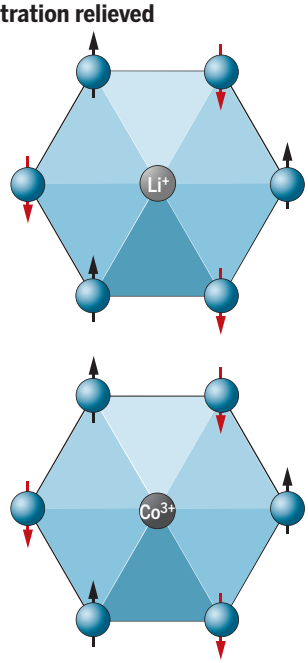
[1]Chemical Sciences and Engineering Division, Argonne National Laboratory, Lemont, IL 60439, USA. [2]Department of Chemical Engineering, Waterloo Institute of Nanotechnology, University of Waterloo, Waterloo, ON N2L 3G1, Canada. Email: junlu@anl.gov

Instability of nickel

Nickel (Ni) as a replacement for cobalt (Co) in lithium (Li) ion battery cathodes suffers from magnetic frustration. Discharging mixes Li ions into the Ni layer, versus just storing them between the oxide layers.



Magnetic frustration relieved



Published by AAAS

When cobalt is introduced into LNO, both during synthesis and delithiation, the intermixing of lithium for nickel is deterred statistically because of the lower nickel content. More importantly, it alleviates magnetic frustration (*4*) because the Co^{3+} cation does not have a magnetic moment and serves as a buffer atom in the transition-metal layer (see the figure). Because added cobalt in LNO easily prevents nickel-lithium mixing and subsequent phase transitions, the desired layered structure forms (*5*).

Directly decreasing cobalt content can be effective in achieving acceptable performance but only to some minimum cobalt fraction. For example, in nickel-rich NMC compositions, thermal stability, which is crucial for avoiding catastrophic failures, as well as cycle stability drastically dropped in comparison to equal nickel-cobalt fractions. The NMC systems with progressively higher nickel content from NMC 111, 532, 622, and 811 (where 111 represent 1 part nickel, 1 part manganese, and 1 part cobalt mass composition, respectively) follow a steady trend in decreasing cycle stability and safety (*6*).

By contrast, partial substitution of cobalt with other elements such as titanium has been shown to produce reasonable performance (*7*). Although other metals can limit lithium-nickel mixing, typically poor kinetics and lower capacities result. Binary composition with ultralow O_2 gas, such as $LiNi_{0.94}Co_{0.06}O_2$, has exhibited severe surface reconstruction when compared to NMC 811 (*8*) that pulverizes the particles after prolonged cycling.

Other systems, such as the lithium- and manganese-rich materials, consist of a mix of 0.5 Li_2MnO_3 and 0.5 NMC. This layered-layered structure offers increased capacity at the cost of severe phase transitions. The defects introduced limit the capacity of cathode and the voltage it can produce, an effect called voltage fade. Moving toward completely cobalt-free systems has led researchers to pursue disordered rock-salt materials in hopes of harnessing their increased capacity through the use of anionic redox couples, such as O^{2-} (*9*). However, anionic redox systems have limited cyclability because of the formation of O_2 gas. Recent work has yielded some pure anionic redox systems where the oxidation state of oxygen is maintained below the superoxide charge of −0.5 (*10*).

In the original $LiNiO_2$ system, the role of cobalt may not have been as critical to performance as initially presumed. Often, its importance is only apparent when compounded by the effects of another element. Some combinations of noncobalt transition metals, such as aluminum, manganese, and magnesium, can outperform the cobalt-containing equivalent, although the levels of performance are still lower than the commercially viable ratios (*11*). Even cobalt-free $LiNiO_2$ showed surprisingly good cycling stability when made under synthesis conditions that carefully controlled temperature, sintering time, and O_2 gas (*12*). A mostly layered $Li_{0.98}Ni_{1.02}O_2$ structure formed without the need for other transition-metal additives. This surprising result reopens the question of the optimal cathode composition and its implications for the effect of magnetic frustration of this system. Simply tuning the composition of cobalt-free systems will likely require the substitution of cobalt with another third transition metal. A brief performance comparison between the various commonly studied cathode materials is given in table S1 in the supplementary materials.

Identifying optimal composition and synthesis conditions of new cathodes will likely require rigorous and extensive factorial experimental design that must incorporate compositions with elemental fractions decreased to doping levels (<1%). Machine learning techniques might decrease the search path for optimal elements and compositions. The complete elimination of cobalt is an important research goal, but lower-cost cathodes with less cobalt must maintain performance. This trade-off will depend on the future supply of cobalt from both mining and recycling. ■

REFERENCES AND NOTES

1. B. K. Sovacool, *Extr. Ind. Soc.* **6**, 915 (2019).
2. J. Zheng *et al.*, *J. Phys. Chem. Lett.* **8**, 5537 (2017).
3. W. Li, J. N. Reimers, J. R. Dahn, *Phys. Rev. B Condens. Matter* **46**, 3236 (1992).
4. Y. Xiao *et al.*, *Nano Energy* **49**, 77 (2018).
5. A. Ueda *et al.*, *J. Electrochem. Soc.* **141**, 2010 (1994).
6. T. Li *et al.*, *Electrochem. Energy Rev.* **2019**, 1 (2019).
7. S. Wolff-Goodrich *et al.*, *Phys. Chem. Chem. Phys.* **17**, 21778 (2015).
8. J. Li, A. Manthiram, *Adv. Energy Mater.* **9**, 1902731 (2019).
9. E. Hu *et al.*, *Nat. Energy* **3**, 690 (2018).
10. Z. Zhu *et al.*, *Nat. Energy* **1**, 16111 (2016).
11. H. Li *et al.*, *J. Electrochem. Soc.* **166**, A429 (2019).
12. H. Li, N. Zhang, J. Li, J. R. Dahn, *J. Electrochem. Soc.* **165**, A2985 (2018).

ACKNOWLEDGMENTS

We thank T. Liu for providing consultation on this topic. The work at Argonne National Laboratory was supported by the U.S. Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy, Vehicle Technologies Office. Argonne National Laboratory is operated for the DOE Office of Science by UChicago Argonne, LLC, under contract no. DE-AC02-06CH11357. M.L. acknowledges financial support from the National Sciences and Engineering Research Council of Canada, University of Waterloo, and Waterloo Institute for Nanotechnology.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/367/6481/979/suppl/DC1

10.1126/science.aba9168

PLANETARY SCIENCE

A deep dive into the abyss

The flyby of the Kuiper Belt object Arrokoth provides quick and tantalizing observations

By **David C. Jewitt**

Since 1992, astronomers have unveiled a vast population of solid bodies in orbit beyond Neptune. Known as the Kuiper Belt, this region of the Solar System is a dynamical fossil preserving a record of the planet-formation epoch (*1*). The belt is also a repository of the Solar System's most primordial material and the long-sought nursery from which most short-period comets originate. Most of what we know about the belt was determined using ground-based telescopes, and studies were limited to objects larger than about 100 km because the smaller ones are too faint to easily detect. Now, 5 years after its flyby of the 2000-km-diameter Kuiper Belt object Pluto (*2*), NASA's New Horizons spacecraft has provided the first close-up look at a small, cold classical Kuiper Belt object. On pages 998, 999, and 1000 of this issue, Spencer *et al.* (*3*), Grundy *et al.* (*4*), and McKinnon *et al.* (*5*) show one of these objects to be lightly cratered, ultrared, and binary, respectively.

The scientific impact of the Kuiper Belt has been huge, in many ways reshaping our ideas about the formation and evolution of the Solar System. For example, and quite incredibly, Kuiper Belt objects are a thousand times more plentiful than the (much closer and more familiar) main-belt asteroids that orbit between Mars and Jupiter. The Kuiper Belt objects' orbital distribution shows that the planets formed closer to the Sun and then migrated outward, a finding with profound dynamical consequences. The objects themselves are divided into dynamically distinct groups, one of which (the so-called cold classicals) appears to be the most primordial population in the Solar System.

The object that New Horizons flew by is known formally as (486958) Arrokoth (provisional designation 2014 MU_{69}). An earlier

Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, CA. Email: jewitt@ucla.edu

Published by AAAS

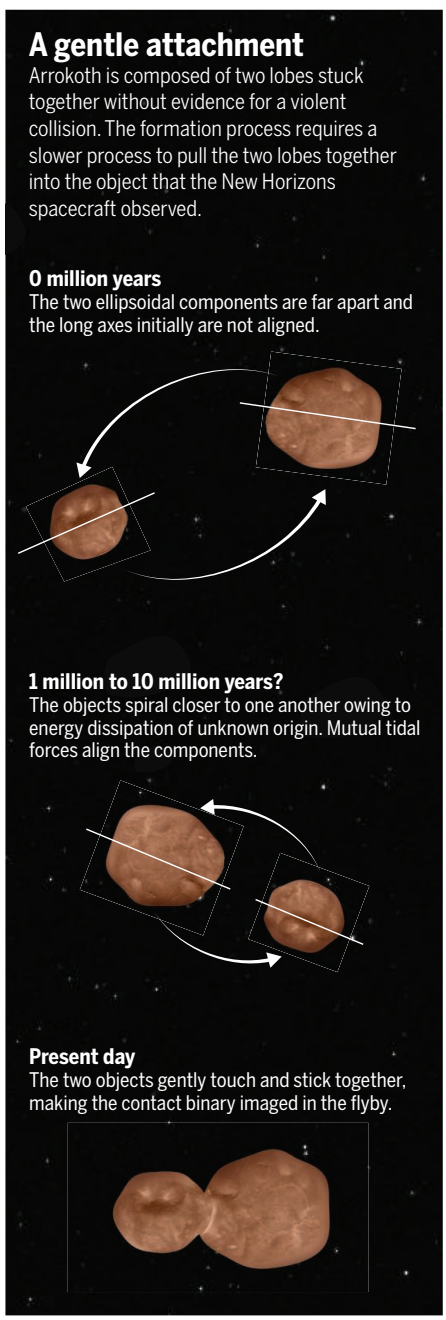
working name, "Ultima Thule," was abandoned after the official name was approved. Flyby observations from a 3500-km close-approach distance provide an image scale as small as 33 m/pixel. The most notable feature of Arrokoth is its overall shape, which appears to consist of two spheroidal, but unequal, lobes in contact, like a giant peanut. Combined, they have a volume equal to that of an 18-km-diameter sphere. As noted by Spencer *et al.*, similar binary structures have been observed before in comets and larger Kuiper Belt objects (*6*). Although binarity in sublimating comets exposed to the Sun could result from nonuniform erosion of an initially single body, the frigid environment of the Kuiper Belt minimizes erosion and renders this explanation unlikely.

The most basic inference, then, is that Arrokoth is the product of a collision between two preexisting bodies. The collision must have been gentle because there is no evidence for compressive deformation at the neck connecting the lobes (see the figure). McKinnon *et al.* infer an impact speed comparable to or smaller than the gravitational escape speed, estimated at a few meters per second (*5*). Low-speed accretion is expected in the young protoplanetary disk, where relative motions are damped by friction, first causing loose binaries to form and then driving them to spiral together (*7*). The modern-day Kuiper Belt has no substantial friction, but the protoplanetary disk was much more densely populated, perhaps creating frictional dissipation from collective gravitational effects or from residual protoplanetary gas (*5*). At the same time, Arrokoth's delicate structure is difficult to reconcile with alternate models (*8*) in which Arrokoth-sized Kuiper Belt objects are fragments of larger objects shattered by energetic collisions.

However, impact craters seen lightly sprinkled across the surface of Arrokoth do provide evidence for smaller, higher-speed collisions. The view is limited by the resolution of the data, and by the high-noon illumination of the surface, such that only a few craters are clear. Although the largest crater, 7 km in diameter, is well resolved, all the others are subkilometer, and most are close to the resolution limit of the images. Within the limitations of the data, larger craters show the bowl-shaped morphology that is typical of impact craters on small asteroids with a depth-to-diameter ratio of 0.1 to 0.2 (*3*). Impacts should affect the surfaces of all Kuiper Belt objects more or less equally, and so crater populations on different objects should look basically the same. This should include the moon of Pluto, the most well-known resident of the Kuiper Belt. Curiously, Arrokoth has a steeper crater size distribution and higher crater density at a given size than does the surface of Pluto's moon, Charon. Unlike on Earth's moon, measuring surface age is not possible, and so the crater density cannot be accurately converted into a cratering flux. Such a flux would be extremely useful in assessing the billion-year-scale evolution of the outer Solar System.

The Kuiper Belt is home to the reddest material in the Solar System. This "ultrared matter" is widespread throughout the belt and is especially abundant in the cold classicals. Ultrared matter is rare or absent interior to the orbit of Saturn, probably because the material is thermodynamically unstable at the higher temperatures found nearer the Sun (*9*). Consistent with this picture, the observations from New Horizons show that the surface of Arrokoth is ultrared. This allows some confirmation of the long-standing suspicion that the color is due to organic materials. Specifically, near-infrared spectra show clear absorption bands due to the alcohol methanol as well as additional unidentified bands. Grundy *et al.* suggest several ways to form methanol, including formation by cosmic-ray irradiation of a simple mixture of water and methane ices (*4*). Although radiation chemistry in the Kuiper Belt is no doubt much more complicated, the latter reaction consumes water, perhaps explaining why New Horizons did not detect it.

New Horizons was a flyby mission. It took more than a decade to advance from concept to launch and another decade to coast to the Kuiper Belt. By contrast, New Horizons acquired the key measurements of Pluto and Arrokoth over encounter periods of just a few days. Having done it once, we can be sure that this is not a particularly efficient or desirable way to investigate the outer Solar System. For future missions, we need to be able to send spacecraft to the Kuiper Belt and keep them there, perhaps by using the gravity of larger Kuiper Belt objects to assist in their capture. A Pluto or Eris orbiter, for example, would allow these intriguing bodies to be studied in stunning geological and geophysical detail. More interesting would be a hopper mission, capable of moving from one Kuiper Belt object to another in much the same way that NASA's Dawn spacecraft moved from Ceres to Vesta using its own ion drive engine. In the Kuiper Belt, where the flux of sunlight is only 0.1% of that on Earth and the distances between objects are truly vast, nuclear rockets are likely necessary to move from place to place with reasonable transit times. Technologically, we could probably do it. Scientific vision and institutional commitment are the extra ingredients needed to make such a mission happen. ■



GRAPHIC: N. DESAI/*SCIENCE*; (PHOTO) NASA/ROMAN TKACHENKO

REFERENCES AND NOTES

1. D. Prialnik, M. A. Barucci, L. Young, *The Trans-Neptunian Solar System* (Elsevier, 2020).
2. S. A. Stern, W. M. Grundy, W. B. McKinnon, H. A. Weaver, L. A. Young, *Annu. Rev. Astron. Astrophys.* **56**, 357 (2018).
3. J. R. Spencer *et al.*, *Science* **367**, eaay3999 (2020).
4. W. M. Grundy *et al.*, *Science* **367**, eaay3705 (2020).
5. W. B. McKinnon *et al.*, *Science* **367**, eaay6620 (2020).
6. S. Sheppard, D. Jewitt, *Astron. J.* **127**, 3023 (2004).
7. P. Goldreich, Y. Lithwick, R. Sari, *Nature* **420**, 643 (2002).
8. A. Morbidelli, H. Rickman, *Astron. Astrophys.* **583**, A43 (2015).
9. D. Jewitt, *Astron. J.* **123**, 1039 (2002).

Published online 13 February 2020
10.1126/science.aba6889

Published by AAAS

MEDICINE

Artificial intelligence in cancer therapy

Artificial intelligence can optimize cancer drug discovery, development, and administration

By **Dean Ho**

Artificial intelligence (AI) approaches have the potential to affect several facets of cancer therapy. These include drug discovery and development and how these drugs are clinically validated and ultimately administered at the point of care, among others. Currently, these processes are expensive and time-consuming. Moreover, therapies often result in variable treatment outcomes between patients. The convergence of AI and cancer therapy has resulted in multiple solutions to address these challenges. AI platforms ranging from machine learning to neural networks can accelerate drug discovery, harness biomarkers to accurately match patients to clinical trials, and truly personalize cancer therapy using only a patient's own data. These advances are indicators that practice-changing cancer therapy empowered by AI may be on the horizon.

AI is already making progress in accelerating drug discovery and has successfully predicted drug behavior by using genomic and chemical data in lieu of large-scale screening assays, which may accelerate drug repurposing (that is, using drugs indicated for one disease to treat other diseases) (*1*). Reinforcement learning, which uses reward and punishment to train algorithms to achieve a desired drug structure, successfully designed a new compound in just 21 days versus conventional timelines of approximately 1 year. Subsequently observed pharmacokinetic properties suggested that threshold drug exposure and efficacy could be attained, supporting further evaluation of the lead compound (*2*). In this study, the generative tensorial reinforcement learning (GENTRL) platform was trained by using a dataset of chemical structures that target the tyrosine kinase discoidin domain receptor 1 (DDR1), which is involved in the progression of multiple cancers. Computer simulations predicted lead compound and receptor binding, aiming to minimize targeting of other tyrosine kinases and enhance DDR1 targeting. Although additional compound optimization is required, this is an important step toward AI-accelerated cancer drug discovery.

AI-enhanced drug discovery represents one stage of a pipeline of processes needed to optimize cancer therapy (see the figure). During traditional drug development, approved and investigational compounds are often codelivered in combination preclinically and clinically to address multiple drug targets, improving treatment efficacy. Subsequent clinical dose escalation studies identify the doses that achieve drug synergy, in which the drug combination has a greater effect together than individually. Unfortunately, issues such as off-target effects can preclude drug approval because of unforeseen toxicity (*3*). Additionally, well-designed compounds delivered at suboptimal doses may limit efficacy. Therefore, optimized combination therapy design simultaneously identifies the best drugs for combination and doses that address the right targets while minimizing toxicity. Testing all possible drug combinations at multiple doses for each drug is virtually impossible. However, AI can overcome this challenge by markedly reducing the number of experiments needed to resolve drug and dose parameters, optimizing combination therapy development.

Achieving drug synergy is a key objective of traditional approaches to design combination therapies to improve efficacy. However, patient responses to combination therapy are highly variable. Computational modeling showed that effective combination therapy can be realized without drug additivity or synergy; a combination of drugs that act independently with favorable efficacy may improve treatment outcomes over synergy-driven combinations (*4*). Modeling tumor growth kinetics and drug sensitivity predictions by using data from clinical trials that recruited large cohorts of patients with different cancer types demonstrated that maximizing independent drug efficacy is a major determinant of treatment response.

AI will play a vital role in designing drug combinations without relying on synergy-based modeling or predicted synergy between different drug targets and pathways. This may markedly increase the pool of drugs considered for treatment and identify unexpected combinations that outperform standards of care. When the dose of each candidate drug is also considered, the possible drug and dose combinations are too extensive for comprehensive validation. However, AI can rapidly resolve large drug and dose parameter spaces. For example, the quadratic phenotypic optimization platform (QPOP) uses quadratic relationships represented by parabolas to visually correlate a set of inputs (e.g., drugs and doses) with optimal outputs (e.g., preclinical tumor reduction with minimal toxicity). This correlation markedly reduces the number of experiments and data needed to identify the drugs and doses that optimize combination therapy design. Furthermore, QPOP is agnostic to disease mechanisms, drug targets, and drug synergy.

The QPOP platform evaluated 14 chemotherapy drugs to treat multiple myeloma (MM) with combination therapy (*5*). The resulting drug combinations—such as decitabine and mitomycin C, which were unexpectedly and agonistically identified by QPOP—markedly improved outcomes in a MM mouse model compared with the clinical standard-of-care drug combinations (*5*). Importantly, neither decitabine nor mitomycin c monotherapy mediated efficacy alone, but their co-administration optimally and synergistically reduced tumor burden. Expanding QPOP to build combinations of targeted therapies and immunotherapies will be an important next step toward treatment strategies beyond chemotherapy.

As oncology drug candidates move into clinical validation, their regulatory approval rates have been reported to be as low as 3.4% (*6*). Recent advances in trial design have used biomarkers, such as genomic alterations, because of their potential as predictive indicators of treatment response to stratify patient recruitment. Including biomarkers in study recruitment has improved patient outcomes compared with traditional stratification information such as pathology or responses to prior treatments (*3*, *7*). Combining patient biomarker data and electronic health records (EHRs) for AI analysis may further affect trial outcomes (*8*). In the SYNERGY-AI study (NCT03452774), virtual tumor boards, which enable clinician teams to provide remote treatment guidance and diagnosis input, and EHRs pair oncology patients with suitable clinical trials. Multiple disease factors outlined by EHRs

The N.1 Institute for Health (N.1), Institute for Digital Medicine (WisDM), Department of Biomedical Engineering, and Department of Pharmacology, National University of Singapore, Singapore. Email: biedh@nus.edu.sg

Published by AAAS

will be correlated with progression-free survival and overall survival to assess the effectiveness of the trial-matching platform. Although many types of data may be useful for stratification, AI will ultimately need both population-scale and individualized data to ensure that patients given therapies, including those designed by AI, have a high likelihood of responding.

AI will also play a pivotal role in how cancer therapy is administered. Maximum tolerated dosing eliminates drug-sensitive tumor cells. However, drug-resistant cells can eventually cause treatment failure. Game theory is being explored to address this challenge, with dose-reduction algorithms competing against the tumor to prevent drug-resistant cells (which have high energy costs) from outnumbering drug-sensitive cells (*9*, *10*). This is known as adaptive therapy and may prolong treatment efficacy by maintaining threshold drug-sensitive cell populations in a tumor to combat drug-resistant cell proliferation. Adaptive therapy was recently used to modify paclitaxel dosing to treat mouse models of breast cancer. Specifically, an adaptive therapy algorithm (AT-1) iteratively reduced paclitaxel dosing after observed tumor size reductions. AT-1 was compared with a fixed-dosage algorithm that halted drug dosing after observed tumor reduction (AT-2) and high-dose standard therapy (*11*). The AT-1 algorithm improved tumor control and survival compared with AT-2 and standard therapy, demonstrating that game theory–driven dosing can improve treatment outcomes over established therapies at high doses.

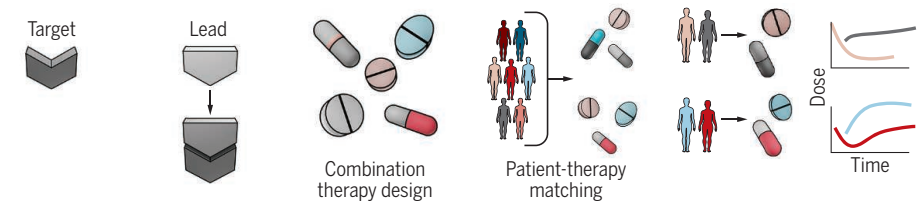
Adaptive therapy was translated into a pilot study for prostate cancer patients on abiraterone hormone therapy (*12*). The adaptive therapy cohort received, on average, 47% of the standard abiraterone dose, and three of the patients received less than 25% of the conventional dose. At the time of reporting, 1 of 11 adaptive therapy participants experienced tumor progression. This led to an estimated median time to progression, using the biomarker prostate-specific antigen (PSA) and radiographic imaging, of no less than 27 months in this cohort, which was superior to 11.1 months (PSA) and 16.5 months (radiographic) for patients on continuous abiraterone therapy (*12*). To maximize the benefits of this platform, personalizing adaptive therapy by use of each patient's response to therapy versus population-based dose adjustment rules will likely be needed.

To further personalize patient-specific combination therapy administration with AI, CURATE.AI, a neural network–derived platform, modulated multidrug dosing using only a patient's data by means of a second-order algebraic algorithm that dynamically related dosing to optimal tumor reduction and safety at any treatment time point (*13*, *14*). In a patient treated with enzalutamide hormone therapy and an investigational bromodomain and extraterminal domain (BET) inhibitor, data such as physician-guided dose variations and corresponding amounts of PSA were used to recommend a 50% reduction of the BET inhibitor dose to increase efficacy. Subsequent dynamic dosing of both drugs resulted in a durable response and halted tumor progression, which was confirmed by computed tomography (CT) imaging (*13*). This study successfully used AI to modulate experimental therapy dosing, demonstrating that dosage guidance, without requiring big data and complex genetic information, markedly enhanced treatment efficacy compared with fixed- and high-dose chemotherapy. Future studies will need to evaluate whether this platform can be implemented with additional classes of disease markers. Also, because of the individualized nature of this platform, it will need to be further evaluated in larger patient cohorts. As such, it is being tested in a clinical trial that involves a large pool of patients with hematological malignancies (NCT03759093).

AI represents a gateway to the next frontier of cancer therapy, reconciling extraordinary amounts and different types of data into actionable therapy. Among its barriers to deployment are its siloed use in narrow segments of the cancer therapy workflow. For example, AI-optimized compounds that are suboptimally combined with other therapies or dosed incorrectly are unlikely to substantially improve patient outcomes. Overcoming this challenge in oncology will require the seamless implementation of AI across the spectrum of discovery, development, and administration. Its potential downstream applications include increasing the resolution of personalized care by tailoring bespoke regimens that integrate multiple therapeutic strategies. For example, AI-optimized radiation therapy dosing, which maintained robust tumor size control, can potentially be combined with AI-driven drug administration (*15*). Ultimately, comprehensively adapting AI into clinical oncology practice may improve drug accessibility and reduce health-care costs. As AI continues to be validated and a path toward widespread practice is identified, its potential to redefine the clinical standards of cancer therapy is becoming evident. ■

Improving multiple aspects of cancer therapy

Cancer therapy involves different stages, including drug discovery, development, and administration. Artificial intelligence (AI) is poised to benefit each stage but is also confronted by challenges that, when overcome, may lead to practice-changing cancer treatment.



Discovery	Development	Administration
Opportunities Minimize off-target effects and toxicity Enhance drug exposure	**Opportunities** Optimize drug and dose selection Match patients to therapies and trials	**Opportunities** Sustained dose optimization Overcoming resistance with game theory
Challenges Identifying optimal targets Properly validating AI-designed drugs	**Challenges** Improving trial outcomes Stratification with the right patient data	**Challenges** More clinical validation needed Use in more cancer types

GRAPHIC: A. KITTERMAN/*SCIENCE*

REFERENCES AND NOTES

1. M. P. Menden *et al.*, *PLOS ONE* **8**, e61318 (2013).
2. A. Zhavoronkov *et al.*, *Nat. Biotechnol.* **37**, 1038 (2019).
3. A. Lin *et al.*, *Sci. Transl. Med.* **11**, eaaw8412 (2019).
4. A. C. Palmer, P. K. Sorger, *Cell* **171**, 1678 (2017).
5. M. B. M. A. Rashid *et al.*, *Sci. Transl. Med.* **10**, eaan0941 (2018).
6. C. H. Wong *et al.*, *Biostatistics* **20**, 273 (2019).
7. S. Harrer *et al.*, *Trends Pharmacol. Sci.* **40**, 577 (2019).
8. A. Rajkomar *et al.*, *NPJ Digit. Med.* **1**, 18 (2018).
9. J. Chmielecki *et al.*, *Sci. Transl. Med.* **3**, 90ra59 (2011).
10. C. Jedeszko *et al.*, *Sci. Transl. Med.* **7**, 282ra250 (2015).
11. P. M. Enriquez-Navas *et al.*, *Sci. Transl. Med.* **8**, 327ra24 (2016).
12. J. Zhang *et al.*, *Nat. Commun.* **8**, 1816 (2017).
13. A. J. Pantuck *et al.*, *Adv. Ther.* **1**, 1800104 (2018).
14. A. Zarrinpar *et al.*, *Sci. Transl. Med.* **8**, 333ra49 (2016).
15. B. Lou *et al.*, *Lancet Digit. Health* **1**, e136 (2019).

ACKNOWLEDGMENTS

D.H. holds pending patents on AI-based personalized medicine. D.H. acknowledges support from AI Singapore (NRF), the National Medical Research Council, and the Ministry of Education.

10.1126/science.aaz3023

Published by AAAS

POLICY FORUM

As demand for data centers rises, energy efficiency improvements to the IT devices and cooling systems they house can keep energy use in check.

PHOTO: GOOGLE

ENERGY

Recalibrating global data center energy-use estimates

Growth in energy use has slowed owing to efficiency gains that smart policies can help maintain in the near term

By **Eric Masanet[1,2], Arman Shehabi[3], Nuoa Lei[1], Sarah Smith[3], Jonathan Koomey[4]**

Data centers represent the information backbone of an increasingly digitalized world. Demand for their services has been rising rapidly (*1*), and data-intensive technologies such as artificial intelligence, smart and connected energy systems, distributed manufacturing systems, and autonomous vehicles promise to increase demand further (*2*). Given that data centers are energy-intensive enterprises, estimated to account for around 1% of worldwide electricity use, these trends have clear implications for global energy demand and must be analyzed rigorously. Several oft-cited yet simplistic analyses claim that the energy used by the world's data centers has doubled over the past decade and that their energy use will triple or even quadruple within the next decade (*3*–*5*). Such estimates contribute to a conventional wisdom (*5*, *6*) that as demand for data center services rises rapidly, so too must their global energy use. But such extrapolations based on recent service demand growth indicators overlook strong countervailing energy efficiency trends that have occurred in parallel (see the first figure). Here, we integrate new data from different sources that have emerged recently and suggest more modest growth in global data center energy use (see the second figure). This provides policy-makers and energy analysts a recalibrated understanding of global data center energy use, its drivers, and near-term efficiency potential.

Assessing implications of growing demand for data centers requires robust understanding of the scale and drivers of global data center energy use that has eluded many policy-makers and energy analysts. The reason for this blind spot is a historical lack of "bottom-up" information on data center types and locations, their information technology (IT) equipment, and their energy efficiency trends. This has led to a sporadic and often contradictory literature on global data center energy use.

Understanding where data center energy use is heading requires considering service demand growth factors alongside myriad equipment, energy efficiency, and market structure factors (see the first figure). Bottom-up analyses tend to best reflect this broad range of factors, generating the most credible historical and near-term energy-use estimates (*7*). Despite several recent national studies (*8*), the latest fully replicable bottom-up estimates of global data center energy use appeared nearly a decade ago. These estimates suggested that the worldwide energy use of data centers had grown from 153 terawatt-hours (TWh) in 2005 to between 203 and 273 TWh by 2010, totaling 1.1 to 1.5% of global electricity use (*9*).

Since 2010, however, the data center landscape has changed dramatically (see the first figure). By 2018, global data center workloads and compute instances had increased more than sixfold, whereas data center internet protocol (IP) traffic had increased by more than 10-fold (*1*). Data center storage capacity has also grown rapidly, increasing by an estimated factor of 25 over the same time period (*1*, *8*). There has been a tendency among analysts to use such service demand trends to simply extrapolate earlier bottom-up energy values, leading to unreliable predictions of current and future global data center energy use (*3*–*5*). They might, for example, scale up previous bottom-up values (e.g., total data center energy use in 2010) on the basis of the growth rate of a service demand indicator (e.g., growth in global IP traffic from 2010 to 2020) to arrive at an estimate of future energy use (e.g., total data center energy use in 2020).

But since 2010, electricity use per computation of a typical volume server—the workhorse of the data center—has dropped by a factor of four, largely owing to processor-efficiency improvements and reductions in idle power (*10*). At the same time, the watts per terabyte of installed storage has dropped by an estimated factor of nine owing to storage-drive density and efficiency gains (*8*). Furthermore, growth in the number of servers has slowed considerably owing to a fivefold increase in the average number of compute instances hosted per server (owing to virtualization), alongside steady reductions in data center power usage effectiveness (PUE, the total amount of energy used by a data center divided by the energy used by its IT equipment). Both of these trends have been largely driven by shifts in compute instances to energy-efficient cloud and "hyperscale" data centers, the largest data center type (*1*, *2*). In the United States—the world's largest data center market—industry-vetted bottom-up analyses of these efficiency trends identified a plateau in national data center en-

[1]McCormick School of Engineering and Applied Science, Northwestern University, Evanston, IL, USA. [2]Bren School of Environmental Science and Management, University of California, Santa Barbara, CA, USA. [3]Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [4]Koomey Analytics, Burlingame, CA, USA.
Email: eric.masanet@northwestern.edu

Published by AAAS

ergy use since 2010, despite rapid increases in demand for U.S. data center services (*11*). We now expand that analysis to the global level and show that strong continued efficiency progress can maintain an energy use plateau for the next few years through proactive policy initiatives and data center energy-management practices. These new bottom-up estimates form the basis of recent global data center energy values utilized by the International Energy Agency (*12*).

The data leveraged here facilitate a more technology-rich and temporally consistent approach than was available previously. Since 2011, analysts at Cisco have published data and outlooks for worldwide server stocks, data center workloads, server virtualization levels, and storage estimates for traditional, cloud, and, most recently, hyperscale data centers (*1*). In a series of reports starting in 2016, Lawrence Berkeley National Laboratory has published energy trend analyses of servers, storage devices, and network devices commonly used within data centers (*8*, *11*, *13*). Analysts have documented the numbers and locations of hyperscale data centers that represent a substantial fraction of global data center compute instances, and major data center operators are increasingly reporting their PUE (*14*).

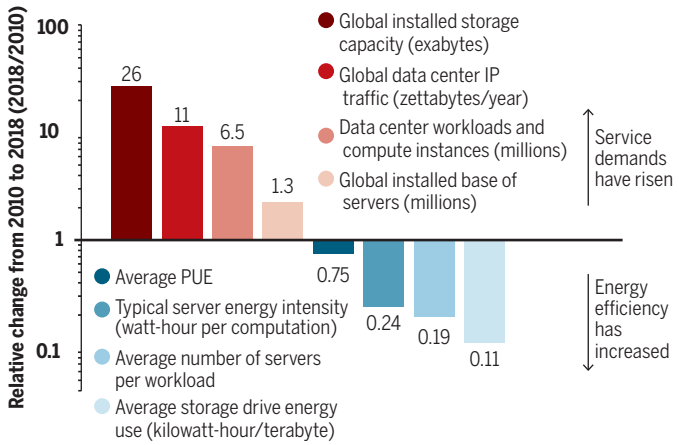
When integrated into a bottom-up modeling framework, these data suggest that, although global data center energy has increased slightly since 2010, growth in energy use has been substantially decoupled from growth in data center compute instances over the same time period (see the second figure, second graph). Moreover, the refined view provided by these new data suggests that global data center energy use in 2010 was around 194 TWh, slightly less than the lower-bound estimate in the 2010 bottom-up study (203 TWh) when fewer data were available (*9*).

In 2018, we estimated that global data center energy use rose to 205 TWh, or around 1% of global electricity consumption. This represents a 6% increase compared with 2010, whereas global data center compute instances increased by 550% over the same time period. Expressed as energy use per compute instance, the energy intensity of global data centers has decreased by 20% annually since 2010, a notable improvement compared with recent annual efficiency gains in other major demand sectors (e.g., aviation and industry), which are an order of magnitude lower (*12*).

The new integrated data illuminate some key technological and structural trends that help explain these large energy intensity improvements (see the first figure and the second figure, second graph). The combination of increased server efficiencies and greater server virtualization (which reduces the amount of server power required for each compute instance) has enabled a sixfold increase in compute instances with only a 25% increase in global server energy use, whereas the combination of increased storage-drive efficiencies and densities has enabled a 25-fold increase in storage capacity with only a threefold increase in global storage energy use. Shifts to faster and more energy-efficient port technologies have enabled a 10-fold increase in data center IP traffic with only modest increases in network device energy use. In sum, although overall energy use of IT devices (servers, storage, and network) has increased from around 92 TWh in 2010 to around 130 TWh in 2018, technological and operational efficiency gains have enabled substantial growth in services with comparatively much smaller growth in energy use.

Notably, the new data also suggest a large decrease in the energy use of data center infrastructure systems (i.e., cooling and power provisioning), enough to mostly offset the growth in total IT device energy use. This decrease is explainable by ongoing shifts in servers away from smaller traditional data centers (79% of compute instances in 2010) and toward larger and more energy-efficient cloud (including hyperscale) data centers (89% of compute instances in 2018) (see the second figure, third graph), which have much lower reported PUE values owing to cutting-edge cooling-system and power-supply efficiencies (*1*, *11*).

Trends in global data center energy-use drivers



PUE, power usage effectiveness; IP, internet protocol.

Yet given ever-growing demand for data center services, how much longer can these current efficiency trends last? Predicting the long-term efficiency limits of IT devices is notoriously difficult, especially in light of potential game-changing technologies such as quantum computing, for which energy use is unclear (*2*). Yet over the near term, market analysts predict that even greater levels of server virtualization are feasible (*1*), and technology studies indicate remaining potential for IT device efficiency gains, including more shifts to low-power storage devices (*8*). On the infrastructure side, world-class hyperscale data centers are already operating with PUEs of 1.1 or lower, which is close to the practical minimum value. Additional structural shifts from smaller traditional data centers to hyperscale data centers are predicted in the near term (*1*), indicating that infrastructure energy use may be dampened even further. Should these trends play out over the next few years, our approach indicates that there is a sufficient energy efficiency resource to absorb the next doubling of data center compute instances that would occur in parallel with a negligible increase in global data center energy use (see the second figure, second graph).

These findings lie in contrast to recent predictions of rapid and unavoidable near-term energy demand growth. Yet the IT industry, data center operators, and policy-makers can't rest on their laurels; diligent efforts will be required to manage possibly sharp energy demand growth once the existing efficiency resource is fully tapped. The next doubling of global data center compute instances may occur within the next 3 to 4 years (*1*).

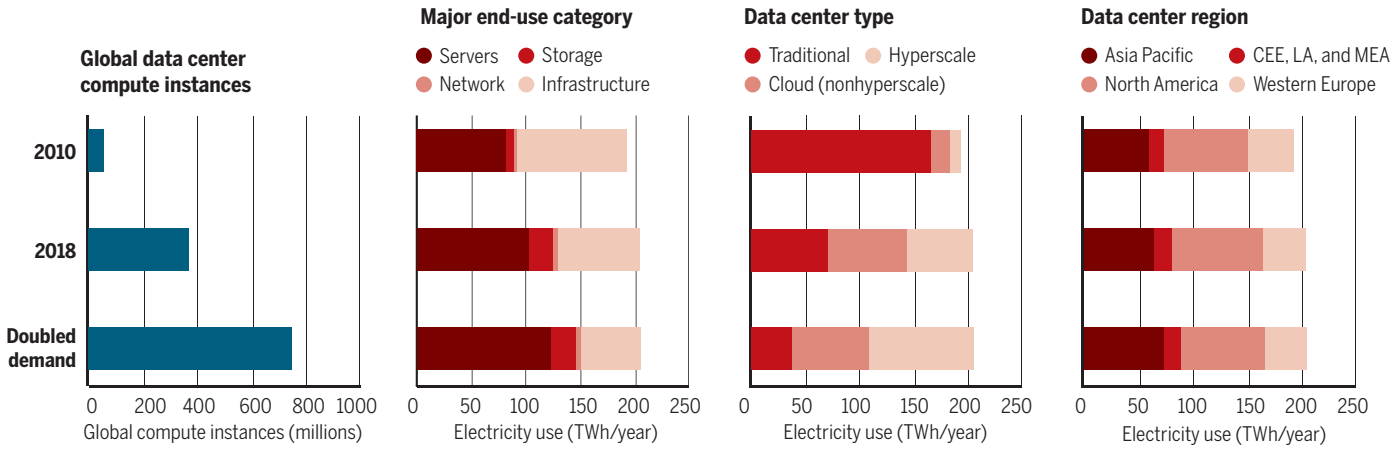
For policy-makers, there are three main areas of action. First, policy support can help data centers seize the remaining efficiency potential of current technology and structural trends. One key strategy includes further strengthening and promotion of efficiency standards such as Energy Star for servers, storage, and network devices while requiring such certifications in public IT procurement programs. Efficiency standards give data center operators access to more efficient IT devices while creating strong market incentives to manufacturers to continue innovating energy-efficient products. To support such standards, greater investments are needed to develop energy efficiency benchmarks for storage and network devices—similar to the Standard Performance Evaluation Corporation's (SPEC's) SPEC Power bench-

GRAPHIC: X. LIU/*SCIENCE*

Published by AAAS

Historical energy usage and projected energy usage under doubled computing demand

Doubled demand (relative to 2018) reflects current efficiency trends continuing alongside predicted growth in compute instances.



CEE, LA, and MEA, Central and Eastern Europe, Latin America, and Middle East and Africa; TWh, terrawatt-hour.

mark for servers—while policy should require that measured performance of all certified IT devices be made public to spur ongoing competition. Another strategy is to incentivize shifts to cloud services when economically and institutionally feasible—for example, through procurement standards and utility rebates—ensuring that future compute instances are delivered by data centers at the cutting edge of energy efficiency. Yet another is to encourage and incentivize continuous reductions in PUE, some of which are attainable through low-cost measures such as improved airflow management and temperature set-point optimization and through vehicles such as subsidized energy efficiency audits and tax credits. These and other proven data center efficiency strategies (*2*, *7*, *8*) can bring about a near-term plateau in energy use, which provides critical time to prepare for the possibility of future energy demand growth. But this time must be used wisely.

Second, investment in new technologies is needed to manage future energy demand growth in the cleanest manner possible once current efficiency trends reach their feasible limits. Strong deployment incentives should be provided to accelerate the pace of renewable energy adoption by data centers, including low-carbon procurement standards and corporate tax credits, so that the carbon intensity of current and future energy demand is reduced substantially (*15*). And greater public funding should be allocated to advancements in computing, data storage, communications, and heat removal technologies that may extend the IT industry's historical efficiency gains well into the future. Key examples include quantum computing, materials for ultrahigh density storage, increased chip specialization, artificial intelligence for computing resource and infrastructure management, and liquid and immersion cooling technologies. However, it is crucial to increase investments immediately to ensure such technologies are economical and scalable in time to prevent a demand surge later this decade, which would also make required renewable capacity additions more challenging.

Third, much greater public data and modeling capacities are required for understanding and monitoring data center energy use and its drivers and for designing and evaluating effective policies. National policy-makers should enact robust data collection and open data repository systems for data center energy use, in much the same way as has been done historically for other demand sectors. Proprietary data concerns can be addressed through data reporting and aggregation protocols, similar to energy data for the industrial sector, which shares many of the same confidentiality concerns (see, for example, the U.S. Manufacturing Energy Consumption Survey). Such efforts are important in all world regions and particularly in Asia, where data center energy use is poised to grow (see the second figure, fourth graph), but reliable data are scarce, especially for China, where data centers are multiplying quickly. In parallel, more public reporting by large data center operators should be encouraged and incentivized (e.g., through efficiency rating systems) for greater energy-use transparency and accountability.

To make full use of these important data, more research funding is needed for developing policy-relevant data center energy models and for model sharing and research community building that can disseminate and ensure best analytical practices. With better data, analysts should also quantify uncertainties in future modeling results, leading to more robust policy decisions. Given the important role data centers will play in future energy systems, the historical dearth of knowledge on their energy use and the mixed signals given to policy-makers by contradictory findings are unacceptable. Global data center energy use is entering a critical transition phase; to ensure a low-carbon and energy-efficient future, we cannot wait another decade for the next reliable bottom-up estimates. ■

REFERENCES AND NOTES

1. Cisco, "Cisco Global Cloud Index: Forecast and methodology, 2016–2021 white paper" (Cisco, document 1513879861264127, 2018).
2. International Energy Agency (IEA), *Digitalization & Energy* (IEA, 2017).
3. L. Belkhir, A. Elmeligi, *J. Clean. Prod.* **177**, 448 (2018).
4. A.S.G. Andrae, T. Edler, *Challenges* **6**, 117 (2015).
5. T. Bawdy, "Global warming: Data centres to consume three times as much energy in next decade, experts warn," *The Independent*, 23 January 2016.
6. N. Jones, *Nature* **561**, 163 (2018).
7. E. Masanet, R. E. Brown, A. Shehabi, J. G. Koomey, B. Nordman, *Proc. IEEE* **99**, 1440 (2011).
8. A. Shehabi *et al.*, "United States data center energy usage report" (Lawrence Berkeley National Laboratory, LBNL-1005775, 2016).
9. J. G. Koomey, "Growth in data center electricity use 2005 to 2010" (Analytics Press for the *New York Times*, 2011).
10. B. Wagner, "Intergenerational energy efficiency of Dell EMC PowerEdge servers" (Dell, DellEMC white paper, 2018).
11. A. Shehabi, S. J. Smith, E. Masanet, J. Koomey, *Environ. Res. Lett.* **13**, 124030 (2018).
12. IEA, "Tracking clean energy progress" (IEA, 2019); www.iea.org/tcep/.
13. H. Fuchs *et al.*, *Energy Effic.* 10.1007/s12053-019-09809-8 (2019).
14. M. Avgerinou, P. Bertoldi, L. Castellazzi, *Energies* **10**, 1470 (2017).
15. E. Masanet, A. Shehabi, J. G. Koomey, *Nat. Clim. Chang.* **3**, 627 (2013).

ACKNOWLEDGMENTS

This material includes work conducted by Lawrence Berkeley National Laboratory (LBNL) with support from the U.S. Department of Energy (DOE) Advanced Manufacturing Office. LBNL is supported by the Office of Science of the DOE and operated under contract grant No. DE-AC02-05CH11231. E.M. and N.L. are grateful for financial support provided by Leslie and Mac McQuown. The global data center analysis modeling file with all data inputs, results, methodological notes, figures, discussion of uncertainties, and sources is available on GitHub (doi: 10.5281/zenodo.3668743).

10.1126/science.aba3758

GRAPHIC: X. LIU/*SCIENCE*

Published by AAAS

ENGINEERING

Out of sight

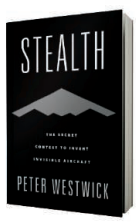
A historian investigates the Cold War competition to create an invisible aircraft

By **F. Robert van der Linden**

Since the beginning of aerial combat, designers have attempted to make aircraft invisible. With each advance, however, other technologists have developed better methods for detecting covert combatants.

Although scientists and engineers worked on both sides of this difficult problem for years, it was not until the 1970s that the creation of an aircraft invisible to radar became more than a mere possibility. In his excellent new book, *Stealth*, Peter Westwick argues that the solution was the product of a host of special circumstances, fortuitous geography, and the aggregation of thousands of innovative and highly skilled individuals in Southern California. This concise, highly readable history of the creation, development, and application of one of the most important technologies of the Cold War brings clarity and a thorough understanding to this complex subject.

Westwick's story examines not the senior leaders of U.S. aerospace and the military but the brilliant and generally unknown individuals in the trenches—the engineers, project managers, and shop floor workers—who conceived of the idea of stealth, identified the problems involved, and found solutions for seemingly intractable challenges. These "men in the middle" worked primarily for either the Lockheed Aircraft Corporation, in Burbank, or the Northrop Corporation, 20 miles away in El Segundo, both of which were challenged to create a stealth bomber by the U.S. Department of Defense in 1979.



Stealth: The Secret Contest to Invent Invisible Aircraft
Peter Westwick
Oxford University Press, 2020. 272 pp.

Tensions between the aeronautical engineers and the electrical engineers arose immediately at both Lockheed and Northrop. The deflection and absorption of radio waves was an electrical problem, yet the aeronautical engineers were loath to surrender any of their stature as the lead designers. Engineers tend to be rational, however, and over time the two parties began to collaborate.

With the fortuitous discovery of the work of Soviet physicist Pyotr Ufimtsev, which described equations for predicting the reflection of magnetic waves from two- and three-dimensional objects, Lockheed began attempts to create an aircraft consisting of radar-deflecting facets. In 1981, after much work and little sleep, the Lockheed F-117A was introduced. This stealth aircraft—armed with newly developed precision-guided munitions and controlled by computers that kept the unstable aircraft flying—finally gave the Pentagon the force multiplier it had been looking for. Lockheed won the Department of Defense's initial contract for the stealth fighter.

Across the Los Angeles basin, Northrop kept working. The U.S. Air Force was looking for a long-range, high-altitude stealth bomber to augment the aging B-52 and the new, but troubled, B-1 bombers. From its work developing Tacit Blue—a high-altitude battlefield surveillance aircraft that was ultimately scrapped—the team had learned that curved surfaces, if properly designed, could deflect radar as well.

Northrop founder John K. "Jack" Northrop had been obsessed with "flying wing" aircraft (fixed-wing planes that lacked a defined fuselage) and for good reason. Pure flying wings have no upright tail surfaces and are extremely aerodynamic, making them an ideal shape for a long-range aircraft that is inherently low-observable. The company had lost interest in the flying wing when the Air Force canceled the YB-49 in 1949, but when Lockheed began to circle back to the design, Northrop mounted a renewed effort, delivering the B-2 flying wing bomber in 1989, much to the delight of its retired founder.

Although stealth technology did not end the Cold War, it helped prevent a nuclear conflict by making a precision stealth attack potentially as destructive as a nuclear strike. It also inspired the creation of technologies that have benefited both civilians and the military in ways not originally foreseen. New aircraft, particularly commercial airliners, are equipped with computer-driven fly-by-wire controls and built from innovative composite materials that provide greater strength and lightness.

Today, the F-117A is retired, but the B-2 flies on, joined by new aircraft and a host of remotely piloted vehicles and missiles. Together, these technologies dominate the skies—a tribute to the creativity and hard work of the countless men and women "in the middle." ■

10.1126/science.aba5372

The reviewer is curator of air transportation and special purpose aircraft at the National Air and Space Museum, Washington, DC 20560, USA. Email: vanderlindenb@si.edu

A pair of F-117 Nighthawks accompany a B-2 Spirit bomber.

PHOTO: U.S. AIR FORCE

Published by AAAS

DEVELOPMENTAL BIOLOGY

When life meets research

The personal and professional collide in a scientist's story of early human development

By **Janet Rossant**

With one phone call, Magdalena Zernicka-Goetz's life as a scientist collided with her personal life in the most dramatic way. As Zernicka-Goetz stood at her desk at the University of Cambridge, a genetic counselor explained that some cells derived from the placenta supporting her fetus carried a serious chromosomal anomaly. But her knowledge of how embryos develop suggested that the fetus might be able to exclude the cells with the genetic abnormality or that it may not contain the damaged DNA at all. Further testing confirmed that the fetus was genetically normal, and her son was born in perfect health. "At that time," however, she writes, "I couldn't possibly be sure."

In *The Dance of Life*, Zernicka-Goetz and science writer Roger Highfield weave Zernicka-Goetz's personal memoir together with an accessible introduction to contemporary mouse and human embryonic research and with a discussion of the clinical, ethical, and societal implications of this research and related areas. This is a lofty goal for a relatively slim volume, and it succeeds better in some parts than in others.

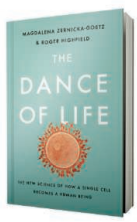
Developmental biologists have studied the progression of the fertilized mammalian egg through its early cleavage divisions to the formation of the 100-cell blastocyst for many years. Such studies have suggested that there is a gradual segregation of cell fate influenced by cell polarity, cell position (inside or outside), and mechanical signals and that the embryo is able to regulate for loss, gain, or rearrangement of cells right up to the blastocyst stage. However, none of these studies really addressed the question of whether there might be asymmetries in the egg or early embryo that could bias later cell fate.

Zernicka-Goetz, who has long been fascinated with patterning in the early embryo, took on this challenge when she began her own research lab at the University of Cambridge. She describes her work on defining early asymmetries in the mouse embryo and their role in informing later development in careful detail, recounting how she used tools such as cell marking, live imaging, and gene manipulation to determine that early blastomeres show a bias toward different regions and cell types of the blastocyst. However, other researchers—using different techniques—found less evidence for early differences, leading to some vigorous debates, as described in the book.

This controversy compelled researchers who had set aside work on the early embryo to reenter the fray, bringing new tools and ideas. And, although it is still not clear what initiates asymmetries after fertilization, it is increasingly clear that by the four-cell stage, there are differences in chromatin modification and transcription factor activity among the cells that, while not permanently specifying cell fate, may bias their future lineage contributions.

The book does not shy away from discussing the moral and ethical implications inherent in such research, tackling prenatal testing, the ongoing quest to create synthetic embryos, and the question of whether human embryos should be used in research. "Although an early embryo is not a person, I believe that it deserves protection, and I fully appreciate that balancing that protection with scientific research is not easy," writes Zernicka-Goetz of the latter issue. Her attitude has been to tread carefully, weighing potential concerns against potential value to humanity. "I believe in taking a measured approach to enable research that is fully consistent with our values."

The authors do their best to describe experiments on the early embryo and stem cell–based embryo models in simple terms, but the book would have been greatly enhanced by the inclusion of some illustrations. The early embryo is truly beautiful, especially when the complexity of gene expression patterns is revealed by fluorescent imaging. Also, some of the science may be hard for the uninitiated to follow. In the fifth chapter, for example, the embryonic cleavage divisions from one cell to four cells are described using a two-tone soccer ball analogy. This takes two full paragraphs to explain a concept that could have been easily conveyed with a simple diagram.



The Dance of Life
Magdalena Zernicka-Goetz and Roger Highfield
Basic Books, 2020. 304 pp.

The most engaging parts of *The Dance of Life* are the personal stories about the trials and tribulations Zernicka-Goetz has faced during her life in science. Scientific disagreements occur, papers get rejected, grants are not funded, and balancing family and work is never easy. Those of us who, like Zernicka-Goetz, are developmental biologists are in the fortunate position of being in a field where many leading scientists are women (no all-male panels for us!), but female scientists still struggle to overcome the persistent biases and societal and institutional barriers that block their progress. At the end of the day however—as Zernicka-Goetz and Highfield so ably show—the thrill of scientific discovery is what keeps us coming back to the bench. ■

10.1126/science.aba2771

The reviewer is president and scientific director of the Gairdner Foundation, Toronto, ON M5G 1L7, Canada, and a senior scientist at the Hospital for Sick Children, University of Toronto, Toronto, ON M5G 0A4, Canada. Email: janet.rossant@sickkids.ca

Public clocks like the 14th-century Wells clock made time a commodity.

PHOTO: ISTOCK.COM/STEVEALLENPHOTO

PODCAST

Transcendence: How Humans Evolved Through Fire, Language, Beauty, and Time
Gaia Vince
Basic Books, 2020. 352 pp.

Four factors—fire, language, beauty, and time—were key to humankind's ascension, argues science writer Gaia Vince. This week on the *Science* podcast, Vince examines our evolution through the lens of deep time, revealing how our genes, culture, and environment combined to create modern humans. sciencemag.org/podcasts

10.1126/science.aba9449

Published by AAAS

Chinese giant salamanders are overstocked in farms but endangered in the wild.

PHOTO: NATURE PICTURE LIBRARY/ALAMY STOCK PHOTO

Edited by **Jennifer Sills**

Giant salamanders: Farmed yet endangered

Chinese giant salamanders (*Andrias davidianus s.l.*) are the largest extant amphibians, attaining a body size of almost 2 meters (*1*). As "living fossils," their common ancestor lived in the Middle Jurassic (*2*). However, their population density has decreased since the 1950s because of habitat loss and overharvesting (*3*). Chinese giant salamanders now appear in Appendix I of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), and the International Union for Conservation of Nature lists them as Critically Endangered (*4*, *5*). Conservation efforts include national reserves, breeding in captivity (*6*), and release back into nature (*7*), but these strategies have proved inadequate. The existence of multiple species (*1*) complicates conservation planning. Moreover, although extremely endangered in nature (*8*), Chinese giant salamanders are overstocked in commercial farms (*9*). Conservation strategies must reflect this contradiction.

Between 1990 and 2010, the potential for high profits drove frequent illegal capture of salamanders in the wild for farm breedstock, decimating wild populations (*9*). Since 2002, in an attempt to facilitate the recovery of wild salamander populations, the government has paid farmers to release more than 270,000 farm-bred individuals (*10*) but has not required genetic or health assessments before release. The mass release could accelerate extinction of some species through genetic homogenization (*1*). Most farms also pump water directly from streams and rivers into their facilities, circulate it within ponds, and then discharge the effluent directly into the wild without wastewater treatment (*9*). This activity may drive transmission of viruses and threaten ecological security (*9*). Given that observations of Chinese giant salamanders remain extremely rare in nature (*3*, *8*), it seems that the release from farms has not succeeded in augmenting the population.

The balancing of conservation and utilization is key to the future of Chinese giant salamanders. Governmental agencies should coordinate unfailing supervision of the commercial market. Wastewater from farms must be treated before release back into nature. Stringent law enforcement must stop commercial farming in or near reserves and end poaching. Coordinated national scientific investigations need to evaluate the status of each species, especially in areas where natural breeding-caves persist (*11*, *12*). All releasing of farm-bred animals should cease until testing confirms disease-free, pure-native species. Ecotourism should be developed to educate and promote the conservation of Chinese giant salamanders and build local pride in them as cultural and biodiversity resources.

Chenqi Lu[1,2], Jing Chai[1], Robert W. Murphy[1,3], Jing Che[1,4]*

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 650223 Kunming, China. [2]Kunming College of Life Science, University of Chinese Academy of Sciences, 650204 Kunming, China. [3]Centre for Biodiversity, Royal Ontario Museum, Toronto, ON M5S 2C6, Canada. [4]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, 650223 Kunming, China.
*Corresponding author. Email: chej@mail.kiz.ac.cn

REFERENCES AND NOTES

1. F. Yan *et al.*, *Curr. Biol.* **28**, R590 (2018).
2. K. Gao, N. H. Shubin, *Nature* **422**, 424 (2003).
3. K. Zhang, X. Wang, W. Wu, Z. Wang, S. Huang, *Biodiv. Sci.* **10**, 291 (2002) [in Chinese].
4. G. Liang, B. R. Geng, E. M. Zhao, "*Andrias davidianus*," *The IUCN Red List of Threatened Species* (2004).
5. CITES, Appendices I, II and III (2019); https://cites.org/eng/app/appendices.php.
6. A. S. Yang, G. J. Liu, *Hunan Fish. Sci. Technol.* **4**, 29 (1978) [in Chinese].
7. Q. H. Luo, Y. Liu, L. Y. Zhang, *Hunan. Biodiv. Sci.* **17**, 310 (2009) [in Chinese].
8. S. T. Turvey *et al.*, *Curr. Biol.* **28**, R592 (2018).
9. A. A. Cunningham *et al.*, *Oryx* **50**, 265 (2016).
10. G. C. Shu *et al.*, paper presented at the 2019 Annual Meeting of Chinese Herpetological Society, Nanchong, China, 18 to 21 October (2019).
11. Z. Q. Liang *et al.*, *Ecol. Evol.* **9**, 3879 (2019).
12. J. Wang, H. X. Zhang, F. Xie, G. Wei, J. P. Jiang, *Asian Herpetol. Res.* **8**, 174 (2017).

10.1126/science.abb2375

The U.S. military is not sustainable

As Australian wildfires raged and youth-led climate movements inspired millions globally to march against climate change, commentators dubbed 2019 "the year the world woke up to the climate crisis" (*1*). However, one of the major contributors to climate change over the course of the past century too often remains overlooked: the U.S. military. Two recent studies demonstrate the scale of U.S. military greenhouse gas emissions, which rivals the emissions of the majority of countries around the world (*2*, *3*). As global leaders prepare to discuss the next phase of international agreements at COP26 in Glasgow and political discourse around sustainable transitions to green economies becomes more mainstream, the United States must reconsider the ecological costs of its military's global operations, including its domestic and global base infrastructure.

The U.S. military's contribution to global climate change and local environmental damage is extensive (*4*). The U.S. military's global greenhouse gas (GHG) emissions amount to 593 million metric tons of CO_2 equivalent from 2010 to 2018, an annual average similar to the annual GHG emission output of 14 million passenger cars. [(*2*), p. 14]. U.S. military operations—such as the use of herbicides during the Vietnam War (*5*) and white phosphorous in Iraq (*6*), and the construction of the global network of military bases (*7*)—have disrupted local ecological systems. Moreover, in May 2019, the Anthropocene Working Group

Published by AAAS

AAAS Travels

Members and friends of AAAS are invited to join AAAS Travels on fascinating trips to all 7 continents!

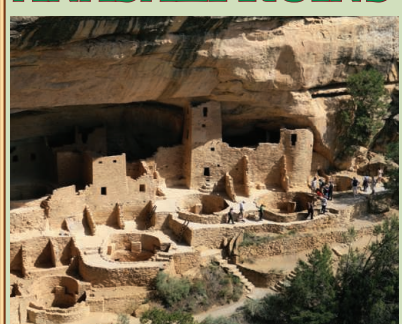
Discover LAPLAND

September 1-11, 2020

Discover the mysteries of Sami culture in Lapland, their exquisite artifacts in museums, reindeer herds, and huskies as you relish Finland and the Aurora Borealis! Enjoy the treasures of the forest during fall color, waterfalls, Finland's deepest gorge, and more!

From $4,420 pp + air.

ANASAZI RUINS



September 11-19, 2020

Explore the archaeological wonders and cultural heritage of the ancient peoples of the Southwest. The impressive Native American heritage, natural habitats, and breathtaking scenery will come alive with leadership by an excellent anthropologist and naturalist.

$3,495 pp twin share + air

For a detailed brochure, please call (800) 252-4910

All prices are per person twin share + air

Betchart Expeditions Inc.
17050 Montebello Rd
Cupertino, California 95014
Email: AAASInfo@betchartexpeditions.com
www.betchartexpeditions.com

under the International Commission on Stratigraphy voted to designate the Anthropocene as a new geological epoch, with the 1945 atomic bomb blasts serving as the strongest anthropogenic marker in the geological record (*8*). Looking forward, the U.S. military continues to invest billions in carbon-intensive preparations for conflicts in areas disproportionately affected by climate change (*9*).

The U.S. military's status quo is inconsistent with the aspirations of policy proposals like the Green New Deal (*10*) and the Paris Climate Accord (*11*). To decarbonize government and private sectors to the greatest possible extent, the United States and countries around the world must reevaluate American geopolitical aspirations and the foreign policy norms that have guided decisions since the Second World War.

Oliver Belcher[1]*, Benjamin Neimark[2], Patrick Bigger[2]
[1]School of Government and International Affairs, Durham University, UK. [2]Lancaster Environment Centre, Lancaster University, UK.
*Corresponding author.
Email: oliver.belcher@durham.ac.uk

REFERENCES AND NOTES

1. E. Kolbert, "What will another decade of climate crisis bring?" *The New Yorker* (2020).
2. N. C. Crawford, "Pentagon fuel use, climate change, and the costs of war" (Watson Institute, Brown University, 2019).
3. O. Belcher, P. Bigger, B. Neimark, C. Kennelly, *Trans. Inst. Br. Geograph.* **45**, 1 (2020).
4. A. Jorgenson, B. Clark, J. Kentor, *Glob. Environ. Politics* **10**, 7 (2010).
5. J. Stellman *et al.*, *Nature* **422**, 681 (2003).
6. P. Hashey, *N. Engl. J. Intl. Comp. Law* **291**, 299 (2011).
7. Y. Woo, *Southeast. Environ. Law J.* **15**, 577 (2007).
8. C. Waters *et al.*, *Earth-Sci. Rev.* **178**, 379 (2019).
9. M. Klare, *All Hell Breaking Loose: The Pentagon's Perspective on Climate Change* (Metropolitan Books, New York, 2019).
10. 116th Congress, H. Res. 109 (www.congress.gov/116/bills/hres109/BILLS-116hres109ih.pdf).
11. The Paris Agreement (https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement).

10.1126/science.abb1173

Brazilian national parks at risk

Brazilian national parks could be seriously threatened if pending legislation [draft bills 984/2019 (*1*) and 61/2013 (*2*)] is approved. These two bills aim to change the federal law [9.985/2000 (*3*)] that created Brazil's national system of conservation units and instead create a new category of conservation unit called "park-road." Beyond this, the bills would reopen the Colono Road, an 18-km-long road closed more than 30 years ago that bisects the famed Iguaçu National Park (*4*) and Brazil's largest remnant of the Atlantic Rainforest, a global biodiversity hotspot (*5*).

"Park-roads"—roads for vehicular traffic that traverse parks—are not currently regulated as a conservation-unit category in Brazil. Brazil's proposed bills are designed to allow roads for traveling through the parks, even if they serve no other purpose, such as allowing access to points of scenic, historic, or scientific importance. Reopening the Colono Road would expose Iguaçu National Park to problems such as poaching, illegal deforestation, wildlife trafficking, and vehicle roadkill of wildlife (*4*). Opening roads would also increase habitat fragmentation, one of the main causes of the contemporary biodiversity crisis (*6*), which alters many aspects of forest ecology and composition (*7*–*11*).

Brazil has more than 70 national parks occupying over 25 million hectares across different biomes (*12*), including the imperiled Atlantic Forest, Cerrado ecosystems, and Amazon rainforest. The proposed bills could set an alarming precedent for opening new roads inside these and other protected areas. Such action could cause irreparable damage to biodiversity and the climate-stabilization services of intact forests (*9*, *10*), which could affect not only Brazil but also the entire planet. Brazilian citizens and decision-makers must consider the consequences of such bills for Brazilian national parks and make their concerns about irreversible environmental impacts known to policymakers.

Renata Ruaro[1]*, William Laurance[2], Roger Paulo Mormul[3]
[1]Federal Technological University of Paraná, Câmpus Curitiba–Sede Ecoville, CEP 81280-340, Curitiba, Paraná, Brazil. [2]Centre for Tropical Environmental and Sustainability Science, College of Science and Engineering, James Cook University, Cairns, QLD 4878, Australia. [3]Ecology of Inland Water Ecosystems, State University of Maringá, CEP 87020-900, Maringá, Paraná, Brazil.
*Corresponding author.
Email: renataruaro_@hotmail.com

REFERENCES AND NOTES

1. Projeto de Lei N°984 (2019); www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2192602 [in Portuguese].
2. Projeto de Lei da Câmara N°61 (2013); https://www25.senado.leg.br/web/atividade/materias/-/materia/114299 [in Portuguese].
3. Lei N° 9.985 (18 July 2000); www.planalto.gov.br/ccivil_03/leis/L9985.htm [in Portuguese].
4. R. A. Ortiz, *Ambientalia* **1**, 141 (2010).
5. World Wildlife Fund, "Guia de Fauna do Parque Nacional do Iguaçu" (2014); https://d3nehc6yl9qzo4.cloudfront.net/downloads/guia_fauna_parna_iguacu.pdf [in Portuguese].
6. W. F. Laurance, I. Arrea, *Science* **358**, 442 (2017).
7. V. Arroyo-Rodríguez *et al.*, *Ecol. Res.* **32**, 81 (2017).
8. D. A. Saunders *et al.*, *Conserv. Biol.* **5**, 18 (1991).
9. W. F. Laurance, *Phil. Trans. Royal Soc. B* **359**, 345 (2004).
10. N. M. Haddad *et al.*, *Sci. Adv.* **1**, e1500052 (2015).
11. D. Lesbarrères, L. Fahrig, *Trends Ecol. Evol.* **27**, 374 (2012).
12. Brasil, MMA/ICMBio, "Ministro anuncia concessões em parques nacionais" (2016). www.icmbio.gov.br/portal/ultimas-noticias/20-geral/8569-ministro-anuncia-concessoes-em-parques [in Portuguese].

10.1126/science.abb0926

Published by AAAS

Claire Fraser introduces AAAS CEO Sudip Parikh at the 2020 AAAS Annual Meeting in Seattle.

PHOTO: ROBB COHEN PHOTOGRAPHY & VIDEO

AAAS President Claire Fraser begins a year-long term

Fraser offers AAAS a career driven by curiosity, collaboration, and discovery

By **Anne Q. Hoy**

Claire Fraser began her scientific research on the banks of upstate New York's Lake George and over the years progressed to a commanding role in the emerging field of microbial genomics.

Throughout an acclaimed career that helped launch the field of microbial genomics, she has somehow reserved time for adventures outside of the scientific realm: sailboat racing when younger and now competitive ballroom dancing, alongside long-standing commitments to the scientific enterprise that include her new role as president of the American Association for the Advancement of Science.

Fraser was a senior at Rensselaer Polytechnic Institute when she took on her first independent research project, examining why and how the ecology of the Lake George region was being disturbed.

Algae blooms had begun to take hold of water surfaces, pesticide runoff was flowing into waterways, and plants and fish were dying at a time when agricultural compounds were increasingly in use.

She collected water and soil samples and, in the laboratory, hunted for bacteria able to break down agricultural pollutants, a process known as bioremediation—now widely used to address many types of water contamination.

Fraser identified a species of bacteria that could be used to counteract the pollutants, but the project's 1-year time limit dictated a premature ending. Still, the introduction to research was rewarding for a soon-to-be college graduate with two targets in sight: a bachelor of science degree in biology and the option of medical school.

"I remember thinking, 'how cool is this?'" said Fraser, now director of the Institute for Genome Sciences at the University of Maryland School of Medicine. "That's what has always driven me: just the sheer joy of knowing that when you set up a new experiment, when you're trying something new, and when it works, you're seeing something for the first time. That's so great."

From the start, Fraser welcomed the all-consuming nature of research and the exhilaration she drew from witnessing even incremental results, factors that convinced her to bypass medical school and pursue what became a globally recognized scientific research career that continues to reverberate through the scientific enterprise today.

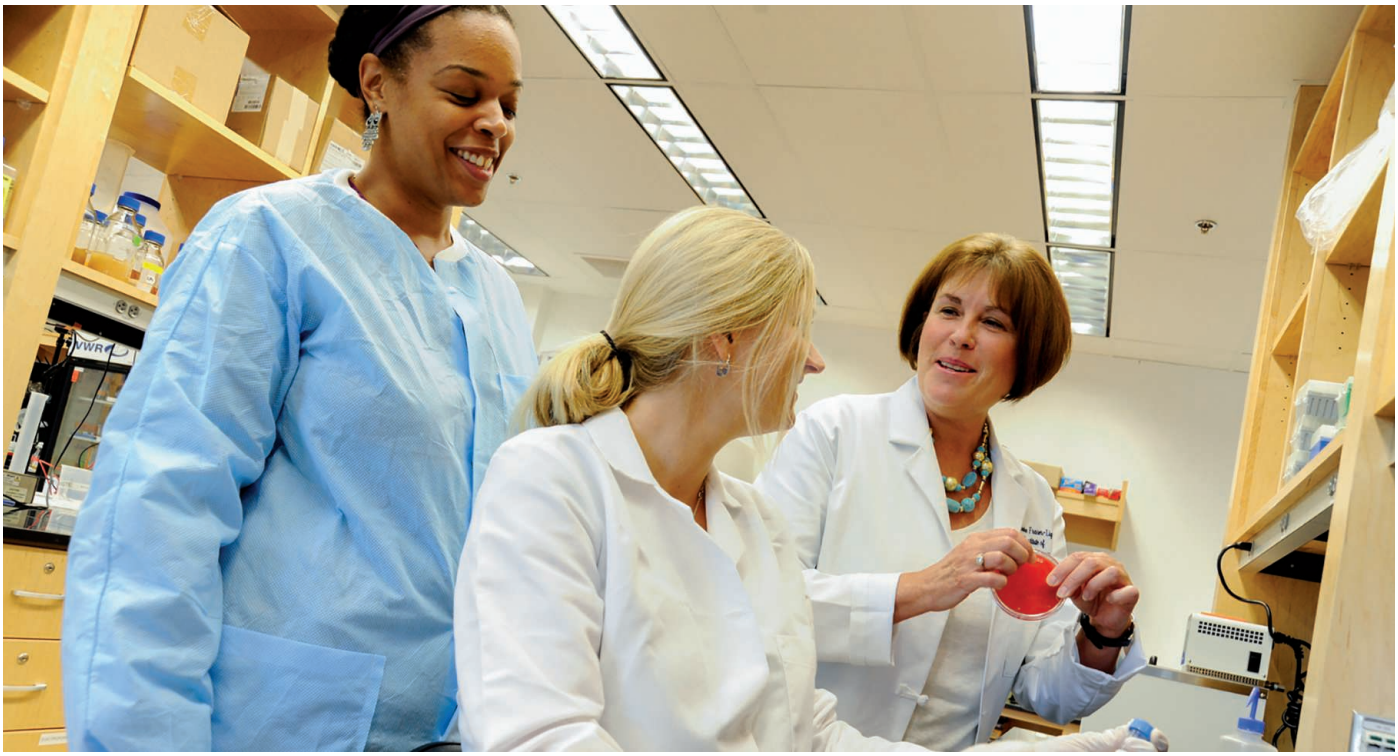
The University of Maryland School of Medicine has described her as someone who knows "as much about bacteria as anyone in the world."

On 17 February, the day following the close of the 2020 AAAS Annual Meeting in Seattle, Washington, Fraser began a 1-year term as president of AAAS, a position that will be followed by another 1-year term as chair of the AAAS Board of Directors in 2021.

Among contributions during her presidency, Fraser will help set the theme of the 2021 AAAS Annual Meeting slated to take place in Phoenix, Arizona. She is also looking forward to assisting AAAS CEO Sudip Parikh, a structural biologist with a Ph.D. in biochemistry, to "formulate a 21st-century strategic vision" for the organization.

Fraser believes, as she put it in a recent University of Maryland video, that one of the highest priorities for the scientific community is to work hard to reverse "the relatively low regard that science and scientists are held in by the general public," and continually demonstrate how science can address such issues as climate change, food and water security, emerging infectious diseases and antibiotic resistance, and more. "That's really my biggest concern

Published by AAAS



Early career research scientists who worked with Fraser praise her mentoring skills, encouragement, and support.

and that's where an organization like AAAS, that cuts across many disciplines, must play an essential role," she said.

Fraser's dedication to the joy of scientific discovery has dominated her work even before she earned her Ph.D. in pharmacology in 1981 at the State University of New York at Buffalo. It also has dominated a scientific career that has combined biology and genomics and helped decipher the genetic instructions of bacteria, plants, and parasites.

Following her Ph.D. work at SUNY Buffalo, Fraser learned the tools of molecular biology in the National Institutes of Health Intramural Program. The program, she said, provided an open environment that encouraged "fearless" scientific pursuits and offered "outstanding training." Her research at NIH was devoted to investigating G protein–coupled receptors, which facilitate cell-to-cell communication and are important pharmaceutical targets.

Eight years later, she helped to establish and eventually became president of The Institute for Genomic Research, TIGR, a renowned private research organization founded by Fraser's first husband, J. Craig Venter, a biochemist and geneticist.

In 1995, Fraser was part of a TIGR research group that was the first to map the genetic code of *Haemophilus influenzae*, a bacterium responsible for respiratory infections and meningitis in babies and young children. The findings were published in the 28 July 1995 issue of *Science* and are often cited as "the first representation of the complete genetic code of a free-living organism," a discovery that helped pioneer the field of microbial genomics.

Fraser enjoys ballroom dancing competitions as much as science.

"We had no guarantee that this was going to work at all, but it did," said Fraser in an interview. "That was really a turning point, when the potential of microbial genomics was revealed with the publication of that landmark paper on the cover of *Science*. Once this first proof of concept was accomplished, all the major U.S. funding agencies jumped on the microbial genomics bandwagon."

Fraser and her TIGR colleagues look back at this period with fond nostalgia, as pioneers in a new field of scientific investigation that would produce what she estimates as "probably close to a hundred thousand bacterial genome sequences in public databases today."

Using all the same tools that were being leveraged to generate bacteria, plant, and parasite sequences, Fraser would be part of a team of investigators who would sequence the genome of Shadow, the oldest of one of her three standard poodles. "He was a wonderful dog," she said, noting that she remains the owner of another trio of standard poodles.

Scientific advances by TIGR led Rita Colwell, then the director of the National Science Foundation, to seek out Fraser and her research team for help with a national security project to sequence anthrax bacterium spores collected from anonymous letters that had been sent to members of Congress and the offices of media companies a week after the 11 September 2001 attacks.

Fraser's research "was very, very productive and very important for national security," said Colwell, adding that she expected nothing less, having worked successfully with Fraser on earlier projects such as sequencing two cholera chromosomes, research that would advance

PHOTOS (TOP TO BOTTOM): UNIVERSITY OF MARYLAND; GIACOMELLO GROUP (ITALY)

Published by AAAS

understanding of the deadly organism.

Fraser also has served on editorial boards of scientific journals and as a member of scientific, private, and public advisory boards, including the National Science Advisory Board for Biosecurity, for which her policy-specific knowledge was sought to inform national security concerns.

Among the numerous, prestigious awards Fraser has received was election as a AAAS Fellow in 2004. A year later, she was recognized as the world's most highly cited microbiological researcher over the previous decade, and she was elected to the National Academy of Medicine in 2011.

TIGR research scientists who worked with Fraser praise the support and freedom she provided them and underscore how her approach helped launch their early research careers.

Elodie Ghedin, now a professor of biology and epidemiology at New York University and a 2011 MacArthur Fellowship recipient, said that Fraser encouraged exploration, free thinking, and teamwork, inspiration that prompted Ghedin to examine unexplored topics during her time at TIGR. Fraser's support contributed to the work that earned Ghedin the MacArthur Fellowship, she said.

"What was incredible with Claire was if you were hungry and enthusiastic and had ideas for things, she would completely facilitate things for you," said Ghedin, describing how Fraser found research funding for a novel investigation Ghedin wanted to pursue. "It was an environment that really led to people carving out their own path. Claire is really a leader…she's hands-off, but at the same time, there when you needed her."

Fraser approaches her personal interests with a vigor equal to the dedication she applies to her scientific work. She spends much of her free time on Nantucket in Massachusetts near where she grew up in the Boston suburb of Saugus, the daughter of "incredibly supportive" parents—her mother an elementary school teacher and father the principal of her local high school.

Over the past 5 years, her embrace of competitive ballroom dancing reveals the depth of her determination even outside science. Fraser trains up to four times a week getting ready for competitions, and she participates in local and international events a few times each year, as her work schedule permits.

She has two closets filled with costumes and drawers stuffed with theatrical makeup and false eyelashes. Her husband, author Jack Kammer, supports her and accompanies her to competitions.

Like science, ballroom dancing is an obsession. It is an activity that fills her with joy. And like science, she knows that what you get out of dancing is only as good as what you put in.

2019 election results

Below are the results of the 2019 election. Terms begin on 17 February 2020.

GENERAL ELECTION

President-Elect: *Susan G. Amara, National Institute of Mental Health/NIMH*

Board of Directors: *Cynthia M. Beall, Case Western Univ.; Alondra Nelson, Social Science Research Council/ I nstitute for Advanced Study*

Committee on Nominations: *Marlene Belfort, Univ. at Albany, the State Univ. of New York; John G. Hildebrand, Univ. of Arizona; Michael Prather, Univ. of California, Irvine; Esther Takeuchi, Stony Brook Univ./Brookhaven National Laboratory*

SECTION ELECTIONS

Agriculture, Food, and Renewable Resources

Chair-Elect: G. Philip Robertson, *Michigan State Univ.*

Member-at-Large of the Section Committee: *Mary Ann Ottinger, Univ. of Houston*

Electorate Nominating Committee: *Sylvie M. Brouder, Purdue Univ.; Catherine W. Ernst, Michigan State Univ.*

Anthropology

Chair-Elect: *George Robert Milner, Pennsylvania State Univ.*

Member-at-Large of the Section Committee: *Leslea J. Hlusko, Univ. of California, Berkeley*

Electorate Nominating Committee: *Lyle W. Konigsberg, Univ. of Illinois at Urbana-Champaign; Jada Benn Torres, Vanderbilt Univ.*

Astronomy

Chair-Elect: *Anneila I. Sargent, California Institute of Technology*

Member-at-Large of the Section Committee: *Kelly Holley-Bockelmann, Vanderbilt Univ.*

Electorate Nominating Committee: *Catharine (Katy) Garmany, NSF's National Optical-Infrared Astronomy Research Laboratory (Retired); Rachel A. Osten, Space Telescope Science Institute*

Atmospheric and Hydrospheric Sciences

Chair-Elect: Susan S. Hubbard, *Lawrence Berkeley National Laboratory*

Member-at-Large of the Section Committee: *Colette L. Heald, Massachusetts Institute of Technology*

Electorate Nominating Committee: *Sarah B. Kapnick, Geophysical Fluid Dynamics Laboratory; Isabella Velicogna, Univ. of California, Irvine*

Council Delegate: *Jennifer Francis, Woods Hole Research Center*

Biological Sciences

Chair-Elect: *Nancy Jean Cox, Vanderbilt Univ. Medical Center*

Member-at-Large of the Section Committee: *Lara M. Kueppers, Univ. of California, Berkeley*

Electorate Nominating Committee: *Daniela Drummond-Barbosa, Johns Hopkins Univ.; Pamela Geyer, Univ. of Iowa*

Council Delegate: *Susan E. Celniker, Lawrence Berkeley National Laboratory; Carolyn M. Teschke, Univ. of Connecticut*

Chemistry

Chair-Elect: *Peter K. Dorhout, Kansas State Univ.*

Member-at-Large of the Section Committee: *Julia Chan, The Univ. of Texas at Dallas*

Electorate Nominating Committee: *Barbara A. Baird, Cornell Univ.; Karen L. Wooley, Texas A&M Univ.*

Dentistry and Oral Health Sciences

Chair-Elect: *Ichiro Nishimura, Univ. of California, Los Angeles*

Member-at-Large of the Section Committee: *Nisha J. D'Silva, Univ. of Michigan School of Dentistry*

Electorate Nominating Committee: *Yang Chai, Univ. of Southern California; Linda M. Kaste, Univ. of Illinois at Chicago College of Dentistry*

Council Delegate: *Catherine E. Ovitt, Univ. of Rochester School of Medicine and Dentistry*

Education

Chair-Elect: *Melanie M. Cooper, Michigan State Univ.*

Member-at-Large of the Section Committee: *Maria Elizabeth Alvarez, El Paso County Community College*

Electorate Nominating Committee: *Jeffery P. Braden, North Carolina State Univ.; Sandra S. West, Texas State Univ.*

Council Delegate: *Cynthia M. Bauerle, James Madison Univ.; Kristin P. Jenkins, BioQUEST Curriculum Consortium, Inc.*

Engineering

Chair-Elect: *Jennifer Sinclair Curtis, Univ. of California, Davis*

Member-at-Large of the Section Committee: *Pramod P. Khargonekar, Univ. of California, Irvine*

Electorate Nominating Committee: *Lynnette D. Madsen*, National Science Foundation (*serving in personal capacity); Georgia Gina Tourassi, Oak Ridge National Laboratory*

Council Delegate: *Jacqueline H. Chen, Sandia National Laboratories*

General Interest in Science and Engineering

Chair-Elect: *Erika C. Shugart, American Society for Cell Biology*

Published by AAAS

Member-at-Large of the Section Committee: *Marilee Long, Colorado State Univ.*

Electorate Nominating Committee: *Sue Nichols, Michigan State Univ.; Jocelyn Steinke, Western Michigan Univ.*

Council Delegate: *JoAnn M. Valenti, Emerita*

Geology and Geography

Chair-Elect: *Suzanne P. Anderson, Univ. of Colorado Boulder*

Member-at-Large of the Section Committee: *Kam-biu Liu, Louisiana State Univ.*

Electorate Nominating Committee: *Luis A. González, Univ. of Kansas; Adina Paytan, Univ. of California, Santa Cruz*

History and Philosophy of Science

Chair-Elect: *Jonathan C. Coopersmith, Texas A&M Univ.*

Member-at-Large of the Section Committee: *Carol E. Cleland, Univ. of Colorado Boulder*

Electorate Nominating Committee: *Kristine C. Harper, Univ. of Copenhagen (Denmark); Susan D. Jones, Univ. of Minnesota*

Industrial Science and Technology

Chair-Elect: *Ernesto E. Marinero, Purdue Univ.*

Member-at-Large of the Section Committee: *Lizhi Sun, Univ. of California, Irvine*

Electorate Nominating Committee: *Vanessa Z. Chan, Univ. of Pennsylvania School of Engineering and Applied Sciences; Anne L. Plant, National Institute of Standards and Technology*

Information, Computing, and Communication

Chair-Elect: *Nancy M. Amato, Univ. of Illinois at Urbana-Champaign*

Member-at-Large of the Section Committee: *Sven Koenig, Univ. of Southern California*

Electorate Nominating Committee: *Julia Gelfand, Univ. of California, Irvine; Dana S. Nau, Univ. of Maryland*

Council Delegate: *Benjamin Kuipers, Univ. of Michigan; Edward D. Lazowska, Univ. of Washington*

Linguistics and Language Science

Chair-Elect: *Randi C. Martin, Rice Univ.*

Member-at-Large of the Section Committee: *Susan E. Brennan, Stony Brook Univ.*

Electorate Nominating Committee: *Jane MacNeil, Red Deer College (Canada); Matthew Wagers, Univ. of California, Santa Cruz*

Council Delegate: *Joan A. Sereno, Univ. of Kansas*

Mathematics

Chair-Elect: *Ken Ono, Univ. of Virginia*

Member-at-Large of the Section Committee: *Michel L. Lapidus, Univ. of California, Riverside*

Electorate Nominating Committee: *Keri A. Kornelson, Univ. of Oklahoma; Judy L. Walker, Univ. of Nebraska-Lincoln*

Medical Sciences

Chair-Elect: *David A. Frank, Dana-Farber Cancer Institute*

Member-at-Large of the Section Committee: *Gabriela K. Popescu, Univ. at Buffalo*

Electorate Nominating Committee: *Lewis A. Chodosh, Perelman School of Medicine at the Univ. of Pennsylvania; Mary Dasso, National Institute of Child Health and Human Development*

Neuroscience

Chair-Elect: *Judith R. Walters, National Institute of Neurological Disorders and Stroke/National Institutes of Health*

Member-at-Large of the Section Committee: *Nina F. Schor, National Institute of Neurological Disorders and Stroke/National Institutes of Health*

Electorate Nominating Committee: *Peggy Mason, Univ. of Chicago; Alcino J. Silva, Univ. of California, Los Angeles*

Council Delegate: *Jeffrey L. Noebels, Baylor College of Medicine*

Pharmaceutical Sciences

Chair-Elect: *Peter J. Houghton, The Univ. of Texas Health Center at San Antonio*

Member-at-Large of the Section Committee: *James T. Dalton, Univ. of Michigan*

Electorate Nominating Committee: *Thomas L. Poulos, Univ. of California, Irvine; Thomas E. Prisinzano, Univ. of Kentucky*

Council Delegate: *Patrick M. Woster, Medical Univ. of South Carolina*

Physics

Chair-Elect: *Susan N. Coppersmith, Univ. of New South Wales (Australia)/ Univ. of Wisconsin-Madison*

Member-at-Large of the Section Committee: *Melissa A. Hines, Cornell Univ.*

Electorate Nominating Committee: *Jonathan L. Feng, Univ. of California, Irvine; Dragana Popovic, National High Magnetic Field Laboratory (NHMFL)/ Florida State Univ.*

Psychology

Chair-Elect: *Deanna M. Barch, Washington Univ.*

Member-at-Large of the Section Committee: *Diane M. Mackie, Univ. of California, Santa Barbara*

Electorate Nominating Committee: *Robert M. Nosofsky, Indiana Univ.; Suparna Rajaram, Stony Brook Univ.*

Social, Economic, and Political Sciences

Chair-Elect: *Mary Frank Fox, Georgia Institute of Technology*

Member-at-Large of the Section Committee: *Constance A. Nathanson, Columbia Univ.; Suzanne M. Rivera, Case Western Reserve Univ.*

Electorate Nominating Committee: *Maria Charles, Univ. of California, Santa Barbara; Virginia Sapiro, Boston Univ.*

Societal Impacts of Science and Engineering

Chair-Elect: *Melanie Roberts, Pacific Northwest National Laboratory*

Member-at-Large of the Section Committee: *Dahlia L. Sokolov, Committee on Science, Space, and Technology, U.S. House of Representatives*

Electorate Nominating Committee: *Evans S. Michelson, Alfred P. Sloan Foundation; Lori A. Perine, Interpretech*

Council Delegate: *Francesca Grifo, U.S. Environmental Protection Agency*

Statistics

Chair-Elect: *Leonard A. Stefanski, North Carolina State Univ.*

Member-at-Large of the Section Committee: *Joseph W. Hogan, Brown Univ.*

Electorate Nominating Committee: *Sharon-Lise Normand, Harvard Medical School; Elizabeth A. Stuart, Johns Hopkins Univ. Bloomberg School of Public Health*

Council Delegate: *Nancy Reid, Univ. of Toronto (Canada)*

Call for nomination of 2020 Fellows

Fellows who are current members of AAAS are invited to nominate members for election as Fellows. A member whose efforts on behalf of the advancement of science or its applications are scientifically or socially distinguished, and who has been a continuous member for the 4-year period leading up to the year of nomination, may by virtue of such meritorious contribution be elected a Fellow by the AAAS Council.

A nomination must be sponsored by three previously elected AAAS Fellows (who are current in their membership), two of whom must have no affiliation with the nominee's institution. Nominations undergo review by the steering groups of the Association's sections (the chair, chair-elect, retiring chair, secretary, and four members-at-large of each section).

Each steering group reviews only those nominations designated for its section. Names of Fellow nominees who are approved by the steering groups are presented to the Council in the fall for election.

Nominations with complete documentation must be received by 29 April 2020. Nominations received after that date or nominations that are incomplete as of the deadline will not move forward.

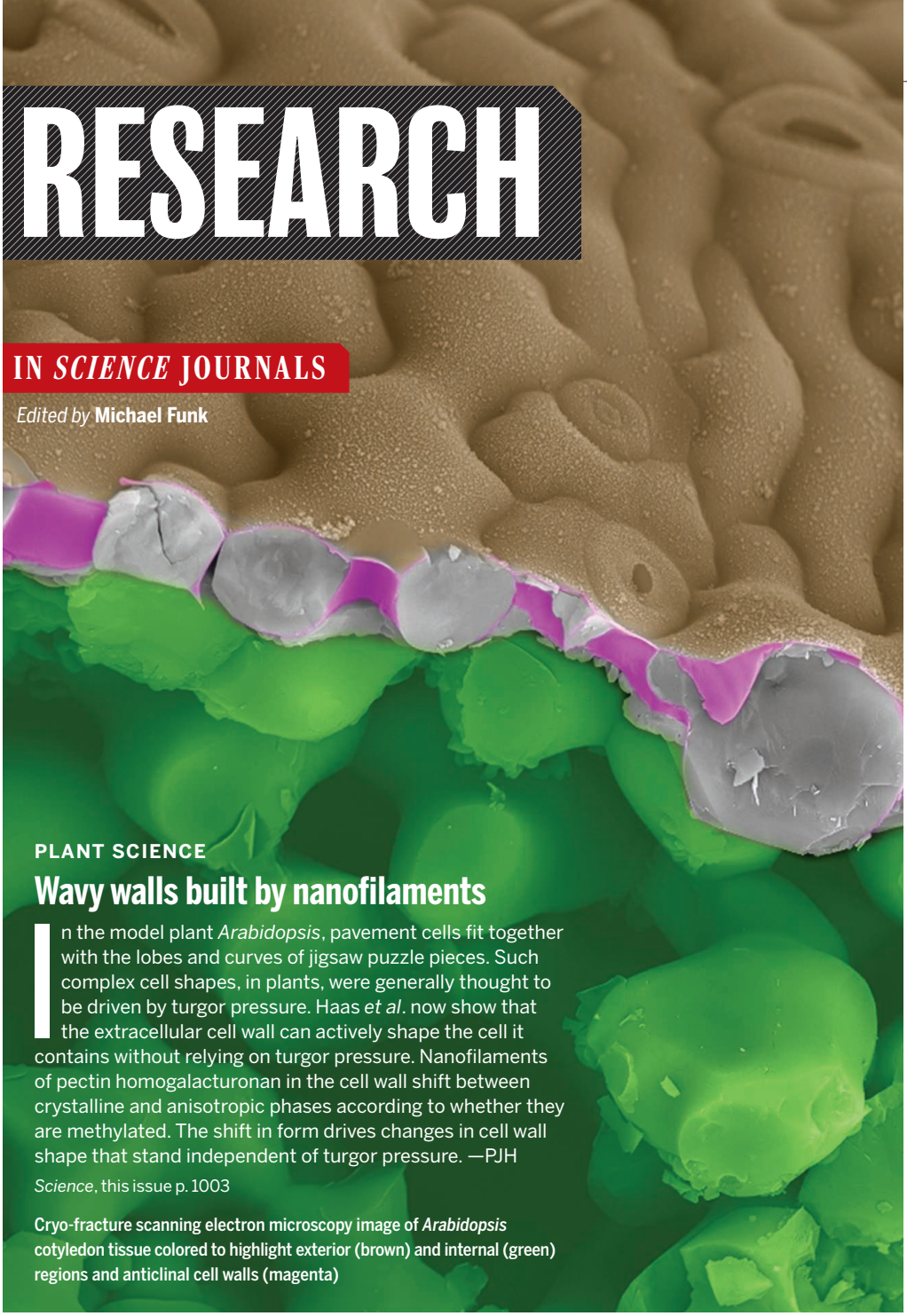
Complete instructions and a copy of the nomination form are available at www.aaas.org/current-nomination-cycle. Questions may be directed to fellownomination@aaas.org.

Published by AAAS

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



PLANT SCIENCE

Wavy walls built by nanofilaments

In the model plant *Arabidopsis*, pavement cells fit together with the lobes and curves of jigsaw puzzle pieces. Such complex cell shapes, in plants, were generally thought to be driven by turgor pressure. Haas *et al*. now show that the extracellular cell wall can actively shape the cell it contains without relying on turgor pressure. Nanofilaments of pectin homogalacturonan in the cell wall shift between crystalline and anisotropic phases according to whether they are methylated. The shift in form drives changes in cell wall shape that stand independent of turgor pressure. —PJH

Science, this issue p. 1003

Cryo-fracture scanning electron microscopy image of *Arabidopsis* cotyledon tissue colored to highlight exterior (brown) and internal (green) regions and anticlinal cell walls (magenta)

IMAGE: HAAS *ET AL*.

NEUROSCIENCE

Speech versus music in the brain

To what extent does the perception of speech and music depend on different mechanisms in the human brain? What is the anatomical basis underlying this specialization? Albouy *et al*. created a corpus of a cappella songs that contain both speech (semantic) and music (melodic) information and degraded each stimulus selectively in either the temporal or spectral domain. Degradation of temporal information impaired speech recognition but not melody recognition, whereas degradation of spectral information impaired melody recognition but not speech recognition. Brain scanning revealed a right-left asymmetry for speech and music. Classification of speech content occurred exclusively in the left auditory cortex, whereas classification of melodic content occurred only in the right auditory cortex. —PRS

Science, this issue p. 1043

ECOLOGICAL SPECIATION

Resisting extinction

Prevailing evolutionary wisdom tells us that ecological differentiation leads to speciation. Whether this pattern can be seen over paleontological time, however, has been difficult to test. Knope *et al*. looked at a dataset of thousands of modern and extinct marine groups and found the relationship to be more complex than expected. Ecological diversification is associated with lower rates of origination, and the taxonomical richness seemingly associated with these groups is due to resistance to extinction. Furthermore, the researchers found that the strong association between ecological differentiation and taxonomic diversity is a recent development shaped by extinction events over time. —SNV

Science, this issue p. 1035

FLUID DYNAMICS

Machine-learning fluid flow

Quantifying fluid flow is relevant to disciplines ranging from geophysics to medicine. Flow can be experimentally visualized using, for example, smoke or contrast agents, but extracting velocity and pressure fields from this information is tricky. Raissi *et al*. developed a machine-learning approach to tackle this problem. Their method exploits the knowledge of Navier-Stokes equations, which govern the dynamics of fluid flow in many scientifically relevant situations. The authors illustrate their approach using examples such as blood flow in an aneurysm. —JS

Science, this issue p. 1026

TUMOR IMMUNOLOGY

Getting a hold on MDSCs

Myeloid-derived suppressor cells (MDSCs) are immune cells that mediate immune suppression and are correlated with progressing cancer. How these cells arise and whether they can be therapeutically targeted akin to exhausted T cells are areas of active investigation. A persistent challenge in studying MDSCs has been the identification of MDSC-specific cell-surface markers that can facilitate their

Published by AAAS

isolation and characterization. Using single-cell RNA sequencing in a mouse model of breast cancer, Alshetaiwi *et al*. defined gene signatures that distinguish MDSCs from other myeloid and granulocytic cells. They identified the protein CD84 to be a robust cell-surface marker for identification of MDSCs in both human and murine breast cancer. Whether their findings can be extended to MDSCs in other cancer settings remains to be seen. —AB

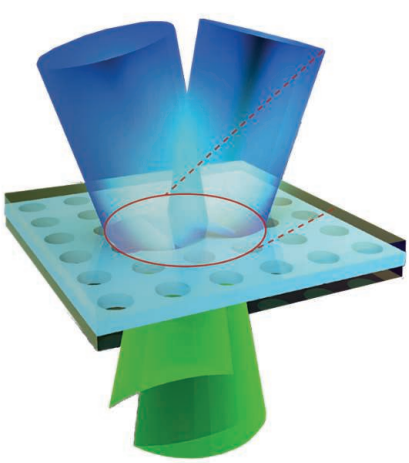
Sci. Immunol. 5, eaay6017 (2020).

OPTICS

Ultrafast vortex microlasers

For applications in ultrafast communication, all-optical switches will require low energy consumption, high speed, a strong modulation ratio, a small footprint, and on-chip integration. Although the small footprint and on-chip integration are accessible, the trade-off between low energy consumption and high speed has been challenging. Huang *et al*. exploited the idea of bound states in the continuum, effectively a high–quality (*Q*) cavity without the physical cavity, to design vortex lasers with highly directional output and single-mode operation. With the trade-off between low energy consumption and high speed now broken, it should be possible to realize ultrafast optical switching that meets all the requirements of modern classic and quantum information. —ISO

Science, this issue p. 1018



Experimental schematic for a vortex laser pumped with two circular beams

SOCIAL SCIENCE

Economic losses from illicit fish trade

Illegal, unreported, and unregulated fishing—and the subsequent illicit fish trade market that follows—have a substantial impact on local and national economies. Sumaila *et al.* estimate that the illicit trade in marine fish catch adds up to between 8 million and 14 million metric tons per year. With suggested gross revenues totaling $9 billion to $17 billion, the researchers estimate potential losses of $26 billion to $50 billion to world economies owing to the diversion of fish from the legitimate trade system. Bold policy and action by both the private and public sectors are needed to mitigate the substantial economic effects of illicit trade in marine fish catch. —JJ

Sci. Adv. 10.1126/sciadv.aaz3801 (2020).

ORGANIC CHEMISTRY

Amines as a gateway to alkyl radicals

In recent years, photoredox catalysis driven by blue light has often been used to oxidize carbon centers adjacent to nitrogen. Constantin *et al*. now show that these aminoalkyl radicals can, in turn, conveniently strip iodine atoms from a variety of alkyl carbons. The new alkyl radicals that result readily undergo deuteration and couplings such as alkylation, allylation, and olefination. The upshot is that simple amines can replace more hazardous conventional reagents such as trialkyltin compounds in the homolytic activation and functionalization of halocarbons. —JSY

Science, this issue p. 1021

IN OTHER JOURNALS

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

CELL BIOLOGY

Exiting without leaving a hole

The endoplasmic reticulum (ER) is an intracellular membranous organelle that is sometimes used to accommodate the assembly of viruses or other large macromolecular particles. How can such entities get out of the ER and into the cytosol without disrupting the ER membranes? During infection, polyomavirus SV40 enters the cytosol en route to the nucleus after trafficking through the ER. Chen *et al*. studied the release of SV40 from the ER into the cytosol in cultured mammalian cells. They found that reticulons—ER proteins known to be involved in the tubular morphology of the ER—help to protect the integrity of the ER during viral exit. This same function appears to be important in the transfer of misfolded proinsulin aggregates directly from the ER to lysosomes. Thus, as is their wont, viruses have hijacked a normal cellular pathway to ensure their successful entry, replication, assembly, and transmission. —SMH

J. Cell Biol. **219**, e201908182 (2020).

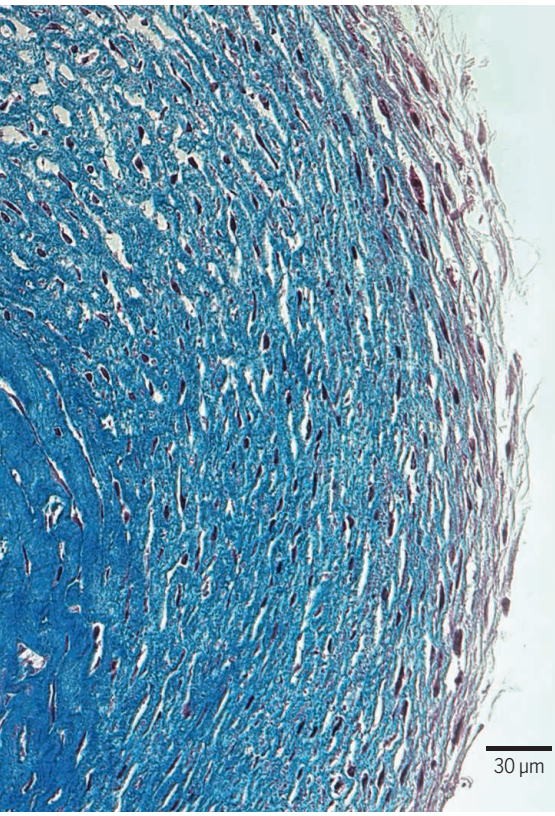
NEUROSCIENCE

Rewiring the adolescent brain

As children go through adolescence, many undergo profound and occasionally disruptive changes in behavior. How does this map onto changes in the developing brain? Using functional magnetic resonance imaging in healthy young people (14 to 26 years old), Váša *et al*. observed two modes in functional connectivity between brain regions. By 14 years of age, conservative regions, often specialized for basic sensory and motor functions, are already strongly connected and consolidate further as an adolescent reaches adulthood. By contrast, disruptive regions, activated by complex tasks, show shifts in connectivity as a person transitions between the ages of 14 and 26. The disruptive maturation of connectivity between cortex and

CREDITS (FROM LEFT): HUANG *ET AL.*; J. LUO *ET AL.*, *CELL STEM CELL* **26**, 251 (2020)

Published by AAAS



Section of a stress-exposed, engineered vascular graft generated from human stem cells and stained by Masson's trichrome

VASCULAR BIOLOGY

Generating strong vascular grafts

When treating injury or vascular disease, autologous or synthetic grafts are useful, but supplies are limited and often suboptimal. Robust tissue-engineered vascular grafts (TEVGs), particularly ones that can withstand mechanical stress, are highly desirable. Luo *et al*. generated vascular smooth muscle cells from human induced pluripotent stem cells and cultured them on biodegradable polyglycolic acid scaffolds exposed to pulsatile radial stress. When implanted into rats, these TEVGs showed high mechanical strength with no dilation, elongation, or wall thickening. Moreover, the TEVGs associated with limited thrombosis and showed no teratoma formation. Further optimization and engineering to limit immune rejection should provide cells and tissues suitable for treating injury or dysfunction of the cardiovascular system. —BAP

Cell Stem Cell **26**, 251 (2020).

subcortex may reflect metabolic remodeling that underpins development of sophisticated adult faculties, such as socializing, mentalizing, and executive skills. —PRS

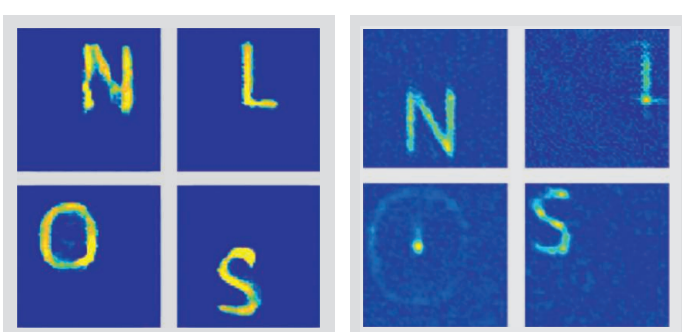
Proc. Natl. Acad. Sci. U.S.A. **117**, 3248 (2020).

COMPUTATIONAL IMAGING

Imaging out-of-sight objects

Seeing around corners or imaging subjects in a room without a direct line of sight might seem like an attribute of a superhero, but by bouncing photons off a wall so that some of them scatter from a hidden object and then capturing some of the ones that make it back to a detector, an image of that object can be reconstructed. However, such non–line-of-sight imaging techniques are often compromised by the resolution and quality of the reconstructed image, limiting existing methods to long acquisition times and small standoff distances. Metzler *et al*. developed a theoretical model to describe correlations within the scattered speckle pattern and combined it with a trained neural network for solving the details of the scattering process. The technique might help improve the quality of non–line-of-sight imaging methods and provide access to real-time, high-resolution imaging capabilities. —ISO

Optica **7**, 63 (2020).



Non–line-of-sight image of an out-of-view target using a convolutional neural network approach (left, this work) is superior to that made using an older hybrid input-output method (right).

CREDIT: C. METZLER *ET AL*., *OPTICA* **7**, 63 (2020)

NEURODEVELOPMENT

Sourcing brain abnormality

Disruptions in the transcription factor TCF4 lie at the root of Pitt-Hopkins syndrome (PTHS), which manifests with intellectual disability and disruptions in brain development. Schoof *et al*. generated mice with inducible and tissue-specific truncation of TCF4 to ensure that gene disruption occurs only in the central nervous system. The truncated protein lacks the DNA-binding domain of TCF4, a mutational hotspot for PTHS. Embryonic mice with this truncated TCF4 showed deficits in differentiation and migration of neural precursor cells. During postnatal development, Cajal-Retzius cells were misplaced and hippocampal development was disrupted. The corpus callosum, which is the fiber tract connecting the brain hemispheres, also developed poorly in these mice. These findings thus map the anatomical abnormalities that underlie this syndrome. —PJH

Eur. J. Neurosci. 10.1111/ejn.14674 (2020).

ASTEROIDS

Asteroid dynamics inside Venus' orbit

Most asteroids reside in the main-belt region between the orbits of Mars and Jupiter. A population of asteroids within the orbit of Venus has been hypothesized, but searching that region is difficult because it can only be observed just before sunrise or just after sunset. The first asteroid with an orbit entirely inside Venus' orbit, designated 2020 AV_2, was discovered on 4 January 2020. Greenstreet used simulations to assess its orbital dynamics and found that the orbit is unstable: Within a million years, the asteroid will either collide with Venus or be scattered onto an Earth-crossing orbit. A nearby 3:2 resonance with Venus could extend the lifetime to a few million years. —KTS

Mon. Not. R Astron. Soc. Lett. **2020**, slaa025 (2020).

NUCLEAR PHYSICS

A wobbly nucleus

Atomic nuclei consist of nucleons—protons and neutrons. The composition of the nucleus affects its overall shape, ranging from spherical to ellipsoidal with three unequal axes. Nuclei with the latter shape—referred to as triaxial—are prone to wobbling, behaving not unlike a spinning top. Sensharma *et al*. studied a particular kind of wobble in a gold nucleus, ^{187}Au. They found that this nucleus exhibits longitudinal wobbling, a phenomenon that has been theorized but not observed experimentally before. —JS

Phys. Rev. Lett. **124**, 052501 (2020).

Published by AAAS

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CANCER

Improving the drug development pipeline

Artificial intelligence (AI) applications are being developed to improve cancer drug discovery and administration. Moreover, AI can improve clinical trial design and therapeutic dosing strategies such as adaptive therapy, whereby patients are treated according to tumor response to avoid resistance. In a Perspective, Ho discusses how AI applications can improve cancer drug development and the challenges to be overcome. —GKA

Science, this issue p. 982

CLINICAL TRIALS

CRISPR takes first steps in humans

CRISPR-Cas9 is a revolutionary gene-editing technology that offers the potential to treat diseases such as cancer, but the effects of CRISPR in patients are currently unknown. Stadtmauer *et al*. report a phase 1 clinical trial to assess the safety and feasibility of CRISPR-Cas9 gene editing in three patients with advanced cancer (see the Perspective by Hamilton and Doudna). They removed immune cells called T lymphocytes from patients and used CRISPR-Cas9 to disrupt three genes (*TRAC*, *TRBC*, and *PDCD1*) with the goal of improving antitumor immunity. A cancer-targeting transgene, NY-ESO-1, was also introduced to recognize tumors. The engineered cells were administered to patients and were well tolerated, with durable engraftment observed for the study duration. These encouraging observations pave the way for future trials to study CRISPR-engineered cancer immunotherapies. —PNK

Science, this issue p. 1001; see also p. 976

MICROBIOTA

Mouse mothers transfer metabolic mode

Obesity and metabolic diseases tend to go together, and humans who become obese are also prone to type 2 diabetes and cardiovascular problems. Starting with the observation that offspring of germ-free mice tended to become obese on high-fat diets, Kimura *et al*. investigated how the presence of the microbiota might be protective in mice (see the Perspective by Ferguson). Short-chain fatty acids (SCFAs) from the microbiota are known to suppress insulin signaling and reduce fat deposition in adipocytes. Further experiments showed that SCFAs in the bloodstream were able to pass from a non–germ-free mother's gut microbiota across the placenta and into the developing embryos. The authors found that in the embryos, the SCFA propionate mediates not only insulin levels through GPR43 signaling but also sympathetic nervous system development through GPR41 signaling. A high-fiber diet promoted propionate production from the maternal microbiota, and maternal antibiotic treatment resulted in obese-prone offspring. —CA

Science, this issue p. 1002; see also p. 978

IMMUNOLOGY

Hefty structures of IgA and IgM complexes

Immunoglobulin M (IgM) and IgA are antibody isotypes that can form higher-order secretory complexes (sIgM and sIgA), which allows them to effectively bind and neutralize antigens with low-affinity repetitive epitopes, such as those found on the surface of many bacteria and viruses. The assembly and transport of these molecules is also dependent on the joining chain (J-chain) and the polymeric immunoglobulin receptor (pIgR) secretory component (SC). The architecture of these complex, multimeric structures has remained elusive. Li *et al*. resolved cryo–electron microscopy structures of the sIgM-Fc pentamer in complex with the J-chain and SC. Using similar techniques, Kumar *et al*. visualized dimeric, tetrameric, and pentameric structures of secretory sIgA-Fc interacting with the J-chain and SC. Both groups report highly similar mechanisms wherein the J-chain serves as a template for antibody oligomerization. An unanticipated, amyloid-like assembly of the oligomerized structure is present in both cases, with the J-chain conferring asymmetry for pIgR binding and transcytosis. These studies may inform structure-based engineering of these molecules for future therapeutic purposes. —STS

Science, this issue p. 1014, p. 1008

BATTERIES

Metastable pathways allow high rates

In batteries that allow for fast charging and discharging, lithium usually forms a solid solution with the anode so that the only limiting factor is the ionic diffusion. However, for a lithium titanate ($Li_4Ti_5O_{12}$) anode, the lithium ions interact with two phases and the diffusion is slow in both, but it still shows high-rate capabilities. Zhang *et al*. used electron energy-loss spectroscopy combined with density functional theory calculations to probe the anomalous behavior. They found that a diffuse interface forms between the starting and ending compositions, $Li_4Ti_5O_{12}$ and $Li_7Ti_5O_{12}$, and this is what allows the lithium ions to travel quickly. —MSL

Science, this issue p. 1030

CELL BIOLOGY

Polymerization regulates hexokinase activity

The yeast hexokinase Glk1 is an actin-fold protein that forms polymers in response to binding its substrates and products. Stoddard *et al*. now show that Glk1 polymers are structurally distinct from actin filaments and suggest that polymerization of Glk1 evolved independently of the polymerization of other actin-like polymers. Glk1 polymerization inhibits its hexokinase activity, and the monomer-polymer equilibrium appears to set a maximum rate for the entire enzyme pool rather than a maximum rate per enzyme. This inhibition was found to be important for cell viability in the context of nutrient shifts, allowing yeast cells to modulate their metabolism rapidly in response to stochastic changes in the environment. —SMH

Science, this issue p. 1039

PRENATAL THERAPIES

When treating at birth is too late

Mucopolysaccharidosis type VII (MPS7) is a rare and severe lysosomal storage disorder that causes dysfunction of multiple organs, including the brain. By the time of birth, the organ damage may already be severe and the fetus may not survive. Thus, the prenatal period provides the most promising opportunity for intervention. Nguyen *et al*. assessed two prenatal approaches—in utero enzyme replacement therapy and in utero hematopoietic stem cell transplantation—and demonstrated the potential of these treatments to improve survival and functional outcomes in a mouse model of MPS7. —YN

Sci. Transl. Med. **12**, eaay8980 (2020).

Published by AAAS

CELL BIOLOGY

Oriented mechanically, not by adhesion

Adherent cells position their spindle poles parallel to the plane of the substrate and orient the spindle within that plane according to cell shape–dependent mechanical forces. Anastasiou *et al*. found that proper spindle positioning in cultured cells was supported by integrin-mediated cell–extracellular matrix adhesion. Adhesion-dependent integrin activation was not required to position spindles parallel to the substrate plane, but force-dependent integrin activation was required to orient spindles in response to mechanical cues generated by adhesion topology. —AV

Sci. Signal. **13**, eaax9940 (2020).

OUTER SOLAR SYSTEM

Examining Arrokoth

The New Horizons spacecraft flew past the Kuiper Belt object (486958) Arrokoth (also known as 2014 MU_{69}) in January 2019. Because of the great distance to the outer Solar System and limited bandwidth, it will take until late 2020 to downlink all the spacecraft's observations back to Earth. Three papers in this issue analyze recently downlinked data, including the highest-resolution images taken during the encounter (see the Perspective by Jewitt). Spencer *et al*. examined Arrokoth's geology and geophysics using stereo imaging, dated the surface using impact craters, and produced a geomorphological map. Grundy *et al*. investigated the composition of the surface using color imaging and spectroscopic data and assessed Arrokoth's thermal emission using microwave radiometry. McKinnon *et al*. used simulations to determine how Arrokoth formed: Two gravitationally bound objects gently spiraled together during the formation of the Solar System. Together, these papers determine the age, composition, and formation process of the most pristine object yet visited by a spacecraft. —KTS

Science, this issue p. 998, p. 999, p. 1000; see also p. 980

Published by AAAS

RESEARCH ARTICLE SUMMARY

OUTER SOLAR SYSTEM

The geology and geophysics of Kuiper Belt object (486958) Arrokoth

J. R. Spencer *et al.*

INTRODUCTION: On 1 January 2019, the New Horizons spacecraft passed 3538 km from Kuiper Belt object (KBO) (486958) Arrokoth. Arrokoth is a contact binary consisting of two distinct lobes, connected by a narrow neck. Its orbital parameters, albedo, and color make Arrokoth a typical cold classical KBO (CCKBO). CCKBOs are the most dynamically and physically primitive population of small Solar System bodies known.

RATIONALE: Since the publication of initial results from the flyby, additional data have been downlinked and analyzed. This paper describes the resulting analysis of Arrokoth's shape, geological evolution, and satellite and ring constraints.

RESULTS: Improved stereo imaging constrains the object's shape and topography and allows us to generate a stereographic terrain model. Typical relief on both lobes (away from the neck region) is ~0.5 km or smaller.

Arrokoth's rotational period is 15.92 ± 0.02 hours, with its rotational pole pointing to right ascension = 317.5 ± 1°, declination = −24.9 ± 1°, J2000 equinox. The object consists of two roughly ellipsoidal lobes with overall dimensions of 36 km by 20 km by 10 km. The maximum dimensions of the two lobes are 20.6 km by 19.9 km by 9.4 km and 15.4 km by 13.8 km by 9.8 km, with uncertainties of 0.5 km by 0.5 km by 2.0 km. The total volume is equal to a sphere of diameter 18.3 ± 1.2 km, and the volume ratio of the two lobes is 1.9 ± 0.5. Global bulk density must be >290 kg m^{-3} if the neck is not in tension. Assuming a bulk density of 500 kg m^{-3}, as measured for comets, the mean surface gravity is ~1 mm s^{-2}, and the compressive strength of the neck must be >2.3 kPa.

The two lobes are closely aligned. The maximum axis of inertia of the large lobe is aligned within <5° of that of the small lobe. The equatorial planes of the two lobes are also almost coincident in space.

The small lobe's surface is marked by complex albedo patterns, often with sinuous margins and no detectable topographic signature, whereas the large lobe's surface is dominated by clusters of low dark hills superposed on brighter, smoother terrain. The large lobe's surface is divided into distinct subunits, which may represent smaller bodies that accreted to form it, though the overall smoothness of the surface, and the youthful appearance of many boundaries, which are sometimes undetectable or cross-cut by clusters of hills, suggest a more complex postformation history. If the subunits did accrete first, the smoothness of their mutual boundaries suggests subsequent accretion of additional material and later reactivation of the boundaries.

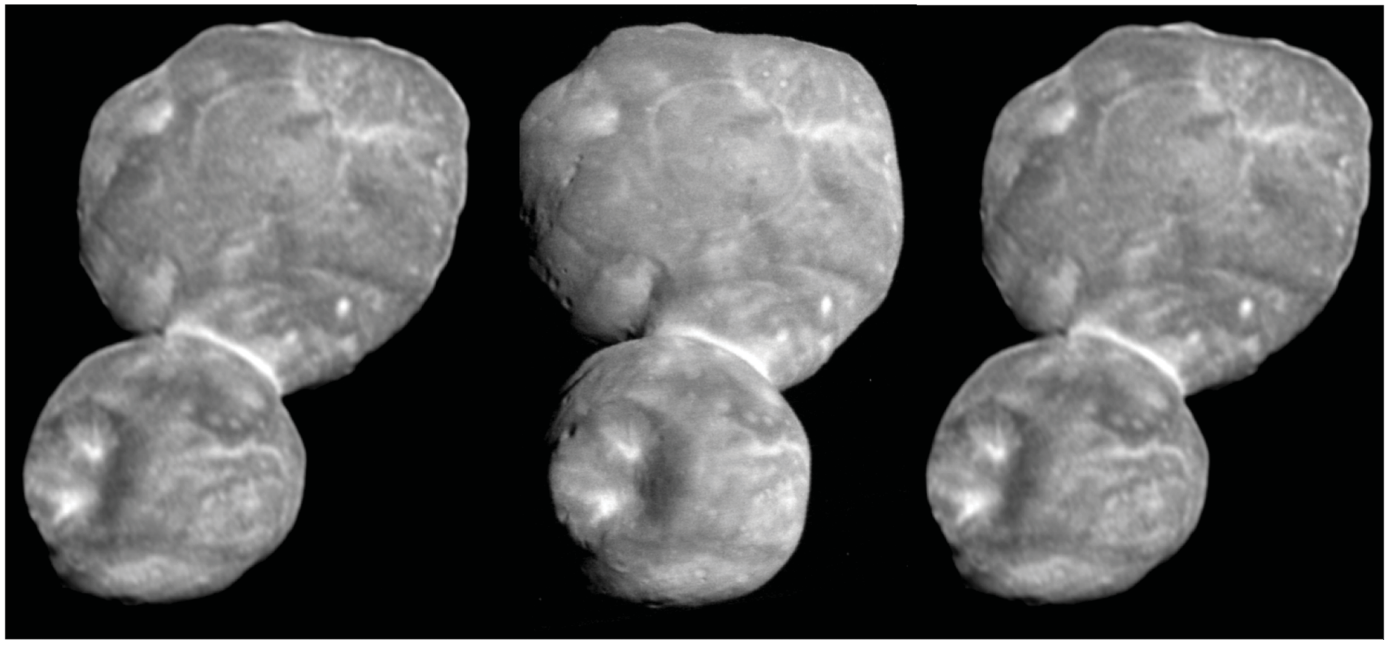
ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aay3999

We identify ~40 possible impact craters on Arrokoth, though only about 10 with high confidence. The largest crater, nicknamed Maryland, is about 7 km in diameter, and the rest are smaller than 1 km. Their size-frequency distribution is consistent with a single power law. Crater densities are lower than on many other small bodies but are consistent with a surface age of >4 billion years. No satellites or rings are detected: Satellite diameter upper limit is 180 m out to 8000-km radius from Arrokoth.

CONCLUSION: Arrokoth's smooth, lightly cratered surface is unlike that of other Solar System bodies and appears to date from the period of planetary accretion. The alignment of its two lobes constrains the processes that formed this contact binary. Because its orbit, albedo, color, and rotation are typical of other CCKBOs, Arrokoth can likely be used to understand the cold classical belt as a whole. ■

The list of authors and their affiliations is available in the full article online.
*Corresponding author. Email: spencer@boulder.swri.edu
Cite this article as J. R. Spencer *et al.*, *Science* **367**, eaay3999 (2020). DOI: 10.1126/science.aay3999



Stereo image pair of Arrokoth. The left and center images can be viewed cross-eyed, or the right and center by direct viewing.

RESEARCH ARTICLE

OUTER SOLAR SYSTEM

The geology and geophysics of Kuiper Belt object (486958) Arrokoth

J. R. Spencer[1]*, S. A. Stern[1], J. M. Moore[2], H. A. Weaver[3], K. N. Singer[1], C. B Olkin[1], A. J. Verbiscer[4], W. B. McKinnon[5], J. Wm. Parker[1], R. A. Beyer[6,2], J. T. Keane[7], T. R. Lauer[8], S. B. Porter[1], O. L. White[6,2], B. J. Buratti[9], M. R. El-Maarry[10,11], C. M. Lisse[3], A. H. Parker[1], H. B. Throop[12], S. J. Robbins[1], O. M. Umurhan[2], R. P. Binzel[13], D. T. Britt[14], M. W. Buie[1], A. F. Cheng[3], D. P. Cruikshank[2], H. A. Elliott[15], G. R. Gladstone[15], W. M. Grundy[16,17], M. E. Hill[3], M. Horanyi[18], D. E. Jennings[19], J. J. Kavelaars[20], I. R. Linscott[21], D. J. McComas[22], R. L. McNutt Jr.[3], S. Protopapa[1], D. C. Reuter[19], P. M. Schenk[23], M. R. Showalter[6], L. A. Young[1], A. M. Zangari[1], A. Y. Abedin[20], C. B. Beddingfield[6], S. D. Benecchi[24], E. Bernardoni[18], C. J. Bierson[25], D. Borncamp[26], V. J. Bray[27], A. L. Chaikin[28], R. D. Dhingra[29], C. Fuentes[30], T. Fuse[31], P. L Gay[24], S. D. J. Gwyn[20], D. P. Hamilton[32], J. D. Hofgartner[9], M. J. Holman[33], A. D. Howard[34], C. J. A. Howett[1], H. Karoji[35], D. E. Kaufmann[1], M. Kinczyk[36], B. H. May[37], M. Mountain[38], M. Pätzold[39], J. M. Petit[40], M. R. Piquette[18], I. N. Reid[41], H. J. Reitsema[42], K. D. Runyon[3], S. S. Sheppard[43], J. A. Stansberry[41], T. Stryk[44], P. Tanga[45], D. J. Tholen[46], D. E. Trilling[17], L. H. Wasserman[16]

The Cold Classical Kuiper Belt, a class of small bodies in undisturbed orbits beyond Neptune, is composed of primitive objects preserving information about Solar System formation. In January 2019, the New Horizons spacecraft flew past one of these objects, the 36-kilometer-long contact binary (486958) Arrokoth (provisional designation 2014 MU_{69}). Images from the flyby show that Arrokoth has no detectable rings, and no satellites (larger than 180 meters in diameter) within a radius of 8000 kilometers. Arrokoth has a lightly cratered, smooth surface with complex geological features, unlike those on previously visited Solar System bodies. The density of impact craters indicates the surface dates from the formation of the Solar System. The two lobes of the contact binary have closely aligned poles and equators, constraining their accretion mechanism.

On 1 January 2019 at 05:33:22 Universal Time (UT) the New Horizons spacecraft flew past the Kuiper Belt object (KBO) (486958) Arrokoth (provisional designation 2014 MU_{69}, previously nicknamed "Ultima Thule"), at a distance of 3538 km (*1*). Arrokoth is a contact binary consisting of two distinct lobes, connected by a narrow neck. On the basis of its orbital semi-major axis, low eccentricity and inclination (*2*), and albedo and color (*1*, *3*), Arrokoth is classified as a member of the dynamically cold, nonresonant cold classical KBO (CCKBO) population and is probably a member of the tight orbital clustering of CCKBOs known as the kernel (*4*). There is no known mechanism for transporting the majority of these objects onto these nearly circular orbits, so they are thought to have formed in situ and remained dynamically undisturbed since the formation of the Solar System. Owing to the low impact rates (*5*) and low temperatures in the Kuiper Belt, CCKBOs are also thought to be physically primitive bodies. Arrokoth's equivalent spherical diameter of 18 km (see below) makes it about 5.5 times smaller in diameter than a known break in the size-frequency distribution of CCKBOs at diameter ~100 km (*6*).

Initial results from this flyby (*1*) were based on early data downlinked from the spacecraft. Since then, additional data have been downlinked, including (i) the highest-resolution images from the flyby, taken with the narrow-angle Long-Range Reconnaissance Imager (LORRI) camera (*7*). These LORRI images have a pixel scale that is four times finer (33 m $pixel^{-1}$) than the 130 m $pixel^{-1}$ of previously available Multicolor Visible Imaging Camera (MVIC) (*8*) images (*1*), though because of smear and a lower signal-to-noise ratio (SNR), the effective resolution of the LORRI images is only about two times better than that of the MVIC images. Other downlinked data include (ii) additional LORRI images from earlier approach epochs, with higher SNR than previously downlinked data; (iii) improved LORRI distant approach rotational coverage, constraining the shape and rotational parameters; and (iv) additional satellite and ring search data from LORRI and MVIC. See (*9*) for image-processing details. We describe Arrokoth's shape, geological evolution, and satellite and ring constraints resulting from these additional data and from continued analysis of all downlinked data.

Stereo imaging

A pair of LORRI images, designated CA04 and CA06 [Fig. 1A and table S1 (*9*)], provides improved stereo imaging to constrain the shape and topography of the close approach hemispheres of the two lobes. A stereographic terrain model derived from these images [data S1 (*9*)], is shown in Fig. 2. Topographic relief in the stereo model is ~0.5 km or less on both lobes (away from the neck region), similar to the 1.0- and 0.5-km relief seen in limb profiles of the large and small lobes, respectively (*1*). The stereo images (Fig. 1A) show additional topographic detail that is visible to the eye but

[1]Southwest Research Institute, Boulder, CO 80302, USA. [2]NASA Ames Research Center, Moffett Field, CA 94035-1000, USA. [3]Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723, USA. [4]Department of Astronomy, University of Virginia, Charlottesville, VA 22904, USA. [5]Department of Earth and Planetary Sciences and McDonnell Center for the Space Sciences, Washington University, St. Louis, MO 63130, USA. [6]SETI Institute, Mountain View, CA 94043, USA. [7]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA. [8]National Science Foundation's National Optical Infrared Astronomy Research Laboratory, Tucson, AZ 26732, USA. [9]Jet Propulsion Laboratory, California Institute of Technology Pasadena, CA 91109, USA. [10]Department of Earth and Planetary Sciences, Birkbeck, University of London, London WC1E 7HX, UK. [11]University College London, Gower St, Bloomsbury, London WC1E 6BT, UK. [12]Independent Consultant, Washington, D.C., USA. [13]Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [14]Department of Physics, University of Central Florida, Orlando, FL 32816, USA. [15]Southwest Research Institute, San Antonio, TX 78238, USA. [16]Lowell Observatory, Flagstaff, AZ 86001, USA. [17]Department of Astronomy and Planetary Science, Northern Arizona University, Flagstaff, AZ, 86011, USA. [18]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO 80303, USA. [19]NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. [20]National Research Council of Canada, Victoria, BC V9E 2E7, Canada. [21]Independent Consultant, Mountain View, CA 94043, USA. [22]Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA. [23]Lunar and Planetary Institute, Houston, TX 77058, USA. [24]Planetary Science Institute, Tucson, AZ 85719, USA. [25]Earth and Planetary Science Department, University of California, Santa Cruz, CA 95064, USA. [26]Decipher Technology Studios, Alexandria, VA 22314, USA. [27]Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ 85721, USA. [28]Independent Science Writer, Arlington, VT 05250, USA. [29]University of Idaho, Moscow, ID 83844, USA. [30]Universidad de Chile, Centro de Astrofísica y Tecnologías Afines, Santiago, Chile. [31]Kashima Space Technology Center, National Institute of Information and Communications Technology, Kashima, Ibaraki 314-8501, Japan. [32]Department of Astronomy, University of Maryland, College Park, MD 20742, USA. [33]Center for Astrophysics, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA. [34]Department of Environmental Sciences, University of Virginia, Charlottesville, VA 22904, USA. [35]National Institutes of Natural Sciences, Tokyo, Japan. [36]Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695, USA. [37]Independent Collaborator, Windlesham GU20 6YW, UK. [38]Association of Universities for Research in Astronomy, Washington, DC 20004, USA. [39]Rheinisches Institut für Umweltforschung an der Universität zu Köln, Cologne 50931, Germany. [40]Institut Univers, Temps-fréquence, Interfaces, Nanostructures, Atmosphère et environnement, Molécules, Unité Mixte de Recherche, Centre National de la Recherche Scientifique, Universite Bourgogne Franche Comte, F-25000 Besancon, France. [41]Space Telescope Science Institute, Baltimore, MD 21218, USA. [42]Independent Consultant, Holland, MI 49424, USA. [43]Department of Terrestrial Magnetism, Carnegie Institution for Science, Washington, DC 20015, USA. [44]Roane State Community College, Oak Ridge, TN 37830, USA. [45]Université Côte d'Azur, Observatoire de la Côte d'Azur, Laboratoire Lagrange/ Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7293, 06304 Nice Cedex 4, France. [46]Institute for Astronomy, University of Hawaii, Honolulu, HI 96822, USA.
***Corresponding author. Email: spencer@boulder.swri.edu**

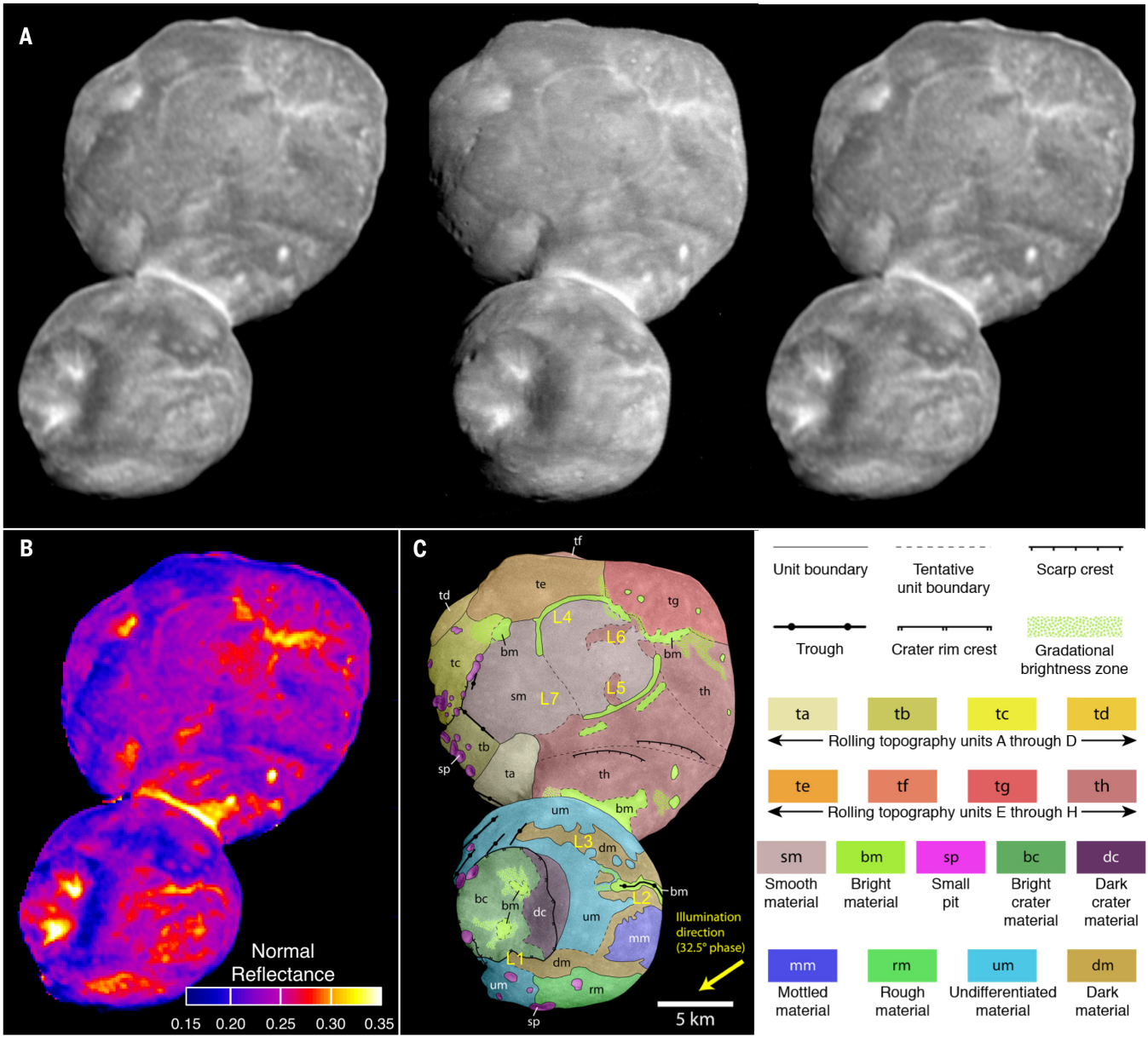


Fig. 1. Mapping of Arrokoth. (**A**) Cross-eyed (left+center) and direct (center+right) stereo pair image of Arrokoth, taken by LORRI. The left and right images are CA04, range = 27,850 km, phase = 12.9°, 138 m pixel^{-1}; the center image is CA06, range = 6634 km, phase = 32.5°, 33 m pixel^{-1}. Both images have been deconvolved to remove the LORRI point-spread function, and motion blur from CA06, to maximize detail (*9*). (**B**) A 0.6-μm normal reflectance map of Arrokoth, based on image CA04. (**C**) Geomorphological map of Arrokoth, overlain on the deconvolved CA06 image. The positive spin axis of Arrokoth is pointing approximately into the page. Yellow labels L1 to L7 identify locations mentioned in the text. Geological units are labeled and colored as shown in the legend.

smaller than the 200-m vertical resolution of the terrain model. Our interpretation is based on both the terrain model and subjective analysis of the stereo pair.

Rotation and global shape modeling

No periodic brightness variation due to rotation was detected in Hubble Space Telescope (HST) photometry before the flyby, with an upper limit amplitude of about 0.15 magnitudes (*10*). Stellar occultations in July 2017 and August 2018 showed that Arrokoth had an elongated, possibly contact-binary shape (*11*). The elongated shape and the low light-curve amplitude implied that Arrokoth's rotational pole was roughly aligned with the direction of the Sun and Earth.

Arrokoth's rotation and global shape are mostly determined from LORRI images taken between 2.2 days before the encounter, when Arrokoth first exceeded 2 pixels in length, and 9 min after encounter, when Arrokoth was last imaged (at high phase angle) as a receding crescent (Fig. 3). Disk-integrated photometry from earlier unresolved LORRI images showed no periodic variations in brightness, with an upper limit amplitude of 0.1 magnitudes (*12*), but were affected by confusion from the dense stellar background. The strongest constraints on the shape model are from a series of approach images with a cadence between 1 hour and 20 min, starting 13.6 hours before closest approach, when Arrokoth subtended 10 pixels in length (Fig. 4A). These images covered 85% of the 15.92-hour rotation period, though only one hemisphere of Arrokoth was visible because of the near-alignment of the rotational pole with both the direction of the Sun and New Horizons' approach direction.

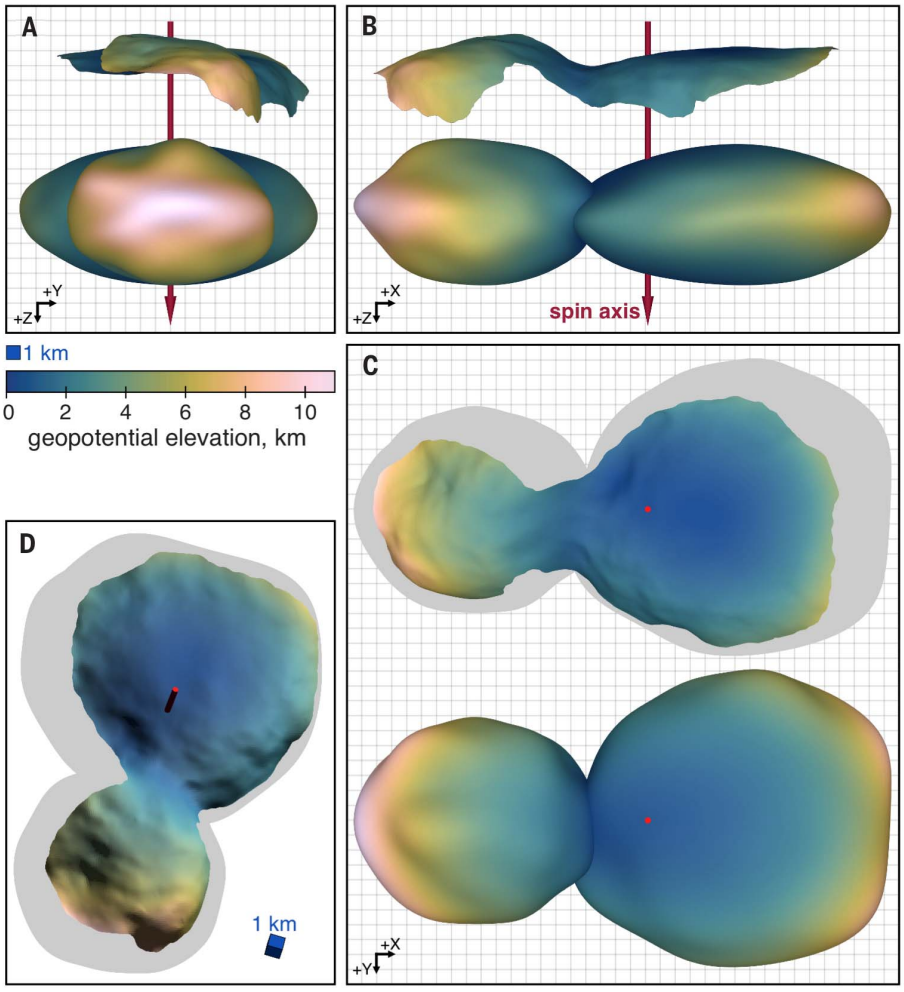


Fig. 2. Stereo and global shape models. (**A** to **C**) Comparison of the stereo shape model of the encounter face (top of each panel) to the global shape model (bottom of each panel), as seen from the −X (small lobe) direction (A), the +Y direction (B), and the south polar (−Z) direction (C). The red arrow shows the orientation and location of the positive spin axis. Each model is colored to show the variation in geopotential across the surface. The stereo model has been trimmed to remove edge effects. (**D**) Stereo model seen from the same geometry as the CA06 observation (Fig. 1A, center), but with different lighting, chosen to highlight the small-scale topography.

Incorporating the additional rotational coverage images now available into the same rotational modeling techniques as before (*1*), the rotational period of Arrokoth is unchanged at 15.92 ± 0.02 hours, but its pole orientation has been refined. The positive rotational pole points to right ascension 317.5 ± 1°, declination −24.9 ± 1° in the J2000 equinox. The rotation rate is within the range of other CCKBOs (*13*–*15*). The resulting obliquity of Arrokoth's pole to its orbit is 99 ± 1°, and the rotational pole is 39 ± 1° from the New Horizons approach vector and 28 ± 1° from the direction from the Sun to Arrokoth during the encounter. The rotational brightness variation implied by the shape model would have a peak-to-peak amplitude of 0.05 magnitudes from New Horizons' approach direction, consistent with the earlier nondetections.

A low-resolution global shape model [data S2 and Movie 1 (*9*)] was produced using all available observations—including the early, distant ones—to refine the model. The high phase angle CA07 observation [Fig. 3 and table S1 (*9*)], of the illuminated double crescent of Arrokoth, provides a constraint on how thick the unilluminated side can be, based on which stars are and are not eclipsed by the object (Fig. 4B). There remain differences between the shape model and the LORRI images in Fig. 4A; e.g., compared to the model, the images show a less indented neck and flatter outer end of the small lobe between December 31 20:38 and January 1 01:12.

The best-fitting global shape model consists of two roughly ellipsoidal lobes with overall dimensions X, Y, and Z of 36 km by 20 km by 10 km. Maximum dimensions of the large and small lobes are 20.6 km by 19.9 km by 9.4 km and 15.4 km by 13.8 km by 9.8 km, respectively. The uncertainty for these dimensions is roughly 0.5 km by 0.5 km by 2.0 km in X, Y, and Z, respectively; it is larger in the Z direction because the flyby imaged little of the +Z (northern) half of the object. The total volume is 3210 ± 650 km^3, equivalent to a sphere of diameter 18.3 ± 1.2 km. This volume is 30% larger than the previous estimate of 2450 ± 720 km^3 (*1*), though consistent within the uncertainties. The larger lobe has a volume equal to a sphere of diameter 15.9 ± 1.0 km, whereas the equivalent diameter for the smaller lobe is 12.9 ± 0.8 km. These values lead to a volume ratio (and mass ratio if densities are equal) of 1.9 ± 0.5.

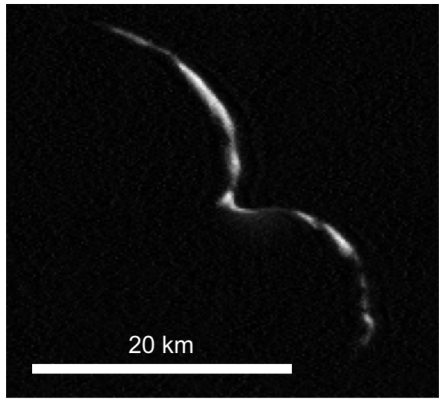


Fig. 3. Arrokoth seen at high phase. New Horizons' last view of Arrokoth (CA07), taken with the LORRI camera 9.4 min after closest approach at phase angle 152.4°, range 8834 km, and resolution 175 m pixel^{-1}. This image has been deconvolved to remove the motion smear visible in Fig. 4B (*9*). The large lobe is in the upper left and the small lobe is in the lower right.

Figure 2 compares the global shape model to the stereo model of the encounter (−Z) side of Arrokoth. There is broad agreement between the two techniques, although the south polar region of the large lobe is flatter in the stereo model, and the neck is smoother (a slope discontinuity at the neck is an intrinsic feature of the global shape model, owing to its dual-lobe nature). We regard the stereo model as more reliable than the global shape model in the south polar and neck regions, because the stereo model incorporates additional information due to the matching of albedo features and because these albedo features can also produce artifacts in the global shape model, which assumes a uniform surface albedo. However, near the limbs, the stereo model performs poorly because foreshortening makes feature matching difficult, whereas the global shape model is well constrained near the limbs.

Gravity modeling

The irregular shape of Arrokoth produces a complex geophysical environment. We calculated

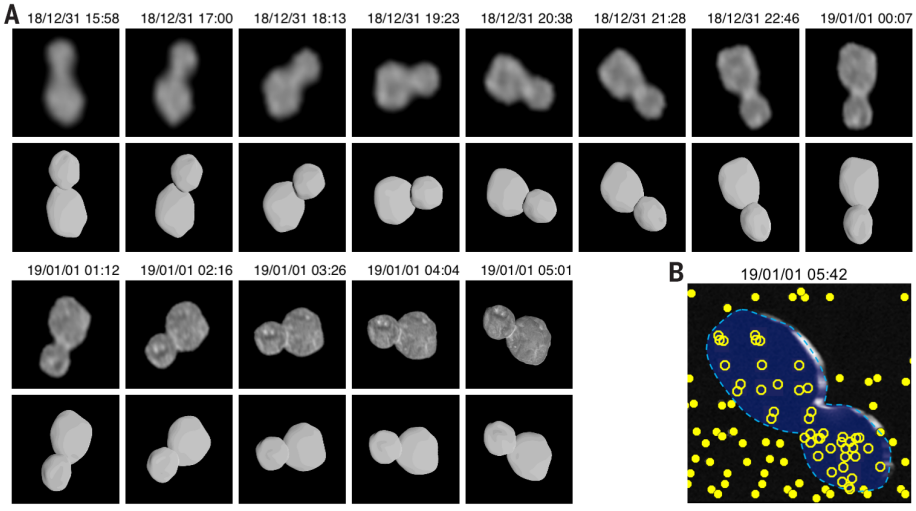
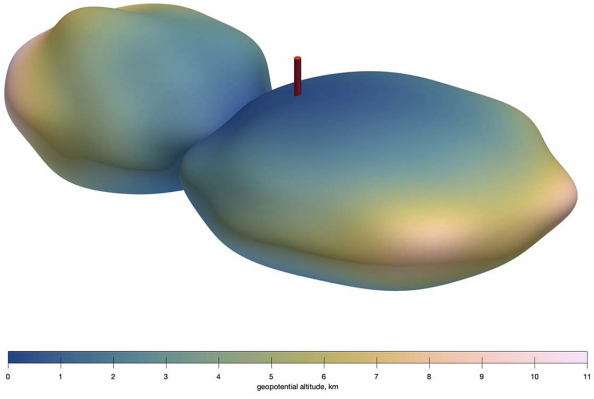


Fig. 4. Shape model compared to LORRI images. (**A**) Deconvolved LORRI approach images of Arrokoth, compared to synthetic images with the same geometry derived from the global shape model. Images have been scaled to a constant frame size of 44 km by 44 km, so become sharper as time progresses and range decreases. Celestial north is up. (**B**) The CA07 departure image, with the silhouette (dark blue) and outline (light blue dashed line) of the shape model superposed. Open and filled yellow dots indicate the locations of occulted and unocculted stars, respectively, in the six-frame CA07 sequence, used to constrain the shape of the unilluminated hemisphere.



Movie 1. Animation of the global shape model of Arrokoth. The model is shown rotating about its true spin axis (red arrow), highlighting the encounter hemisphere. The model is colored to show geopotential altitude and is obliquely illuminated.

Arrokoth's geopotential (the sum of the gravitational and rotational potentials in a body-fixed reference frame) using the low-resolution global shape model, the 15.92-hour rotation period, and an assumed bulk density. In the absence of spacecraft gravity measurements or detected satellites, the density of Arrokoth is not directly constrained. However, if the neck of Arrokoth is assumed to have no tensile strength, the density must be >290 kg m^{-3}, or the rotation would overcome the mutual gravity of the two lobes, causing them to separate. We assume a nominal bulk density of 500 kg m^{-3}, similar to the measured densities of cometary nuclei [e.g., comet 67P/Churyumov-Gerasimenko (*16*)], which leads to a mean surface gravity of ~1 mm s^{-2}. If this density is correct, the requirement for the two lobes to support each other against their mutual gravity over their ~28 km^2 contact area implies a compressive strength (accounting for centrifugal force) of >2.3 kPa.

Figure 2 uses color to show the geopotential altitude, calculated by dividing the geopotential by the total acceleration, which represents elevation with respect to a gravitational equipotential surface (*17*). The geopotential is calculated from the global shape model, then evaluated on the surfaces of the global shape model and the stereo model [with positions matched to the global shape model (*9*)]. This approach results in slight inaccuracies in the geopotential calculated across the stereo model, as there are regions where the stereo model rises above or below the surface of the global shape model. We focus on general trends that are robust to the uncertainties in the shape model. The geopotential is highest at the distal ends and equator and decreases with increasing latitude on each lobe, reaching a global minimum at the neck. For an assumed density of 500 kg m^{-3}, surface slopes [derivatives of the geopotential (*17*)] are generally gentle (<20°) and slope downward to higher latitudes and into the neck region (fig. S1). If material can flow downslope, then it will collect at higher latitudes and in the neck region. The stereo model shows that the neck is relatively smooth compared to its sharp appearance in the global shape model, with shallow slopes. The global shape model shows slopes of >30° at the neck, but this steepness is in part an artifact of the global model's treatment of Arrokoth as two separate overlapping bodies.

The configuration of the two lobes of Arrokoth has implications for its formation and evolution (*1*, *18*). Using the same assumptions as above, we calculate the principal axes of inertia for the two lobes by dividing the model at the narrowest point of the neck. This confirms that the large lobe's highest moment of inertia axis is aligned within <5° of its small lobe counterpart, and the equatorial planes of the two bodies are also almost coincident in space, with the estimated center of mass of the small lobe displaced only 0.2 km from the equatorial plane of the large lobe.

Surface units

Figure 1B shows a map of 0.6-μm normal reflectance (*19*). The map is derived from the high-SNR CA04 image, using a merger of the global and stereo shape models to determine illumination at each point, and an assumed lunar-like photometric function, which has no limb darkening at zero phase (*20*). The normal reflectance is equal to the geometric albedo of a body covered in material with that location's photometric properties. Arrokoth's mean 0.6-μm normal reflectance, and thus its geometric albedo, is 0.23. The mean and standard deviation of the normal reflectance are 0.230 and 0.035, respectively, for the large lobe, and 0.228 and 0.043, respectively, for the small lobe.

We have also produced (*9*) an updated geological unit map of Arrokoth (Fig. 1C) that supersedes the previous preliminary map (*1*). This mapping is physiographic in nature and is not intended to rigorously convey stratigraphic relations between units. The small and large lobes have distinctly different surface appearances, so we mapped their surface units separately and describe them separately below.

Small lobe

This lobe is dominated by a large depression (informally named Maryland), which is very likely to be an impact crater (*1*). The projected crater rim measures ~6.7 km by 6 km across in the image plane, with its longer axis roughly aligned with the principal axis of Arrokoth. The ellipticity might be due to foreshortening, in which case Maryland could be circular with a diameter of 6.7 km. Stereo measurements show that the deepest well-determined point in Maryland is 0.51 km below a plane defined by the rim, or 1.3 km below the surface of a sphere with the small lobe's mean radius, giving a depth/diameter ratio of 0.08 to 0.19. This depth/diameter ratio is similar to that of craters on other bodies with gravities similar to Arrokoth's ~1 mm s^{-2}, including asteroids Šteins [~0.12, 0.8 to 1.3 mm s^{-2} (*21*)] and Eros [~0.13, 2.4 to 5.5 mm s^{-2} (*22*)], though these bodies are composed of different materials and may have different porosities. Stereo imaging (Fig. 1A) reveals that the part of its rim furthest from the large lobe features a promontory protruding into the crater (marked L1 in Fig. 1C), at an elevation similar to the rest of the rim, which is not a common feature of impact crater rims.

Albedo patterns across the small lobe are complex. There are two patches of bright material (unit bm) within Maryland, which show discrete boundaries near the crater bottom, and fade toward the crater rim. Straddling the Maryland rim on the side opposite the bright patches is discrete, dark crater rim material (unit dc), which contrasts with the brighter terrain (unit bc) that forms the remainder of the crater interior. Elsewhere on the small lobe, discrete morphological units have albedo variations of almost a factor of 2 (Fig. 1B). The rough terrain at the distal end of the small lobe (unit rm) forms a facet that is relatively flat compared to the overall curvature of the surface and is brighter than its immediate surroundings. The low illumination angle on this facet reveals a rough surface texture at a scale of a few hundred meters, apparently mostly composed of sub-kilometer pits, with one prominent ~340-m-diameter pit (marked 27 in Fig. 6A) that resembles a small, fresh, bowl-shaped impact crater. Another nearby mottled bright unit (mm) may be similar, but it is seen at a higher illumination angle so topographic roughness is not apparent, and it has a distinctly crenulated and angular margin relative to that of unit rm (L2 in Fig. 1C).

Dark material surrounding the mm unit seems to be part of a discrete unit, designated *dm*, that wraps around much of the remainder of the observable surface of the small lobe—this material is the darkest on Arrokoth, with minimum 0.6-μm reflectance of 0.18. In places (L3 in Fig. 1C), it has a boundary with pointed and angular protrusions and rounded indentations, which may indicate material erosion and removal due to scarp retreat (*1*). Near L3 in Fig. 1C, there are also bright circular patches within the dark material. Running down the center of the principal mapped outcrop of dark material is a sinuous unit of bright material (unit bm), which stereo observations show occupies a V-shaped trough. The rest of the surface of the small lobe is nondescript at the available lighting and resolution and has been mapped as undifferentiated material (unit um). Crossing the undifferentiated material near the terminator between Maryland and the large lobe are a series of roughly parallel troughs, which are reminiscent of structural troughs seen on other similar-sized bodies—for instance, asteroid Eros (*23*, *24*), Saturn satellites Epimetheus and Pandora (*25*), and the Martian satellite Phobos (*26*).

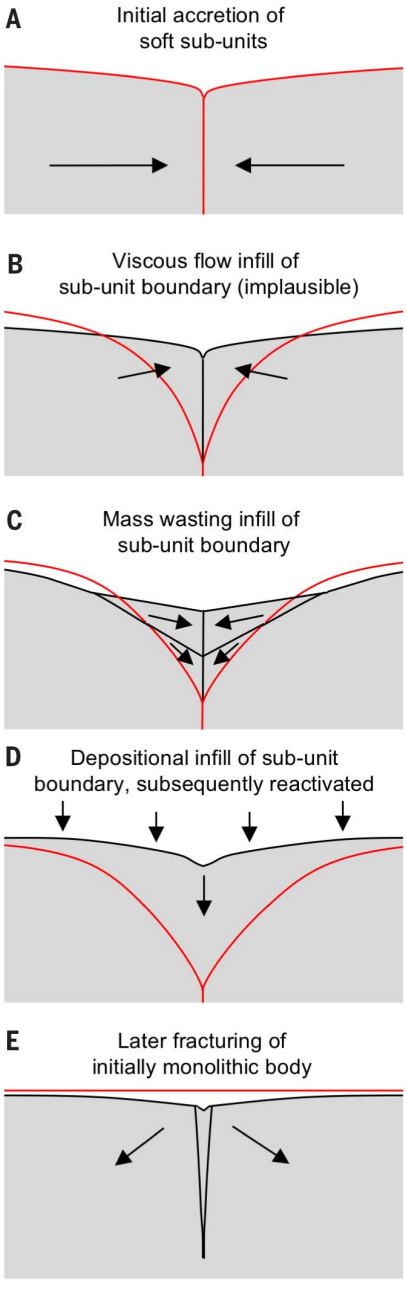


Fig. 5. Possible explanations for the appearance of the boundaries between terrain subunits on the large lobe. The original surface (shown in red) is modified by the processes labeled in each panel. We consider options (D) and (E) to be most consistent with the available evidence; see text for discussion.

Our data confirm that the bright neck region connecting the two lobes has a diffuse margin at least on the large lobe side, but extreme foreshortening makes it difficult to characterize its margin on the small lobe side.

Large lobe

The larger lobe is very different in appearance from the small lobe. Previous analysis (*1*) mapped the large lobe as composed of a series of roughly equal-sized, discretely bounded, rolling topographic units. We interpret some of these units and their boundaries differently, though confirm the discrete nature of many of the units (ta through tg). Those near the terminator, ta–td, are distinctive, being brighter than adjacent units (Fig. 1B) [though ta is noticeably less red than the others (*3*)], and are clearly separated from the rest of the large lobe by a common, continuous scarp or trough and chain of pits. Units tg and th appear more mottled than adjacent units, and stereo imaging of these suggests that their surface consists of dark ridges and hills surrounded by brighter low terrain.

The rest of the large lobe is occupied by smooth material (unit sm) of moderate albedo, transected by a series of distinctive bright linear features (unit bm), some of which form an incomplete annulus. In some areas (e.g., L4 in Fig. 1C), the inner margin of the annulus appears sharply bounded, possibly with an outward-facing scarp, whereas the outer margin is more diffuse. Stereo observations (Figs. 1A and 2D) show that terrain within the annulus is flatter than the undulating surface of the rest of the visible portion of the large lobe and suggest that the annulus occupies a shallow trough. At the boundary between units tg and sm, the annulus appears to be interrupted by diffuse bright material, which may be superimposed upon it. In two places, L5 and L6 in Fig. 1C, dark hills appear to extend into the sm unit. At L5 in Fig. 1C, these hills seem to be an extension, cut by the bm annulus, of similar hills on unit th. We discuss the possible origin of these features below.

Geological interpretation

Our data, particularly the stereo images, confirm that the brighter material on both lobes occurs preferentially in depressions. The

brightest material on the large lobe (the possible crater numbered 17 in Fig. 6A), on the small lobe (bright features 42 and 43 in Fig. 6A), and in the bright collar between the two lobes all have normal 0.6-μm reflectance values near 0.37, suggesting that the bright material has similar chemical and physical properties in all these regions. The most extensive bright region, the bright collar in the topographic low of the neck region, may be simply the largest-scale example of a general process that creates bright low-lying material across Arrokoth. As previously proposed (*1*), loose, poorly consolidated, likely fine-grained bright material may move downslope and accumulate in depressions, which would imply that bright material is more mobile than dark material on Arrokoth. The complex albedo patterns on the small lobe, and their crenulated margins, may result from the exposure and differential erosion of multiple lighter and darker layers oriented roughly parallel to its surface, though independent topographic information is of insufficient quality to confirm this explanation.

It was previously proposed (*1*) that the large lobe might be composed of smaller subunits that accreted separately. However, the improved imagery and topography raise issues with this interpretation. First, the central bm annulus, enclosing what was mapped as a discrete subunit in (*1*), appears to be younger than some other surface features, and not an unmodified primordial boundary, for the following reasons: (i) The annulus is incomplete, with no discernable topographic feature or textural change in the gap region where it is missing (L7 in Fig. 1C)—for this reason we map a continuous unit, sm, across this gap; (ii) even where the annulus is conspicuous, it cuts across flat terrain for most of its length; and (iii) dark hills found on the th and sm subunits appear to form a continuous physiographic unit cut by the annulus (at L5 in Fig. 1C), and (iv) the partially concentric nature of the annulus suggests a structural basis, not greatly obscured by subsequent deposition. Second, though other proposed subunits are distinguishable by differing surface textures, albedos, and modest topographic inflections or other surface features, the overall shape of the large lobe is smooth and undulating. There are no major topographic discontinuities between the subunits comparable to that between the two lobes, as would be expected if the subunits had a similar internal strength to the lobes as a whole. Erosion and alteration over the past 4.5 billion years (Ga) (see below) are likely to have modified the optical surface and the uppermost few meters (*27*) but probably do not explain the smoothness seen at the >30-m scale of the New Horizons imaging resolution.

Some possible explanations for the appearance of the annulus and other subunit boundaries are illustrated in Fig. 5. The subunits may have been soft enough at the time of merger that they conformed to each other's shapes on contact (*1*, *28*, *29*) (Fig. 5A), though no evidence for impact deformation is seen. For such deformation to take place at the time, the shear strength of the merging components must have been no more than 2 kPa, the ram pressure of an impacting body assuming a merger velocity of 1 to 2 m s^{-1} and a material density of 500 kg m^{-3}. The possibility that subunits flowed viscously as a result of gravity after contact while still soft (Fig. 5B) can be discounted, because such flow would require an implausibly low shear strength of ~100 Pa. Erosion and downslope movement (mass wasting) may have filled in original gaps between the subunits (Fig. 5C), though there is an absence of obvious boundaries (except perhaps at the the tg/sm contact) between material transported by mass wasting and in situ material. The fact that mass wasting has not filled the much larger depression between the two lobes also implies that any major mass wasting process must have ceased before the merger of the two lobes. The original discontinuities may have been buried by subsequent accretion or redistribution of surface material (Fig. 5D). The boundaries would then need to be reactivated in some way to still be visible on the surface, possibly by collapse into subsurface voids or degassing of volatiles such as N_2 or CO, which may explain the trough-like appearance of parts of the bm annulus, and the troughs and pit chains seen at low illumination angle between the ta – td subunits and the rest of the larger lobe. However, it's not clear how burial could preserve different surface textures for the different subunits. Alternatively, the large lobe may be monolithic, and the visible boundaries may be secondary features (Fig. 5E), e.g., produced by subsequent

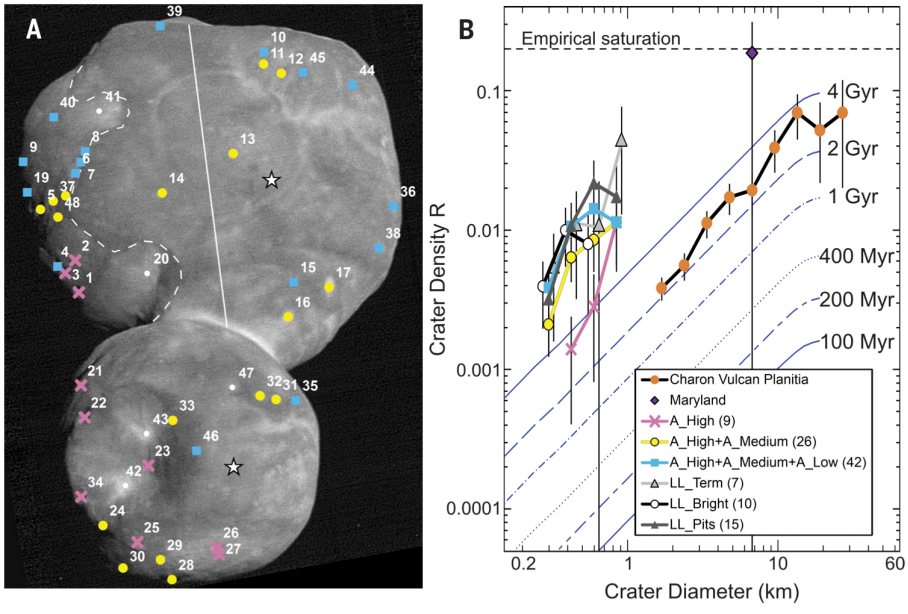


Fig. 6. Craters and Pits on Arrokoth. (**A**) Locations of features considered for crater analysis; numbers refer to crater listings in data S3. Color denotes confidence class: pink, high confidence (A_High); yellow, medium confidence (A_Medium); light blue, low confidence (A_Low). Features indicated in white are considered to be highly unlikely to be of impact origin and are not included in the crater statistics. The solid white line splits the large lobe into regions with differing lighting conditions, a more obliquely illuminated region with more visible depressions (LL_Pits, left) and a more vertically illuminated region with bright spots (LL_Bright, right). The white dashed curve delineates the boundary of combined geologic units ta, td, tc, and td (LL_Term), considered together for crater density determination. The star symbols indicate the planetocentric subsolar point on each lobe according to the shape model. Lighting direction is shown in Fig. 1C. (**B**). The size-frequency distribution of craters on Arrokoth for each crater subgroup and region described in the text and shown in (A) and (*9*). The yellow curve includes both high- and medium-confidence classes, and the light blue curve includes all confidence classes. Parenthetical numbers are the total number of craters and pits in each category. The Arrokoth crater data are compared to crater densities on Charon's Vulcan Planitia (*39*) without diameter adjustments for gravity or velocity scaling, and to predictions based on an impactor flux model for six different ages of surfaces on Arrokoth and gravity regime scaling [blue curves with different line styles (*5*)]. The LL_Term and LL_Bright distributions are offset horizontally by ±9% for clarity. The empirical saturation line refers to a D^{-3} differential power law distribution (*72*). Myr, million year; Gyr, billion year.

fracturing. For the annulus, we consider the evidence to be most consistent with scenarios D and E in Fig. 5. However, in any of these cases, the processes that produced the distinctive surface textural contrasts between the units, in particular the patches of dark hills and ridges, are unknown.

Pits and craters

In addition to the 7-km-diameter probable impact crater Maryland, scattered across the body of Arrokoth are numerous roughly circular subkilometer bright patches and pits, though even if these are mostly impact craters, the crater density is relatively low compared to many other small bodies (*1*) (fig. S2). The bright patches are generally seen in areas that have high illumination angle and are away from the terminator. Some of these patches appear in stereo imaging (Fig. 1A) to occupy depressions. These may be equivalent to the pits seen in low–illumination angle areas near the terminator (unit sp, Fig. 1C): These pits might also feature bright material on their floors that is invisible because of the unfavorable lighting.

We have classified these bright patches and pits to reflect our confidence that they are impact craters, based on the morphology expected for either fresh or degraded impact craters (*9*) (supplementary text), as determined by multiple independent investigators. Crater candidates and their classifications are listed in data S3 and shown in Fig. 6A. Our criteria included the spatial arrangement of the potential craters and their relationship to other geologic features. For instance, as noted above, a chain of pits that is coincident with a scarp on the boundary between units tc and sm possibly originated through surface collapse rather than impact (*1*). For a fresh crater formed on a flat and smooth surface, a crater rim is expected to be close to circular and raised above the surrounding terrain [unless the terrain is substantially porous (*30*)], though image resolution does not always allow identification of a raised rim. The interior shape of a crater is expected to be bowl-like with a depth/diameter ratio typically not higher than ~0.2 (*31*). The predicted modal impact velocity onto Arrokoth is ~300 m s^{-1} (*5*), which is sufficient to form craters with typical morphologies (see supplementary text). In the case of Arrokoth, the lowest-velocity impacts ($\lesssim$20 m s^{-1}) are unlikely to leave conspicuous depressions, but these impacts are expected to be a small fraction of the total (*5*). The formation of a crater on a slope or modification by later geologic processes (such as mass wasting or a subsequent fault near the crater) may also alter the crater's appearance.

Potential small craters were subdivided in three ways (Fig. 6A) (*9*): (i) All pits and bright patches were subdivided based on our confidence that they are impact craters; (ii) features on the large lobe were subdivided into pits nearer the terminator, and bright patches away from the terminator, as shown in Fig. 6A; and (iii) a combination of geologic units—ta, tb, tc, and td, designated "LL_Term" as they are on the large lobe terminator (Fig. 6A)—was analyzed separately, because the entire combined unit has low-angle lighting optimal for crater identification. These subdivisions yielded a range of plausible crater densities, shown in Fig. 6B as a crater relative- or R plot (*9*). Overall R values for each dataset are somewhat uncertain as they depend on the areas used for each distribution, and densities are lower if uncertain craters are excluded. The resulting uncertainty range of crater densities is less than a factor of 10 in each diameter bin in Fig. 6B.

Besides Maryland, all other possible impact features are 1 km in diameter or smaller. Although the diameter gap between Maryland and second-largest crater on Arrokoth is large, the gap does not strongly disfavor a single power-law size distribution for the craters. We tested a model crater population with a power-law size distribution with slope $q = -2$ against the observed Arrokoth craters in the combined "A_High" and "A_Medium" categories. The resulting Anderson-Darling statistic indicates no substantial disagreement between the model and observed sample, with a significance level of $p \leq 17\%$.

Our analysis shows that Arrokoth appears to be only modestly cratered, relative to heavily cratered small objects like Phobos (fig. S2), and there are some areas on Arrokoth where very few, if any, potential craters exist, in particular the part of the large lobe between the dashed and solid white lines in Fig. 6A.

The age of the surface can be estimated from the observed crater density. We converted impact flux estimates for Arrokoth to crater densities corresponding to several surface ages (*5*) and show these in Fig. 6B. The resulting age estimates are uncertain, given the uncertainty in identifying which craters are impact generated, and because the model curves shift on the basis of the crater scaling parameters used. Scaling in the strength regime, as opposed to scaling in the gravity regime assumed here (*5*), could in principle reduce the sizes of craters produced, if the surface strength of Arrokoth were sufficiently high. The expected strengths of porous cometary surfaces are, however, generally low enough [~1 kPa or less (*32*)] that the observed craters on Arrokoth should have formed in the gravity regime. By contrast, accounting for the additional cratering in an early but brief dynamical instability phase in the outer Solar System (*33*) would shift the model curves in Fig. 6B upward, although possibly by no more than a factor of 2 (*5*). Low relative densities of small craters are also observed on near-Earth asteroids and are conventionally explained as being due to seismic shaking from larger impacts or surface evolution due to changes in spin state (*34*–*36*). However, Arrokoth's spin state is likely to have evolved only very slowly (*18*), there do not appear to be sufficient impacts to act as effective seismic sources, and Arrokoth's likely high porosity would make seismic energy propagation highly inefficient. Overall, despite the paucity of craters on its surface, the observed crater density is consistent with a crater retention age of greater than ~4 billion years. The visible surface at the scale of the LORRI image resolution thus plausibly dates from the end of Solar System accretion.

Though the diameters of observed craters on Arrokoth (apart from Maryland) are smaller than those measured in the Pluto system, the slopes of the Arrokoth and Pluto system craters are consistent given the small number statistics. Using approximate Bayesian computation forward-modeling methods (*37*, *38*), we estimated the posterior probability density functions for the parameters of independent truncated power-law crater size–frequency distribution models for Arrokoth's and Charon's (*39*) observed crater populations (for craters <10 km in diameter, below the break in slope observed on Charon). We then conducted the same analysis for a model with a common slope, q, between the two populations, but a separate offset. The mean slope $q = -1.8^{+0.4}_{-0.6}$ for Charon alone, $q = -2.3^{+0.6}_{-0.6}$ for Arrokoth alone, and $q = -2.0^{+0.4}_{-0.3}$ for the joint set (95% confidence). However, as seen in Fig. 6B, crater density on Arrokoth is higher than would be obtained from an extrapolation of the Charon slope and density to subkilometer craters.

Satellites and rings

Before the Arrokoth flyby, constraints on the prevalence of satellites and rings around sub–100-km-diameter Kuiper Belt objects were limited. Larger CCKBOs are frequently members of orbiting binary pairs (*40*). Satellites with a primary/secondary brightness ratio larger than 20 have not been found for KBOs smaller than 500 km in diameter (*41*), though this is likely in part due to observational biases. By contrast, satellites with high primary/secondary brightness ratio are common around large KBOs in non-CCKBO populations. The presence or absence of satellites provides a constraint on formation of the Arrokoth contact binary (e.g., a satellite could potentially remove angular momentum from the central body). At least two known asteroid contact binaries have small satellites: The large Trojan asteroid Hektor has a satellite that orbits at only 5 times the primary radius and has a diameter of 5% of the primary (*42*), and the large bilobed main-belt asteroid Kleopatra has two known satellites orbiting at 8 and 12 times the primary

radius, with diameters 6% that of the primary (*43*).

New Horizons conducted a nested series of satellite searches with the LORRI camera during its approach to Arrokoth, using stacks of many images taken using 4 by 4 pixel binning to increase sensitivity and reduce data volume. Our dataset allows a deeper and broader search than previously reported (*1*, *9*). No satellites have been found. We can exclude satellites larger than 100 to 180 m in diameter (~0.5% the diameter of the primary) on orbits ranging from Arrokoth's surface to 8000-km radius, and <300 m in diameter throughout most of the Hill sphere (the region within which a moon could be gravitationally bound to Arrokoth), assuming albedos similar to that of Arrokoth itself (Fig. 7). Satellites analogous to those of Hektor and Kleopatra can thus be excluded.

The prevalence of rings around small KBOs is poorly constrained, but they are known to exist around Chariklo (*44*), Haumea (*45*), and perhaps Chiron (*46*). We searched for rings and dust clouds within the Arrokoth environment at all phases of the encounter. The LORRI satellite searches on approach, discussed above, constrained backscattered light due to any ring or dust clouds to $I/F \lesssim 2 \times 10^{-7}$ (*19*) at 11° phase for a 10-km-wide ring, assuming neutral colors (*1*). This limit is fainter than Jupiter's main ring [$I/F = 7 \times 10^{-7}$ at 11° phase (*47*)]. We also conducted dedicated ring searches in forward-scattered light after closest approach, using images taken 1.7 to 2.3 hours after closest approach at a phase angle of 168°, covering radii up to 6000 km from Arrokoth. The MVIC instrument, which has better rejection of scattered sunlight than LORRI, was used in its panchromatic framing mode, with total exposure times of 30 s. Reduction and analysis followed methodologies used for similar Pluto data (*48*). No rings or dust structures were detected, with an upper limit I/F of ~1.5×10^{-6} for structures wider than about 10 km in Arrokoth's equatorial plane (fig. S4). Any ring around Arrokoth is thus also fainter in forward scattering than Jupiter's main ring [$I/F = 4 \times 10^{-6}$ at this phase angle (*47*)].

New Horizons' Student Dust Detector (SDC) instrument (*49*) detected no signals above the noise threshold within ±5 days of the Arrokoth encounter, implying that there were no impacts by dust particles >1.6 μm in radius, giving a 90% confidence upper limit of 3×10^7 particles km^{-2}. For 10% albedo, this is equivalent to an I/F limit of 3×10^{-11}, even more constraining than the optical limit, for particles of this size or larger along the spacecraft trajectory.

Comparison to other KBOs, and to possible captured KBOs

Though most other known CCKBOs are larger than Arrokoth, owing to observational biases, Arrokoth appears typical of CCKBOs using the few metrics that can be directly compared. Arrokoth's 0.6-μm geometric albedo, 0.23, is within the known range of other CCKBOs (*50*). Rotational lightcurves suggest that up to 25% of larger CCKBOs could be contact binaries like Arrokoth (*13*), though contact binaries appear to be more abundant, up to 50%, in the Plutino population (*51*). Arrokoth's color is also typical of CCKBOs (*1*, *3*).

Many irregular satellites of the giant planets may be captured KBOs, but only three have resolved spacecraft images. Neptune's satellite Triton, with a diameter of 2700 km, is far too large and active to be a useful comparison body to Arrokoth. Neptune's smaller irregular satellite Nereid, 170 km in diameter, has a geometric albedo of 0.16 to 0.20, similar to Arrokoth's, but is neutral in color (*52*). Saturn's 210-km-diameter irregular satellite Phoebe [possibly a captured Kuiper Belt object (*53*), though perhaps instead a captured C-type asteroid (*54*, *55*)], is darker [geometric albedo 0.08 (*56*)] and less red (*57*), and has a completely different surface appearance, dominated entirely by impact features (*58*). If Phoebe ever resembled Arrokoth, it has been drastically altered by subsequent evolution.

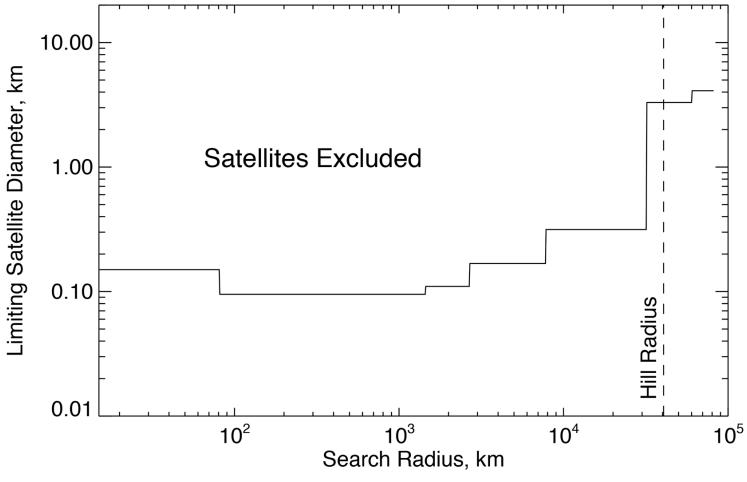


Fig. 7. Upper limits on possible satellites of Arrokoth. Excluded regions are plotted as a function of radius from the primary center of mass. The limits assume a satellite with photometric properties similar to those of Arrokoth itself. Gravitationally bound objects must lie within the Hill radius (dashed line), which is calculated assuming Arrokoth has a density of 500 kg m^{-3}.

Comparison to Jupiter family comets

A class of objects previously explored by spacecraft that may be analogous to Arrokoth in ultimate origin are the Jupiter family comets (JFCs). These differ from Arrokoth in three major respects: (i) Provenance: The vast majority of these bodies likely originated in the Kuiper belt, but from a different family of KBOs: the population of "scattered KBOs," which likely originated closer to the Sun than Arrokoth, and whose orbits are strongly perturbed by gravitational interactions with Neptune (*59*). (ii) Size: The effective spherical diameters of the JFC nuclei visited by spacecraft are 3 to 18 times smaller than that of Arrokoth. (iii) Thermal history: JFCs have experienced intense solar heating, which has heavily modified their surfaces. By comparing the properties of Arrokoth and JFC nuclei, we can explore the effects of these differences.

The JFC nuclei visited by spacecraft have diverse shapes and surfaces (Fig. 8, fig. S3, and table S3). Comets 19P, 67P, and 103P appear to be highly elongated bilobate objects, suggesting the merger of two distinct bodies, as has been proposed for Arrokoth (*1*, *18*), though for comets it is also possible that thermal evolution has generated this shape [e.g., (*60*)]. Except for 67P, whose bulk density is 538 ± 1 kg m^{-3} (*16*), the densities of the other JFC nuclei are uncertain by a factor of 2 or more, but all are consistent with ~500 kg m^{-3} (*61*), which implies average bulk porosities of ~50 to 80%. Arrokoth's density is likely greater than 290 kg m^{-3} (see above), and thus at least consistent with those of JFC nuclei. The rotation period of Arrokoth is similar to those measured for 67P and 103P and falls well within the range measured for the JFC population (*62*), though JFC rotation is known to be affected by cometary activity (*63*).

The JFC nuclei listed in table S3 are much darker than Arrokoth, with ~3 to 5 times smaller geometric albedos. If the JFC nuclei once had higher albedos in their nascent state in the Kuiper belt, then the darkening of their surfaces might be associated with cometary activity while the JFCs are in the inner Solar System. Most surface features on JFC nuclei have been attributed to cometary activity [e.g.,

(*64*, *65*)]. Generally, the surfaces of JFC nuclei can be divided into "smooth" and "rough" (or "mottled") regions, with the rough terrains associated with a preponderance of pits and depressions or mounds and hills (*66*, *67*). The smooth regions of JFCs are generally brighter than average and are often associated with topographic lows, suggesting accumulation by small grains that scatter light more efficiently than the average surface, as we proposed for Arrokoth above. However, on comets, the fallback of grains ejected by sublimation is likely to contribute to smooth terrains (*68*), and this is less likely to be important on Arrokoth where evidence for sublimation erosion is limited to the pit chains of possible sublimation origin, and tentative evidence for scarp retreat on the small lobe, as mentioned above.

Whereas the large (multikilometer) scale bilobate morphology of Arrokoth is similar to that of four out of the six comets listed in table S3 (see also Fig. 8 and fig. S3), the finer surface textures are not. JFCs imaged at the same resolution as Arrokoth show fewer impact craters than Arrokoth (*64*), consistent with these comets having highly erosional surfaces. They may lose their surfaces at ~0.5 to 1.0 m per orbit (*69*) with 5- to 10-year orbital periods, so small pits will be removed within a few thousand years. They also show a much rougher surface texture at the 50- to 100-m scale, consistent with sublimation erosion and loss of most of the erosional debris.

Conclusions

Our dataset from the New Horizons flyby of Arrokoth provides a more complete picture of the physical nature of this object. Images taken on approach show that although both components of Arrokoth are flattened, the flattening is less extreme than initially inferred (*1*), and the two components have a larger volume ratio, 1.9 ± 0.5, than previous estimates. Stereo topography and the highest-resolution imaging taken during the flyby show that the large lobe is very flat on the encounter hemisphere. If the large lobe is composed of multiple components that accreted separately, as previously proposed (*1*), the topographic signature of the boundaries between the components would be expected to be large initially, if the subunits were mechanically similar to the two present lobes at the time of their coming into contact (*18*). The observed flatness of the large lobe shows that any such discontinuities have been subdued, and in some cases, eliminated entirely. If subsequent deposition subdued the boundaries, postdepositional processes must be invoked to explain why many of the boundaries are still visible as differences in surface texture or as linear albedo features. Alternatively, the large lobe may be a monolithic body, and the apparent division into subunits may be due entirely to secondary processes. Multiple processes, including impacts, have reworked the surfaces of both lobes after their formation, producing the fissures, small dark hills, and sinuous albedo boundaries seen in the images.

Crater densities on Arrokoth are low but consistent with a surface age of >4 Ga, owing to the expected low cratering rates in the CCKB, even if only craters with the highest confidence of being impact features are included in the counts. This dates the surface as plausibly from the end of Solar System accretion. Crater size–frequency distribution slopes for <1-km craters on Arrokoth are poorly constrained, but are consistent with the slopes seen for 2- to 15-km craters in the Pluto system (*39*), suggesting that the shallow size-frequency

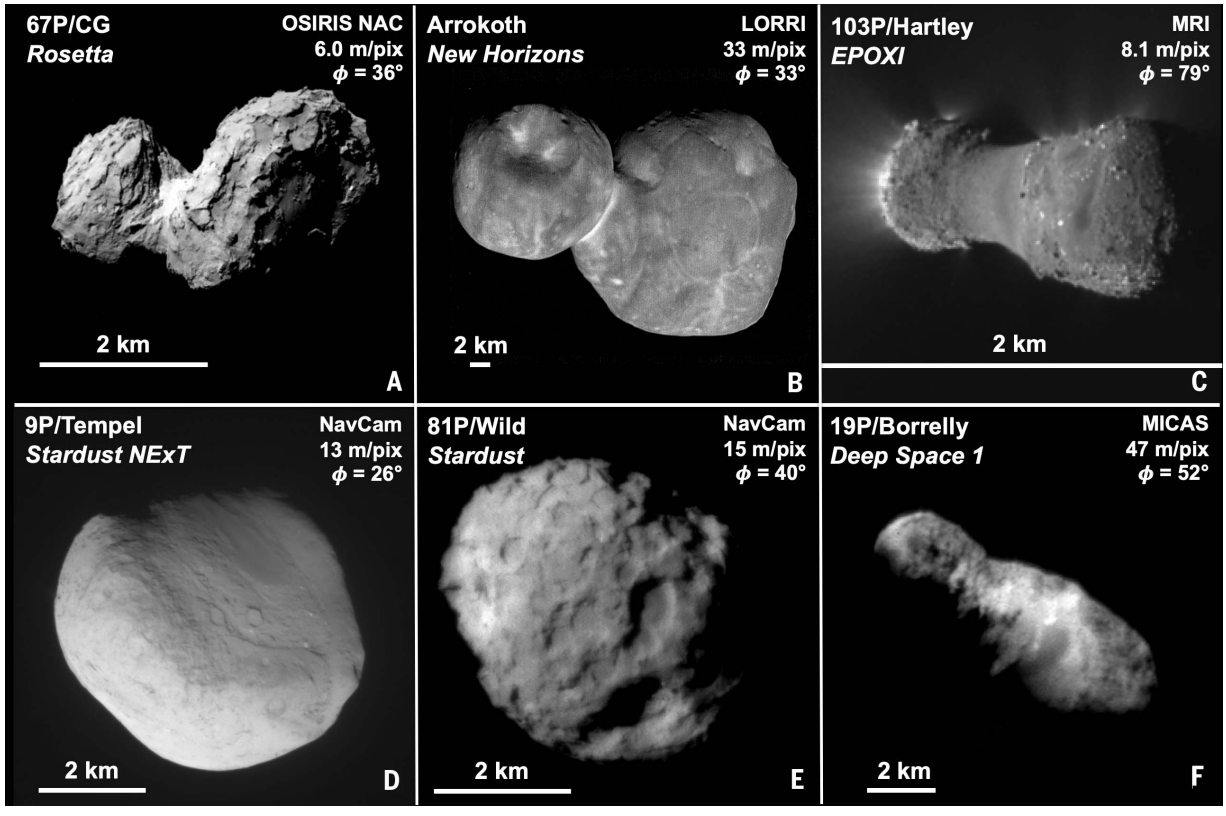


Fig. 8. Comparison of JFC nuclei to Arrokoth. The images of JFC nuclei have phase angles similar to those of the highest-resolution image of Arrokoth, except for 103P, which was only observed at much higher phase angles. (**A**) Rosetta image of 67P/Churyumov–Gerasimenko (*73*). (**B**) New Horizons image of Arrokoth (this paper). (**C**) Extrasolar Planet Observation and Characterization–Deep Impact Extended Investigation (EPOXI) image of 103P/Hartley (*74*). (**D**) Stardust image of 9P/Tempel (*75*). (**E**) Stardust image of 81P/Wild (*76*). (**F**) Deep Space 1 image of 19P/Borrelly (*77*, *78*). [Credit: NASA/JPL] Each frame is scaled so that the body nearly fills it, with the true relative sizes of each body indicated by the scale bars. Arrokoth is much larger than these comets. Figure S3 shows the equivalent images scaled to the same linear resolutions.

distribution for 0.2- to 2-km-diameter KBO impactors (*39*) may persist down to smaller sizes.

Arrokoth is unlike other small bodies visited by spacecraft. The surfaces of comets are dominated by volatile loss and sublimation erosion driven by the thermal energy inputs, owing to their position in the inner Solar System. The surfaces of asteroids are dominated by high-energy impacts. As a result, asteroid surfaces are primarily rubble or impact ejecta. In both cases, the dominant energy environment (thermal and impact) is driving the surface morphology. Arrokoth's surface is probably a consequence of its presence in the CCKB, where there is much less energy input. The very small relative velocities in this dynamical population result in few impacts, and those that do occur have very slow impact velocities. Without strong energy inputs, either from solar radiation or impacts, we expect the surface of Arrokoth to be dominated by low-level energy inputs from interstellar, solar, and micrometeorite energy sources at slow rates, likely extending to just a few meters' depth (*27*). It is this low-energy environment that has allowed its surface to be preserved for 4 billion years.

Arrokoth appears to be a typical CCKBO, to the extent that we can compare it to others, so it can be used to understand the cold classical belt as a whole. The bilobed nature of Arrokoth might be common in the Kuiper Belt and could indicate that the bilobed shape of many comet nuclei is a primordial feature. In addition, Arrokoth appears to be a direct product of accretion rather than a collisional fragment and is much smaller than the ~100-km diameter of the break in slope of the size-frequency distribution of CCKBOs (*6*, *70*). These facts are consistent with the break in slope being a primordial feature, as predicted by streaming instability models (*71*). Arrokoth's appearance is much less consistent with the break in slope being a result of later destruction of small CCKBOs by collisions, a hypothesis also inconsistent with the observed deficit of small craters in the Pluto system (*39*).

REFERENCES AND NOTES

1. S. A. Stern *et al*., Initial results from the New Horizons exploration of 2014 MU_{69}, a small Kuiper Belt object. *Science* **364**, eaaw9771 (2019). doi: 10.1126/science.aaw9771; pmid: 31097641
2. S. B. Porter *et al*., High-precision orbit fitting and uncertainty analysis of (486958) 2014 MU69. *Astron. J.* **156**, 20 (2018). doi: 10.3847/1538-3881/aac2e1
3. W. M. Grundy *et al*., Color, composition, and thermal environment of Kuiper Belt object (486958) Arrokoth. *Science* 10.1126/science.aay3705 (2020).
4. J.-M. Petit *et al*., The Canada-France Ecliptic Plane Survey - full data release: The orbital structure of the Kuiper belt. *Astron. J.* **142**, 131 (2011). doi: 10.1088/0004-6256/142/4/131
5. S. Greenstreet, B. Gladman, W. B. McKinnon, J. J. Kavelaars, K. N. Singer, Crater density predictions for New Horizons flyby target 2014 MU69. *Astrophys. J.* **872**, L5 (2019). doi: 10.3847/2041-8213/ab01db
6. S. Greenstreet, B. Gladman, W. B. McKinnon, Impact and cratering rates onto Pluto. *Icarus* **258**, 267–288 (2015). doi: 10.1016/j.icarus.2015.05.026
7. A. F. Cheng *et al*., Long-Range Reconnaissance Imager on New Horizons. *Space Sci. Rev.* **140**, 189–215 (2008). doi: 10.1007/s11214-007-9271-6
8. D. C. Reuter *et al*., Ralph: A Visible/Infrared Imager for the New Horizons Pluto/Kuiper Belt mission. *Space Sci. Rev.* **140**, 129–154 (2008). doi: 10.1007/s11214-008-9375-7
9. Materials and methods are available as supplementary materials.
10. S. D. Benecchi *et al*., The HST lightcurve of (486958) 2014 MU69. *Icarus* **334**, 11–21 (2019). doi: 10.1016/j.icarus.2019.01.023
11. M. W. Buie *et al*., Size and shape constraints of (486958) Arrokoth from stellar occultations (2020). https://arxiv.org/abs/2001.00125.
12. A. M. Zangari, *et al.*, The mysterious missing light curve of (486958) 2014 MU69, a bi-lobate contact binary visited by New Horizons. *Lunar Planet. Sci. Conf.* 3007 (2019).
13. A. Thirouin, S. S. Sheppard, Light curves and rotational properties of the pristine Cold Classical Kuiper Belt objects. *Astron. J.* **157**, 228 (2019). doi: 10.3847/1538-3881/ab18a9
14. S. B. Porter *et al*., New Horizons distant observations of Cold Classical KBOs. AAS/Division for Planetary Sciences Meeting Abstracts 509.07 (2018).
15. A. J. Verbiscer *et al*., Phase curves from the Kuiper belt: Photometric properties of "distant" KBOs observed by New Horizons. *Astron. J.* **158**, 123 (2019). doi: 10.3847/1538-3881/ab3211
16. F. Preusker *et al*., The global meter-level shape model of comet 67P/Churyumov-Gerasimenko. *Astron. Astrophys.* **607**, L1 (2017). doi: 10.1051/0004-6361/201731798
17. D. J. Scheeres, *Orbital Motion in Strongly Perturbed Environments: Applications to Asteroids, Comet and Planetary Satellite Orbiters* (Springer-Praxis, Chichester, 2012).
18. W. B. McKinnon *et al*., The solar nebula origin of (486958) Arrokoth, a primordial contact binary in the Kuiper Belt. *Science* 10.1126/science.aay6620 (2020).
19. Normal reflectance is the *I/F* (where *I* is the scattered intensity from the surface and πF is the solar flux at the distance of the scattering surface; also called the bidirectional reflectance) when the incident and emission angles are both zero.
20. B. Buratti, J. Veverka, Voyager photometry of Europa. *Icarus* **55**, 93–110 (1983). doi: 10.1016/0019-1035(83)90053-2
21. S. Besse, P. Lamy, L. Jorda, S. Marchi, C. Barbieri, Identification and physical properties of craters on Asteroid (2867) Steins. *Icarus* **221**, 1119–1129 (2012). doi: 10.1016/j.icarus.2012.08.008
22. M. S. Robinson, P. C. Thomas, J. Veverka, S. L. Murchie, B. B. Wilcox, The geology of 433 Eros. *Meteorit. Planet. Sci.* **37**, 1651–1684 (2002). doi: 10.1111/j.1945-5100.2002.tb01157.x
23. L. Prockter *et al*., Surface expressions of structural features on Eros. *Icarus* **155**, 75–93 (2002). doi: 10.1006/icar.2001.6770
24. D. L. Buczkowski, O. S. Barnouin-Jha, L. M. Prockter, 433 Eros lineaments: Global mapping and analysis. *Icarus* **193**, 39–52 (2008). doi: 10.1016/j.icarus.2007.06.028
25. S. J. Morrison, P. C. Thomas, M. S. Tiscareno, J. A. Burns, J. Veverka, Grooves on small Saturnian satellites and other objects: Characteristics and significance. *Icarus* **204**, 262–270 (2009). doi: 10.1016/j.icarus.2009.06.003
26. T. A. Hurford *et al*., Tidal disruption of Phobos as the cause of surface fractures. *J. Geophys. Res. Planets* **121**, 1054–1065 (2016). doi: 10.1002/2015JE004943
27. S. A. Stern, The evolution of comets in the Oort cloud and Kuiper belt. *Nature* **424**, 639–642 (2003). doi: 10.1038/nature01725; pmid: 12904784
28. M. J. S. Belton *et al*., The internal structure of Jupiter family cometary nuclei from Deep Impact observations: The "talps" or "layered pile" model. *Icarus* **187**, 332–344 (2007). doi: 10.1016/j.icarus.2006.09.005
29. M. Jutzi, E. Asphaug, The shape and structure of cometary nuclei as a result of low-velocity accretion. *Science* **348**, 1355–1358 (2015). 10.1126/science.aaa4747 pmid: 26022415
30. K. R. Housen, W. J. Sweet, K. A. Holsapple, Impacts into porous asteroids. *Icarus* **300**, 72–96 (2018). doi: 10.1016/j.icarus.2017.08.019
31. S. J. Robbins *et al*., Measuring impact crater depth throughout the solar system. *Meteorit. Planet. Sci.* **53**, 583–637 (2018). doi: 10.1111/maps.12956
32. K. A. Holsapple, K. R. Housen, A crater and its ejecta: An interpretation of Deep Impact. *Icarus* **191**, 586–597 (2007). doi: 10.1016/j.icarus.2006.08.035
33. D. Nesvorný, Dynamical evolution of the early Solar System. *Annu. Rev. Astron. Astrophys.* **56**, 137–174 (2018). doi: 10.1146/annurev-astro-081817-052028
34. S. Sugita *et al*., The geomorphology, color, and thermal properties of Ryugu: Implications for parent-body processes. *Science* **364**, eaaw0422 (2019). doi: 10.1126/science.aaw0422; pmid: 30890587
35. K. J. Walsh *et al*., Craters, boulders and regolith of (101955) Bennu. *Nat. Geosci.* **12**, 242–246 (2019). doi:10.1038/s41561-019-0326-6
36. D. J. Scheeres *et al*., The dynamic geophysical environment of (101955) Bennu based on OSIRIS-REx measurements. *Nat. Astron.* **3**, 352–361 (2019). doi: 10.1038/s41550-019-0721-3
37. A. H. Parker, The intrinsic Neptune Trojan orbit distribution: Implications for the primordial disk and planet migration. *Icarus* **247**, 112–125 (2015). doi: 10.1016/j.icarus.2014.09.043
38. S. Mazrouei, R. R. Ghent, W. F. Bottke, A. H. Parker, T. M. Gernon, Earth and Moon impact flux increased at the end of the Paleozoic. *Science* **363**, 253–257 (2019). doi: 10.1126/science.aar4058; pmid: 30655437
39. K. N. Singer *et al*., Impact craters on Pluto and Charon indicate a deficit of small Kuiper belt objects. *Science* **363**, 955–959 (2019). doi: 10.1126/science.aap8628; pmid: 30819958
40. K. S. Noll, W. M. Grundy, D. Nesvorný, A. Thirouin, "Transneptunian binaries" in *The Trans-Neptunian Solar System*, D. Prialnik, M. A. Barucci, L. Young, Eds. (Elsevier, 2019), pp. 205-224.
41. W. R. Johnston, Binary Minor Planets Compilation V3.0. urn: nasa:pds:ast_binary_parameters_compilation:3.0. NASA Planetary Data System, (2019), https://sbn.psi.edu/pds/resource/binmp.html.
42. F. Marchis *et al*., The puzzling mutual orbit of the binary Trojan asteroid (624) Hektor. *Astrophys. J.* **783**, L37 (2014). doi: 10.1088/2041-8205/783/2/L37
43. P. Descamps *et al*., Triplicity and physical characteristics of Asteroid (216) Kleopatra. *Icarus* **211**, 1022–1033 (2011). doi: 10.1016/j.icarus.2010.11.016
44. F. Braga-Ribas *et al*., A ring system detected around the Centaur (10199) Chariklo. *Nature* **508**, 72–75 (2014). doi: 10.1038/nature13155; pmid: 24670644
45. J. L. Ortiz *et al*., The size, shape, density and ring of the dwarf planet Haumea from a stellar occultation. *Nature* **550**, 219–223 (2017). doi: 10.1038/nature24051; pmid: 29022593
46. J. L. Ortiz *et al*., Possible ring material around centaur (2060) Chiron. *Astron. Astrophys.* **576**, A18 (2015). doi: 10.1051/0004-6361/201424461
47. H. B. Throop *et al*., The Jovian rings: New results derived from Cassini, Galileo, Voyager, and Earth-based observations. *Icarus* **172**, 59–77 (2004). doi: 10.1016/j.icarus.2003.12.020
48. T. R. Lauer *et al*., The New Horizons and Hubble Space Telescope search for rings, dust, and debris in the Pluto-Charon system. *Icarus* **301**, 155–172 (2018). doi: 10.1016/j.icarus.2017.09.033
49. M. Horányi *et al*., The Student Dust Counter on the New Horizons Mission. *Space Sci. Rev.* **140**, 387–402 (2008). doi: 10.1007/s11214-007-9250-y
50. P. Lacerda *et al*., The albedo-color diversity of transneptunian objects. *Astrophys. J.* **793**, L2 (2014). doi: 10.1088/2041-8205/793/1/L2
51. A. Thirouin, S. S. Sheppard, The Plutino population: An abundance of contact binaries. *Astron. J.* **155**, 248 (2018). doi: 10.3847/1538-3881/aac0ff
52. P. Thomas, J. Veverka, P. Helfenstein, Voyager observations of Nereid. *J. Geophys. Res.* **96** (S01), 19253 (1991). doi: 10.1029/91JA01735
53. T. V. Johnson, J. I. Lunine, Saturn's moon Phoebe as a captured body from the outer Solar System. *Nature* **435**, 69–71 (2005). doi: 10.1038/nature03384; pmid: 15875015
54. W. K. Hartmann, A satellite-asteroid mystery and a possible early flux of scattered C-class asteroids. *Icarus* **71**, 57–68 (1987). doi: 10.1016/0019-1035(87)90162-X
55. J. Castillo-Rogez, P. Vernazza, K. Walsh, Geophysical evidence that Saturn's Moon Phoebe originated from a C-type asteroid reservoir. *Mon. Not. R. Astron. Soc.* **486**, 538–543 (2019). doi: 10.1093/mnras/stz786
56. D. P. Simonelli *et al*., Phoebe: Albedo map and photometric properties. *Icarus* **138**, 249–258 (1999). doi: 10.1006/icar.1999.6077
57. B. J. Buratti, M. D. Hicks, K. A. Tryka, M. S. Sittig, R. L. Newburn, High-resolution 0.33–0.92 mm spectra of

Iapetus, Hyperion, Phoebe, Rhea, Dione, and D-type asteroids: How are they related? *Icarus* **155**, 375–381 (2002). doi: 10.1006/icar.2001.6730
58. C. C. Porco *et al*., Cassini Imaging Science: Initial results on Phoebe and Iapetus. *Science* **307**, 1237–1242 (2005). doi: 10.1126/science.1107981; pmid: 15731440
59. M. Duncan, H. Levison, L. Dones, "Dynamical evolution of ecliptic comets". In *Comets II*, M. C. Festou, H. U. Keller, H. A. Weaver Eds. (University of Arizona Press, Tucson, 193, 2004).
60. D. E. Vavilov, S. Eggl, Y. D. Medvedev, P. B. Zatitskiy, Shape evolution of cometary nuclei via anisotropic mass loss. *Astron. Astrophys.* **622**, L5 (2019). doi: 10.1051/0004-6361/201834806
61. O. Groussin *et al*., The thermal, mechanical, structural, and dielectric properties of cometary nuclei after Rosetta. *Space Sci. Rev.* **215**, 29 (2019). doi: 10.1007/s11214-019-0594-x
62. R. Kokotanekova *et al*., Rotation of cometary nuclei: New light curves and an update of the ensemble properties of Jupiter-family comets. *Mon. Not. R. Astron. Soc.* **471**, 2974–3007 (2017). doi: 10.1093/mnras/stx1716
63. D. Bodewits, T. L. Farnham, M. S. P. Kelley, M. M. Knight, A rapid decrease in the rotation rate of comet 41P/Tuttle-Giacobini-Kresák. *Nature* **553**, 186–188 (2018). doi: 10.1038/nature25150; pmid: 29323296
64. J. Sunshine, N. Thomas, M. R. El-Maarry, T. L. Farnham, Evidence for geologic processes on comets. *J. Geophys. Res. Planets* **121**, 2194–2210 (2016). doi: 10.1002/2016JE005119
65. M. R. El-Maarry *et al*., Surface changes on comet 67P/Churyumov-Gerasimenko suggest a more active past. *Science* **355**, 1392–1395 (2017). doi: 10.1126/science.aak9384; pmid: 28325842
66. P. C. Thomas *et al*., Shape, density, and geology of the nucleus of Comet 103P/Hartley 2. *Icarus* **222**, 550–558 (2013). doi: 10.1016/j.icarus.2012.05.034
67. D. T. Britt *et al*., The morphology and surface processes of Comet 19/P Borrelly. *Icarus* **167**, 45–53 (2004). doi: 10.1016/j.icarus.2003.09.004
68. N. Thomas *et al*., Redistribution of particles across the nucleus of comet 67P/Churyumov-Gerasimenko. *Astron. Astrophys.* **583**, A17 (2015). doi: 10.1051/0004-6361/201526049
69. D. Prialnik, J. Benkhoff, M. Podolak, "Modeling the structure and activity of comet nuclei" in *Comets II*, M. C. Festou, H. U. Keller, H. A. Weaver Eds. (Univ. of Arizona Press, Tucson, 2004), pp. 359–387.
70. W. C. Fraser, M. E. Brown, A. Morbidelli, A. Parker, K. Batygin, The absolute magnitude distribution of Kuiper Belt Objects. *Astrophys. J.* **782**, 100 (2014). doi: 10.1088/0004-637X/782/2/100
71. R. Li, A. N. Youdin, J. B. Simon, Demographics of planetesimals formed by the streaming instability. *Astrophys. J.* **885**, 69 (2019). doi: 10.3847/1538-4357/ab480d
72. H. J. Melosh, *Impact Cratering: A Geologic Process* (Oxford Univ. Press, Oxford, 1989).
73. H. Sierks *et al*., On the nucleus structure and activity of comet 67P/Churyumov-Gerasimenko. *Science* **347**, aaa1044 (2015). doi: 10.1126/science.aaa1044; pmid: 25613897
74. M. F. A'Hearn *et al*., EPOXI at comet Hartley 2. *Science* **332**, 1396–1400 (2011). doi: 10.1126/science.1204054; pmid: 21680835
75. J. Veverka *et al*., Return to Comet Tempel 1: Overview of Stardust-NExT results. *Icarus* **222**, 424–435 (2013). doi: 10.1016/j.icarus.2012.03.034
76. D. E. Brownlee *et al*., Surface of young Jupiter family comet 81P/Wild 2: View from the Stardust Spacecraft. *Science* **304**, 1764–1769 (2004). doi: 10.1126/science.1097899; pmid: 15205524
77. L. A. Soderblom *et al*., Observations of comet 19P/Borrelly by the miniature integrated camera and spectrometer aboard Deep Space 1. *Science* **296**, 1087–1091 (2002). doi: 10.1126/science.1069527; pmid: 11934989
78. B. J. Buratti *et al*., Deep Space 1 photometry of the nucleus of Comet 19P/Borrelly. *Icarus* **167**, 16–29 (2004). doi: 10.1016/j.icarus.2003.05.002
79. J. R. Spencer *et al*., Data archive for Spencer et al. 2020, The geology and geophysics of Kuiper Belt object (486958) Arrokoth, Science, figshare (2020); https://doi.org/10.6084/m9.figshare.11485443.

ACKNOWLEDGMENTS

We thank all who contributed to the success of the New Horizons flyby of Arrokoth, and in particular the National Astronomical Observatory of Japan's Subaru Telescope, the Carnegie Observatory's Magellan Telescopes, the Canada-France-Hawaii Telescope, the NASA Hubble Space Telescope, the Harvard-Smithsonian Center for Astrophysics, the Massachusetts Institute of Technology, Northern Arizona University, the University of Hawaii, the Hertzberg Institute for Astrophysics, and NASA, for their support of the search campaign that led to its discovery. We also are indebted to the Hubble Space Telescope and the European Space Agency's Gaia mission for their key roles in the precise orbit determination required to enable the successful flyby. **Funding:** Supported by NASA's New Horizons project under contracts NASW-02008 and NAS5- 97271/TaskOrder30. J.J.K. was supported by the National Research Council of Canada and the National Science and Engineering Research Council of Canada. **Author contributions:** The primary contributors to each section were as follows: Stereo and shape modeling: S.B.P., R.A.B., P.M.S., A.M.Z., M.W.B., and T.R.L. Geophysical analysis: J.T.K., O.M.U., and W.B.M. Photometric analysis: B.J.B., J.D.H., A.J.V., and S.D.B. Geological mapping and interpretation: J.M.M., O.L.W., R.A.B., O.M.U., J.R.S., and T.R.L. Crater analysis: K.N.S., S.J.R., K.D.R., P.M.S., and A.H.P. Satellite and ring search: J.R.S., S.B.P., M.W.B., M.R.S., T.R.L., H.B.T., A.J.V., A.M.Z., W.M.G., D.P.H., E.B., and D.E.K. Astronomical context: H.A.W., D.T.B., C.M.L., M.R.E.-M., J.J.K., and W.B.M. Paper compilation and synthesis: J.R.S. S.A.S. is the principal investigator of the New Horizons mission and reviewed this manuscript. All other authors participated in mission planning, initial discovery and tracking of Arrokoth, science data reduction or analysis, or provided inputs and critique to this manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability**: All images, spacecraft data, and the shape model used in this paper are available at figshare (*79*). Additional fully calibrated New Horizons Arrokoth data and higher-order data products will be released by the NASA Planetary Data System in a series of stages in 2020 and 2021, owing to the time required to fully downlink and calibrate the dataset.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/eaay3999/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S4
Tables S1 to S3
References (*80*–*90*)
Data S1 to S3

14 June 2019; accepted 27 January 2020
Published online 13 February 2020
10.1126/science.aay3999

RESEARCH ARTICLE SUMMARY

OUTER SOLAR SYSTEM

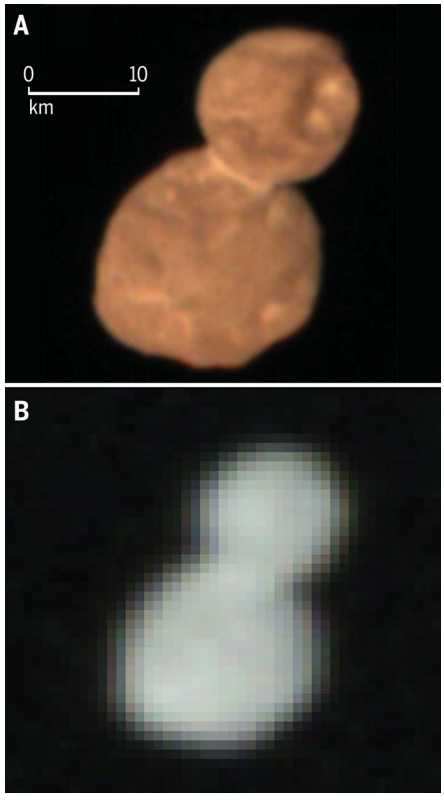
Color, composition, and thermal environment of Kuiper Belt object (486958) Arrokoth

W. M. Grundy*, M. K. Bird, D. T. Britt, J. C. Cook, D. P. Cruikshank, C. J. A. Howett, S. Krijt, I. R. Linscott, C. B. Olkin, A. H. Parker, S. Protopapa, M. Ruaud, O. M. Umurhan, L. A. Young, C. M. Dalle Ore, J. J. Kavelaars, J. T. Keane, Y. J. Pendleton, S. B. Porter, F. Scipioni, J. R. Spencer, S. A. Stern, A. J. Verbiscer, H. A. Weaver, R. P. Binzel, M. W. Buie, B. J. Buratti, A. Cheng, A. M. Earle, H. A. Elliott, L. Gabasova, G. R. Gladstone, M. E. Hill, M. Horanyi, D. E. Jennings, A. W. Lunsford, D. J. McComas, W. B. McKinnon, R. L. McNutt Jr., J. M. Moore, J. W. Parker, E. Quirico, D. C. Reuter, P. M. Schenk, B. Schmitt, M. R. Showalter, K. N. Singer, G. E. Weigle II, A. M. Zangari

INTRODUCTION: The New Horizons spacecraft flew past (486958) Arrokoth (provisional designation 2014 MU_{69}) on 1 January 2019. Arrokoth is a member of the subclass of trans-neptunian or Kuiper Belt objects (KBOs), known as the cold classical KBOs (CCKBOs). Most KBOs formed in a disk of planetesimals that extended to about 30 AU from the Sun. Neptune eventually disrupted that disk by migrating outward through it, with the migration halted by the sparseness of the disk beyond 30 AU. That event eliminated most members of the planetesimal disk, but a minority were emplaced into dynamically excited orbits in the present-day Kuiper belt. CCKBOs differ from those objects in having formed well beyond the 30-AU edge of the main planetesimal disk. They remain approximately where they formed, on low-inclination, near-circular orbits between 42 and 47 AU from the Sun, relicts of the early Solar System. Their distributions of colors, albedos, sizes, and binarity differ from those of the more excited KBOs.

RATIONALE: Initial results from the exploration of Arrokoth were published previously. More data have since been received from the spacecraft, allowing a more detailed analysis. We analyze a high–spatial resolution color imaging observation, near-infrared spectral imaging, and microwave radiometry of Arrokoth. The infrared spectral data have been processed to compensate for the changing range and scale during the observation. Our multiple scattering radiative transfer models provide compositional constraints from the infrared spectral imagery. Microwave thermal radiometry at 4.2-cm wavelength is combined with heat transport models that account for the bilobate shape of Arrokoth and for self-radiation.

RESULTS: At visual wavelengths, Arrokoth's reflectance rises toward longer wavelengths. This red coloration is typical of the broader CCKBO population that has been studied using telescopic observations. Color differences across the surface of Arrokoth correspond to geological features. These color differences are subtle, with deviations of just a few percent around the prevailing red coloration. Some of the color variations are associated with albedo markings, such as the bright neck between the two lobes, bright splotches associated with a large pit or crater on the smaller lobe, and poorly resolved small bright spots. Methanol ice (CH_3OH) and complex organic tholins dominate the near-infrared reflectance spectrum, with H_2O ice contributing little or no detectable absorption. At the 4.2-cm microwave wavelength of New Horizons' radio system, Arrokoth's winter night side glows with an average brightness temperature of 29 ± 5 K. This emission probably emerges from below the cold winter surface, at depths where warmth from the previous summer lingers. Our models show that self-radiation more than compensates for self-shadowing in the neck region between the two lobes, resulting in warmer temperatures in that region, by up to a few kelvin.



Visible and near-infrared wavelength views of Arrokoth. (**A**) Blue, green, and red channels show wavelengths 0.40 to 0.55 μm, 0.54 to 0.70 μm, and 0.78 to 0.98 μm, respectively. (**B**) These channels show wavelengths 1.2 to 1.6 μm, 1.6 to 2.0 μm, and 2.0 to 2.5 μm, respectively.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aay3705

CONCLUSION: The nearly uniform coloration across Arrokoth is consistent with expectations for an object that accreted too quickly for the composition of the available nebular solids to have changed during the course of its accretion. Radiolysis and photolysis from long exposure to space radiation would be expected to result in a dark, space-weathered surface veneer that is distinct from the more pristine interior, but there is little evidence for such a coating, perhaps because radiolytically processed material is eroded away faster than it accumulates. The abundance of CH_3OH ice and apparent scarcity of H_2O ice appear to be signatures of a distinct environment in the cold, dust-shaded midplane of the outer nebula during formation of the Solar System. In this region, temperatures would have been low enough that volatile CO and CH_4 could freeze onto dust grains, enabling production of CH_3OH and perhaps also destruction of H_2O. When the nebular dust dissipated some time after Arrokoth's formation, exposure to sunlight would have raised its temperature, rapidly driving off condensed CO and CH_4. The temperature has remained too cold to crystallize amorphous H_2O. Volatile species may remain trapped in amorphous H_2O ice within Arrokoth's interior, but the infrared spectrum shows little evidence for such ice at the surface. Although the neck region gets slightly warmer than the rest of Arrokoth's surface, this effect is small relative to the winter-summer temperature contrast and is thus unlikely to account for the distinctly higher albedo and slightly less red material that is seen there. A more plausible explanation for the neck's albedo and color contrasts involves texture changes induced by the merger of the two lobes or subsequent downslope movement of material there. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: w.grundy@lowell.edu
Cite this article as W. M. Grundy *et al.*, *Science* **367**, eaay3705 (2020). DOI: 10.1126/science.aay3705

RESEARCH ARTICLE

OUTER SOLAR SYSTEM

Color, composition, and thermal environment of Kuiper Belt object (486958) Arrokoth

W. M. Grundy[1,2]*, M. K. Bird[3,4], D. T. Britt[5], J. C. Cook[6], D. P. Cruikshank[7], C. J. A. Howett[8], S. Krijt[9], I. R. Linscott[10], C. B. Olkin[8], A. H. Parker[8], S. Protopapa[8], M. Ruaud[7], O. M. Umurhan[7,11], L. A. Young[8], C. M. Dalle Ore[7,11], J. J. Kavelaars[12,13], J. T. Keane[14], Y. J. Pendleton[7], S. B. Porter[8], F. Scipioni[7,11], J. R. Spencer[8], S. A. Stern[8], A. J. Verbiscer[15], H. A. Weaver[16], R. P. Binzel[17], M. W. Buie[8], B. J. Buratti[18], A. Cheng[16], A. M. Earle[17], H. A. Elliott[19], L. Gabasova[20], G. R. Gladstone[19], M. E. Hill[16], M. Horanyi[21], D. E. Jennings[22], A. W. Lunsford[22], D. J. McComas[23], W. B. McKinnon[24], R. L. McNutt Jr.[16], J. M. Moore[7], J. W. Parker[8], E. Quirico[20], D. C. Reuter[22], P. M. Schenk[25], B. Schmitt[20], M. R. Showalter[11], K. N. Singer[8], G. E. Weigle II[26], A. M. Zangari[8]

The outer Solar System object (486958) Arrokoth (provisional designation 2014 MU_{69}) has been largely undisturbed since its formation. We studied its surface composition using data collected by the New Horizons spacecraft. Methanol ice is present along with organic material, which may have formed through irradiation of simple molecules. Water ice was not detected. This composition indicates hydrogenation of carbon monoxide–rich ice and/or energetic processing of methane condensed on water ice grains in the cold, outer edge of the early Solar System. There are only small regional variations in color and spectra across the surface, which suggests that Arrokoth formed from a homogeneous or well-mixed reservoir of solids. Microwave thermal emission from the winter night side is consistent with a mean brightness temperature of 29 ± 5 kelvin.

The New Horizons spacecraft flew past (486958) Arrokoth at the beginning of 2019 (*1*). Arrokoth rotates with a 15.9-hour period, about a spin axis inclined 99.3° to the pole of its 298-year orbit, at a mean distance from the Sun of 44.2 AU (*2*, *3*). Its near-circular orbit, with a mean eccentricity of 0.03 and inclination of 2.4° to the plane of the Solar System, makes it a Kuiper Belt object (KBO) and specifically a member of the "kernel" subpopulation of the cold classical KBOs (CCKBOs) (*4*). The origins and properties of CCKBOs are distinct from those of KBOs on more excited orbits, which are thought to have formed closer to the Sun before being perturbed outward by migrating giant planets early in Solar System history (*5*). CCKBOs still orbit where they formed in the protoplanetary nebula (the accretion disk of gas and dust around the young Sun). CCKBOs have a high fraction of binary objects (*6*), a uniformly red color distribution (*7*, *8*), a size-frequency distribution deficient in large objects (*9*, *10*), and higher albedos (*11*, *12*) than other KBOs. These properties arise from the environment at the outermost edge of the protoplanetary nebula, from a distinct history of subsequent evolution of CCKBOs relative to other KBOs, or of some combination of these two. Arrokoth provides a record of the process of forming planetesimals—the first generation of gravitationally bound bodies—that has been minimally altered by subsequent processes such as heating and impactor bombardment (*3*). Its distinctive bilobed, 35-km-long shape with few impact craters favors formation via rapid gravitational collapse, rather than scenarios involving more gradual accretion via piecewise agglomeration of dust particles to assemble incrementally larger aggregates (*13*). We study Arrokoth's color, composition, and thermal environment using data from the New Horizons flyby and discuss the resulting implications for its formation and subsequent evolution.

[1]Lowell Observatory, Flagstaff, AZ 86001, USA. [2]Department of Astronomy and Planetary Science, Northern Arizona University, Flagstaff, AZ 86011, USA. [3]Argelander-Institut für Astronomie, University of Bonn, D-53121 Bonn, Germany. [4]Rheinisches Institut für Umweltforschung, Universität zu Köln, 50931 Cologne, Germany. [5]University of Central Florida, Orlando, FL 32816, USA. [6]Pinhead Institute, Telluride, CO 81435, USA. [7]NASA Ames Research Center, Moffett Field, CA 94035, USA. [8]Southwest Research Institute, Boulder, CO 80302, USA. [9]Steward Observatory, University of Arizona, Tucson, AZ 85719, USA. [10]Stanford University, Stanford, CA 94305, USA. [11]Carl Sagan Center, SETI Institute, Mountain View, CA 94043, USA. [12]National Research Council, Victoria, BC V9E 2E7, Canada. [13]Department of Physics and Astronomy, University of Victoria, Victoria, BC V8W 2Y2, Canada. [14]California Institute of Technology, Pasadena, CA 91125, USA. [15]University of Virginia, Charlottesville, VA 22904, USA. [16]Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723, USA. [17]Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [18]NASA Jet Propulsion Laboratory, La Cañada Flintridge, CA 91011, USA. [19]Southwest Research Institute, San Antonio, TX 78238, USA. [20]Institut de Planétologie et d'Astrophysique de Grenoble, Centre National de la Recherche Scientifique, Université Grenoble Alpes, Grenoble, France. [21]University of Colorado, Boulder, CO 80309, USA. [22]NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. [23]Princeton University, Princeton, NJ 08544, USA. [24]Washington University, St. Louis, MO 63130, USA. [25]Lunar and Planetary Institute, Houston, TX 77058, USA. [26]Big Head Endian LLC, Leawood, KS 67019, USA.
*Corresponding author. Email: w.grundy@lowell.edu

Instruments and data

New Horizons encountered Arrokoth when it was 43.28 AU from the Sun, collecting data with a suite of scientific instruments. Color and compositional remote sensing data were provided by the Ralph color camera and imaging spectrometer, sensitive to wavelengths between 0.4 and 2.5 μm (*14*). Over this wavelength range, all light observed from Arrokoth is reflected sunlight, with the wavelength dependence of the reflectance indicative of surface composition and texture. Ralph's two focal planes share a single 75-mm aperture telescope using a dichroic beamsplitter. The Multispectral Visible Imaging Camera (MVIC) provides panchromatic and color imaging in four color filters: BLUE (400 to 550 nm), RED (540 to 700 nm), NIR (780 to 975 nm), and CH4 (860 to 910 nm) (*15*). The highest–spatial resolution MVIC color observation of Arrokoth, designated as CA05, was obtained on 1 January 2019 at 05:14 UTC (coordinated universal time), from a range of 17,200 km, at an image scale of 340 m per pixel and phase angle of 15.5°. This provides more spatial detail than the CA02 MVIC color scan at 860 m per pixel (*1*).

Ralph's Linear Etalon Infrared Spectral Array (LEISA) images its target scene through a linear variable filter covering wavelengths from 1.2 to 2.5 μm at a spectral resolving power of about 240. Frames are recorded while the spacecraft scans LEISA's field of view across the scene; this enables each location to be captured at each wavelength of the filter. The highest–spatial resolution LEISA observation, designated as CA04, was executed around 04:58 UTC, shortly before the CA05 MVIC observation, from a phase angle of 12.6° and a mean range of 31,000 km, resulting in a mean image scale of 1.9 km per pixel (*15*).

New Horizons' panchromatic Long-Range Reconnaissance Imager [LORRI (*16*)] is co-aligned with Ralph and can record images while the spacecraft is scanning for a Ralph observation. Such LORRI observations, referred to as "riders," are limited to short integration times to minimize image smear from scan motion, but multiple images can be recorded and combined in postprocessing, providing for longer effective integration times (*3*). LORRI rider observations were obtained during both the CA04 and CA05 observations, providing higher–spatial resolution context images for the Ralph observations.

New Horizons' Radio Science Experiment [REX (*17*)] was used to observe thermal emission in the X-band (4.2-cm wavelength, 7.2 GHz) from Arrokoth's Sun-oriented face on approach and then from its anti–Sun-oriented face on departure. The two REX observations, designated as CA03 and CA08, respectively, were obtained on 1 January at mean times of 04:34 and 05:52 UTC, phase angles of 11.9° and 162.0°, and ranges of 52,000 and 16,700 km. At those

distances, Arrokoth was unresolved, appearing much smaller than the 1.2° width (at 3 dB) of the high-gain antenna beam. Two independent receivers recorded the radio flux density in different polarizations at a sampling rate of 10 Hz. The REX A receiver recorded right circularly polarized flux while REX B recorded left circularly polarized flux.

Visible-wavelength

The CA02 MVIC color scan (*1*) had shown Arrokoth to be red but revealed little spatial variation in color. The higher-resolution CA05 observation allows us to better quantify Arrokoth's regional color variations. Figure 1 compares the CA05 color image with the contemporaneous LORRI panchromatic rider image. Color slopes, computed by fitting a linear model to the MVIC BLUE, RED, and NIR filter reflectance data, are shown in Fig. 1C. All of Arrokoth's surface is red in color, with a mean color slope of 27% rise per 100 nm relative (at 550 nm). This quantification of color slopes is commonly used for KBOs, being convenient for comparison of colors obtained using different filter sets (*18*). Even in the higher spatial resolution of the CA05 observation, the color distribution is largely uniform across the observed face of Arrokoth, with a standard deviation in slope of only ±2.7% per 100 nm.

Subtle regional color differences correspond to specific geological and albedo features discussed in a companion paper (*3*). The smaller lobe (SL) appears slightly redder on average than the larger lobe (LL) [28 ± 2% average slope versus 27 ± 2% for LL, where the ±2 values represent the variance across each lobe, rather than the uncertainty in the measurement of the mean slopes, which is much smaller for averages over many pixels (*15*)]. That difference appears to be mostly due to the redder rim (color slope 30 ± 2%) of a 6-km-diameter depression on SL, a possible impact crater informally designated as Maryland (MD; all place names are informal). Statistically significant color differences tend to be on similarly small (or smaller) spatial scales. Several slightly less red regions appear as blue in the color scale used in Fig. 1C. These include the brighter neck region where the two lobes intersect (25 ± 1% slope) and several regions on LL. Two regions that were not resolved in the earlier color data are Louisiana, a depression near the neck (23 ± 2% slope), and North Dakota, a linear depression or groove (24 ± 1% slope), labeled in Fig. 1A. Bright material (bm) in the geological map (*3*) is in some places more and in others less red than average, suggesting that that unit is composed of two or more distinct materials. Another depressed region (dr) on LL is slightly redder than the average (29 ± 2% slope). Some bright spots (sp) appear to have distinct colors as well, although they are not all the same; some are a little redder than average while others are a little less red. The lack of a consistent color pattern for these spots suggests that they may have resulted from delivery of diverse material in impacts rather than by impact excavation of a uniform subsurface material. However, the nature of these spots remains ambiguous (*3*).

We performed a principal components analysis (PCA) of the color data (Fig. 2). This analysis projects the data into an orthogonal basis set, with the first axis corresponding to the axis of maximum variance within the data. The second axis corresponds to the maximum remaining variance after collapsing the data along the first axis, and so forth. Principal component 1 (PC1, Fig. 2A) corresponds with variations in brightness due to lighting and albedo, accounting for 97% of the total variance in the data. The corresponding eigenvector is flat across all filters. PC2 corresponds to redness (Fig. 2B), closely resembling the color slope map in Fig. 1C, and accounts for 1.8% of the total variance. PC3 and PC4 correspond to contrasts between the NIR and CH4 filters and to spectral curvature between BLUE, RED, and NIR filters, respectively. They account for only 1% of the variance between them, much of it due to image noise rather than real color variations across the surface of Arrokoth.

Red coloration on planetary bodies is often attributed to the presence of tholins (*19*). These are a broad class of refractory macromolecular polymer-like organic solids, commonly produced in laboratory simulations of energetic radiation acting on various combinations of simpler molecules (*20*–*22*). The precursors can be in gaseous form (*23*) or frozen solid (*24*, *25*). Figure 3 compares the color of Arrokoth

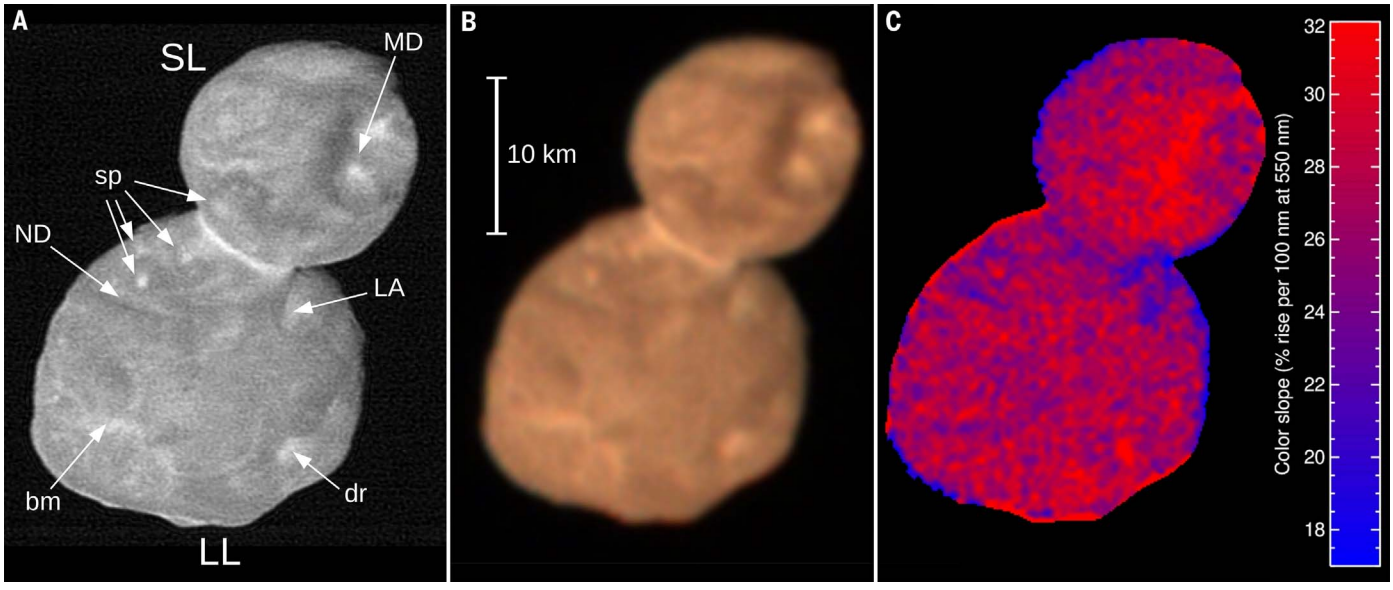


Fig. 1. The CA05 color observation of (486958) Arrokoth. (**A**) LORRI panchromatic context image obtained as a rider during the CA05 observation; the geometry is nearly identical to the MVIC observation, but with a finer spatial scale of 83 m pixel^{-1}. Abbreviations and informal feature names (clockwise from top): SL, smaller lobe; MD, Maryland; LA, Louisiana; dr, depressed region; LL, larger lobe; bm, bright material; ND, North Dakota; sp, bright spots. These features can be seen more clearly in higher-resolution LORRI images [figure 1 of (*3*)]. (**B**) Color observation CA05, at a spatial scale of 340 m pixel^{-1}. The BLUE filter (400 to 550 nm) is displayed in blue, RED filter (540 to 700 nm) in green, and NIR filter (780 to 975 nm) in red. (**C**) Color slope map obtained by fitting a linear model to the BLUE, RED, and NIR reflectance values.

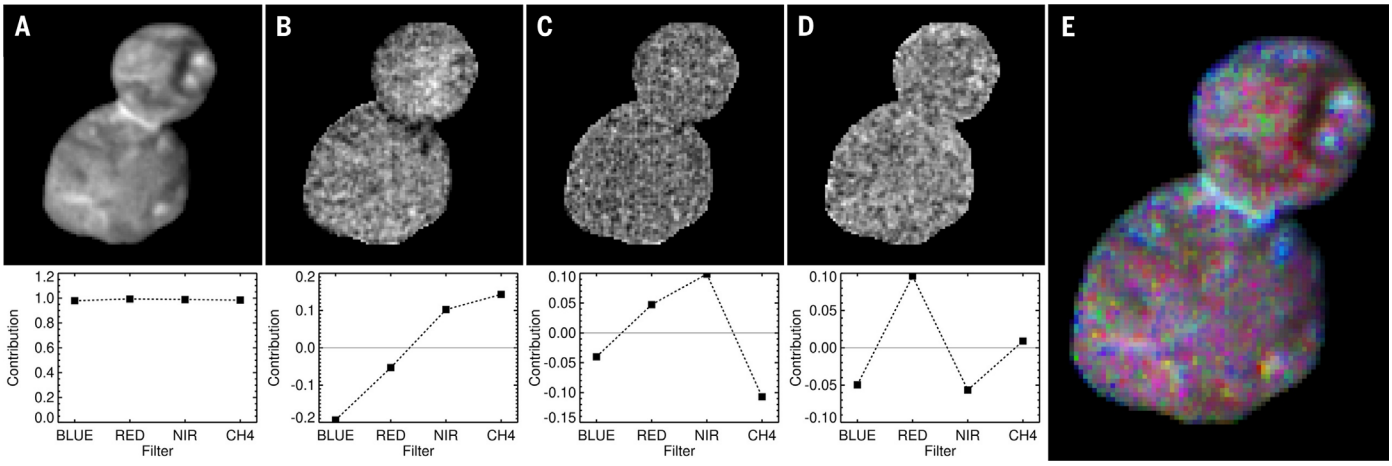


Fig. 2. Principal components analysis of the CA05 color data. (**A** to **D**) Principal components images (PCs) 1 through 4, with the corresponding eigenvector for each. The images show the regional distribution of the spectral characteristic indicated by the eigenvector. For example, the PC2 eigenvector shows a red slope (B). The corresponding image shows high values where Arrokoth is reddest and low values where it is least red. The *y* axes of the eigenvector plots give the contribution of each filter to its respective PC. (**E**) False-color image highlighting color contrasts across Arrokoth with PC2, PC3, and PC4 displayed in the red, green, and blue channels, respectively, with shading from the PC1 image.

with that of other KBOs and related populations. Arrokoth's color slope is consistent with other CCKBOs (*8*). It is less red than the reddest KBOs and the red group of Centaurs (*26*), whose red coloration is generally interpreted as due to tholins (*19*). Arrokoth is much more red than the gray Centaurs and various classes of asteroids (*27*), including the red D-type asteroids that dominate the Jupiter Trojan population (*28*). Using the Sloan *g*, *r*, and *z* filters, KBO colors can be placed on a color-color plot of *g*-*r* and *r*-*z* color differences, which has revealed a distinction between CCKBOs and more dynamically excited KBO populations (*8*). The latter appear to follow two color tracks (*29*) (Fig. 3B) while the CCKBOs cluster below and to the right, owing to their red slopes becoming less steep at *z*-band wavelengths (0.82 to 0.96 μm). Arrokoth's red slope continues into MVIC's NIR band (0.78 to 1 μm). After converting to Sloan colors, Arrokoth lies above the main clump of CCKBO colors in Fig. 3B, although this is consistent with the overall CCKBO color distribution.

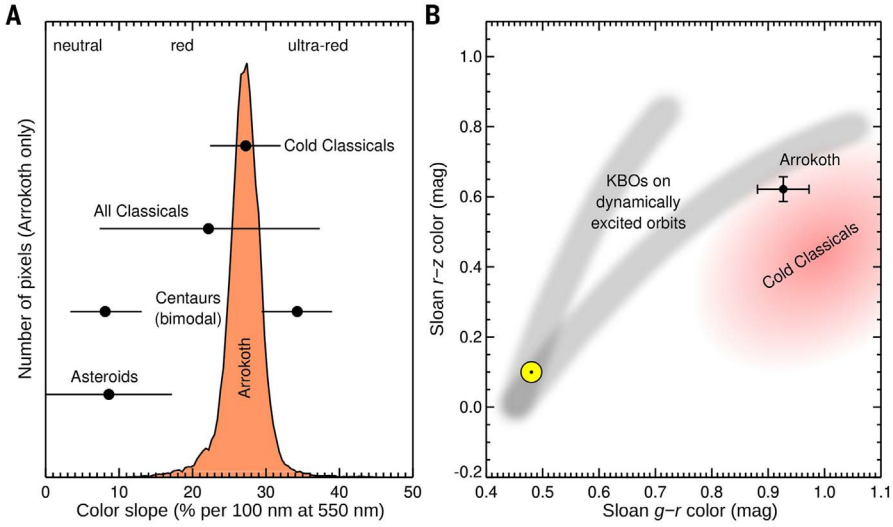


Fig. 3. Comparison of Arrokoth's color with other populations. (**A**) Histogram of Arrokoth color slopes shown in Fig. 1C. Typical color ranges for other Solar System small-body populations are indicated via horizontal bars [e.g., (*8*, *18*, *26*–*29*)]. (**B**) Color-color plot using Sloan *g*, *r*, and *z* filters, with the additional wavelength providing sensitivity to spectral curvature in addition to slope. In this color-color space, KBOs form three distinct zones (*8*). The gray tracks are populated by KBOs on dynamically excited (higher inclination and/or eccentricity) orbits; CCKBOs occupy the red zone. Arrokoth lies above the center of the CCKBO clump, because its red slope declines less into the *z*-band than the average for CCKBOs. The yellow Sun symbol indicates solar color, corresponding to neutral spectral reflectance at visible wavelengths.

Near-infrared spectral reflectance

The LEISA data were processed into a spectral cube, with spatial dimensions along two axes and wavelength along the third axis (*15*). The cube-building algorithm we adopted accounts for changes in spacecraft range over the course of the CA04 LEISA scan [unlike (*1*)]. The spatial resolution of the LEISA data is considerably coarser than that of the corresponding LORRI rider (Fig. 4), and the signal/noise ratio is considerably lower. Noise in the LEISA data is dominated by instrumental effects, which are as large as the signal from sunlit areas on Arrokoth, so spectral features can only be revealed by averaging over multiple pixels and/or wavelengths. The spatially averaged spectrum (Fig. 4C) lacks the strong absorption features that were seen in the Pluto system [e.g., (*30*)]. The red color slope seen in the visible flattens with increasing wavelength to become neutral by ~1.5 μm. We constructed Hapke reflectance models (*15*, *31*) for various combinations of potential surface components. The data support inclusion of amorphous carbon and tholins in the models. Combinations of these materials match the overall albedo and spectral shape. Although we used tholins made in conditions simulating the atmosphere of the moon Titan (*20*), the data do not support singling out any of the various tholins with published optical constants. None have been made under simulated outer nebular conditions, so they may not be particularly good analogs for the tholins on Arrokoth. Amorphous carbon has no diagnostic spectral features in this wavelength

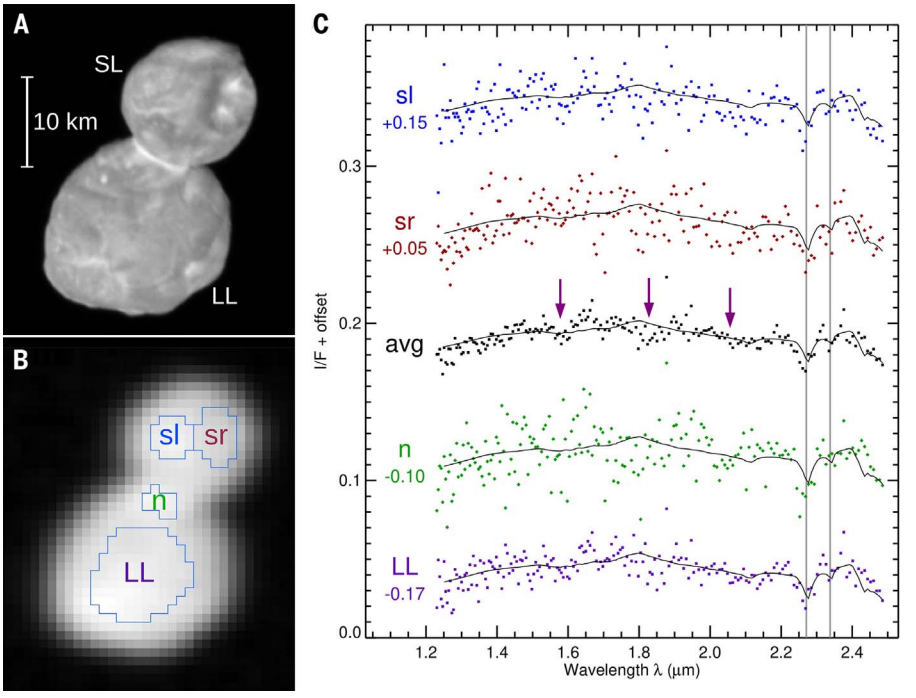


Fig. 4. The CA04 infrared spectral observation of Arrokoth. (**A**) The LORRI CA04 rider image for context, with a spatial scale of 138 m pixel^{-1}. (**B**) LEISA image with regions of interest (ROIs) indicated, with a much coarser mean spatial scale of 1.9 km pixel^{-1}. (**C**) Average and ROI spectra (points) compared with the Hapke model fitted to the average spectrum (black curves). Vertical gray lines indicate the wavelengths of the two strongest CH_3OH ice absorptions. Purple arrows mark other possible features discussed in the text. I/F is defined as the ratio of the bidirectional reflectance to that of a perfectly diffusely scattering surface illuminated normally (*31*).

range, so it cannot be specifically identified. Any dark, spectrally neutral material would be equally consistent with the data.

Shallow absorption bands can be discerned near 1.5 to 1.6 μm, 1.8 μm, 2.0 to 2.1 μm, 2.27 μm, and 2.34 μm (Fig. 4C). Inclusion of ices of methanol (CH_3OH), water (H_2O), and ammonia (NH_3) enables the models to match many of these features, but not the one at 1.8 μm. With the low signal/noise ratio of the LEISA spectrum, even in the global average, we must consider the information content and limit the model free parameters to those that can be statistically justified. Of the molecular ices we tried, only the addition of CH_3OH produces a sufficiently large improvement in χ^2 to constitute a confident detection (*15*). This does not preclude the presence of H_2O or NH_3 ices, but the available data do not provide statistically significant evidence for their presence. Adding more than a trace of them to the models makes χ^2 worse, but the increase can be minimized with large grain sizes that limit the projected area of H_2O or NH_3 grains to a small fraction of the total. Small amounts of other ices, such as H_2CO, CO_2, or C_2H_6, could also be compatible with the data, as could the silicate and metallic phases that have been seen in comets and interplanetary dust particles. The 1.8-μm feature in Arrokoth's spectrum is not matched by available ices or tholins and remains unidentified. It could be an artifact.

To search for spectral contrasts across the surface of Arrokoth, we selected several regions of interest (ROIs), as shown in Fig. 4B. These include LL, the brighter neck region (n), and a pair of ROIs on the left and right sides of SL: sr, which incorporates the redder material on the rim of Maryland, and sl, which represents portions of SL unrelated to Maryland. Spectra of these regions are shown in Fig. 4C. In averaging over fewer pixels, the signal/noise ratios in the ROI spectra are correspondingly poorer than the global average. However, the ROI spectra all look very similar to the average, and fitting Hapke models shows that tholin, carbon, and CH_3OH are favored, without statistically significant evidence for H_2O or NH_3 ices, just as with the average.

Our confidence in the detection of CH_3OH is increased by the appearance of two distinct absorption bands of methanol ice: one band at 2.271 μm, attributed to ($\nu_1 + \nu_{11}$) or ($\nu_1 + \nu_7$) vibrational combination modes, and another at 2.338 μm, attributed to ($\nu_1 + \nu_8$) (*32*). These spectral characteristics have been seen in Earth-based spectra of the Centaur 5145 Pholus (*33*) and the resonant KBO (55638) 2002 VE_{95} (*34*). Additional, weaker CH_3OH absorption bands at 1.6 and 2.1 μm are not visible in Arrokoth's spectrum. In the case of Pholus, the spectrum from 0.45 to 2.45 μm was fitted with a radiative transfer model incorporating solid CH_3OH and H_2O, in addition to an iron-bearing olivine (forsterite Fo 82) and tholin (*33*), similar to our models for Arrokoth, except that H_2O ice is not required in the Arrokoth models.

To assess spectral contrasts in a way that does not depend on multiple scattering models, we performed a PCA on the LEISA cube, with results shown in Fig. 5. Because of the low signal/noise ratio, we first binned the wavelengths down to 28 channels, producing an effective spectral resolving power of 39. We also discarded pixels along the edge where jitter during the scan is prone to producing artifacts. As with the MVIC colors, PC1 is sensitive to the overall light variation from shading and albedo, with the eigenvector flat across all wavelengths (Fig. 5A). PC1 accounts for 72% of the total variance in the LEISA data. PC2 captures only 2.3% of the variance, but the eigenvector shows pronounced dips around 1.5 and 2 μm, where H_2O ice has its strongest absorption bands within LEISA's spectral range; this finding suggests that regional variations in H_2O ice absorption could be the next most prominent source of spectral variance across the surface of Arrokoth, perhaps being most abundant around Maryland crater. However, the absence of strong H_2O absorptions in any of our extracted spectra (including the sr ROI that covers this region) reduces confidence in that conclusion. Subsequent PCs account for even lower fractions of the total variance. The lack of spatial coherence in the images, coupled with eigenvectors that are not suggestive of absorption by likely surface constituents, suggests that the higher PCs are responding mostly to instrumental noise rather than to signal in the LEISA data.

Thermal environment

The CA03 REX observation was performed on approach, observing Arrokoth's day side; the CA08 observation was done after closest approach, looking back at Arrokoth's night side. The microwave sky background is shown in Fig. 6A (*35*, *36*). The CA03 observation was performed with a fixed staring geometry. A later observation of the same field was obtained for background subtraction, but the system antenna temperature drifts over time, making it difficult to separate out the flux from Arrokoth. The CA08 night-side observation was obtained under more favorable geometry, from a closer range, and with the antenna scanned along the uncertainty ellipse for Arrokoth's location. Scanning instead of staring facilitated calibration against the later background observation despite the drift in system response. The flux measurements are shown Fig. 6B. The radiometric signal was converted to radio brightness temperature

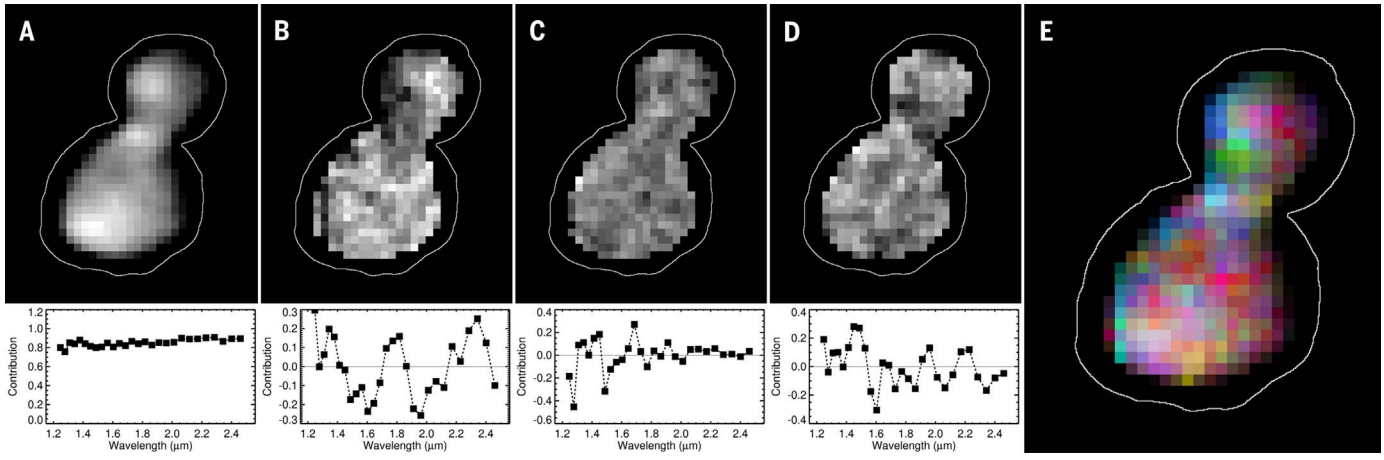


Fig. 5. Principal components analysis of the CA04 infrared spectral imaging data. (**A** to **E**) Same as Fig. 2, but for the LEISA observation in Fig. 4. Edge pixels have been removed; the white outlines show the full extent of Arrokoth.

using procedures developed earlier in the mission, accounting for the solid angle subtended by Arrokoth within the antenna beam (*15*, *17*, *36*, *37*). Accounting for the 414.4-km^2 cross section of Arrokoth and noise from instrument and background, we obtain a mean brightness temperature T_B, averaged across the night-side visible face of Arrokoth, of T_B = 29 ± 5 K, which is within the range of brightness temperatures estimated from an earlier analysis (*1*). To translate that brightness temperature to a kinetic temperature requires knowledge of the X-band emissivity of Arrokoth's surface, which is not known but most likely lies in the range 0.7 to 0.9 (*36*).

Thermophysical models were used to assess the implications of this T_B measurement (*15*). For each surface element of the three-dimensional (3D) shape model (*3*), we balanced radiative losses and thermal conduction against insolation (received sunlight) and also reradiation from other parts of Arrokoth's surface visible from that location. Accounting for self-shadowing and surface reradiation makes this modeling inherently global in scope (*38*). Subsurface thermal evolution was simulated with a 1D thermal diffusion prescription (*15*). For simplicity, we assume that Arrokoth's obliquity does not precess and that its orbit is circular, with a semimajor axis of 44.2 AU and period of 298 years. At this distance, the incident solar radiation flux $F_\odot$ is 0.7 W m^{-2}. Given Arrokoth's 99.3° obliquity (*2*), seasonal effects are strong. We determined the subsolar latitude along approximately 300 equally spaced temporal nodes over one orbital period (*15*). During the New Horizons flyby, the subsolar latitude was approximately −62°. At each orbital node, the daily averaged (15.9-hour period) solar insolation was calculated, accounting for self-shadowing. With these diurnally averaged insolation profiles, we determined the surface temperature on every element over the course of an orbit. We assumed that the subsurface thermal response is in the time-asymptotic limit, meaning there is no net gain or loss of thermal energy into or out of the interior over the course of one orbit (*39*). This assumption requires heat from radioactive decay inside Arrokoth to be negligible and requires the interior to have reached thermal equilibrium with the Sun over the course of the lifetime of the Solar System (see below).

We assume that the low–bond albedo [A_B = 0.06 (*1*, *3*)] surface of Arrokoth is characterized by a very low thermal inertia (Γ = 2.5 J m^{-2} s$^{-1/2}$ K^{-1}) typical of loosely consolidated granular material, as inferred from infrared observations of KBOs (*40*). The thermal inertia is given by $\Gamma \equiv \sqrt{k\rho C_p}$, where k is thermal conductivity, ρ is density, and C_p is specific heat at constant pressure. Arrokoth's bulk density must be at least 290 kg m^{-3} (*3*) and densities of small KBOs and comet nuclei tend to be higher than that, but generally less than 1000 kg m^{-3} (*41*, *42*). The density near the surface that matters for Arrokoth's thermal response is even more uncertain and could differ substantially from the bulk density. We assume a generic density of ρ = 500 kg m^{-3} (*3*) and that C_p = 350 J kg^{-1} K^{-1}. Under these assumptions, the corresponding thermal conductivity is 3.6 × 10^{-5} W m^{-1} K^{-1}, very low relative to values determined for surfaces in the inner Solar System. These values correspond to a seasonal thermal skin depth λ = 0.55 m, where $\lambda \equiv \sqrt{k\tau_s(\rho C_p 2\pi)^{-1}}$ and τ_s is the 298-year seasonal period (*43*). This value is similar to the electrical skin depth (*36*). The subsolar equilibrium temperature T_* = 58 K was obtained from $\varepsilon\sigma T_*^4 = (1 - A_B)F_\odot$, where ε is emissivity [commonly assumed to be 0.9; e.g., (*40*, *43*)], σ is the Stefan-Boltzmann constant (5.67 × 10^{-8} W m^{-2} K^{-4}), and $F_\odot$ is the solar flux at 44.2 AU. The thermal parameter $\Theta \equiv \Gamma\omega_0^{1/2}\sigma^{-1}T_*^{-3}$ characterizes the efficiency of the energy transport rate (per K) of thermal conduction across one seasonal skin depth relative to radiative losses (*43*), which is about $\Theta \equiv 0.02$ for Arrokoth. Such a small value of Θ indicates that the surface layers are highly insulating, leading to extreme variations in surface temperature over the course of a year. Winter surface temperatures are much lower than peak summer temperatures. The low conductivity might only pertain to a surficial layer. If deeper below the surface the texture is more compacted with greater granular contact, the conductivity would likely be higher. We can estimate the body's thermalization time scale by calculating the thermal wave propagation time ($t_{thermalization} \equiv 2\pi\rho C_p R^2 k^{-1}$) across a length scale R corresponding to the characteristic radius of the short axes of both lobes (~3.5 km). For values of k greater than 10^{-4} W m^{-1} K^{-1}, this time scale is less than the age of the Solar System, supporting our assumption of a time-asymptotic state.

Figure 7A shows the insolation averaged over an orbit from two viewing positions. The flattened shape and high obliquity lead to the equator receiving less energy on average (~0.1 W m^{-2}) than the poles (~0.2 W m^{-2}). Owing to self-shadowing, the neck region generally receives less energy than the equatorial zone (ranging from 0.06 to 0.08 W m^{-2}). Figure 7B shows the additional radiation received from thermal emission from other parts of Arrokoth itself, again averaged over an orbit. Our model indicates that the neck region is warmed by this trapping process, receiving about 0.025 to 0.04 W m^{-2} from self-reradiation, which partially offsets the effect of shadowing. Figure 7C shows the orbital average of the warming due to self-reradiation. The neck region experiences the greatest amount of relative warming, in the range of ~1 to 3 K. Maryland crater also receives enhanced thermal reradiation, but the relative warming in

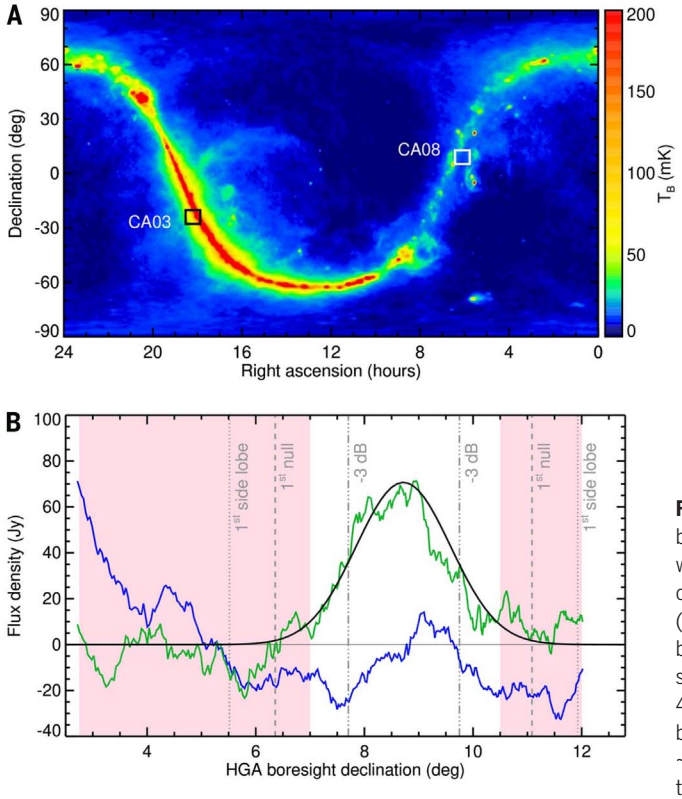


Fig. 6. Microwave radiometry of Arrokoth. (**A**) 7.1-GHz microwave sky background based on an all-sky radio map (*35*) sampled at REX's 4.2-cm wavelength and smoothed to REX's 1.2° beam width (*36*), indicating the locations of the CA03 and CA08 observations. [Reproduced from (*36*) with permission] (**B**) Observed flux during the CA08 scan in green, with the later background scan in blue. Shaded areas were used to calibrate the two observations for background subtraction. The black curve is a model response for a T_B = 30 K source with the 414-km^2 projected area of Arrokoth. (**C**) CA07 LORRI image (*3*) obtained 10 min before the mid-time of the REX scan, at nearly identical lighting geometry but a ~10° shift in viewing geometry, showing more of the lit crescent than was visible at the time of the CA08 REX observation.

that region is slight, about 0.5 K. Figure 7D shows the predicted observable surface temperature at the time of the New Horizons encounter: Typical temperatures are in the range of 50 to 57 K near the poles of each lobe, falling to ~40 K near the equator. Parts of the neck region that are not shaded at times of high subsolar latitude (–62° at the time of the encounter) stand out as having among the warmest surface temperatures during the encounter, as high as 60 K.

Figure 7E shows the predicted surface temperature at the time and viewing geometry of the CA08 REX observation, during which the sub-spacecraft point was latitude +44°, longitude 78°E. From this orientation, the model global averaged surface temperature is 16.1 K. Although the surface temperature within the bulk of the body's winter night side is predicted to be in the range of 12 to 14 K, the contribution to the average from the viewable part of the lit equatorial region (top of Fig. 7E; T ~ 40 to 55 K) raises the average temperature only slightly. The observed thermal emission seen by REX yields a much warmer mean brightness temperature of 29 K. If the model is correct, this discrepancy implies that the 4.2-cm radiation emerges primarily from the warmer subsurface, which is consistent with the expectation that the 4.2-cm thermal emission samples many wavelengths into the surface (*36*). Higher thermal inertias than we have assumed would also permit warmer winter surface temperatures (*15*).

Implications for formation

The distribution of orbits in the present-day Kuiper belt was strongly influenced by an outward migration of Neptune early in Solar System history, resulting from dynamical interaction between Neptune and the disk of planetesimals [e.g., (*44*)]. Neptune's migration stopped at 30 AU from the Sun, indicating a break in the distribution of planetesimals in the disk, beyond which there was insufficient mass to drive Neptune's migration further outward (*45*). Most KBOs that have been studied spectroscopically [e.g., (*46*, *47*)] are not CCKBOs; they originated in the denser planetesimal disk from inside 30 AU. Likewise, most comets that approach the Sun and Earth, and therefore can be studied in detail, likely do not sample the outer planetesimal disk from beyond 30 AU. Arrokoth may contain a record of conditions in the outer part of the nebula where it formed. Constraints include the evidence for methanol ice and the lack of evidence for water ice, which is unlike the high abundance of H_2O in many outer Solar System bodies and interstellar grains.

CCKBOs appear to have formed in the outer solar nebula through the gravitational collapse of pebble-size particles, concentrated aerodynamically (*13*, *48*). In this scenario, microscopic dust grains coagulate into larger particles (*49*, *50*). As particles approach pebble sizes, they decouple from the gas, causing them to spend most of their time near the cold disk midplane and allowing them to become concentrated in dense clumps that can gravitationally collapse into planetesimals (*51*, *52*). The (original) bulk composition of CCKBOs should reflect the makeup of the solids present in the midplane of the solar nebula at the time and location of their formation. It remains unclear, however, how long the dust coagulation phase lasted and/or how far pebbles were able to move radially inward [e.g., (*53*)] before they formed planetesimals.

When the solar nebula formed, the chemical composition of the ices present in the outer regions was set by a combination of inheritance from the parent molecular cloud and chemistry taking place during formation of the disk. In the midplane of the outer disk, the resulting CH_3OH/H_2O ratio on cold grain surfaces likely did not exceed a few percent (*54*). During the subsequent disk evolution, spatial variations in physical conditions such as temperature, density, and radiation environment, coupled with ongoing chemistry (*55*) and transport or mixing processes (*56*), result in gas-phase and ice compositions changing over time.

In regions where it was cold enough for highly volatile CO to freeze as ice onto grains (*57*), methanol could be formed through successive addition of hydrogen atoms to CO ice. Both interstellar and outer nebular environments are potential settings for this chemistry (*58*–*60*). Before the loss of nebular gas and dust, the midplane of the disk was shaded from direct sunlight and extremely cold, favoring condensation of CO in its outermost

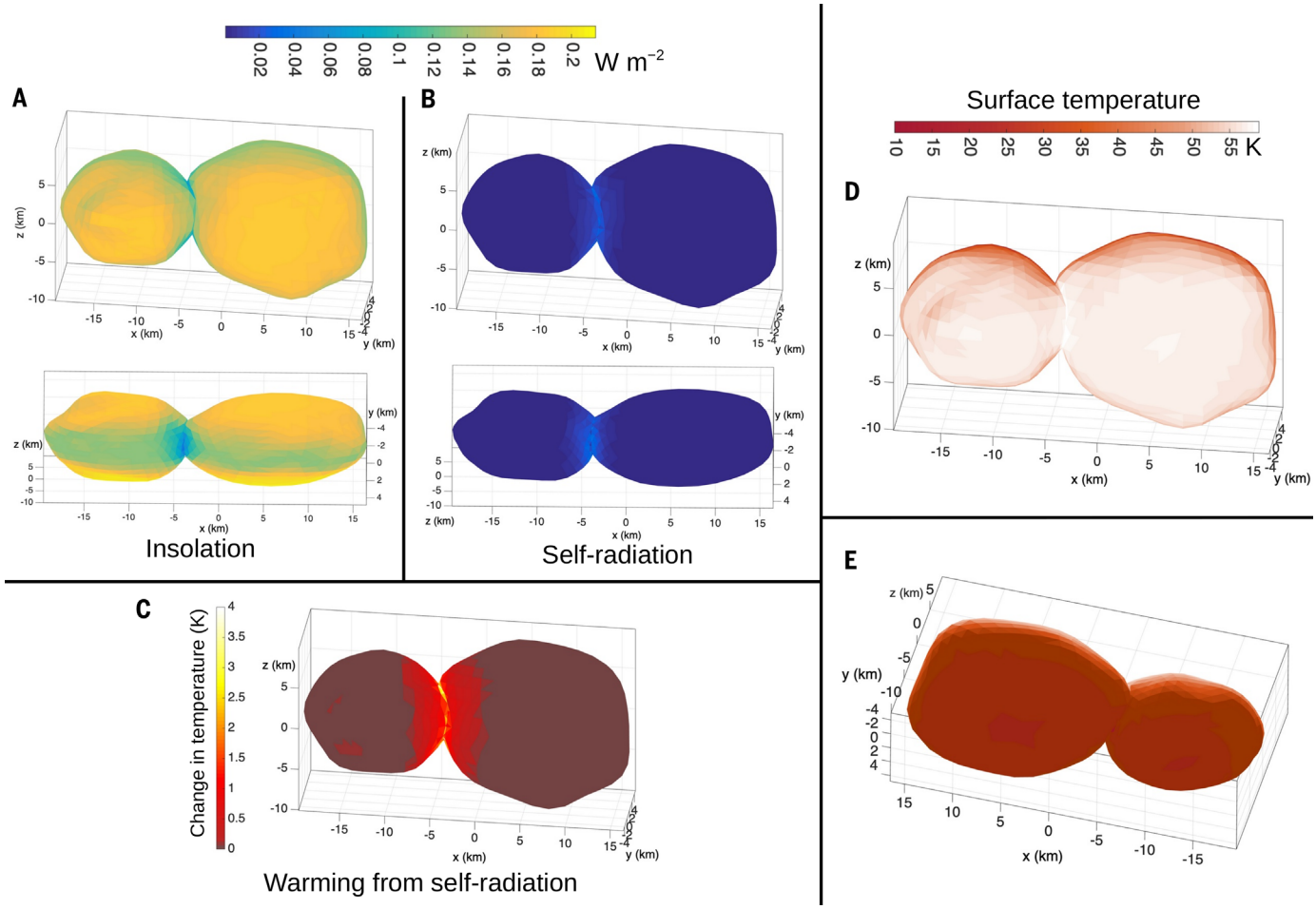


Fig. 7. Models of insolation and temperature across the surface of Arrokoth. (**A**) Insolation averaged over Arrokoth's orbit viewed from two different orientations. (**B**) Reradiation received from other portions of the shape on the same scale and from the same orientations as (A), also averaged over the orbit. The color scale applies to (A) and (B). (**C**) Seasonally averaged additional warming resulting from reradiation is shown as a temperature increase above that which would have been computed in the absence of self-radiation. (**D**) Summer day-side temperature distribution as seen from New Horizons during LORRI observation CA06 (*3*). (**E**) Winter night-side temperature distribution as seen from New Horizons during the REX CA08 scan. In all panels, the *x*, *y*, and *z* axes correspond to the principal axes of inertia with the origin at the center of mass (*13*).

portions. Simulations of protoplanetary disks indicate that methanol can be produced in this way (consuming CO) on time scales of ~1 million years at Arrokoth's distance from the Sun (*61*). Intermediate steps include formation of formaldehyde (H_2CO) and radicals (e.g., CH_3O). Radiolytic destruction of CH_3OH can produce H_2CO, but the band at 2.27 μm remains prominent even after irradiation (*62*).

Another potential CH_3OH formation mechanism involves radiolysis of mixed H_2O and CH_4 ices (*63*–*66*). Again, low temperatures consistent with the shaded midplane are required for CH_4 to be frozen onto grains, although CH_4 is not quite as volatile as CO. If such radiolytic production occurs with an excess of CH_4, the H_2O could be consumed, providing a possible explanation for the lack of evidence for H_2O at Arrokoth. Such a radiolytic process would also efficiently form simple hydrocarbons such as C_2H_4 and C_2H_6, which are known to be precursors of complex organic tholins [e.g., (*20*, *67*)].

Gas-phase methanol has been detected at low abundance in protoplanetary disks around nearby stars (*68*). This is consistent with methanol ice formation on grain surfaces, with a small fraction subsequently released to the gas through nonthermal desorption mechanisms [e.g., (*60*, *69*)]. This would also be consistent with the observation that many of these disks are also depleted in gaseous CO (*70*, *71*), requiring a combination of sequestration on pebbles and chemical processing (*61*, *72*, *73*).

Although H_2O was not detected on Arrokoth, it could be present but somehow masked or hidden from view, such as by materials produced through radiolysis or photolysis of CH_3OH ice and perhaps other undetected precursor materials. Preferential removal of H_2O ice from the uppermost surface by a process such as sputtering is another possibility, although it is unclear that H_2O should be more susceptible to such removal than CH_3OH is. Spectra of some larger KBOs also lack H_2O absorption features (*46*, *47*). H_2O ice absorption is considerably weaker in the spectrum of Arrokoth than seen on Pholus and (55638) 2002 VE_{95}, the two other objects with strong CH_3OH signatures, but those objects are much larger than Arrokoth (*74*, *75*). They likely formed in the closer, more densely populated planetesimal disk originally inside 30 AU, as did other, large KBOs where strong H_2O ice signatures have been detected spectroscopically. It is hard to envision a mechanism that preferentially masks or removes H_2O from the surface of Arrokoth but not the H_2O on these other objects. Their contrasting compositions suggest that the observed surface composition of these bodies is reflective of their bulk compositions, and that Arrokoth's composition is

distinct from those of planetesimals that formed closer to the Sun. A contrast in planetesimal composition driven by nebular chemistry enabled by CO and/or CH_4 frozen on grains may be connected to the transition at 30 AU that halted Neptune's outward migration at that distance.

Although regional variations in tholin and ice abundance could cause albedo, color, and spectral variations, the subtle variations that are seen at Arrokoth do not require compositional differences. Reflectance also depends on mechanical properties such as particle size distribution and degree of compaction (*31*). The merger of the two lobes (*13*) could have mechanically modified the material in the neck region. After the formation of Arrokoth from the nebula, low-speed impacts of residual debris could locally modify surface textures, which might account for some of the spots with slightly contrasting colors and albedos.

Surface and interior evolution

The surface features of comets are dominated by geologically rapid volatile loss and sublimation erosion, whereas the surfaces of larger asteroids are dominated by high-energy impacts. In contrast, Arrokoth and the CCKBOs are distinct in inhabiting an environment with very little energy input from interstellar, solar, and micrometeorite sources that require long time scales to modify the surface. Depending on the thermal parameters, surface temperatures range from as low as 10 to 20 K in winter to 50 to 60 K in summer, with the neck region getting at most a few degrees warmer because of self-radiation. Summer surface temperatures are warm enough to drive off volatiles such as CO, CH_4, and N_2, but are not warm enough to crystallize amorphous H_2O ice or to sublimate it. We therefore expect little thermally driven evolution of the surface, except early in Arrokoth's history when the volatile ices would have been lost soon after nebular dust cleared, allowing sunlight to illuminate the surface. Galactic and solar energetic photons and charged particles can break bonds and drive chemical reactions that produce refractory macromolecular tholins (*19*, *24*, *25*). Photolysis and radiolysis of solid methanol also produce formaldehyde, which may subsequently polymerize (*76*, *77*). Although formaldehyde polymer (paraformaldehyde) shows some spectral structure in the 2- to 2.5-μm region, it does not match the bands seen in Pholus (*33*) or Arrokoth. Macromolecular tholins seen on Arrokoth's surface could derive from three potential sources: (i) They could have originated in the pre-solar gas and dust cloud. (ii) They could arise from photolysis and radiolysis of hydrocarbons and associated nitrogen- and oxygen-bearing components in the nebula, especially where material is transported to regions near the surface of the nebula exposed to radiation from the forming Sun or other astrophysical sources (*56*). (iii) Radiolysis and photolysis of Arrokoth's surface components could produce tholins, as discussed above. All these sources likely contributed to Arrokoth's inventory of complex organics, with the products of the first two mechanisms being distributed throughout the body, whereas the products of the third mechanism should only occur at the surface. The flux of Solar System and interstellar micrometeorites at Arrokoth's location is highly uncertain (*78*, *79*), but such bombardment could produce up to several meters of mechanical erosion over the age of the Solar System (*80*, *81*). If this erosion operates faster than space weathering by energetic radiation, the visible surface could be representative of the deep interior. If it is slower, the surface should accumulate a lag deposit enriched in more refractory materials through loss or destruction of more volatile and fragile molecules.

It is not obvious from the encounter data whether a distinct surface veneer exists on Arrokoth. Albedo and color contrasts corresponding to ancient features such as the neck suggest that such contrasts are not quickly masked by a space weathering processes. If compositionally distinct interior material was exposed at the neck, it might weather differently and thus maintain a contrasting appearance, but the LEISA data show no evidence for a distinct composition in the neck region. The warmer thermal environment of the neck would be another potential reason for distinct evolution there, but the temperature difference is too small to produce outcomes that differ substantially from the rest of the surface. No obviously fresh craters expose distinct-looking interior material (with the possible exception of Maryland), and color trends do not appear to correspond to downslope transport, which is generally from equators to the poles of the lobes, and ultimately to the neck (*3*). Less-altered interior material should be preferentially exposed at high elevations, but we do not see obvious color differences in high-standing regions along the equators. Brighter material in the neck and in Maryland may accumulate in topographic lows, suggestive of textural rather than compositional contrasts. Among the CCKBO population, the diversity of colors coupled with similar colors of different sized components of binaries have been used to argue against the importance of size-dependent factors such as the balance between erosion and space weathering in altering their surface colors (*82*).

The evolution of Arrokoth's interior is shaped by energy inputs that are even smaller than at its surface. Subsequent to the loss of shading from nebular dust, insolation would have raised Arrokoth's equilibrium temperature. As that thermal wave slowly propagated inward, highly volatile species such as N_2, CO, and CH_4 would have become unstable, at least as condensed ices. Early outgassing of these species may have produced what appear to be collapse or outgassing pits at the boundaries of terrain units (*1*, *3*). Such features may be analogous to pits or sinkholes on comet 67P/Churyumov-Gerasimenko [e.g., (*83*, *84*)]. Localization of the pits to certain regions may arise from variable permeability of surface deposits that would favor volatiles escaping through weaker zones at unit boundaries. However, the equilibrium temperature in Arrokoth's interior would never have been high enough for amorphous ice to crystallize and expel its payload of trapped volatiles [e.g., (*85*, *86*)]. Apart from loss of these volatile species, Arrokoth's interior may have undergone little alteration or processing since accretion, and could thus preserve many characteristics of the original accretion such as layering [as observed on comets (*87*, *88*)], very high porosity, and an intimate mixture of nebular ices, organics, and silicate dust grains.

REFERENCES AND NOTES

1. S. A. Stern *et al.*, Initial results from the New Horizons exploration of 2014 MU_{69}, a small Kuiper Belt object. *Science* **364**, eaaw9771 (2019). doi: 10.1126/science.aaw9771; pmid: 31097641
2. S. B. Porter *et al.*, High-precision orbit fitting and uncertainty analysis of (486958) 2014 MU_{69}. *Astron. J.* **156**, 20 (2018). doi: 10.3847/1538-3881/aac2e1
3. J. R. Spencer *et al.*, The geology and geophysics of Kuiper Belt object (486958) Arrokoth. *Science* **367**, 10.1126/science.aay3999 (2020).
4. J.-M. Petit *et al.*, The Canada-France Ecliptic Plane Survey - full data release: The orbital structure of the Kuiper belt. *Astron. J.* **142**, 131 (2011). doi: 10.1088/0004-6256/142/4/131
5. H. F. Levison, A. Morbidelli, C. Van Laerhoven, R. Gomes, K. Tsiganis, Origin of the structure of the Kuiper belt during a dynamical instability in the orbits of Uranus and Neptune. *Icarus* **196**, 258–273 (2008). doi: 10.1016/j.icarus.2007.11.035
6. K. S. Noll, W. M. Grundy, D. C. Stephens, H. F. Levison, S. D. Kern, Evidence for two populations of classical transneptunian objects: The strong inclination dependence of classical binaries. *Icarus* **194**, 758–768 (2008). doi: 10.1016/j.icarus.2007.10.022
7. S. C. Tegler, W. Romanishin, Extremely red Kuiper-belt objects in near-circular orbits beyond 40 AU. *Nature* **407**, 979–981 (2000). doi: 10.1038/35039572; pmid: 11069171
8. R. E. Pike *et al.*, Col-OSSOS: Z-band photometry reveals three distinct TNO surface types. *Astron. J.* **154**, 101 (2017). doi: 10.3847/1538-3881/aa83b1
9. H. F. Levison, S. A. Stern, On the size dependence of the inclination distribution of the main Kuiper Belt. *Astron. J.* **121**, 1730–1735 (2001). doi: 10.1086/319420
10. M. E. Schwamb, M. E. Brown, W. C. Fraser, The small numbers of large Kuiper belt objects. *Astron. J.* **147**, 2 (2014). doi: 10.1088/0004-6256/147/1/2
11. M. J. Brucker *et al.*, High albedos of low inclination Classical Kuiper belt objects. *Icarus* **201**, 284–294 (2009). doi: 10.1016/j.icarus.2008.12.040
12. E. Vilenius *et al.*, "TNOs are Cool": A survey of the trans-neptunian region X. Analysis of classical Kuiper belt objects from Herschel and Spitzer observations. *Astron. Astrophys.* **564**, A35 (2014). doi: 10.1051/0004-6361/201322416
13. W. B. McKinnon *et al.*, The solar nebula origin of (486958) Arrokoth, a primordial contact binary in the Kuiper Belt. *Science* **367**, 10.1126/science.aay6620 (2020).
14. D. C. Reuter *et al.*, Ralph: A visible/infrared imager for the New Horizons Pluto/Kuiper belt mission. *Space Sci. Rev.* **140**, 129–154 (2008). doi: 10.1007/s11214-008-9375-7
15. See supplementary materials.
16. A. F. Cheng *et al.*, Long-Range Reconnaissance Imager on New Horizons. *Space Sci. Rev.* **140**, 189–215 (2008). doi: 10.1007/s11214-007-9271-6

17. G. L. Tyler *et al*., The New Horizons radio science experiment (REX). *Space Sci. Rev.* **140**, 217–259 (2008). doi: 10.1007/s11214-007-9302-3
18. O. R. Hainaut, A. C. Delsanti, Colors of minor bodies in the outer Solar System: A statistical analysis. *Astron. Astrophys.* **389**, 641–664 (2002). doi: 10.1051/0004-6361:20020431
19. D. P. Cruikshank, H. Imanaka, C. M. Dalle Ore, Tholins as coloring agents on outer Solar System bodies. *Adv. Space Res.* **36**, 178–183 (2005). doi: 10.1016/j.asr.2005.07.026
20. B. N. Khare *et al*., Optical constants of organic tholins produced in a simulated Titanian atmosphere: From soft X-ray to microwave frequencies. *Icarus* **60**, 127–134 (1984). doi: 10.1016/0019-1035(84)90142-8
21. B. N. Khare *et al*., Production and optical constants of ice tholin from charged particle irradiation of (1:6) C_2H_6/H_2O at 77 K. *Icarus* **103**, 290–300 (1993). doi: 10.1006/icar.1993.1071
22. B. N. Khare *et al*., Optical constants of Triton tholin: Prelimary results. *Bull. Am. Astron. Soc.* **26**, 1176–1177 (1994).
23. H. Imanaka *et al*., Laboratory experiments of Titan tholin formed in cold plasma at various pressures: Implications for nitrogen-containing polycyclic aromatic compounds in Titan haze. *Icarus* **168**, 344–366 (2004). doi: 10.1016/j.icarus.2003.12.014
24. C. K. Materese *et al*., Ice chemistry on outer Solar System bodies: Carboxylic acids, nitriles, and urea detected in refractory residues produced from the UV photolysis of N_2:CH_4:CO-containing ices. *Astrophys. J.* **788**, 111 (2014). doi: 10.1088/0004-637X/788/2/111
25. C. K. Materese, D. P. Cruikshank, S. A. Sandford, H. Imanaka, M. Nuevo, Ice chemistry on outer Solar System bodies: Electron radiolysis of N_2-, CH_4-, and CO-containing ices. *Astrophys. J.* **812**, 150 (2015). doi: 10.1088/0004-637X/812/2/150
26. N. Peixinho, A. Delsanti, A. Guilbert-Lepoutre, R. Gafeira, P. Lacerda, The bimodal colors of Centaurs and small Kuiper belt objects. *Astron. Astrophys.* **546**, A86 (2012). doi: 10.1051/0004-6361/201219057
27. F. E. DeMeo, B. Carry, The taxonomic distribution of asteroids from multi-filter all-sky photometric surveys. *Icarus* **226**, 723–741 (2013). doi: 10.1016/j.icarus.2013.06.027
28. G. M. Szabó, Z. Ivezić, M. Jurić, R. Lupton, The properties of jovian Trojan asteroids listed in SDSS Moving Object Catalogue 3. *Mon. Not. R. Astron. Soc.* **377**, 1393–1406 (2007). doi: 10.1111/j.1365-2966.2007.11687.x
29. W. C. Fraser, M. E. Brown, The Hubble Wide Field Camera 3 test of surfaces in the outer Solar System: The compositional classes of the Kuiper belt. *Astrophys. J.* **749**, 33 (2012). doi: 10.1088/0004-637X/749/1/33
30. B. Schmitt *et al*., Physical state and distribution of materials at the surface of Pluto from New Horizons LEISA imaging spectrometer. *Icarus* **287**, 229–260 (2017). doi: 10.1016/j.icarus.2016.12.025
31. B. Hapke, *Theory of Reflectance and Emittance Spectroscopy* (Cambridge Univ. Press, ed. 2, 2012).
32. S. A. Sandford, L. J. Allamandola, Condensation and vaporization studies of CH_3OH and NH_3 ices: Major implications for astrochemistry. *Astrophys. J.* **417**, 815–825 (1993). doi: 10.1086/173362; pmid: 11540092
33. D. P. Cruikshank *et al*., The composition of Centaur 5145 Pholus. *Icarus* **135**, 389–407 (1998). doi: 10.1006/icar.1998.5997
34. M. A. Barucci, F. Merlin, E. Dotto, A. Doressoundiram, C. de Bergh, TNO surface ices: Observations of the TNO 55638 (2002 VE_{95}) and analysis of the population's spectral properties. *Astron. Astrophys.* **455**, 725–730 (2006). doi: 10.1051/0004-6361:20064951
35. H. Zheng *et al*., An improved model of diffuse galactic radio emission from 10 MHz to 5 THz. *Mon. Not. R. Astron. Soc.* **464**, 3486–3497 (2017). doi: 10.1093/mnras/stw2525
36. M. K. Bird *et al*., Radio thermal emission from Pluto and Charon during the New Horizons encounter. *Icarus* **322**, 192–209 (2019). doi: 10.1016/j.icarus.2019.01.004
37. I. R. Linscott *et al*., Radiometer Calibration at 4.2 cm on New Horizons. Stanford Radioscience Report No. 17-06-0 (2017); https://sbn.astro.umd.edu/holdings/nh-j-rex-2-jupiter-v2.0/document/nh_rex_radiometer_calib_v4p7.pdf.
38. D. Vokrouhlický, W. F. Bottke, S. R. Chesley, D. J. Scheeres, T. S. Statler, in *Asteroids IV*, P. Michel, F. E. DeMeo, W. F. Bottke, Eds. (Univ. of Arizona Press, 2015), pp. 509–531.
39. O. L. White, O. M. Umurhan, J. W. Moore, A. D. Howard, Modeling of ice pinnacle formation on Callisto. *J. Geophys. Res.* **121**, 21–45 (2016). doi: 10.1002/2015JE004846
40. E. Lellouch *et al*., "TNOs are Cool": A survey of the trans-neptunian region IX. Thermal properties of Kuiper belt objects and Centaurs from combined Herschel and Spitzer observations. *Astron. Astrophys.* **557**, A60 (2013). doi: 10.1051/0004-6361/201322047
41. M. Pätzold *et al*., A homogeneous nucleus for comet 67P/Churyumov-Gerasimenko from its gravity field. *Nature* **530**, 63–65 (2016). doi: 10.1038/nature16535; pmid: 26842054
42. C. J. Bierson, F. Nimmo, Using the density of Kuiper belt objects to constrain their composition and formation history. *Icarus* **326**, 10–17 (2019). doi: 10.1016/j.icarus.2019.01.027
43. J. R. Spencer, A rough-surface thermophysical model for airless planets. *Icarus* **83**, 27–38 (1990). doi: 10.1016/0019-1035(90)90004-S
44. R. S. Gomes, A. Morbidelli, H. F. Levison, Planetary migration in a planetesimal disk: Why did Neptune stop at 30 au. *Icarus* **170**, 492–507 (2004). doi: 10.1016/j.icarus.2004.03.011
45. A. Shannon, Y. Wu, Y. Lithwick, Forming the cold classical Kuiper belt in a light disk. *Astrophys. J.* **818**, 175 (2016). doi: 10.3847/0004-637X/818/2/175
46. K. M. Barkume, M. E. Brown, E. L. Schaller, Near-infrared spectra of Centaurs and Kuiper belt objects. *Astron. J.* **135**, 55–67 (2008). doi: 10.1088/0004-6256/135/1/55
47. M. A. Barucci *et al*., New insights on ices in Centaur and transneptunian populations. *Icarus* **214**, 297–307 (2011). doi: 10.1016/j.icarus.2011.04.019
48. D. Nesvorný, R. Li, A. Youdin, J. B. Simon, W. M. Grundy, Trans-Neptunian binaries as evidence for planetesimal formation by the streaming instability. *Nat. Astron.* **3**, 808–812 (2019). doi: 10.1016/j.icarus.2011.04.019
49. C. Dominik, A. G. G. M. Tielens, The physics of dust coagulation and the structure of dust aggregates in space. *Astrophys. J.* **480**, 647–673 (1997). doi: 10.1086/303996
50. J. Blum, G. Wurm, The growth mechanisms of macroscopic bodies in protoplanetary disks. *Annu. Rev. Astron. Astrophys.* **46**, 21–56 (2008). doi: 10.1146/annurev.astro.46.060407.145152
51. A. N. Youdin, J. Goodman, Streaming instabilities in protoplanetary disks. *Astrophys. J.* **620**, 459–469 (2005). doi: 10.1086/426895
52. A. Johansen *et al*., Rapid planetesimal formation in turbulent circumstellar disks. *Nature* **448**, 1022–1025 (2007). doi: 10.1038/nature06086; pmid: 17728751
53. S. J. Weidenschilling, Aerodynamics of solid bodies in the solar nebula. *Mon. Not. R. Astron. Soc.* **180**, 57–70 (1977). doi: 10.1093/mnras/180.2.57
54. M. N. Drozdovskaya *et al*., Cometary ices in forming protoplanetary disc midplanes. *Mon. Not. R. Astron. Soc.* **462**, 977–993 (2016). doi: 10.1093/mnras/stw1632
55. T. Henning, D. Semenov, Chemistry in protoplanetary disks. *Chem. Rev.* **113**, 9016–9042 (2013). doi: 10.1021/cr400128p; pmid: 24191756
56. F. J. Ciesla, S. A. Sandford, Organic synthesis via irradiation and warming of ice grains in the solar nebula. *Science* **336**, 452–454 (2012). doi: 10.1126/science.1217291; pmid: 22461502
57. K. I. Öberg, R. Murray-Clay, E. A. Bergin, The effect of snowlines on C/O in planetary atmospheres. *Astrophys. J.* **743**, L16 (2011). doi: 10.1088/2041-8205/743/1/L16
58. A. G. G. M. Tielens, W. Hagen, Model calculations of the molecular composition of interstellar grain mantles. *Astron. Astrophys.* **114**, 245–260 (1982).
59. H. M. Cuppen, E. F. van Dishoeck, E. Herbst, A. G. G. M. Tielens, Microscopic simulation of methanol and formaldehyde ice formation in cold dense cores. *Astron. Astrophys.* **508**, 275–287 (2009). doi: 10.1051/0004-6361/200913119
60. M. Ruaud, U. Gorti, A three-phase approach to grain surface chemistry in protoplanetary disks: Gas, ice surfaces, and ice mantles of dust grains. *Astrophys. J.* **885**, 146 (2019). doi: 10.3847/1538-4357/ab4996
61. A. D. Bosman, C. Walsh, E. F. van Dishoeck, CO destruction in protoplanetary disk midplanes: Inside versus outside the CO snow surface. *Astron. Astrophys.* **618**, A182 (2018). doi: 10.1051/0004-6361/201833497
62. R. Brunetto, G. A. Baratta, M. Domingo, G. Strazzulla, Reflectance and transmittance spectra (2.2-2.4 μm) of ion irradiated methanol. *Icarus* **175**, 226–232 (2005). doi: 10.1016/j.icarus.2004.10.011
63. M. H. Moore, R. L. Hudson, Infrared study of ion-irradiated water-ice mixtures with hydrocarbons relevant to comets. *Icarus* **135**, 518–527 (1998). doi: 10.1006/icar.1998.5996
64. A. Wada, N. Mochizuki, K. Hiraoka, Methanol formation from electron-irradiated mixed H_2O/CH_4 ice at 10 K. *Astrophys. J.* **644**, 300–306 (2006). doi: 10.1086/503380
65. R. Hodyss, P. V. Johnson, J. V. Stern, J. D. Goguen, I. Kanik, Photochemistry of methane–water ices. *Icarus* **200**, 338–342 (2009). doi: 10.1016/j.icarus.2008.10.024
66. M. P. Pearce *et al*., Formation of methanol from methane and water in an electrical discharge. *Phys. Chem. Chem. Phys.* **14**, 3444–3449 (2012). doi: 10.1039/c2cp22135g; pmid: 22307542
67. G. D. McDonald *et al*., Production and chemical analysis of cometary ice tholins. *Icarus* **122**, 107–117 (1996). doi: 10.1006/icar.1996.0112
68. C. Walsh *et al*., First detection of gas-phase methanol in a protoplanetary disk. *Astrophys. J.* **823**, L10 (2016). doi: 10.3847/2041-8205/823/1/L10
69. C. Walsh, S. Vissapragada, H. McGee, in *Astrochemistry VII - Through the Cosmos from Galaxies to Planets, Proceedings IAU Symposium No. 332*, M. Cunningham, T. Millar, Y. Aikawa, Eds. (International Astronomical Union, 2018), pp. 395–402.
70. E. A. Bergin *et al*., An old disk still capable of forming a planetary system. *Nature* **493**, 644–646 (2013). doi: 10.1038/nature11805; pmid: 23364742
71. M. K. McClure *et al*., Mass measurements in protoplanetary disks from hydrogen deuteride. *Astrophys. J.* **831**, 167 (2016). doi: 10.3847/0004-637X/831/2/167
72. S. Krijt, K. R. Schwarz, E. A. Bergin, F. J. Ciesla, Transport of CO in protoplanetary disks: Consequences of pebble formation, settling, and radial drift. *Astrophys. J.* **864**, 78 (2018). doi: 10.3847/1538-4357/aad69b
73. K. Zhang, E. A. Bergin, K. Schwarz, S. Krijt, F. Ciesla, Systematic variations of CO gas abundance with radius in gas-rich protoplanetary disks. *Astrophys. J.* **883**, 98 (2019). doi: 10.3847/1538-4357/ab38b9
74. S. C. Tegler *et al*., The period of rotation, shape, density, and homogeneous surface color of the Centaur 5145 Pholus. *Icarus* **175**, 390–396 (2005). doi: 10.1016/j.icarus.2004.12.011
75. M. A. Barucci *et al*., The extra red plutino (55638) 2002 VE_{95}. *Astron. Astrophys.* **539**, A152 (2012). doi: 10.1051/0004-6361/201118505
76. M. P. Bernstein, S. A. Sandford, L. J. Allamandola, S. Chang, M. A. Scharberg, Organic compounds produced by photolysis of realistic interstellar and cometary ice analogs containing methanol. *Astrophys. J.* **454**, 327–344 (1995). doi: 10.1086/176485
77. T. Butscher *et al*., Radical-assisted polymerization in interstellar ice analogues: Formyl radical and polyoxymethylene. *Mon. Not. R. Astron. Soc.* **486**, 1953–1963 (2019). doi: 10.1093/mnras/stz879
78. M. Piquette *et al*., Student Dust Counter: Status report at 38 au. *Icarus* **321**, 116–125 (2019). doi: 10.1016/j.icarus.2018.11.012
79. A. R. Poppe *et al*., Constraining the Solar System's debris disk with in situ New Horizons measurements from the Edgeworth-Kuiper belt. *Astrophys. J.* **881**, L12 (2019). doi: 10.3847/2041-8213/ab322a
80. S. A. Stern, ISM-induced erosion and gas-dynamical drag in the Oort cloud. *Icarus* **84**, 447–466 (1990). doi: 10.1016/0019-1035(90)90048-E
81. S. A. Stern, The evolution of comets in the Oort cloud and Kuiper belt. *Nature* **424**, 639–642 (2003). doi: 10.1038/nature01725; pmid: 12904784
82. S. D. Benecchi *et al*., The correlated colors of transneptunian binaries. *Icarus* **200**, 292–303 (2009). doi: 10.1016/j.icarus.2008.10.025
83. O. Mousis *et al*., Pits formation from volatile outgassing on 67P/Churyumov-Gerasimenko. *Astrophys. J.* **814**, L5 (2015). doi: 10.1088/2041-8205/814/1/L5
84. J. B. Vincent *et al*., Large heterogeneities in comet 67P as revealed by active pits from sinkhole collapse. *Nature* **523**, 63–66 (2015). doi: 10.1038/nature14564; pmid: 26135448
85. P. Ayotte *et al*., Effect of porosity on the adsorption, desorption, trapping, and release of volatile gases by amorphous solid water. *J. Geophys. Res.* **106**, 33387–33392 (2001). doi: 10.1029/2000JE001362
86. A. Bar-Nun, G. Notesco, T. Owen, Trapping of N_2, CO and Ar in amorphous ice: Application to comets. *Icarus* **190**, 655–659 (2007). doi: 10.1016/j.icarus.2007.03.021
87. M. Massironi *et al*., Two independent and primitive envelopes of the bilobate nucleus of comet 67P. *Nature* **526**, 402–405 (2015). doi: 10.1038/nature15511; pmid: 26416730
88. L. Penasa *et al*., A three-dimensional modelling of the layered structure of comet 67P/Churyumov-Gerasimenko. *Mon. Not. R. Astron. Soc.* **469** (suppl. 2), S741–S754 (2017). doi: 10.1093/mnras/stx2899

ACKNOWLEDGMENTS

We gratefully thank the many hundreds of people who worked together on the New Horizons team. Their hard work enabled the encounter. **Funding:** Supported by NASA's New Horizons project via contracts NASW-02008 and NAS5-97271/TaskOrder30; the National Research Council of Canada (J.J.K.); a Hubble Fellowship, program number HST-HF2-51394.002 provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy Inc. under NASA contract NAS5-26555 (S.K.); the NASA Astrobiology Institute under Cooperative Agreement Notice NNH13ZDA017C issued through the Science Mission Directorate (M.R.); the French Centre National d'Etudes Spatiales (CNES) (B.S., L.G., and E.Q.); and NASA Astrophysics Theory Program Grant NNX17AK59G (O.M.U.). **Author contributions:** W.M.G. led authorship of the manuscript. W.M.G., C.J.A.H., C.B.O., A.H.P., and S.P. analyzed the MVIC color data and wrote that section. W.M.G., A.H.P., S.P., J.C.C., and D.P.C. analyzed the LEISA spectral data and wrote that section. M.K.B. and I.R.L. analyzed the REX data and wrote that section. O.M.U. did the thermal modeling and wrote that section, with input from W.M.G., C.J.A.H., and L.A.Y. W.M.G., D.P.C., D.T.B., S.K., and M.R. wrote the implications section. C.M.D.O., J.J.K., J.T.K., Y.J.P., S.B.P., F.S., J.R.S., S.A.S., A.J.V., and H.A.W. contributed to data analysis and development of scientific ideas. All authors played roles in experimental design and/or provided feedback and insights on drafts of the manuscript. **Competing interests:** We declare no competing interests. **Data and materials availability:** All images, spacecraft data, and the shape model used in this paper are available at https://doi.org/10.6084/m9.figshare.c.4819110 and at https://doi.org/10.6084/m9.figshare.11485443. The three Charon observations used in flat-fielding the LEISA data are available at https://pds-smallbodies.astro.umd.edu/holdings/nh-p-leisa-3-pluto-v3.0/data/20150714_029914/lsb_0299146219_0x54b_sci.fit, https://pds-smallbodies.astro.umd.edu/holdings/nh-p-leisa-3-pluto-v3.0/data/20150714_029917/lsb_0299171308_0x53c_sci.fit, and https://pds-smallbodies.astro.umd.edu/holdings/nh-p-leisa-3-pluto-v3.0/data/20150714_029917/lsb_0299175509_0x53c_sci.fit. Additional fully calibrated New Horizons data and higher-order data products will be released by the NASA Planetary Data System at https://pds-smallbodies.astro.umd.edu/data_sb/missions/newhorizons/index.shtml in a series of stages in 2020 and 2021, due to the time required to fully downlink and calibrate the data.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/eaay3705/suppl/DC1
Materials and Methods
Fig. S1
Data S1
References (*89*–*98*)

12 June 2019; accepted 22 January 2020
Published online 13 February 2020
10.1126/science.aay3705

RESEARCH ARTICLE SUMMARY

OUTER SOLAR SYSTEM

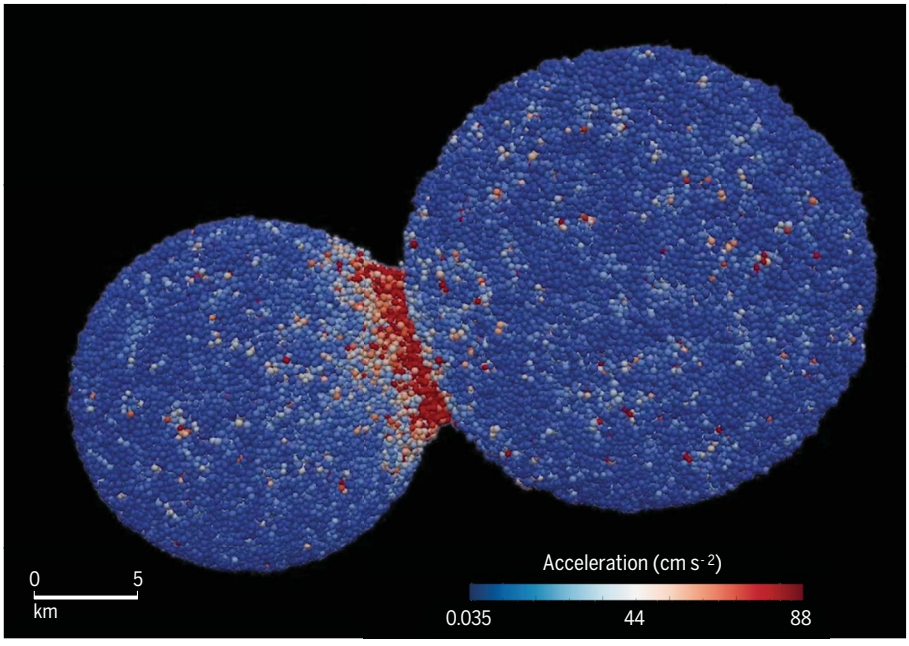
The solar nebula origin of (486958) Arrokoth, a primordial contact binary in the Kuiper Belt

W. B. McKinnon*, D. C. Richardson, J. C. Marohnic, J. T. Keane, W. M. Grundy, D. P. Hamilton, D. Nesvorný, O. M. Umurhan, T. R. Lauer, K. N. Singer, S. A. Stern, H. A. Weaver, J. R. Spencer, M. W. Buie, J. M. Moore, J. J. Kavelaars, C. M. Lisse, X. Mao, A. H. Parker, S. B. Porter, M. R. Showalter, C. B. Olkin, D. P. Cruikshank, H. A. Elliott, G. R. Gladstone, J. Wm. Parker, A. J. Verbiscer, L. A. Young, the New Horizons Science Team

INTRODUCTION: The close flyby of the Kuiper Belt object (486958) Arrokoth (formerly 2014 MU_{69}) by NASA's New Horizons spacecraft revealed details of the body's structure, geology, and composition. Arrokoth is a member of the cold classical component of the Kuiper Belt, a population of dwarf planets and smaller bodies thought to be only modestly dynamically or collisionally disturbed, unlike the asteroids of the inner Solar System, comets, or other groups of Kuiper Belt objects. Data from this flyby provides the opportunity to observe the results of primordial planetesimal accretion, largely unobscured by later geological or dynamical processes.

RATIONALE: Planetesimal formation is an unsolved problem in planetary science. Many mechanisms have been proposed in which small solid particles (dust and pebbles) agglomerate into planetesimals and ultimately into planets. The flyby of Arrokoth provides data that constrain planetesimal formation theories and allow us to construct models of Arrokoth's specific physical characteristics. The accretion processes that operated in the cold classical region of the Kuiper Belt during the formation of the Solar System are expected to have also occurred elsewhere in the protosolar nebula.

Arrokoth is a contact binary about 35 km long composed of two unequally sized lobes. Each lobe is flattened or lenticular in shape, and the planes of flattening of both (determined from their principal axes) are closely aligned, to within 5°. The smaller lobe is slightly oblong, with its long axis pointing down the long axis of the binary as a whole (to within 5°). The surface and overall structure of Arrokoth do not display any obvious signs of catastrophic or subcatastrophic collision, and the join or neck between the two lobes is narrow. Each lobe is compositionally similar to within the precision of spectral measurements.



Simulated maximum accelerations experienced by particles during low-velocity impact of two spheres, approximating the scale of the two lobes of Arrokoth. Spheres are modeled as granular aggregates with bulk densities of 500 kg m^{-3} and an impact speed of 2.9 m s^{-1} at a tangent angle of 80°; such gentle collisional conditions are necessary to preserve Arrokoth's overall undamaged shape. Extreme reds and blues correspond to the greatest and least maximum accelerations experienced, respectively. The maximum disturbance is concentrated in the narrow contact area, or neck, between the two bodies.

RESULTS: We show that stresses in the neck region today are compatible with the structural integrity of Arrokoth for densities (several 100 kg m^{-3}) and material strengths (a few kilopascals) similar to those observed in comets, but at mass scales ~1000 times the mass of typical cometary nuclei. We performed numerical simulations of collisions between two bodies on the scale of the two lobes of Arrokoth, assuming those density and strength parameters. We found that impacts at or greater than their mutual escape speed (a few meters per second) would have been highly damaging. The close geometric alignment of the lobes is highly unlikely to the be the result of a chance collision alone but can be readily understood as the result of tidal evolution of a tight, co-orbiting binary. This requires a mechanism to extract angular momentum from the binary orbit, causing the orbit to shrink, and the two components to gently merge.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aay6620

Numerical models show that overdense concentrations of particles in the protosolar gas nebula can become gravitationally unstable and collapse to form planetesimals. The angular momentum in the simulated pebble clouds is high enough that formation of co-orbiting binaries is efficient and with binary characteristics that are a good match to binaries observed in the Kuiper Belt today. We examined a range of mechanisms to extract or transfer angular momentum from a co-orbiting binary and drive an ultimate merger, including mutual tides, tidal effects of the Sun (Kozai-Lidov oscillations), collisions with smaller Kuiper Belt objects, the ejection of third bodies, asymmetric radiation forces, and gas drag. We found that for bodies the size of Arrokoth, gas drag may be most effective in this merger process over the lifetime of the protosolar nebula.

CONCLUSION: We show that models of Arrokoth's formation and evolution support accretion of the binary through the gravitational collapse of an overdense pebble cloud in the presence of protosolar nebular gas, either as a contact binary initially or as a co-orbiting binary that later inspiraled and gently merged. Similar accretional processes and binary planetesimal formation likely occurred throughout the early Solar System. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: mckinnon@wustl.edu
Cite this article as W. B. McKinnon *et al.*, *Science* **367**, eaay6620 (2019). DOI: 10.1126/science.aay6620

RESEARCH ARTICLE

OUTER SOLAR SYSTEM

The solar nebula origin of (486958) Arrokoth, a primordial contact binary in the Kuiper Belt

W. B. McKinnon[1]*, D. C. Richardson[2], J. C. Marohnic[2], J. T. Keane[3], W. M. Grundy[4,5], D. P. Hamilton[2], D. Nesvorný[6], O. M. Umurhan[7,8], T. R. Lauer[9], K. N. Singer[6], S. A. Stern[6], H. A. Weaver[10], J. R. Spencer[11], M. W. Buie[6], J. M. Moore[7], J. J. Kavelaars[11], C. M. Lisse[10], X. Mao[1], A. H. Parker[6], S. B. Porter[6], M. R. Showalter[8], C. B. Olkin[6], D. P. Cruikshank[7], H. A. Elliott[12,13], G. R. Gladstone[12], J. Wm. Parker[6], A. J. Verbiscer[14], L. A. Young[6], the New Horizons Science Team†

The New Horizons spacecraft's encounter with the cold classical Kuiper Belt object (486958) Arrokoth (provisional designation 2014 MU_{69}) revealed a contact-binary planetesimal. We investigated how Arrokoth formed and found that it is the product of a gentle, low-speed merger in the early Solar System. Its two lenticular lobes suggest low-velocity accumulation of numerous smaller planetesimals within a gravitationally collapsing cloud of solid particles. The geometric alignment of the lobes indicates that they were a co-orbiting binary that experienced angular momentum loss and subsequent merger, possibly because of dynamical friction and collisions within the cloud or later gas drag. Arrokoth's contact-binary shape was preserved by the benign dynamical and collisional environment of the cold classical Kuiper Belt and therefore informs the accretion processes that operated in the early Solar System.

After its encounter with Pluto in 2015 (*1*), the New Horizons spacecraft continued further into the Kuiper Belt (*2*). This included a flyby of (486958) Arrokoth (provisional designation 2014 MU_{69}, also informally known as Ultima Thule), which had been discovered in a dedicated Hubble Space Telescope campaign (*3*). Arrokoth's orbit has a semimajor axis $a_{\odot}$ = 44.2 astronomical units (AU), eccentricity e = 0.037, and inclination i = 2.54°, making it a member of the cold classical Kuiper Belt (CCKB), a reservoir of mainly small bodies on dynamically cold orbits—those with low-to-moderate e and low i (typically $i < 5°$)—in the outer Solar System (*4*). CCKB objects have a steeper size-frequency distribution, higher binary fraction, higher albedos, and redder optical colors than those of the dynamically hot and Neptune-resonant populations of the Kuiper Belt, implying a distinct formation mechanism and/or evolutionary history (*4*). CCKB objects are thought to have formed in place and remained largely undisturbed by the migration of the Solar System's giant planets (*4*–*6*), making them unperturbed remnants of the original protoplanetary disk.

The encounter showed that Arrokoth is a bilobed object, consisting of two discrete, quasi-ellipsoidal lobes (equivalent spherical diameters of 15.9 and 12.9 km, respectively) joined at a narrow contact area or "neck" (Fig. 1) (*7*, *8*). We interpret this geometric, cojoined object as a contact binary: two formerly separate objects that have gravitated toward each other until they touch. The larger lobe (hereafter LL) is more oblate than the smaller lobe (hereafter SL) (*8*). Arrokoth rotates with a 15.92-hour period at an obliquity of 99° (the angle between its rotation axis and heliocentric orbital plane). The short axes of both lobes are aligned to within a few degrees of each other and with the spin axis of the body as a whole (*8*). The average visible and near-infrared colors of both lobes are indistinguishable from each other (*9*). Near-infrared spectral absorptions on both lobes indicate the presence of methanol ice—a common, relatively (for an ice) thermally stable component of cometary bodies and extrasolar protoplanetary disks (*10*). The very red optical colors of both lobes are similar to that of other CCKB objects (*9*) and consistent with space weathering of simple ices to produce organic compounds, although other sources of reddening are also possible [such as iron and sulfur compounds (*9*)]. LL and SL both appear to be intact, or at least little disturbed, with no obvious morphological signs of a violent or energetic merger (*7*, *8*).

We examined the implications of these findings for the planetesimal formation process within the Kuiper Belt, which might be broadly applicable throughout the primordial Solar System. We focused on binary formation in the outer Solar System, which appears to have been common in the Kuiper Belt, on the basis of the abundance of binaries detected there in telescopic surveys (*11*, *12*). A related issue is the formation of the Kuiper Belt itself, its dynamical components—including the CCKB subpopulation—and the relationship between Kuiper Belt objects (KBOs) and short period comets (*4*). Many cometary nuclei are bilobate, but because cometary surfaces and shapes have been strongly affected by solar heating—causing sublimation, mass loss, and splitting—and the disruptive effects of close planetary encounters, it is not clear whether comets' bilobate shapes are a primordial characteristic or acquired during later evolution (*13*–*16*).

The CCKB

Most of the bodies in the Kuiper Belt are hypothesized to have been scattered and dynamically emplaced as Neptune slowly migrated outward through a massive [~15 to 30 Earth mass ($M_{\oplus}$)] planetesimal disk that extended from ~20 to 30 AU, outside the (then) compact orbits of the giant planets (*4*). CCKB objects are part of the nonresonant classical Kuiper Belt located farther out, today between 42 and 47 AU. CCKB objects have low dynamical excitation and physical properties distinct from the rest of the belt so are thought to have accreted in situ or in nearby orbits (*17*–*19*). The surface density of planetesimals that built the CCKB objects, in a disk that must have extended well beyond 30 AU, was insufficient for Neptune to continue its migration past that point (*4*, *20*).

The gravitational instability (GI) accretion mechanism posits that locally concentrated, gravitationally bound clouds of small (millimeter to decimeter) solid particles (the latter termed "pebbles") form in either the thick midplane of the protosolar nebular disk or in over-dense regions generated by a collective aerodynamic phenomenon called the streaming instability (SI) (*21*, *22*). These concentrations then collapsed directly into objects tens to hundreds of kilometers in diameter, on time scales $\lesssim 10^3$ years in the outer Solar System (*21*–*24*). GI after the SI has been shown to be viable in laminar, low-viscosity (called low-α) disks, including those with a low overall surface mass density appropriate to the CCKB (*25*, *26*). Such GI models predict planetesimal formation times, velocity distributions, collisional evolution, obliquities, and binary characteristics that differ from alternative hierarchical coagulation (HC) models, in which successive

[1]Department of Earth and Planetary Sciences and McDonnell Center for the Space Sciences, Washington University in St. Louis, St. Louis, MO 63130, USA. [2]Department of Astronomy, University of Maryland, College Park, MD 20742, USA. [3]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA. [4]Lowell Observatory, Flagstaff, AZ 86001, USA. [5]Department of Astronomy and Planetary Science, Northern Arizona University, Flagstaff, AZ 86011, USA. [6]Division of Space Science and Engineering, Southwest Research Institute, Boulder, CO 80302, USA. [7]NASA Ames Research Center, Space Science Division, Moffett Field, CA 94035, USA. [8]SETI Institute, Mountain View, CA 94043, USA. [9]National Optical-Infrared Astronomy Research Laboratory, National Science Foundation, Tucson, AZ 85726, USA. [10]Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723, USA. [11]National Research Council of Canada, Victoria, BC V9E 2E7, Canada. [12]Division of Space Science and Engineering, Southwest Research Institute, San Antonio, TX 78238, USA. [13]Department of Physics and Astronomy, University of Texas, San Antonio, TX 78249, USA. [14]Department of Astronomy, University of Virginia, Charlottesville, VA 22904, USA.
***Corresponding author. Email: mckinnon@wustl.edu**
†New Horizons Science Team members and affiliations are listed in the supplementary materials.

two-body collisions lead to the gradual accretion of larger and larger objects (supplementary text) (*27*, *28*).

Dynamic characteristics from shape and rotation

The shapes of the large and small lobes are approximately ellipsoidal, 20.6 by 19.9 by 9.4 km and 15.4 by 13.8 by 9.8 km, respectively, with a combined equivalent spherical diameter of 18.3 km (*8*). The gravitational acceleration g that would be produced by the equivalent sphere is $0.0013 \times (\rho/500\ \mathrm{kg\ m^{-3}})\ \mathrm{m\ s^{-2}}$, and the equivalent escape speed is $4.8 \times (\rho/500\ \mathrm{kg\ m^{-3}})^{1/2}\ \mathrm{m\ s^{-1}}$, where ρ is the bulk density. We adopted 500 kg m^{-3} as our fiducial density, on the basis of the median densities of cometary nuclei, including 67P/Churyumov-Gerasimenko, whose density has been precisely measured as 527 ± 7 kg m^{-3} [table 1 in (*29*)] (*30*). The effective surface gravity (including the effects of rotation) across the surface of Arrokoth is shown in Fig. 1A.

If the two lobes are of equal density, the center of mass of Arrokoth is within the body of LL, and the separation between the centers of mass of the two lobes is 17.2 km [figure 2 in (*8*)]. The spin-synchronous orbit period of two barely touching lobes, behaving as gravitational point masses, is $12.1 \times (500\ \mathrm{kg\ m^{-3}}/\rho)^{1/2}$ hours. If Arrokoth formed as a binary pair that spiraled inward, then Arrokoth's spin must have slowed from this more rapid rotation to its observed 15.92-hour period, unless Arrokoth is substantially less dense than 500 kg m^{-3}. The observed rotation period of 15.92 hours would match the spin-synchronous orbit period for two barely touching, equal-density lobes if $\rho \approx 290\ \mathrm{kg\ m^{-3}}$ (*8*). Explicitly treating the lobes as ellipsoids increases their mutual gravitational attraction and lowers the limiting density to ≈250 kg m^{-3}.

The range of tensile and compressive strengths plausible for porous, structurally comet-like bodies (*29*, *31*) also broadens the permissible density range. The gravitationally induced stress at the neck, either compressional or tensile, is shown in Fig. 2 as a function of the assumed bulk density. The contact area between the two lobes is ~23 km^2 (*8*). We calculated the gravitational attraction between the lobes from the external gravitational potential of a homogeneous triaxial ellipsoid (*32*), integrated over the mass distribution of the other ellipsoid. For the observed lobe principal axes, the attraction increases by 12.5% over a point (or spherical) approximation. A bulk density greater than ~250 kg m^{-3} would imply that the neck is in compression, but even for the highest comet cohesion of 10 kPa (*33*, *34*), the density must remain <1250 kg m^{-3} or the compressive strength (which is related to the cohesion) would be exceeded and the neck region would collapse under Arrokoth's self-gravity. For a more plausible, nominal bulk cometary tensile

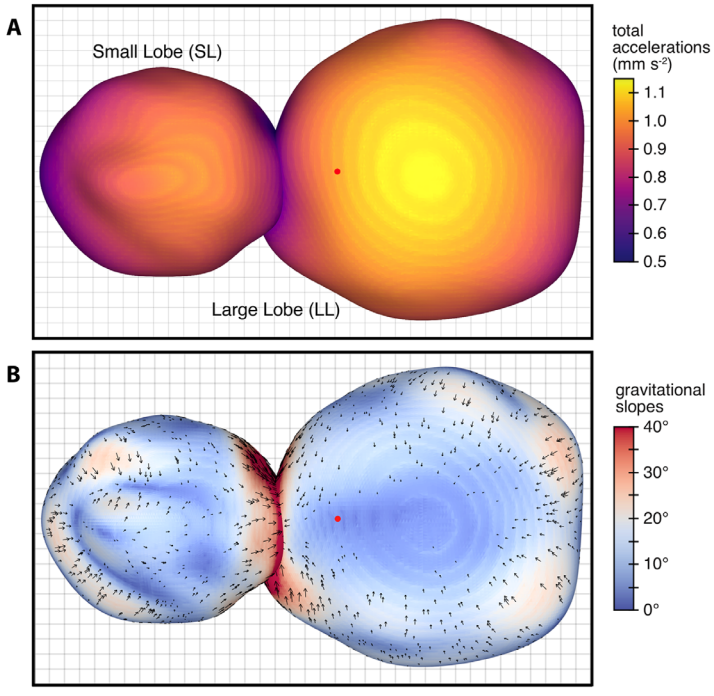


Fig. 1. Dynamic geophysical environment at the surface of Arrokoth is determined by its gravity and rotation. (**A**) Effective surface gravity in Arrokoth's rotating frame, overlain on the shape model (*8*). (**B**) Gravitational slope, the difference between the local effective gravity vector and the surface normal to the global shape model. Arrows indicate the tilt of the local gravitational slope; the steepest slopes occur in or near the neck region [figure S1 in (*8*)]. In both (A) and (B), a uniform density of 500 kg m^{-3} is assumed for both lobes; the red dot indicates the center of mass and rotation axis. The background grid is in 1-km intervals.

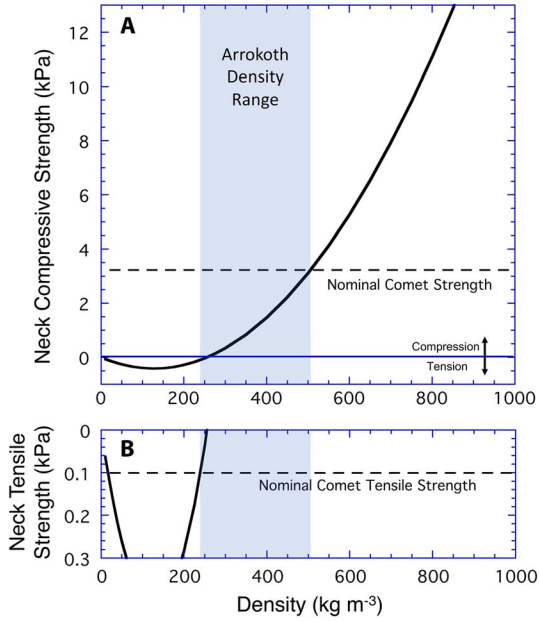


Fig. 2. The compressive or tensile stress supported at Arrokoth's neck depends on the body's density. (**A**) The solid blue line separates the unconfined compression and tension regimes. For bulk densities $\rho \gtrsim 250\ \mathrm{kg\ m^{-3}}$, the neck is in compression. For a nominal cometary cohesion of 1 kPa (dashed black line) and internal friction angle of 30° (*33*), the upper limit density of Arrokoth is ~500 kg m^{-3}; otherwise, the neck region would collapse. Greater strengths are compatible with greater bulk densities. (**B**) An expanded scale. For $\rho \lesssim 250\ \mathrm{kg\ m^{-3}}$, the neck is in tension. For a nominal cometary tensile strength of 100 Pa (dashed black line) (*33*), the lower limit density of Arrokoth remains close to 250 kg m^{-3}. Much lower densities ($\lesssim$50 kg m^{-3}), for which the forces between the lobes vanish, are not considered physical. Strength estimates scale inversely with the assumed contact area [we adopted 23 km^2 (*8*)].

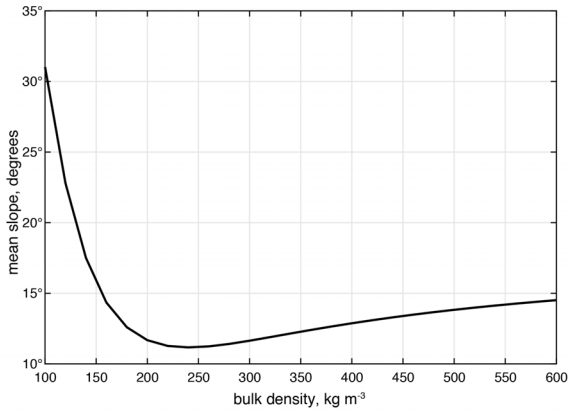


Fig. 3. The mean gravitational slope of Arrokoth as a function of assumed bulk density. The minimum mean slope occurs for a bulk density of ~240 kg m^{-3} (compare with Fig. 1B, which assumes ρ = 500 kg m^{-3}). If Arrokoth's topography behaves similarly to that of asteroids and cometary nuclei (*39*), this may be the approximate density of Arrokoth. The minimum is quite broad, however, which is consistent with a range of densities considered appropriate to cometary nuclei (*29*).

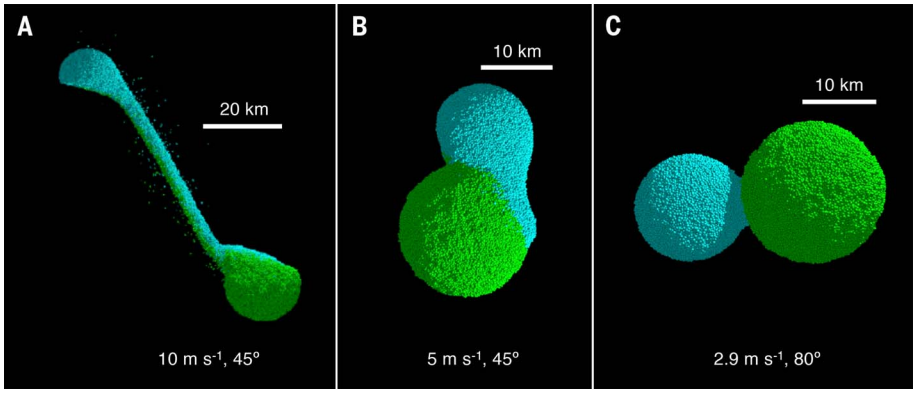


Fig. 4. Numerical N-body calculations of collisions between spherical bodies of the scale and approximate mass ratio of the LL and SL lobes of Arrokoth. The larger lobe (LL) is represented by green particles, and the smaller lobe (SL) is represented by blue particles. A bulk density of 500 kg m^{-3} is assumed for both bodies. (**A**) At a collision speed of 10 m s^{-1} and a moderately oblique angle, the impact severely disrupts both bodies, leaving a long bridge of material stretched between them. As the simulation progresses, this connection breaks as SL moves farther from LL and ultimately escapes. Movie 1 shows an animated version. (**B**) At 5 m s^{-1} and for the same impact angle of 45°, the impact creates a contact binary, but with an asymmetric, thick neck and a lopsided SL. Movie 2 shows an animated version. (**C**) At 2.9 m s^{-1} and an oblique impact angle of 80°, both lobes remain intact, and the contact area between them forms a well-defined, narrow neck. Movie 3 shows an animated version. Interparticle friction between the particles is assumed in all cases; in (A) and (B), the interparticle cohesion is 1 kPa (a value thought typical for comet-like bodies), and zero cohesion is assumed in (C). No initial spin is assumed in (A) and (B), whereas the lobes in (C) are set to rotate synchronously before collision.

strength of 100 Pa and cohesion of 1 kPa (implying a frictional, bulk compressive strength of ~3 kPa) (*34*), the bulk density of Arrokoth must lie in the range of ~250 to 500 kg m^{-3} to explain the lack of observed faulting or distortion of the neck region (Fig. 2).

Arrokoth must possess some internal strength, otherwise it would collapse to a more spherical shape. The surface slopes with respect to the local gravity vector (Fig. 1B) are generally less than the angle of repose (maximum slope) for loose, granular material (~30 to 40°) (*35*), so the overall shapes of LL and SL can be maintained by frictional strength alone. Near the neck, these slopes sometimes exceed 35° to 40°, so these over-steepened surfaces must be held together by finite cohesion (for $\rho \gtrsim 500$ kg m^{-3}). The minimum cohesion c necessary to stabilize an inclined layer of thickness h is given by

$$c/\rho g h = \tan\theta - \tan\phi \quad (1)$$

where θ is the local slope, and ϕ is the internal friction angle. An over-steepened thickness h ~ 1 to 2 km (Fig. 1B), ρ = 500 kg m^{-3}, g = 10^{-3} m s^{-2}, θ ~ 40° to 45°, and a geologically typical $\phi \approx 30°$, implies c ~ 100 to 400 Pa (*36*). This minimum strength is very low by terrestrial standards but similar to the gravitational stresses in other low-gravity small-body environments and the interparticle forces in granular materials (such as electrostatic and van der Waals) (*37*, *38*).

The distribution of gravitational slopes may provide additional constraints on the bulk density of small Solar System bodies (*39*). If an object possesses a sufficiently mobile regolith (surface fragmental layer)—one able to overcome its intrinsic cohesion—then the surface of the body may gradually erode and/or adjust (for example, because of impact-induced seismicity) to a state of maximum topographic stability and lowest internal stress (*39*). The distribution of slopes can therefore be related to the bulk density (subject to the aforementioned caveats). For Arrokoth's shape, there is a broad minimum in gravitational slope between bulk densities of ~200 and 300 kg m^{-3} (Fig. 3), lending additional support to the inference that the density of Arrokoth may be <500 kg m^{-3}. If so, Arrokoth would have to be a highly porous body, given its inferred composition (*9*). Conversely, the surface of Arrokoth is only lightly cratered, so the generation of regolith and surface mobility may be inefficient [or only locally efficient, such as on subkilometer scales, corresponding to the small-scale pitting observed (*8*)]. No other KBOs or cometary nuclei have confirmed densities this low, although such values have been suggested in some cases (*29*).

Merger speed constraints

LL and SL must have merged at a very low velocity (*7*, *8*). Previous numerical simulations of collisions of kilometer-scale (comet-like) porous icy aggregates (*15*, *33*) imply that when extrapolated to bodies the size of LL and SL, closing velocities no greater than their mutual escape speed (several meters per second or less) and an oblique strike are likely required to preserve the shape of a contact binary with a narrow neck. CCKB objects—even with their low-e, low-i orbits—currently have a median mutual impact speed of ~300 m s^{-1}, which is some 100 times greater (*40*). Thus, heliocentric impacts between bodies similar to LL and SL could not have formed Arrokoth (*7*). However, we must consider the impact velocities that would have prevailed during the early Solar System.

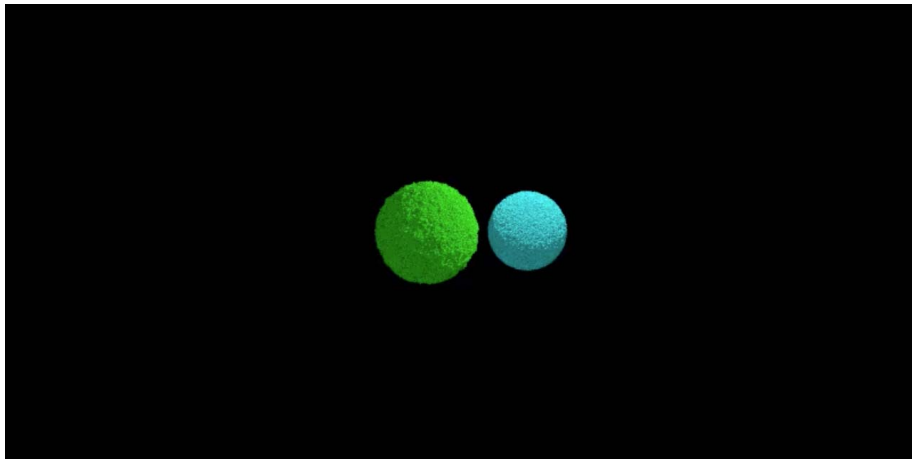
Arrokoth is an order of magnitude larger in size than typical comets (*8*). Therefore, we performed a series of numerical experiments—modeling the collisions of bodies of the appropriate scale, density, and strength characteristics—using a soft-sphere version of the PKDGRAV N-body code (*41*, *42*) to constrain Arrokoth's

formation. This code uses a discrete element method to model the collisions of granular aggregates at slow speeds and low energies, incorporating interparticle cohesion and frictional contact forces (*43*). We focused on velocities near the escape speed from the binary, which we modeled as two spheres for simplicity and to focus on mechanical outcomes, such as the extent of distortion or disruption. The results are shown in Fig. 4.

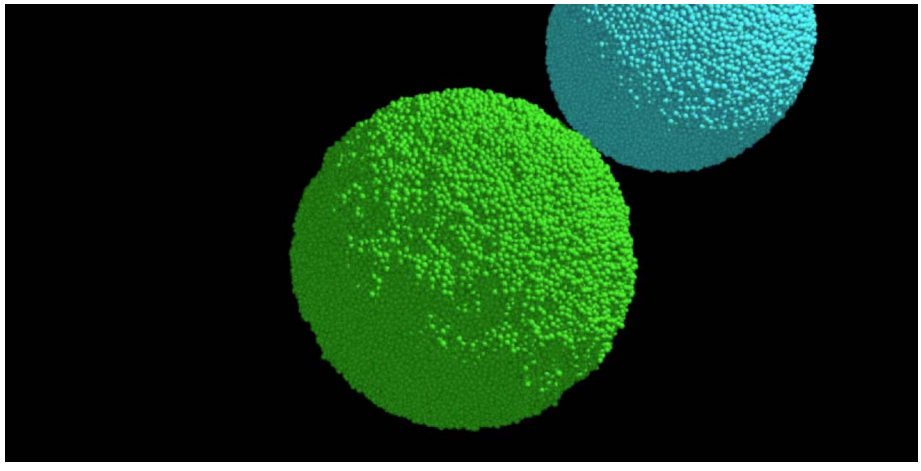
Oblique impacts at 10 m s^{-1}, much greater than escape speed, do not lead to mergers (Fig. 4A and Movie 1) but instead shear or slice off sections of one or both bodies. The collision or merger speed of LL and SL falling from infinity (assuming an initial velocity $\upsilon_\infty = 0$) would have been $\approx 3.5 \times (\rho/500\ \text{kg m}^{-3})^{1/2}$ m s^{-1}. Even oblique impacts at 5 m s^{-1}, slightly higher than the escape speed of 4.3 m s^{-1} for the spheres in the simulations, lead to distortion and merging incompatible with Arrokoth's shape (Fig. 4B and Movie 2).

These results are essentially insensitive to the impact angle assumed. Varying the impact angle from 45° to 85° (measured with respect to the vertical at the impact point) for 10 m s^{-1} collisions changes the amount of damage at the contact regions and the extent of planar shearing, but in all cases, the two bodies remain unbound. At 5 m s^{-1}, the 45° simulation (Fig. 4B) is the only one that produced a final configuration remotely resembling the present-day Arrokoth. In this case, we are left with a mostly intact larger lobe but a lopsided smaller lobe and a neck that is much thicker than observed today (Fig. 4B). At 65° and 5 m s^{-1}, there is again substantial damage to the smaller lobe. At higher angles to the vertical, collisions are grazing, and because the system is initially (marginally) unbound, the simulation ends before the ultimate outcome (escape or recollision).

Only at much lower collision velocities, substantially less than the mutual escape speed, and at an oblique angle, do the outcomes of our simulations begin to resemble Arrokoth (Fig. 4C and Movie 3). Movie 4 shows the maximum surface accelerations experienced by particles in the simulation shown in Fig. 4C. The disruption induced in this gentle merger (the normal velocity component is 0.5 m s^{-1}) is confined to the neck region and more severely affects the smaller lobe portion of the neck. For a bulk density of 500 kg m^{-3}, varying the interparticle cohesion values over a plausible range [100 Pa to 10 kPa (*34*)] likely has only a modest effect on the gentle merger outcomes; the lobes should remain intact. Our numerical models show that the merger speed of Arrokoth's LL and SL was likely sufficiently slow that the two bodies were already gravitationally bound to each other before the collision. We estimate an upper limit on the vertical closing speeds of 4 m s^{-1}.



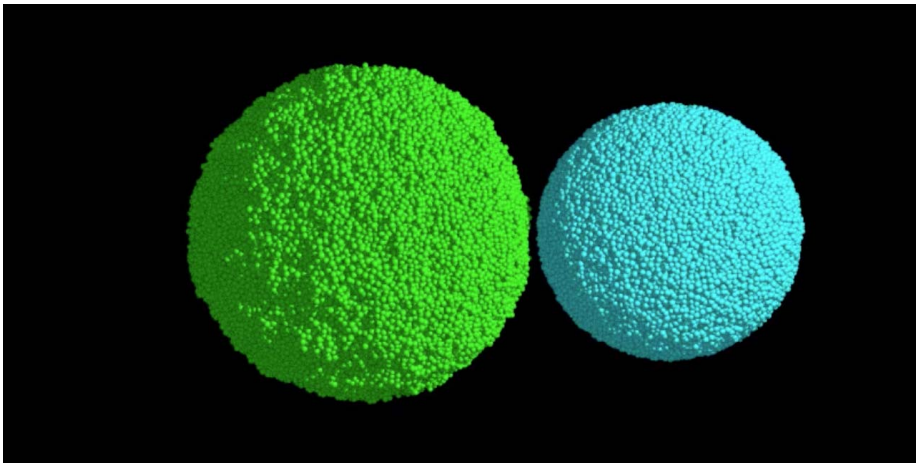
Movie 1. Animated version of Fig. 4A. Merger of spherical components at 10 m s^{-1} and impact angle 45°, using a rubble-pile model with 198,010 particles total. Material parameters correspond to rough surfaces, with a friction angle of ~40° and a cohesion of 1 kPa. For this model, there was no initial component rotation. Particle color indicates body of origin; green particles are from the large lobe (representing LL), whereas blue particles are from the small lobe (representing SL).



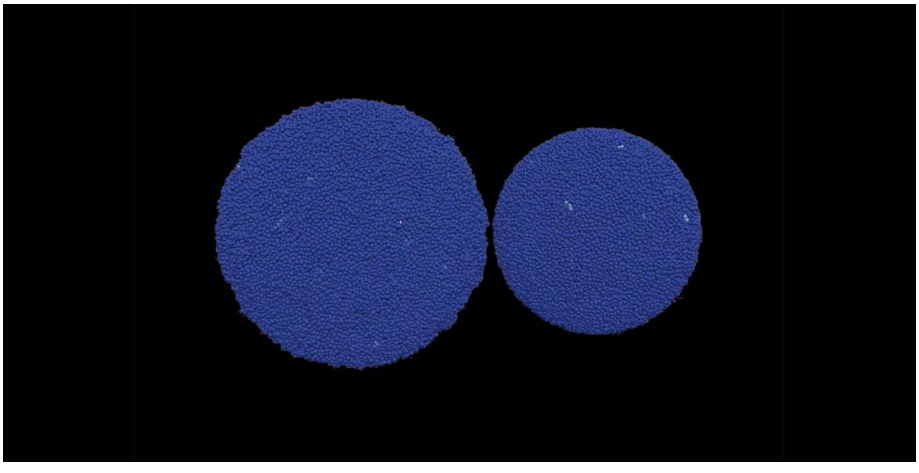
Movie 2. Animated version of Fig. 4B. Merger of spherical components at 5 m s^{-1} and impact angle 45° using a rubble-pile model. Particle size and density and material parameters are identical to the simulations in Movie 1. For this model, there was no initial component rotation. Particle color indicates body of origin.

By way of comparison, in the described hierarchical collisional accretion scenarios, merger speeds scale with the escape speeds of the largest accreting bodies (*44*). For the cold classical region, encounter speeds could have been low initially but would have well exceeded our limit as the planetesimal population evolved. Our numerical models thus show it is unlikely that Arrokoth's shape could be the result of a merger of two independent heliocentric planetesimals, at any nontrivial level of dynamical excitement of the parent planetesimal swarm, unless the two lobes were much stronger (more structurally cohesive) than is usually assumed for comet-like bodies.

More generally, bilobate shapes can also be formed in catastrophic or subcatastrophic collisions. In this scenario, a contact binary results from the merger of collapsing adjacent ejecta streams after a high-speed catastrophic disruption of a parent body (*15*, *45*). Such a scenario may possibly lead to a Arrokoth-like shape, if the two components first collapsed into separate bodies that then slowly came into contact. This scenario would erase any record of the precursor bodies and in principle also permits formation of Arrokoth later in Solar System history. However, the bilobate shapes formed by these models do not resemble Arrokoth because the lobes are not flattened, and the merged components are unaligned and/or highly distorted (*15*, *45*). The CCKB has not experienced strong collisional evolution (*4*, *5*), making disruption of large parent bodies rare (*40*). We conclude that Arrokoth's shape and appearance are more



Movie 3. Animated version of Fig. 4C. Merger of spherical components at 2.9 m s^{-1} and impact angle 80° using a rubble-pile model. Particle size and density and material parameters are identical to the simulations in Movies 1 and 2. For this model, each component has an initial spin period of 9.2 hours in the same sense as the orbit, in order to produce synchronous rotation. Particle color indicates body of origin.



Movie 4. Peak accelerations during the merger of spherical components at 2.9 m s^{-1} and an impact angle of 80°. This is the same simulation as in Movie 3, but now particle colors correspond to the maximum acceleration experienced by each particle up to the time shown. Darkest blues correspond to 3.5×10^{-4} m s^{-2}, and darkest reds correspond to 8.8×10^{-1} m s^{-2}, on a linear scale. Although some disturbance is experienced by loose surface particles globally (each sphere is settled individually before they experience each other's gravity), the maximum disturbance during the simulation is concentrated in the narrow contact area, or "neck," between the two bodies.

likely the result of the low-velocity merger of two bodies that were already gravitationally bound.

Binary formation scenarios

We next consider how the gravitationally bound binary formed before the merger. Mechanisms that have been proposed for the formation of close binaries among the small asteroids may be relevant to this process. Those mechanisms are not primordial in nature but involve later collisions or spin-up followed by rotational fission or mass shedding at the equator, owing to asymmetric solar radiation forces (*46*). Given the very low crater density on Arrokoth (*8*) and its distance from the Sun, however, these do not appear to be promising explanations (although we will return to these points). Secondary satellites produced from these processes are generally much smaller than the primary (*46*) as well, unlike the similar sizes of SL and LL.

The prevalence of binaries in the Kuiper Belt, and especially among the CCKB objects (*11*, *12*, *47*), has prompted theoretical examination of possible binary formation mechanisms specific to the Kuiper Belt (*48*–*50*). Most of these mechanisms operate at the Hill radius (R_{Hill}), the spatial limit of a body's gravitational influence in solar orbit (~4×10^4 km for Arrokoth) (*51*). For example, it has been proposed that binary KBOs could form from the chance interaction of two KBOs within the Hill sphere of a third body, leaving the two permanently bound, or that dynamical friction (multiple gravitational energy and momentum exchanges) with a large number of smaller heliocentric particles could allow two passing KBOs to become bound (*52*). These mechanisms rely on energy exchanges or dissipation and thus are most effective when KBO encounter speeds are low, within an order of magnitude of the Hill speed (~2 to 3 cm s^{-1} for Arrokoth) (*51*). Such low encounter speeds favor binding in the outer regions of the Hill sphere for retrograde orbits or about half that distance for prograde orbits (*53*). These mechanisms thus geometrically favor the production of retrograde binaries, sometimes strongly so (*54*), but observations show that prograde binaries are more common than retrograde (*55*). Such chance encounters of KBOs would produce some binaries with different color characteristics within each pair. This also disagrees with observations, which show that KBO binaries all share the same colors (*12*, *47*, *56*). We therefore discount these models in favor of a binary formation mechanism that produces both bodies from a compositionally uniform region of the protosolar nebula.

An alternative binary formation mechanism posits swarms of locally concentrated solids in the protoplanetary disk that collapse under self-gravity. The swarms of particles could form as concentrations produced by the SI, in which the drag felt by solid particles orbiting in a gas disk leads to a back-reaction and spontaneous concentration of the particles into massive filaments and clumps (Fig. 5A), which can then gravitationally collapse (*22*–*24*, *57*). The collapse mechanics have been simulated for the formation of larger, 100-km-class Kuiper Belt binaries (*58*). That work simulated bound particle clumps in three dimensions with the PKDGRAV N-body code, including collisions and assuming perfect merging (100% sticking). Rotating particle clumps in (*58*) typically collapse to form binaries or higher multiple planetesimal systems (Fig. 5B). The mechanism produces binaries with a broad range of separations and eccentricities, depending on the initial swarm mass and angular momentum [figure 5 in (*58*)]. The resulting binary orbital parameters are consistent with observations of binaries in the classical Kuiper Belt (*11*, *12*, *58*), including the approximately equal radius ratios of the binary components (and of Arrokoth) (*47*, *59*). We also expect such binaries to have matching component colors because they formed from the same material.

The angular momentum vector orientations of collapsing particle clouds have been estimated (*60*). That work fully simulated vertically stratified three-dimensional hydrodynamical SI [following (*61*, *62*)], identified gravitationally bound clumps of solid particles (Fig. 5A),

and determined the obliquity of the corresponding angular momentum vectors. The total angular momentum stored in a particle clump typically exceeds the maximum possible for a compact object of the same solid mass (set by the rotation speed at which it would break up) (*60*). Therefore, as a particle clump contracts and speeds up, it must either shed mass and angular momentum or form binary or higher-multiple systems. The simulations (*60*) did not reach binary formation [unlike (*58*)] owing to computational resolution limits. The resulting angular momentum vectors of the gravitationally bound clumps, however, span a range from prograde to retrograde, with a strong preference for prograde over retrograde rotation (*60*). This is consistent with observations of KBO binaries (*55*), even with the broad range of obliquity produced by the inherently stochastic, turbulent nature of the clump collapse process (*58*). In addition, previous results (*47*, *58*) indicate that a fraction of the non–binary-forming solids in a contracting clump are expelled from the clump into the general nebular population (Fig. 5B). These accretional products and any surviving, unaccreted pebbles are then available for further cycles of concentration because of the SI (or other mechanisms).

Possible mechanisms to produce particle density enhancements in the outer protosolar nebula, acting individually or together, and which could have led to GI, include the SI, photoevaporation, pressure bumps or traps, and volatile-ice lines (supplementary text) (*57*, *63*). Clumping because of the SI in particular is consistent with the mass function of the CCKB and Arrokoth, specifically (supplementary text).

If Arrokoth initially formed as a co-orbiting binary, a subsequent step of orbit contraction is required in which angular momentum is lost, ultimately resulting in a binary merger. For a gravitationally collapsing pebble cloud (*58*), such a merger may happen directly if the angular momentum density is low enough. In a higher-mass cloud, or in one with higher angular momentum density, a smaller-mass binary, co-orbiting or not, may be expelled from the collapsing cloud (Fig. 5B). The shape and alignment of Arrokoth's lobes constrain the nature of any orbital contraction.

Lobe shape and alignment

The global, contact binary shape of Arrokoth (Fig. 1) (*7*, *8*) is reminiscent of co-orbiting Roche ellipsoids in close contact. Roche ellipsoids are the equilibrium shapes of rotating homogeneous fluid masses distorted by the tidal action of a nearby more massive body (supplementary text) (*64*). However, the flattened shapes of the observed lobes (axis ratio ~1/2 for LL and ~2/3 for SL) do not match a Roche ellipsoid because the less massive lobe should be more oblate than the more massive one, even when considering higher-order gravity terms and internal friction (*65*). The present-day shape of Arrokoth does not conform to an equipotential surface at any uniform density or rotation rate (*8*).

The generally ellipsoidal to lenticular shapes of Arrokoth's two lobes, and their general smoothness at scales resolved by the available images (*7*, *8*), nevertheless resemble equilibrium figures, perhaps obtained in the past. It is possible that the flattened shapes of both lobes were acquired as they rapidly accreted in a pebble cloud undergoing gravitational collapse, as described above. The spin rates necessary to reach the observed flattened shapes would have been higher than Arrokoth's spin today, but not by a large margin. For low-density (250 kg m^{-3}) strengthless oblate bodies (Maclaurin spheroids) (*64*), the rotation periods of LL and SL would need to have been ~12 and 14 hours, respectively [these values scale as $\rho^{-1/2}$ (*64*)], compared with the current rotation rate of 15.9 hours. The process(es) that collapsed the co-orbiting binary could have potentially slowed the spin of the individual lobes by this amount.

Regardless of the origin of the shapes of the two lobes (supplemental text), the close alignment of their principal axes (Fig. 6) (*7*, *8*) is unlikely to be due to chance alone. The short axes (which we designate as *c* axes) of LL and SL are closely aligned, to within 5°, a value set by systematic uncertainties in the shape models (*8*). The long *a* and intermediate *b* axes are aligned as well, but the *a* and *b* axes of LL are similar in length (20.6 ± 0.5 and 19.9 ± 0.5 km, respectively; 1σ uncertainties), so that the alignment of SL's *a* axis with the long axis of the body as a whole (also within 5°) is more meaningful (Fig. 6A). These angles are small enough to be considered in sequence. The *c* axis of one lobe must lie within a cone of half-angle 5° with respect to the *c* axis of the other {[1 – cos(5°)] = 0.0038}; with that orientation fixed,

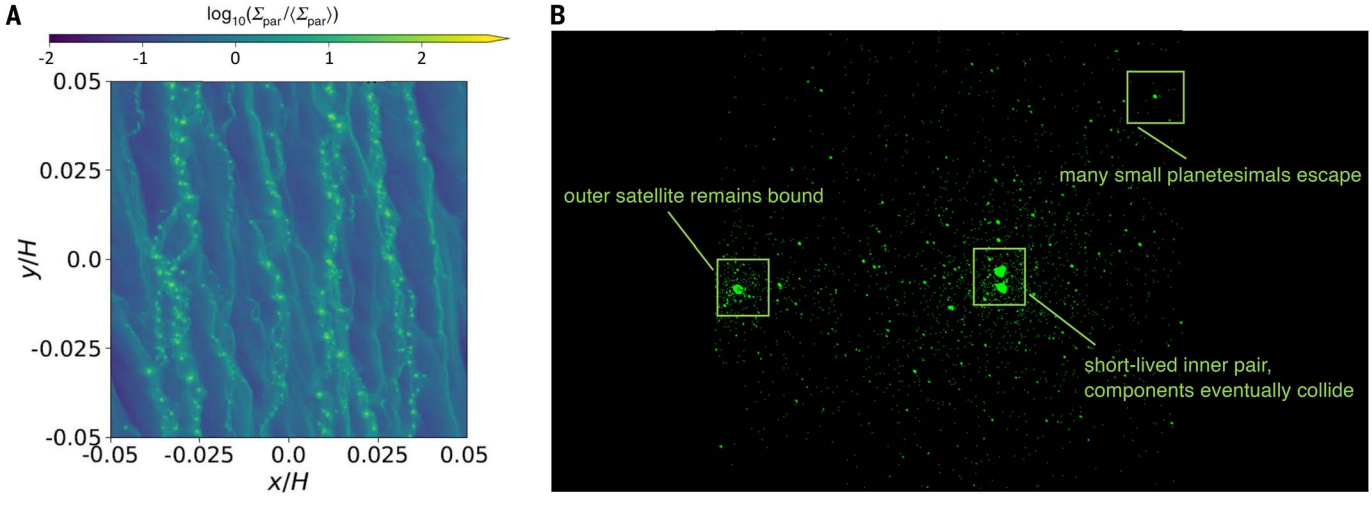


Fig. 5. Possible initial stages in the formation of a contact binary in the Kuiper Belt, illustrated with numerical models. (**A**) Overdense particle concentrations in the protosolar nebula self-amplify by the SI, which then leads to GI and collapse into finer-scale knots. A snapshot from a numerical simulation in (*60*) illustrates vertically integrated particle density, Σ_{par}, viewed perpendicular to the nebular midplane, relative to the initially uniform surface density, $\langle\Sigma_{par}\rangle$; lighter colors indicate greater particle density, *H* is nebular scale height, and 0.02*H* is the initial particle scale height [adapted from (*60*), reproduced with permission]. (**B**) Outcomes of an example collapsing, gravitationally unstable particle cloud, from N-body simulations [modified from (*58*); copyright AAS, reproduced with permission]. Arrokoth may have formed as a binary planetesimal in such a collapsing particle cloud, either as a contact or, more likely, a co-orbiting binary.

the a axis of SL must lie within 5° of the long axis of the body as a whole (10/180 = 0.056). The joint probability of both aligning through chance is ~0.0038 × 0.056 = 2 × 10^{-4}.

We infer that before their final merger, the LL and SL lobes were already aligned. That would be consistent with tidal evolution of a close binary because alignment reduces the total energy of the system. Full spin-orbit synchronism (tidal locking) is not required, however. Two irregular bodies rotating asynchronously while their mutual semimajor axis slowly shrank (by any mechanism) would necessarily first contact each other along their long axes, perhaps repeatedly, if the orbits were circular (the same outcome is likely but not guaranteed for elliptical orbits). Ultimately, mechanical dissipation of rotational kinetic energy while in contact would cause their long axes to come to rest in alignment (or nearly so) (supplementary text).

Regardless of whether the a and b axes of Arrokoth were aligned before the merger, the c axes must have been. Merger, even a slow merger, from an arbitrary direction is very unlikely. Chaotic rotation (tumbling) of either lobe, owing to an eccentric orbit premerger (*66*, *67*), is highly unlikely to produce this alignment. The LL and SL spin poles and their mutual orbit normal vector were most likely close to co-aligned before a merger, which is consistent with mutual tidal dissipation. This may also be a common (although not exclusive) outcome of binary formation in a gravitationally unstable pebble cloud (*58*), with its shared angular momentum. Although previous work focused on wide binaries (*58*), it is possible to form binaries with a range of orbital separations, including those much closer to contact, or already contacting, within the collapsing cloud.

Binary merger mechanisms

The GI formation mechanism produces a high fraction of binary KBOs, but as discussed above, the resulting angular momentum of each binary may have been greater, perhaps much greater, than that of the current Arrokoth system. We considered several, non–mutually exclusive mechanisms that might drain angular momentum from the system over all or part of its 4.5-billion-year lifetime.

Kozai-Lidov cycling

In a system with three (or more) bodies with differing orbital inclinations, the Kozai-Lidov effect (*68*) causes oscillations of the orbit's eccentricity and inclination. We focused on the Sun as the third body and the Kozai-Lidov cycles of the orbits of LL and SL about each other. In this case, the angular momentum component of the binary perpendicular to the heliocentric orbital plane is conserved. On time scales much longer than Arrokoth's 298-year heliocentric orbital period [following (*68*, *69*), ~10^5 year × $(a/1000\ \mathrm{km})^{-3/2}$, where a is the assumed LL-SL semimajor axis], highly inclined, near-circular orbits can transition to and from low-inclination, highly eccentric orbits. During periods of high eccentricity, the binary objects pass closer to one another and so have stronger tidal interactions (*70*, *71*). If the eccentricity becomes sufficiently high, the objects could undergo grazing collisions that would substantially alter the balance between orbital and rotational angular momentum and efficiently dissipate kinetic energy. High-eccentricity phases also cause objects to spend most of the time near their maximum separation (apoapse), where they are more susceptible to perturbation by unbound bodies passing through the system.

Solar tides are weak in the Kuiper Belt, and the Kozai-Lidov cycles occur slowly. Solar tides are not important except for wide binaries because the tides owing to nonspherical shapes can dominate the dynamics of closer binaries. For Arrokoth in particular, solar perturbations would only dominate at binary semimajor axes a > 1000 km (~100 LL radii) (*72*). If Kozai-Lidov oscillations had affected Arrokoth, we expect that the merged body (in most cases) would have a lower obliquity than the observed 99° because the tidal interactions or collisions at high orbital eccentricity would have tended to lock in the low inclinations that correspond to the highest eccentricities (*69*–*71*).

An alternative possibility is that Arrokoth was once a triple system and that the third body was in an inclined orbit with respect to the then LL-SL binary. For suitable orbital parameters, this third body could have driven Kozai-Lidov oscillations of the inner binary. Hierarchical triple systems do exist among small KBOs [such as 47171 Lempo (*73*)] and are a common outcome of simulations (*58*). However, because there is no specific evidence of a lost third body, we did not consider this hypothesis in greater detail.

YORP and BYORP

Interaction with sunlight can affect the angular momentum of small bodies in two main ways: the Yarkovsky-O'Keefe-Radzievskii-Paddack (YORP) effect, which alters the spin rate and obliquity of a single object, and binary YORP (BYORP), which changes the size and shape of a binary's orbit (*74*, *75*). Both mechanisms arise from the asymmetric scattering and thermal reemission of sunlight from the surfaces of irregular bodies, and both can either increase or decrease the angular momentum of the system and alter its vector direction (*74*, *75*).

BYORP can in principle drain the angular momentum of a binary near-Earth asteroid, provided one or both members of the binary are spinning synchronously (*76*). BYORP requires 10^4 to 10^5 years to alter the orbit of a 150-m-radius, synchronously rotating satellite of a 500-m-radius primary, both with density

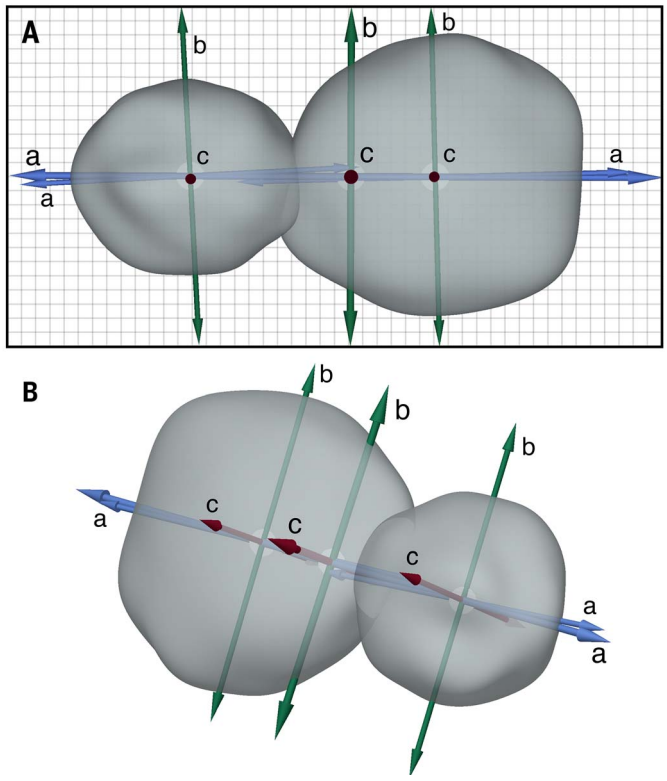


Fig. 6. The inertial axes of Arrokoth and its two lobes are aligned. (**A**) Viewed down the spin axis, arrows indicate the maximum (*c*, or red), intermediate (*b*, or green), and minimum (*a*, or blue) principal axes of inertia for each lobe (thin vectors) and the body as a whole (thick vectors). Vectors originate from the center of mass of each component. Background grid is in 1-km intervals. (**B**) Oblique view of the same, matching the geometry of the CA06 image (*8*). Alignment of the maximum principal axes of inertia, and of SL's minimum principal axis of inertia with that of Arrokoth as a whole, is unlikely to be due to chance alone.

1750 kg m^{-3}, assuming a satellite orbital radius of 4 primary radii and a primary orbit at 1 AU (*76*). Scaling that result to parameters (size, distance, density, and mass ratio) appropriate to Arrokoth (*77*), we obtain a time scale of a few billion to a few tens of billon years, a span that includes the age of the Solar System. Thus, the two components of Arrokoth, if initially separated by a few LL radii, could in principle be driven by BYORP radiation forces alone into a gentle merger of the type needed to account for the narrow neck connecting the two bodies, albeit late in Solar System history (*78*).

YORP accelerations (unlike BYORP) have been detected for several asteroids (table S1). Asteroid (1862) Apollo, at 1.5 km across and orbiting at 1.5 AU, exhibits the largest measured YORP coefficient (*Y*), with an estimated YORP spin doubling time scale of ~5 × 10^5 years for an assumed density of 2500 kg m^{-3} (table S1). Accounting for the larger size of Arrokoth, its greater distance from the Sun, and its much lower density (*77*) lengthens this time scale to 7.5 ×10^9 years. This exceeds the age of the Solar System, and adopting any of the other (lower) *Y* values in table S1 would imply even longer time scales for Arrokoth, by up to two orders of magnitude. However, a more rapidly spinning contact binary in the Kuiper Belt, such as might be produced by BYORP, could plausibly be slowed by subsequent YORP torques over the age of the Solar System (for example, from a 12-hour period to its present 15.9 hours).

The shape and surface properties of Arrokoth and any obliquity variations could change the strength and sign of the YORP torques (*79*–*82*). Both YORP torques perpendicular to the spin axis and BYORP torques perpendicular to the orbit normal vector are minimized at high obliquities and binary inclinations, respectively (*76*, *81*). Arrokoth's high obliquity is therefore less likely to have changed much over the age of the Solar System because of radiation effects, even if its spin has.

Tides

Tides could have contributed to Arrokoth's orbital evolution when the two lobes were close, including establishing the synchronous spin-orbit locking necessary for BYORP torques to have been effective. We calculated the spin-down time scale for SL attributable to tides from LL using standard methods (*83*, *84*). Tidal evolution from the breakup spin limit (*85*) to its present value would take (~65 to 650) × (a/100 km)^6 million years, assuming a circular orbit and adopting a bulk density of 500 kg m^{-3} for both lobes and tidal dissipation parameters for SL (*84*, *86*). If the Arrokoth binary originally formed within 100 to 200 km (≲25 LL radii), or a was driven below that limit by other processes, tides would have dominated. For a < 50 km, tidal synchronization of SL's spin would have been rapid (≲1 million to 10 million years).

On their own, tides between LL and SL do not shed angular momentum but redistribute it among the individually rotating lobes (including aligning their spin and orbital angular momenta). If LL were rotating more slowly than SL's mean motion, for example, tides would act to shrink the binary orbit, and at the moment of tidally induced contact, the overall rotation rate of the merged binary would jump abruptly. A more slowly rotating LL could have resulted from YORP torques, which affect individual binary components even when the components are not rotating synchronously. Tidal interactions within a hierarchical triple system could also have led to angular momentum exchange [as noted above (*68*)] and loss from the system if the more distant member of the triple ultimately escapes to heliocentric orbit.

Collisions

Bilobate comets, such as 67P, have led to the suggestion that mutually orbiting binaries in the Kuiper Belt may have their binary orbital angular momentum altered by repeated impacts with smaller, heliocentric planetesimals, resulting in a contact binary (*87*). Impacts can both bind or unbind a binary—with the binary's orbital angular momentum executing a random walk—so binding to coalescence is only probable (although by no more than ~30%) for very close binaries (*87*). The heliocentric impactor flux in the CCKB object region is estimated to be (and to have always been) low and deficient in smaller, subkilometer-scale bodies (*8*, *40*), making this mechanism unlikely.

Arrokoth's low crater density (*7*, *8*) also makes impacts an unlikely candidate for collapsing the pair's orbit. Only formation of the largest impact crater, informally named Maryland (*8*), could have substantially affected the angular momentum of Arrokoth. Assuming an impactor diameter of ~1 km (1/7 the diameter of Maryland), an impact speed of 300 m s^{-1} [typical for Arrokoth impactors (*40*)], an impact angle of 45°, and an optimistic impact orientation (a velocity vector in Arrokoth's equatorial plane), Arrokoth's total angular momentum only changes by ~10% if a was ~100 km at the time of the impact. The transfer of linear impactor momentum to binary angular momentum scales as $a^{1/2}$, so the formation of Maryland could have had a stronger effect if Arrokoth formed originally as a wide binary. All other observed impact craters are much smaller.

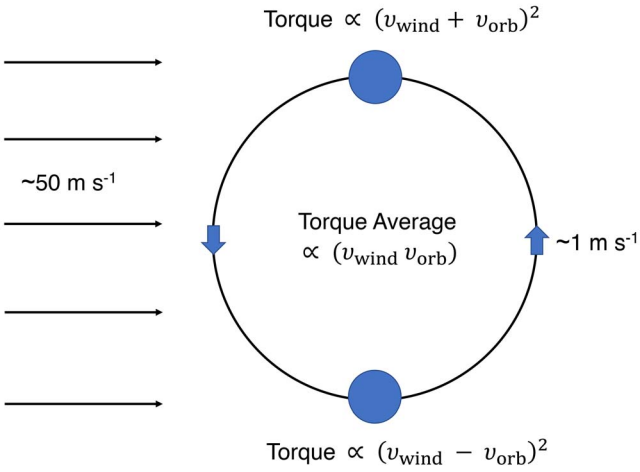


Fig. 7. Illustration of the protosolar nebula headwind interacting with a co-orbiting equal-mass binary. The averaged torque is proportional to the product of the lobe orbital velocity and the differential velocity between the nebular gas and the binary's center of mass about the Sun.

Gas drag

Drag may have been exerted on the binary by protosolar nebular gas. Within a collapsing pebble cloud, the mean collision time is shorter than the gas-drag stopping time (the time it takes for a pebble's linear momentum to drop by a factor of *e*) (*58*). This implies that binary formation and dynamics during GI are dominated by collisions and dynamical friction, not intracloud gas dynamics. Once the unaccreted cloud remnant disperses, however, the binary is subject to gas drag forces for as long as the gas in the protosolar nebula persists at Arrokoth's heliocentric distance (*88*).

The momentum flux (owing to gas drag) imparted by an ambient gas to an orbiting binary yields a stopping time of $t_{stop} \sim \rho R/(\rho_{gas} v_{orb})$ (*89*), where R is the mean radius of either the primary or secondary and ρ_{gas} is the gas density, assuming a drag coefficient C_D of ~1, which is appropriate to fully turbulent drag. Adopting a characteristic midplane ρ_{gas} at 44.2 AU of 1 × 10^{-10} kg m^{-3} (*90*) and an initial semimajor axis for Arrokoth of 100 km yields an orbital speed v_{orb} ~1 m s^{-1} and stopping times of ~500 million years for ρ = 500 kg m^{-3} and an average R = 7 km (with gas drag acting on each lobe). This is much longer than any plausible lifetime for the protosolar nebula, likely no

more than 10 million years (*90*, *91*), so it may seem that ambient gas had little effect on Arrokoth's later evolution.

The gas drag environment experienced by Arrokoth is, however, likely to have been more complex than that simple calculation. The nebular gas at Arrokoth's distance from the Sun would have been moving at speeds slower than the equivalent Keplerian orbit owing to the pressure gradient in the nebula (*88*, *92*)

$$\Omega^2 r = \Omega_K^2 r + \frac{1}{\rho}\frac{\partial P}{\partial r} \quad (2)$$

where Ω is the angular velocity of the gas, Ω_K is the Keplerian angular velocity due to the Sun's gravity, P is the gas pressure, and r is the heliocentric distance. Because Arrokoth itself orbits the Sun at Keplerian speed, it will feel a headwind [at velocity $\upsilon_{wind} = r(\Omega_K - \Omega)$], which we estimated [from (*90*)] to be ~50 m/s, which is about 1% of the Keplerian speed. This gas velocity determines the drag regime at Arrokoth, irrespective of the binary's orientation, and couples to the slower velocity of the co-orbiting binary. As the Arrokoth binary orbits in this nebular wind (Fig. 7), each of its lobes will alternately feel accelerating and decelerating torques; time averaged over both the binary's mutual and heliocentric orbital periods, the difference is proportional to $\upsilon_{orb}\upsilon_{wind}$, resulting in a modified stopping time (time to reduce the binary's angular momentum by a factor of e)

$$t_{stop,wind} \sim \frac{\rho R}{C_D \rho_{gas} \upsilon_{wind}} \quad (3)$$

where the drag coefficient C_D is now explicitly included (and the high obliquity of Arrokoth is included as well).

The kinematic viscosity (η) of solar nebula gas, for the above midplane conditions, is ~10^5 m^2 s^{-1} (*93*), which in turn implies Reynolds numbers $Re \equiv 2R\upsilon_{wind}/\eta \sim 15$ for Arrokoth. This puts Arrokoth into the intermediate drag regime (*92*, *94*), with corresponding C_D values of $24Re^{-0.6} \sim 5$ to 10 for its two, nonspherical lobes. Combined with the wind-speed dependence in the time-averaged torque, the gas-drag stopping time from Eq. 3 (a measure of the binary merger time scale) decreases by a factor of ~250 to 500, to ~1 million to 2 million years for Arrokoth. Such time scales are commensurate with the short lifetimes of protoplanetary gas disks (*95*, *96*). Alternative protosolar nebula models (*88*, *94*, *97*) yield comparable time scales.

Headwind-coupled gas drag may therefore have been the dominant mechanism that drove the merger of small Kuiper Belt binaries such as Arrokoth. In the intermediate-Re drag regime, the merger time scales as $\rho R^{1.6}$ (*92*, *94*), so smaller binaries (for example, those similar to comet 67P in scale) would have evolved to become contact binaries even more rapidly. The effects of gas drag do not cease once the contact binary forms, although the geometry of the drag interaction becomes more complicated. Low-inclination binaries would shrink faster than high-inclination binaries (by a factor of ~$\pi/2$), all other things being equal, because the headwind is always edge-on to their mutual orbits. This leads us to predict that for a given distance from the Sun, the physical sizes of low-inclination contact binaries extend to larger scales than those of high-inclination contact binaries. There may also be a complementary excess of more distant co-orbiting binaries at high inclinations.

We adopted a specific nebular density profile (*90*) above because it is consistent with the initial compact giant planet configuration and outer planetesimal disk thought to have been present in the early Solar System (*4*). This profile was designed to represent the protoplanetary nebula at the time of planetesimal formation. It also assumes that the nebula (gas and solids) does not end abruptly at ~30 AU but gradually declines in surface density to satisfy the constraint that Neptune's outward migration ceases at that distance (*20*). If the gas nebula was instead highly attenuated in the CCKB region, a gas drag–driven binary merger would have been ineffective. Because the characteristics of Arrokoth indicate planetesimal formation through the SI or a related collective instability, we nevertheless conclude that there must have been sufficient gas and, at least locally or intermittently, sufficiently high solid-to-gas ratios for planetesimal-forming instabilities to occur.

Arrokoth's story

Numerous mechanisms have been proposed to produce macroscopic bodies from small particles in the protosolar nebula. The New Horizons encounter with Arrokoth has allowed those mechanisms to be tested with close observation of a primitive planetesimal.

Arrokoth is a contact binary (*7*), which is consistent with being a primordial planetesimal (*7*–*9*). There is no evidence of heliocentric, high-speed collisional evolution or any catastrophic (or even a subcatastrophic) impact during its lifetime. Its shape is not consistent with hierarchical accretion of independent, heliocentric planetesimals because initially slow collisions would have eventually become catastrophic. Instead, we conclude that its two lobes (LL and SL) came together at low velocity, at no more than a few meters per second and possibly much more slowly.

Binary formation is a theoretically predicted common outcome in protoplanetary disks when swarms of locally concentrated solids (pebble clouds) collapse under self-gravity, which plausibly explains the high fraction of binaries among cold classical KBOs (*58*). Cold classical KBO binaries exhibit a range of binary orbital separations, down to the presently observable limit [~1000 km (*47*)]. Numerical modeling indicates that tighter or contact binaries could form in a collapsing pebble cloud. The prominence of bilobate shapes among the short-period comets, which are derived from the scattered disk component of the Kuiper Belt, suggests (but does not require) that there is a process that collapses Kuiper Belt binary orbits (*87*). The alignment of the principal axes of the LL and SL lobes indicates tidal coupling between two co-orbiting bodies, before their final merger.

Our examination of various mechanisms to drive binary mergers in the Kuiper Belt indicates the potentially dominant role of gas drag while the protosolar nebula is still present. We find this process to be effective because in a gas nebula with a radial pressure gradient, the velocity of the gas deviates from the heliocentric Keplerian velocity of the binary. The headwind that the binary feels couples to the motion of the binary pair about its own center of mass. The resulting viscous gas drag can collapse Arrokoth-scale co-orbiting binaries—as well as smaller, cometary-scale binaries—within the few-million-years lifetime of the protosolar gas nebula.

The presence of substantial nebular gas in the region of the cold classical Kuiper Belt does not conflict with the low planetesimal mass density in the same region (*4*). Gas drag drift of small particles can cause large-scale depletion of the solids in the cold classical region (*92*, *94*). Enough solid mass must nevertheless have remained to build the cold classical KBOs, a population that has likely dynamically lost only a few times its present mass over Solar System history (*19*). Collective gravitational instabilities in the presence of nebular gas can produce a planetesimal population from such a low solid mass density. Similar accretional processes may have occurred elsewhere in the early Solar System.

REFERENCES AND NOTES

1. S. A. Stern, W. M. Grundy, W. B. McKinnon, H. A. Weaver, L. A. Young, The Pluto system after New Horizons. *Annu. Rev. Astron. Astrophys.* **56**, 357–392 (2018). doi: 10.1146/annurev-astro-081817-051935
2. S. A. Stern, H. A. Weaver, J. R. Spencer, H. A. Elliott, The New Horizons Kuiper Belt Extended Mission. *Space Sci. Rev.* **214**, 77 (2018). doi: 10.1007/s11214-018-0507-4
3. S. B. Porter *et al.*, High-precision orbit fitting and uncertainty analysis of (486958) 2014 MU69. *Astron. J.* **156**, 20 (2018). doi: 10.3847/1538-3881/aac2e1
4. D. Nesvorný, Dynamical evolution of the early Solar System. *Annu. Rev. Astron. Astrophys.* **56**, 137–174 (2018). doi: 10.1146/annurev-astro-081817-052028
5. A. H. Parker, J. J. Kavelaars, Destruction of binary minor planets during Neptune scattering. *Astrophys. J.* **722**, L204–L208 (2010). doi: 10.1088/2041-8205/722/2/L204
6. J. M. Moore *et al.*, Great expectations: Plans and predictions for New Horizons encounter with Kuiper Belt Object 2014 MU69 ("Ultima Thule"). *Geophys. Res. Lett.* **45**, 8111–8120 (2018). doi: 10.1029/2018GL078996
7. S. A. Stern *et al.*, Initial results from the New Horizons exploration of 2014 MU_{69}, a small Kuiper Belt object. *Science* **364**, eaaw9771 (2019). doi: 10.1126/science.aaw9771; pmid: 31097641

8. J. R. Spencer *et al.*, The geology and geophysics of Kuiper Belt object (486958) Arrokoth. *Science* **367**, eaay3999 (2020).
9. W. M. Grundy *et al.*, Color, composition, and thermal environment of Kuiper belt object (486958) Arrokoth. *Science* **367**, eaay3705 (2020).
10. M. J. Mumma, S. B. Charnley, The chemical composition of comets—Emerging taxonomies and natal heritage. *Annu. Rev. Astron. Astrophys.* **49**, 471–524 (2011). doi: 10.1146/annurev-astro-081309-130811
11. K. S. Noll, W. M. Grundy, E. I. Chiang, J.-L. Margot, S. D. Kern, in *The Solar System Beyond Neptune*, M. A. Barucci, H. Boehnhardt, D. P. Cruikshank, A. Morbidelli, Eds. (Univ. Arizona Press, 2008), pp. 345–364.
12. K. S. Noll, W. M. Grundy, D. Nesvorný, A. Thirouin, in *The Trans-Neptunian Solar System*, D. Prialnic, A. Barucci, L.A. Young, Eds. (Elsevier, 2020), 205–224.
13. A. Morbidelli, H. Rickman, Comets as collisional fragments of a primordial planetesimal disk. *Astron. Astrophys.* **583**, A43 (2016). doi: 10.1051/0004-6361/201526116
14. B. J. R. Davidsson *et al.*, The primordial nucleus of comet 67P/Churyumov-Gerasimenko. *Astron. Astrophys.* **592**, A63 (2016). doi: 10.1051/0004-6361/201526968
15. M. Jutzi, W. Benz, Formation of bi-lobed shapes by subcatastrophic collisions. A late origin of comet 67P's structure. *Astron. Astrophys.* **597**, A62 (2017). doi: 10.1051/0004-6361/201628964
16. M. Hirabayashi *et al.*, Fission and reconfiguration of bilobate comets as revealed by 67P/Churyumov-Gerasimenko. *Nature* **534**, 352–355 (2016). doi: 10.1038/nature17670; pmid: 27281196
17. K. Batygin, M. E. Brown, W. C. Fraser, Retention of a primordial cold classical Kuiper Belt in an instability-driven model of solar system formation. *Astrophys. J.* **738**, 13 (2011). doi: 10.1088/0004-637X/738/1/13
18. R. I. Dawson, R. A. Murray-Clay, Neptune's wild days: Constraints from the eccentricity distribution of the classical Kuiper belt. *Astrophys. J.* **750**, 43 (2012). doi: 10.1088/0004-637X/750/1/43
19. D. Nesvorný, Jumping Neptune can explain the Kuiper Belt kernel. *Astron. J.* **150**, 68 (2015). doi: 10.1088/0004-6256/150/3/68
20. R. S. Gomes, A. Morbidelli, H. F. Levison, Planetary migration in a planetesimal disk: Why did Neptune stop at 30 AU? *Icarus* **170**, 492–507 (2004). doi: 10.1016/j.icarus.2004.03.011
21. J. N. Cuzzi, R. C. Hogan, K. Shariff, Toward planetesimals: Dense chondrule clumps in the protoplanetary nebula. *Astrophys. J.* **687**, 1432–1447 (2008). doi: 10.1086/591239
22. A. N. Youdin, J. Goodman, Streaming instabilities in protoplanetary disks. *Astrophys. J.* **620**, 459–469 (2005). doi: 10.1086/426895
23. A. Johansen *et al.*, Rapid planetesimal formation in turbulent circumstellar disks. *Nature* **448**, 1022–1025 (2007). doi: 10.1038/nature06086; pmid: 17728751
24. A. Johansen, M. M. Low, P. Lacerda, M. Bizzarro, Growth of asteroids, planetary embryos, and Kuiper belt objects by chondrule accretion. *Sci. Adv.* **1**, e1500109 (2015). doi: 10.1126/sciadv.1500109; pmid: 26601169
25. J. B. Simon, P. J. Armitage, R. Li, A. N. Youdin, The mass and size distribution of planetesimals formed by the streaming instability. I. The role of self-gravity. *Astrophys. J.* **822**, 55 (2016). doi: 10.3847/0004-637X/822/1/55
26. C. P. Abod *et al.*, The mass and size distribution of planetesimals formed by the streaming instability. II. The effect of the radial gas pressure gradient. *Astrophys. J.* **883**, 192 (2019). doi: 10.3847/1538-4357/ab40a3
27. S. C. Kenyon, B. C. Bromley, Coagulation calculations of icy planet formation at 15–150 AU: A correlation between the maximum radius and the slope of the size distribution for trans-Neptunian objects. *Astron. J.* **143**, 63 (2012). doi: 10.1088/0004-6256/143/3/63
28. A. Shannon, Y. Wu, Y. Lithwick, Forming the cold classical Kuiper belt in a light disk. *Astrophys. J.* **818**, 175 (2016). doi: 10.3847/0004-637X/818/2/175
29. G. Groussin *et al.*, The thermal, mechanical, structural, and dielectric properties of cometary nuclei after Rosetta. *Space Sci. Rev.* **215**, 29 (2019). doi: 10.1007/s11214-019-0594-x
30. Short-period, Jupiter-family comets such as 67P are derived mostly from the scattered disk component of the Kuiper Belt (*4*, *14*), which is a distinct, dynamically hot Kuiper Belt population not directly related to CCKB objects. The cold classical region did contribute to the original scattered disk population (*17*, *19*), but we expect this contribution to have been minor compared with the population of the scattered disk as a whole.
31. K. A. Holsapple, K. R. Housen, A crater and its ejecta: An interpretation of Deep Impact. *Icarus* **191**, 586–597 (2007). doi: 10.1016/j.icarus.2006.08.035
32. W. D. MacMillan, *The Theory of the Potential* (McGraw-Hill, 1930).
33. M. Jutzi, E. Asphaug, The shape and structure of cometary nuclei as a result of low-velocity accretion. *Science* **348**, 1355–1358 (2015). doi: 10.1126/science.aaa4747; pmid: 26022415
34. M. Jutzi, W. Benz, A. Toliou, A. Morbidelli, R. Brasser, How primordial is the structure of comet 67P? Combined collisional and dynamical models suggest a late formation. *Astron. Astrophys.* **597**, A61 (2017). doi: 10.1051/0004-6361/201628963
35. T. W. Lambe, R. V. Whitman, *Soil Mechanics* (Wiley, ed. 1, 1969).
36. The steepest slopes are found on one of SL's shoulders, coincident with a prominent trough [figure 1 in (*8*)], possibly a sign of incipient slope failure.
37. C. Matonti *et al.*, Bilobate comet morphology and internal structure controlled by shear deformation. *Nat. Geosci.* **12**, 157–162 (2019). doi: 10.1038/s41561-019-0307-9
38. The strength of the near-surface (<1 m deep) lunar regolith is estimated to be ~1 kPa (*98*), which is dynamically equivalent to several kilometers depth on Arrokoth.
39. J. E. Richardson, K. J. Graves, A. W. Harris, T. J. Bowling, Small body shapes and spins reveal a prevailing state of maximum topographic stability. *Icarus* **329**, 207–221 (2019). doi: 10.1016/j.icarus.2019.03.027
40. S. Greenstreet, B. Gladman, W. B. McKinnon, J. J. Kavelaars, K. N. Singer, Crater density predictions for New Horizons flyby target 2014 MU69. *Astrophys. J.* **872**, L5 (2019). doi: 10.3847/2041-8213/ab01db
41. J. G. Stadel, thesis, University of Washington, Seattle, WA (2001).
42. S. R. Schwartz, D. C. Richardson, P. Michel, An implementation of the soft-sphere discrete element method in a high-performance parallel gravity tree-code. *Granul. Matter* **14**, 363–380 (2012). doi: 10.1007/s10035-012-0346-z
43. Materials and methods are available as supplementary materials.
44. V. S. Safronov, *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets* [transl. NASA TTF-667] (Nauka, Moscow, 1972).
45. S. R. Schwartz *et al.*, Catastrophic disruptions as the origin of bilobate comets. *Nature Astron.* **2**, 379–382 (2018). doi: 10.1038/s41550-018-0395-2
46. J.-L. Margot, P. Pravec, P. Taylor, B. Carry, S. Jacobson, in *Asteroids IV*, P. Michel, F. E. DeMeo, W. F. Bottke, Eds. (Univ. Arizona Press, 2015), pp. 355–374.
47. W. C. Fraser *et al.*, All planetesimals born near the Kuiper belt formed as binaries. *Nature Astron.* **1**, 0088 (2017). doi: 10.1038/s41550-017-0088
48. P. Goldreich, Y. Lithwick, R. Sari, Formation of Kuiper-belt binaries by dynamical friction and three-body encounters. *Nature* **420**, 643–646 (2002). doi: 10.1038/nature01227; pmid: 12478286
49. S. A. Astakhov, E. A. Lee, D. Farrelly, Formation of Kuiper-belt binaries through multiple chaotic scattering encounters with low-mass intruders. *Mon. Not. R. Astron. Soc.* **360**, 401–415 (2005). doi: 10.1111/j.1365-2966.2005.09072.x
50. S. Weidenschilling, On the origin of binary transneptunian objects. *Icarus* **160**, 212–215 (2002). doi: 10.1006/icar.2002.6952
51. For Arrokoth, $R_{Hill} = a_{\odot}(m/3M_{\odot})^{1/3} \sim 4 \times 10^4$ km, where $a_{\odot}$ is Arrokoth's heliocentric distance, m is the mass of Arrokoth (1.6×10^{15} kg for $\rho = 500$ kg m^{-3}), and $M_{\odot}$ is the mass of the Sun. The Hill speed, the Keplerian orbital shear velocity at R_{Hill}, is given by $v_{Hill} = \Omega_K R_{Hill}$ (where Ω_K denotes the orbital frequency of the Keplerian orbit in question), which is ~2 to 3 cm s^{-1} for Arrokoth.
52. H. E. Schlichting, R. Sari, Formation of Kuiper Belt binaries. *Astrophys. J.* **673**, 1218–1224 (2008). doi: 10.1086/524930
53. D. P. Hamilton, J. A. Burns, Orbital stability zones about asteroids. *Icarus* **92**, 118–131 (1991). doi: 10.1016/0019-1035(91)90039-V
54. H. E. Schlichting, R. Sari, The ratio of retrograde to prograde orbits: A test for Kuiper Belt binary formation theories. *Astrophys. J.* **686**, 741–747 (2008). doi: 10.1086/591073
55. W. M. Grundy *et al.*, Mutual orbit orientations of transneptunian binaries. *Icarus* **334**, 62–78 (2019). doi: 10.1016/j.icarus.2019.03.035
56. S. D. Benecchi *et al.*, The correlated colors of transneptunian binaries. *Icarus* **200**, 292–303 (2009). doi: 10.1016/j.icarus.2008.10.025
57. A. Johansen *et al.*, in *Protostars and Planets VI*, H. Beuther, R. S. Klessen, C. P. Dullemond, T. Henning, Eds. (Univ. Arizona Press, 2014), pp. 547–570.
58. D. Nesvorný, A. N. Youdin, D. C. Richardson, Formation of Kuiper belt binaries by gravitational collapse. *Astron. J.* **140**, 785–793 (2010). doi: 10.1088/0004-6256/140/3/785
59. K. Noll, W. Grundy, S. Benecchi, H. Levison, The relative sizes of transneptunian binaries: Evidence for different populations from a homogeneous data set. EPSC-DPS Joint Meeting 2011, 2-7 October 2011 in Nantes, France; https://meetingorganizer.copernicus.org/EPSC-DPS2011/EPSC-DPS2011-1029.pdf.
60. D. Nesvorný, R. Li, A. N. Youdin, J. B. Simon, W. M. Grundy, Trans-Neptunian binaries as evidence for planetesimal formation by the streaming instability. *Nature Astron.* **3**, 808–812 (2019). doi: 10.1038/s41550-019-0806-z
61. J. B. Simon, P. J. Armitage, A. N. Youdin, R. Li, Evidence for universality in the initial planetesimal mass function. *Astrophys. J.* **847**, L12–L17 (2017). doi: 10.3847/2041-8213/aa8c79
62. R. Li, A. N. Youdin, J. B. Simon, On the numerical robustness of the streaming instability: Particle concentration and gas dynamics in protoplanetary disks. *Astrophys. J.* **862**, 14–29 (2018). doi: 10.3847/1538-4357/aaca99
63. D. Carrera, U. Gorti, A. Johansen, M. B. Davies, Planetesimal formation by the streaming instability in a photoevaporating disk. *Astrophys. J.* **839**, 16 (2017). doi: 10.3847/1538-4357/aa6932
64. S. Chandrasekhar, *Ellipsoidal Figures of Equilibrium* (Yale Univ. Press, 1969).
65. P. Descamps, Roche figures of doubly synchronous asteroids. *Planet. Space Sci.* **56**, 1839–1846 (2008). doi: 10.1016/j.pss.2008.02.040
66. M. Ćuk, D. Nesvorný, Orbital evolution of small binary asteroids. *Icarus* **207**, 732–743 (2010). doi: 10.1016/j.icarus.2009.12.005
67. J. Fang, J.-L. Margot, Near-earth binaries and triples: Origin and evolution of spin-orbital properties. *Astron. J.* **143**, 24 (2012). doi: 10.1088/0004-6256/143/1/24
68. S. Naoz, The eccentric Kozai-Lidov effect and its applications. *Annu. Rev. Astron. Astrophys.* **54**, 441–489 (2016). doi: 10.1146/annurev-astro-081915-023315
69. L. A. M. Benner, W. B. McKinnon, Orbital behavior of captured satellites: The effect of solar gravity on Triton's post-capture orbit. *Icarus* **114**, 1–20 (1989). doi: 10.1006/icar.1995.1039
70. H. B. Perets, S. Naoz, Kozai cycles, tidal friction, and the dynamical evolution of binary minor planets. *Astrophys. J.* **699**, L17–L21 (2009). doi: 10.1088/0004-637X/699/1/L17
71. S. B. Porter, W. M. Grundy, KCTF evolution of trans-neptunian binaries: Connecting formation to observation. *Icarus* **220**, 947–957 (2012). doi: 10.1016/j.icarus.2012.06.034
72. We estimate a gravitational quadrupole $J_2 \sim 0.14$ for LL from its shape (*8*), and the semimajor axis where the dynamics transition from shape-driven precession to solar (Kozai-Lidov) influence is $(2J_2M_{LL}R_{LL}^2a_{\odot}^3/M_{\odot})^{1/5}$, where M_{LL} and R_{LL} are the mass and spherical equivalent radius of the large lobe, respectively (*67*).
73. S. D. Benecchi, K. S. Noll, W. M. Grundy, H. F. Levison, (47171) 1999 TC_{36}, A transneptunian triple. *Icarus* **207**, 978–991 (2010). doi: 10.1016/j.icarus.2009.12.017
74. K. J. Walsh, S. A. Jacobson, in *Asteroids IV*, P. Michel, F. E. DeMeo, W. F. Bottke, Eds. (Univ. Arizona Press, 2015), pp. 375–393.
75. D. Vokrouhlický, W. F. Bottke, S. R. Chesley, D. J. Scheeres, T. S. Statler, in *Asteroids IV*, P. Michel, F. E. DeMeo, W. F. Bottke, Eds. (Univ. Arizona Press, 2015), pp. 509–531.
76. M. Ćuk, J. A. Burns, Effects of thermal radiation on the dynamics of binary NEAs. *Icarus* **176**, 418–431 (2005). doi: 10.1016/j.icarus.2005.02.001
77. S. A. Jacobson, F. Marzari, A. Rossi, D. J. Scheeres, Matching asteroid population characteristics with a model constructed from the YORP-induced rotational fission hypothesis. *Icarus* **277**, 381–394 (2016). doi: 10.1016/j.icarus.2016.05.032
78. BYORP tends to reset outwardly migrating satellites so that they turn and migrate inward (*76*), making an ultimate merger of the two bodies the most likely outcome, as long as the secondary can reestablish its synchronous spin. This more complicated evolution depends on increasing orbital

eccentricity as the binary orbit expands, which has been questioned (67).
79. O. Golubov, Yu. N. Krugly, Tangential component of the YORP effect. *Astrophys. J.* **752**, L11 (2012). doi: 10.1088/2041-8205/752/1/L11
80. P. Ševeček, M. Brož, D. Čapek, J. Ďurech, The thermal emission from boulders on (25143) Itokawa and general implications for the YORP effect. *Mon. Not. R. Astron. Soc.* **450**, 2104–2115 (2015). doi: 10.1093/mnras/stv738
81. D. P. Rubincam, Radiative spin-up and spin-down of small asteroids. *Icarus* **148**, 2–11 (2000). doi: 10.1006/icar.2000.6485
82. M. C. Nolan *et al*., Detection of rotational acceleration of Bennu using HST light curve observations. *Geophys. Res. Lett.* **46**, 1956–1962 (2019). doi: 10.1029/2018GL080658
83. C. D. Murray, S. F. Dermott, *Solar System Dynamics* (Cambridge Univ. Press, 2000).
84. P. Goldreich, R. Sari, Tidal evolution of rubble piles. *Astrophys. J.* **691**, 54–60 (2009). doi: 10.1088/0004-637X/691/1/54
85. The critical rotation rate (or surface orbital frequency) of a rigid sphere is given by $\sqrt{GM/R^3}$, where M and R are its mass and radius, respectively. This approximately corresponds to the maximum specific angular momentum of a rotating, strengthless body (a Jacobi ellipsoid) (*99*).
86. The second-degree Love number k_2 for a porous, granular SL is given by ~10^{-5} × SL's effective radius in kilometers (*84*), and the tidal dissipation factor Q for SL is taken to be 10 to 100.
87. D. Nesvorný, J. Parker, D. Vokrouhlický, Bi-lobed shape of comet 67P from a collapsed binary. *Astron. J.* **155**, 246 (2018). doi: 10.3847/1538-3881/aac01f
88. H. B. Perets, R. A. Murray-Clay, Wind-shearing in gaseous protoplanetary disks and the evolution of binary planetesimals. *Astrophys. J.* **733**, 56 (2011). doi: 10.1088/0004-637X/733/1/56
89. F. L. Whipple, in *From Plasma to Planet, Proc. Twenty-First Nobel Symposium*, A. Evlius, Ed. (Wiley Interscience, 1972), pp. 211–232.
90. S. J. Desch, Mass distribution and planet formation in the solar nebula. *Astrophys. J.* **671**, 878–893 (2007). doi: 10.1086/522825
91. H. Wang *et al*., Lifetime of the solar nebula constrained by meteorite paleomagnetism. *Science* **355**, 623–627 (2017). doi: 10.1126/science.aaf5043; pmid: 28183977
92. S. J. Weidenschilling, Aerodynamics of solid bodies in the solar nebula. *Mon. Not. R. Astron. Soc.* **180**, 57–70 (1977). doi: 10.1093/mnras/180.2.57
93. The mean free path in the nebular gas is $\lambda = 1/n\sigma_H$, where σ_H is the collisional cross section of H_2 (2×10^{-19} m^2) and the number density $n \equiv \rho_g/\mu m_H$ (with m_H the mass of a hydrogen atom and $\mu = 2.3$ for solar composition gas). For the midplane density quoted in the text, we found $\lambda \sim 0.17$ km, which puts Arrokoth's gas drag interactions into the fluid (Stokes-like) regime. The kinematic viscosity is then $\lambda \times$ sound speed, which for cold, 30 K nebular gas (*90*, *100*) is 7×10^4 m^2 s^{-1}, independent of ρ_g.
94. P. R. Estrada, J. N. Cuzzi, D. Morgan, Global modeling of nebulae with particle growth, drift, and evaporation fronts. I. Methodology and typical results. *Astrophys. J.* **818**, 200 (2016). doi: 10.3847/0004-637X/818/2/200
95. F. C. Adams, D. Hollenbach, G. Laughlin, U. Gorti, Photoevaporation of circumstellar disks due to external far-ultraviolet radiation in stellar aggregates. *Astrophys. J.* **611**, 360–379 (2004). doi: 10.1086/421989
96. Total, upper limit merger times, if Arrokoth's original orbit extended to its outer Hill sphere (~10^4 km), would have been a few times longer.
97. C. Hayashi, Structure of the solar nebula, growth and decay of magnetic fields and effects of magnetic and turbulent viscosities on the nebula. *Prog. Theor. Phys. Suppl.* **70**, 35–53 (1981). doi: 10.1143/PTPS.70.35
98. J. K. Mitchell, L. G. Bromwell, W. D. Carrier III, N. C. Costes, R. F. Scott, Soil mechanical properties at the Apollo 14 site. *J. Geophys. Res.* **77**, 5641–5664 (1972). doi: 10.1029/JB077i029p05641
99. P. Descamps, F. Marchis, Angular momentum of binary asteroids: Implications for their possible origin. *Icarus* **193**, 74–84 (2008). doi: 10.1016/j.icarus.2007.07.024
100. E. I. Chiang, P. Goldreich, Spectral energy distributions of T Tauri stars with passive circumstellar disks. *Astrophys. J.* **490**, 368–376 (1997). doi: 10.1086/304869

ACKNOWLEDGMENTS

These results would not have been possible without NASA and the efforts by the New Horizons search, occultation, encounter, and navigation teams to discover, locate, and precisely rendezvous with Arrokoth. We thank the reviewers for their perceptive comments and E. Asphaug and M. Jutzi for informative discussions. **Funding:** This research was supported by NASA's New Horizons project through contracts NASW-02008 and NAS5-97271/TaskOrder30. J.C.M. and D.C.R. were supported by NASA Solar System Workings grant NNX15AH90G, and J.C.M. was supported by a University of Maryland Graduate School Research and Scholarship Award. J.J.K. was supported by the National Research Council of Canada. **Author contributions:** W.B.M. led the study and wrote the paper with D.C.R., J.C.M., J.T.K., D.P.H., O.M.U., and W.M.G., with inputs from S.A.S., D.N., T.R.L., K.N.S., and H.A.W.; J.C.M. and D.C.R. performed the PKDGRAV collisional simulations; S.B.P. led the development of the shape model; X.M. calculated the ellipsoidal gravity; W.B.M., J.T.K., J.C.M., and D.C.R. produced the figures; M.W.B., J.R.S., M.R.S., C.M.L., J.J.K., J.W.P., and A.H.P. contributed to the discussion; D.P.C., H.A.E., G.R.G., and J.M.M. lead the New Horizons Science Theme Teams; and S.A.S., H.A.W., J.W.P., C.B.O., K.N.S., A.J.V., and L.A.Y. are lead scientists of the New Horizons project. The entire New Horizons Science Team contributed to the success of the Arrokoth encounter. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Executables of our modified PKDGRAV code, along with input and output files for the results presented in this paper, are available at https://doi.org/10.6084/m9.figshare.11653167.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/eaay6620/suppl/DC1
Team Members and Affiliations
Materials and Methods
Supplementary Text
Table S1
References (*101*–*118*)

8 July 2019; accepted 27 January 2020
Published online 13 February 2020
10.1126/science.aay6620

RESEARCH ARTICLE SUMMARY

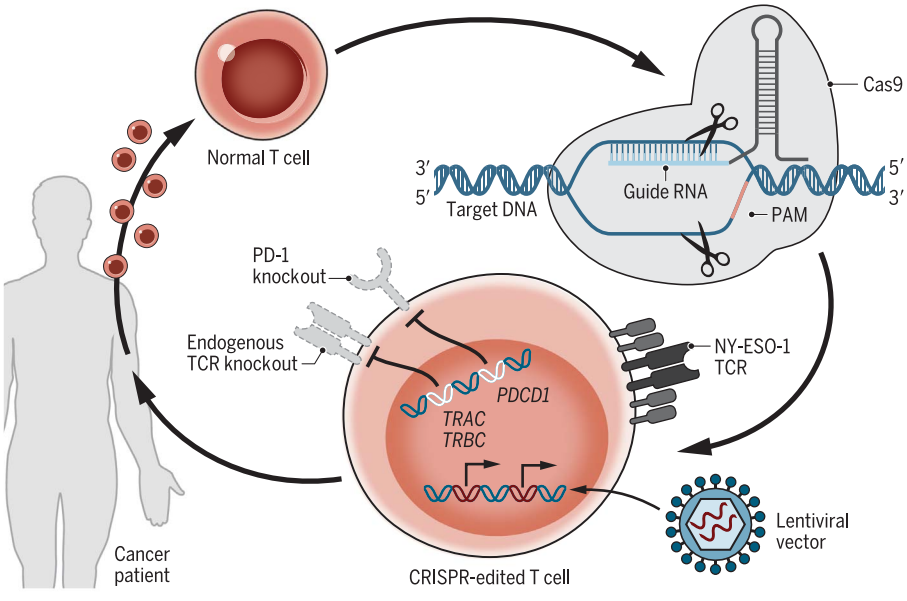
CLINICAL TRIALS

CRISPR-engineered T cells in patients with refractory cancer

Edward A. Stadtmauer*†, Joseph A. Fraietta*, Megan M. Davis, Adam D. Cohen, Kristy L. Weber, Eric Lancaster, Patricia A. Mangan, Irina Kulikovskaya, Minnal Gupta, Fang Chen, Lifeng Tian, Vanessa E. Gonzalez, Jun Xu, In-young Jung, J. Joseph Melenhorst, Gabriela Plesa, Joanne Shea, Tina Matlawski, Amanda Cervini, Avery L. Gaymon, Stephanie Desjardins, Anne Lamontagne, January Salas-Mckee, Andrew Fesnak, Donald L. Siegel, Bruce L. Levine, Julie K. Jadlowsky, Regina M. Young, Anne Chew, Wei-Ting Hwang, Elizabeth O. Hexner, Beatriz M. Carreno, Christopher L. Nobles, Frederic D. Bushman, Kevin R. Parker, Yanyan Qi, Ansuman T. Satpathy, Howard Y. Chang, Yangbing Zhao, Simon F. Lacey*, Carl H. June*†

INTRODUCTION: Most cancers are recognized and attacked by the immune system but can progress owing to tumor-mediated immunosuppression and immune evasion mechanisms. The infusion of ex vivo engineered T cells, termed adoptive T cell therapy, can increase the natural antitumor immune response of the patient. Gene therapy to redirect immune specificity combined with genome editing has the potential to improve the efficacy and increase the safety of engineered T cells. CRISPR coupled with CRISPR-associated protein 9 (Cas9) endonuclease is a powerful gene-editing technology that potentially allows the ability to target multiple genes in T cells to improve cancer immunotherapy.

RATIONALE: Our first-in-human, phase 1 clinical trial (clinicaltrials.gov; trial NCT03399448) was designed to test the safety and feasibility of multiplex CRISPR-Cas9 gene editing of T cells from patients with advanced, refractory cancer. A limitation of adoptively transferred T cell efficacy has been the induction of T cell dysfunction or exhaustion. We hypothesized that removing the endogenous T cell receptor (TCR) and the immune checkpoint molecule programmed cell death protein 1 (PD-1) would improve the function and persistence of engineered T cells. In addition, the removal of PD-1 has the potential to improve safety and reduce toxicity that can be caused by autoimmunity.



CRISPR-Cas9 engineering of T cells in cancer patients. T cells (center) were isolated from the blood of a patient with cancer. CRISPR-Cas9 ribonuclear protein complexes loaded with three sgRNAs were electroporated into the normal T cells, resulting in gene editing of the *TRAC*, *TRBC1*, *TRBC2*, and *PDCD1* (encoding PD-1) loci. The cells were then transduced with a lentiviral vector to express a TCR specific for the cancer-testis antigens NY-ESO-1 and LAGE-1 (right). The engineered T cells were then returned to the patient by intravenous infusion, and patients were monitored to determine safety and feasibility. PAM, protospacer adjacent motif.

A synthetic, cancer-specific TCR transgene (NY-ESO-1) was also introduced to recognize tumor cells. In vivo tracking and persistence of the engineered T cells were monitored to determine if the cells could persist after CRISPR-Cas9 modifications.

RESULTS: Four cell products were manufactured at clinical scale, and three patients (two with advanced refractory myeloma and one with metastatic sarcoma) were infused. The editing efficiency was consistent in all four products and varied as a function of the single guide RNA (sgRNA), with highest efficiency observed for the TCR α chain gene (*TRAC*) and lowest efficiency for the TCR β chain gene (*TRBC*). The mutations induced by CRISPR-Cas9 were highly specific for the targeted loci; however, rare off-target edits were observed. Single-cell RNA sequencing of the infused CRISPR-engineered T cells revealed that ~30% of cells had no detectable mutations, whereas ~40% had a single mutation and ~20 and ~10% of the engineered T cells were double mutated and triple mutated, respectively, at the target sequences. The edited T cells engrafted in all three patients at stable levels for at least 9 months. The persistence of the T cells expressing the engineered TCR was much more durable than in three previous clinical trials during which T cells were infused that retained expression of the endogenous TCR and endogenous PD-1. There were no clinical toxicities associated with the engineered T cells. Chromosomal translocations were observed in vitro during cell manufacturing, and these decreased over time after infusion into patients. Biopsies of bone marrow and tumor showed trafficking of T cells to the sites of tumor in all three patients. Although tumor biopsies revealed residual tumor, in both patients with myeloma, there was a reduction in the target antigens NY-ESO-1 and/or LAGE-1. This result is consistent with an on-target effect of the engineered T cells, resulting in tumor evasion.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aba7365

CONCLUSION: Preliminary results from this pilot trial demonstrate that multiplex human genome engineering is safe and feasible using CRISPR-Cas9. The extended persistence of the engineered T cells indicates that preexisting immune responses to Cas9 do not appear to present a barrier to the implementation of this promising technology. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: edward.stadtmauer@pennmedicine.upenn.edu (E.A.S.); cjune@upenn.edu (C.H.J.)
Cite this article as E. A. Stadtmauer *et al.*, *Science* **367**, eaba7365 (2020). DOI: 10.1126/science.aba7365

RESEARCH ARTICLE

CLINICAL TRIALS

CRISPR-engineered T cells in patients with refractory cancer

Edward A. Stadtmauer[1,2]*†, Joseph A. Fraietta[2,3,4,5,6]*, Megan M. Davis[5,6], Adam D. Cohen[1,2], Kristy L. Weber[2,7], Eric Lancaster[8], Patricia A. Mangan[1], Irina Kulikovskaya[5], Minnal Gupta[5], Fang Chen[5], Lifeng Tian[5], Vanessa E. Gonzalez[5], Jun Xu[5], In-young Jung[4,5], J. Joseph Melenhorst[3,5,6], Gabriela Plesa[5], Joanne Shea[5], Tina Matlawski[5], Amanda Cervini[5], Avery L. Gaymon[5], Stephanie Desjardins[5], Anne Lamontagne[5], January Salas-Mckee[5], Andrew Fesnak[5,6], Donald L. Siegel[5,6], Bruce L. Levine[5,6], Julie K. Jadlowsky[5], Regina M. Young[5], Anne Chew[5], Wei-Ting Hwang[9], Elizabeth O. Hexner[1,2], Beatriz M. Carreno[3,5,6], Christopher L. Nobles[4], Frederic D. Bushman[4], Kevin R. Parker[10], Yanyan Qi[11], Ansuman T. Satpathy[10,11], Howard Y. Chang[10,12], Yangbing Zhao[5,6], Simon F. Lacey[5,6]*, Carl H. June[2,3,5,6]*†

CRISPR-Cas9 gene editing provides a powerful tool to enhance the natural ability of human T cells to fight cancer. We report a first-in-human phase 1 clinical trial to test the safety and feasibility of multiplex CRISPR-Cas9 editing to engineer T cells in three patients with refractory cancer. Two genes encoding the endogenous T cell receptor (TCR) chains, TCRα (*TRAC*) and TCRβ (*TRBC*), were deleted in T cells to reduce TCR mispairing and to enhance the expression of a synthetic, cancer-specific TCR transgene (NY-ESO-1). Removal of a third gene encoding programmed cell death protein 1 (PD-1; *PDCD1*), was performed to improve antitumor immunity. Adoptive transfer of engineered T cells into patients resulted in durable engraftment with edits at all three genomic loci. Although chromosomal translocations were detected, the frequency decreased over time. Modified T cells persisted for up to 9 months, suggesting that immunogenicity is minimal under these conditions and demonstrating the feasibility of CRISPR gene editing for cancer immunotherapy.

Gene editing offers the potential to correct DNA mutations and may offer promise to treat or eliminate countless human genetic diseases. The goal of gene editing is to change the DNA of cells with single–base pair precision. The principle was first demonstrated in mammalian cells when it was shown that expression of a rare cutting endonuclease to create double-strand DNA breaks resulted in repair by homologous and nonhomologous recombination (*1*). A variety of engineered nucleases were then developed to increase efficiency and enable potential therapeutic applications, including zinc finger nucleases, homing endonucleases, transcription activator–like effector nucleases, and CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats associated with Cas9 endonuclease) (*2*). The first pilot human trials using genome editing were conducted in patients with HIV/AIDS and targeted the white blood cell protein CCR5, with the goal of mutating the *CCR5* gene by nonhomologous recombination and thereby inducing resistance to HIV infection (*3*, *4*). The incorporation of multiple guide sequences in CRISPR-Cas9 permits, in principle, multiplex genome engineering at several sites within a mammalian genome (*5*–*9*). The ability of CRISPR to facilitate efficient multiplex genome editing has greatly expanded the scope of possible targeted genetic manipulations, enabling new possibilities such as simultaneous deletion or insertion of multiple DNA sequences in a single round of mutagenesis. The prospect of using CRISPR engineering to treat a host of diseases, such as inherited blood disorders and blindness, is moving closer to reality.

Recent advances in CRISPR-Cas9 technology have also permitted efficient DNA modifications in human T cells, which holds great promise for enhancing the efficacy of cancer therapy. T lymphocytes are specialized immune cells that are largely at the core of the modern-day cancer immunotherapy revolution. The T cell receptor (TCR) complex is located on the surface of T cells and is central for initiating successful antitumor responses by recognizing foreign antigens and peptides bound to major histocompatibility complex molecules. One of the most promising areas of cancer immunotherapy involves adoptive cell therapy, whereby the patient's own T cells are genetically engineered to express a synthetic (transgenic) TCR that can specifically detect and kill tumor cells. Recent studies have shown safety and promising efficacy of such adoptive T cell transfer approaches using transgenic TCRs specific for the immunogenic NY-ESO-1 tumor antigen in patients with myeloma, melanoma, and sarcoma (*10*–*12*). One limitation of this approach is that the transgenic TCR has been shown to mispair and/or compete for expression with the α and β chains of the endogenous TCR (*13*–*15*). Mispairing of the therapeutic TCR α and β chains with endogenous α and β chains reduces therapeutic TCR cell surface expression and potentially generates self-reactive TCRs.

A further shortcoming of adoptively transferred T cells has been the induction of T cell dysfunction or exhaustion leading to reduced efficacy (*16*). Programmed cell death protein 1 (PD-1)–deficient allogeneic mouse T cells with transgenic TCRs showed enhanced responses to alloantigens, indicating that the PD-1 protein on T cells plays a negative regulatory role in antigen responses that are likely to be cell intrinsic (*17*). The adoptive transfer of PD-1–deficient T cells in mice with chronic lymphocytic choriomeningitis virus infection initially leads to enhanced cytotoxicity and later to enhanced accumulation of terminally differentiated T cells (*18*). Antibody blockade of PD-1, or disruption or knockdown of the gene encoding PD-1 (i.e., *PDCD1*), improved chimeric antigen receptor (CAR) or TCR T cell–mediated killing of tumor cells in vitro and enhanced clearance of PD-1 ligand–positive (PD-L1$^+$) tumor xenografts in vivo (*19*–*23*). In preclinical studies, we and others found that CRISPR-Cas9–mediated disruption of *PDCD1* in human T cells transduced with a CAR increased antitumor efficacy in tumor xenografts (*24*–*26*). Adoptive transfer of transgenic TCR T cells specific for the cancer antigen NY-ESO-1, in combination with a monoclonal antibody targeting PD-1, enhanced antitumor efficacy in mice (*27*). We therefore designed a first-in-human, phase 1 human clinical trial to test the safety and feasibility of multiplex CRISPR-Cas9 genome editing for a synthetic biology cancer immunotherapy application. We chose to target endogenous *TRAC*, *TRBC*, and *PDCD1*

[1]Division of Hematology-Oncology, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [2]Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [3]Parker Institute for Cancer Immunotherapy, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [4]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [5]Center for Cellular Immunotherapies, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [6]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [7]Department of Orthopaedic Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [8]Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [9]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [10]Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA. [11]Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA. [12]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: edward.stadtmauer@pennmedicine.upenn.edu (E.A.S.); cjune@upenn.edu (C.H.J.)

on T cells to increase the safety and efficacy profile of NY-ESO-1 TCR–expressing engineered cells. In principle, this strategy allowed us to increase exogenous TCR expression and reduce the potential for mixed heterodimer formation (i.e., by deleting the α and β TCR domain genes *TRAC* and *TRBC*, respectively) and to limit the development of T cell exhaustion, which can be triggered by the checkpoint ligands PD-L1 and PD-L2 (i.e., by deleting *PDCD1*).

Results

Clinical protocol

The phase 1 human trial (clinicaltrials.gov; trial NCT03399448) was designed to assess the safety and feasibility of infusing autologous NY-ESO-1 TCR–engineered T cells in patients after CRISPR-Cas9 editing of the *TRAC*, *TRBC*, and *PDCD1* loci. During the manufacturing process, cells were taken out of the cancer patient, engineered, and then infused back into the individual. The genetically engineered T cell product was termed "NYCE" (NY-ESO-1–transduced CRISPR 3X edited cells) and is referred to as NYCE hereafter. During clinical development of the protocol, we elected to use a TCR rather than a CAR because the incidence of cytokine release syndrome is generally less prevalent using TCRs (*11*). In principle, this allowed a more discriminating assessment of whether gene editing with Cas9 was potentially immunogenic or toxic when compared with the baseline low level of adverse events observed in our previous clinical trial targeting NY-ESO-1 with transgenic TCRs (*11*). The autologous T cells were engineered by lentiviral transduction to express an HLA-A2*0201–restricted TCR specific for the SLLMWITQC peptide in NY-ESO-1 and LAGE-1. The manufacturing process, vector design, and clinical protocol for NYCE T cells are described in the materials and methods and are depicted schematically (figs. S1 and S2). Of the six patients who were initially enrolled, four patients had successfully engineered T cells that were subjected to detailed release criteria testing as specified in the U.S. Food and Drug Administration (FDA)–accepted Investigational New Drug application (table S1) (see fig. S3 for the consort diagram). Of the four patients with cell products available, one patient assigned unique patient number (UPN) 27 experienced rapid clinical progression and was no longer eligible for infusion owing to the inability to meet protocol-mandated safety criteria (see supplementary materials). Of the three patients who were infused with CRISPR-Cas9–engineered T cells, two patients had refractory advanced myeloma and one patient had a refractory metastatic sarcoma not responding to multiple prior therapies (Table 1). The patients were given lymphodepleting chemotherapy with cyclophosphamide and fludarabine on days −5 to −3 (i.e., before administration with CRISPR-Cas9–engineered T cells) and a single infusion of 1×10^8 manufactured CRISPR-Cas9–engineered T cells per kilogram on day 0 of the protocol (fig. S2). No cytokines were administered to the patients.

Characteristics of infused CRISPR-Cas9–engineered T cell products

The T cell product was manufactured by electroporation of ribonucleoprotein complexes (RNPs) comprising recombinant Cas9 loaded with equimolar mixtures of single guide RNA (sgRNA) for *TRAC*, *TRBC*, and *PDCD1* followed by lentiviral transduction of the transgenic TCR (Fig. 1A). All products were expanded to $>1 \times 10^{10}$ T cells by the time of harvest (Fig. 1B). The transgenic TCR could be detected by

Table 1. Patient demographics and date of engineered T cell infusion. MM, multiple myeloma; BM, bone marrow; XRT, radiation therapy; ASCT, autologous hematopoietic stem cell transplant; ND, not done.

Subject ID (UPN) and infusion date	Sex and age	Diagnosis	Clinical sites	Prior therapy	Prior transplant or surgery	LAGE-1*, NY-ESO-1*, NY-ESO-1**
UPN35 7 January 2019	Female 66 years	Immunoglobulin G kappa MM 2008	BM, lytic bone lesions	Lenalidomide, pomalidomide, bortezomib, carfilzomib, daratumumab, panobinostat, etc. (eight lines of therapy; see supplementary materials)	Three ASCTs	Positive, positive, negative
UPN39 18 March 2019	Male 66 years	Myxoid and round cell liposarcoma 2012	Abdominal and pelvic masses	Doxorubicin, ifosfamide, XRT 60 gray, trabectedin, gemcitabine, taxol, XRT	Resection and debulking twice, left nephrectomy and partial sigmoid resection	ND, ND, positive
UPN07 5 August 2019	Female 62 years	Kappa light chain MM 2009	BM, lytic bone lesions	Lenalidomide, pomalidomide, bortezomib, carfilzomib, daratumumab, anti-CD38 immunoconjugate (six lines of therapy; see supplementary materials)	Two ASCTs	Positive, positive, negative

*qPCR **Immunohistochemistry

flow cytometric staining for Vβ8.1 or dextramer staining, ranging from 2 to 7% of T cells in the final product (Fig. 1C). The frequency of editing, as determined by digital polymerase chain reaction (PCR), varied according to the sgRNA and was about 45% for *TRAC*, 15% for *TRBC*, and 20% for *PDCD1* (Fig. 1D). Final product transduction efficiency, CD4:CD8 ratio, and dosing are shown in table S2.

The potency of the final engineered T cells was assessed by coculture with HLA-A2$^+$ tumor cells engineered to express NY-ESO-1 (Fig. 2A). The engineered T cells had potent antigen-specific cytotoxicity over a wide range of effector-to–target cell ratios. Interestingly, the cells treated with CRISPR-Cas9 were more cytotoxic than control cells transduced with the TCR but electroporated without CRISPR-Cas9 (i.e., cells that retained endogenous TCR). This is consistent with previous findings in mouse T cells, when a transgenic TCR was inserted into the endogenous locus, ablating expression of the endogenous TCR (*15*). Further studies will be required to determine if PD-1 knockout contributes to the increased potency afforded by knockout of the endogenous TCR.

We developed a sensitive immunoassay for detection of *Streptococcus pyogenes* Cas9 protein and quantified Cas9 early in the manufacturing process, showing declining levels that were <0.75 fg per cell in the harvested final product (Fig. 2C). Using a competitive fluorescence enzyme-linked immunosorbent assay (ELISA) screen, we found that healthy donors have humoral reactivity to Cas9 in serum (data not shown) and T cells (Fig. 2E), confirming previous reports (*28*–*30*). Interestingly, we found that the three patients tested at a variety of time points after infusion of the engineered T cells did not develop humoral responses to Cas9. The lack of immunization to Cas9 is consistent with the extended persistence of the infused cells (Fig. 3) and could be a consequence of the low content of Cas9 in the infused product and/or to the immunodeficiency in the patients as a result of their extensive previous treatment histories (Table 1).

Engraftment and persistence of infused CRISPR-Cas9–engineered T cells in cancer patients

Three patients with advanced, refractory cancer were given infusions of the CRISPR-Cas9–engineered T cells. The infusions were well tolerated, with no serious adverse events (Table 2); importantly, there were no cases of cytokine release syndrome, which is a potentially life-threatening systemic inflammatory response that has been associated with cancer immunotherapies (*31*). All three patients

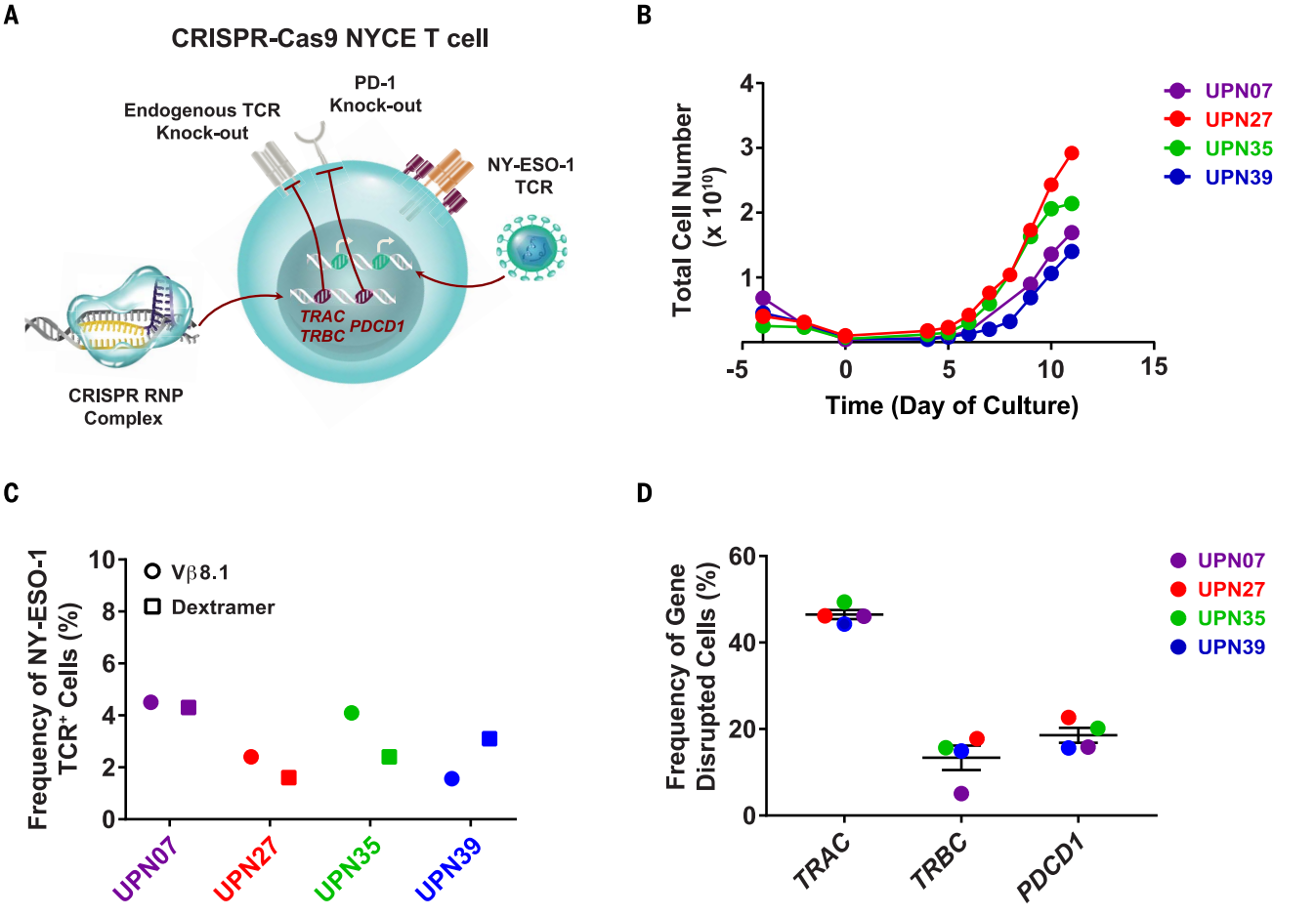


Fig. 1. Feasibility of CRISPR-Cas9 NYCE T cell engineering. (**A**) Schematic representation of CRISPR-Cas9 NYCE T cells. (**B**) Large-scale expansion of NYCE T cells. Autologous T cells were transfected with Cas9 protein complexed with sgRNAs (RNP complex) against *TRAC*, *TRBC* (i.e., endogenous TCR deletion), and *PDCD1* (i.e., PD-1 deletion) and subsequently transduced with a lentiviral vector to express a transgenic NY-ESO-1 cancer-specific TCR. Cells were expanded in dynamic culture for 8 to 12 days. On the final day of culture, NYCE T cells were harvested and cryopreserved in infusible medium. The total number of enriched T cells during culture is plotted for all four subjects (UPN07, UPN27, UPN35, and UPN39). (**C**) NY-ESO-1 TCR transduction efficiency was determined in harvested infusion products by flow cytometry. Data are gated on live CD3-expressing and Vβ8.1- or dextramer-positive lymphocytes and further gated on CD4-positive and/or CD8-positive cells. (**D**) The frequencies of *TRAC*, *TRBC*, and *PDCD1* gene-disrupted total cells in NYCE infusion products were measured using chip-based digital PCR. All data are representative of at least two independent experiments. Error bars represent mean ± SEM.

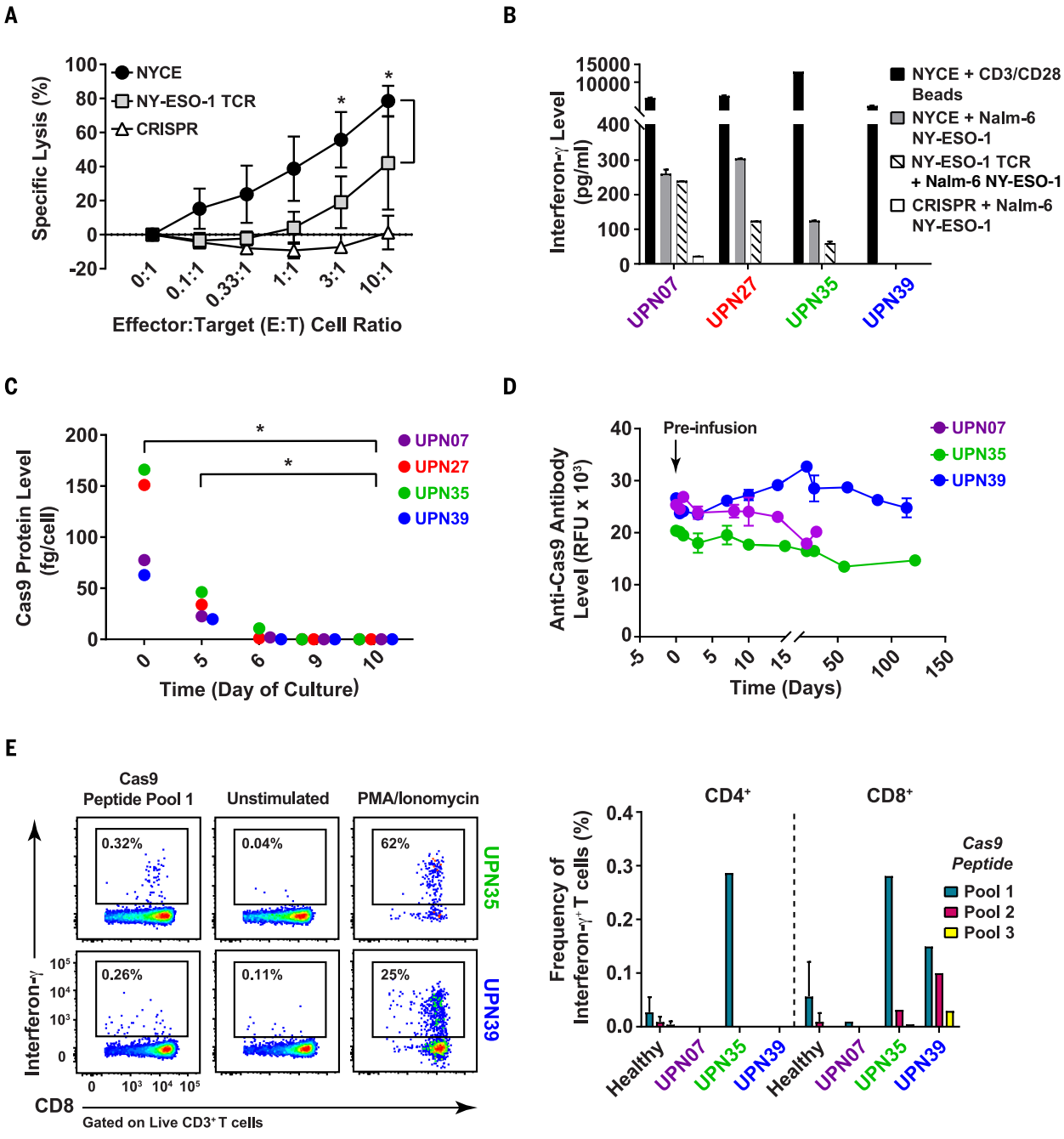


Fig. 2. Potency and immunogenicity of CRISPR-Cas9 engineered T cells. (**A**) Cytotoxicity of NYCE T cells cocultured with HLA-A*0201–positive Nalm-6 tumor cells engineered to express NY-ESO-1 and luciferase. Patient T cells transduced with the NY-ESO-1 TCR without CRISPR-Cas9 editing (NY-ESO-1 TCR) and untransduced T cells with CRISPR-Cas9 editing of *TRAC*, *TRBC*, and *PDCD1* (labeled CRISPR) were included as controls (n = 4 patient T cell infusion products). Asterisks indicate statistical significance determined by paired Student's t tests between groups (*P < 0.05). Error bars represent SEM. (**B**) Levels of soluble interferon-γ produced by patient NYCE T cell infusion products (labeled NYCE) after a 24-hour coculture with anti-CD3 and anti-CD28 antibody-coated beads or NY-ESO-1–expressing Nalm-6 target cells. Patient NY-ESO-1 TCR–transduced T cells (NY-ESO-1 TCR) and untransduced, CRISPR-Cas9–edited T cells (labeled CRISPR) served as controls. Error bars represent SEM. (**C**) Quantification of residual Cas9 protein in NYCE T cell infusion products in clinical-scale manufacturing is shown over time. Asterisks indicate statistical significance determined by paired Student's t tests between time points (*P < 0.05). (**D**) Results from the fluorescence-based indirect ELISA screen performed to detect antibodies against Cas9 protein in the sera of three patients treated with NYCE T cells. Each dot represents the amount of anti-Cas9 signals detected in patient serum before T cell infusion (indicated by a vertical black arrow) and at various time points after NYCE T cell transfer. RFU, relative fluorescent units. (**E**) Immunoreactive Cas9-specific T cells in baseline patient leukapheresis samples were detected. Representative flow cytometry plots (left) from two patients whose T cells were positive for interferon-γ in response to Cas9 peptide stimulation. Unstimulated T cells treated with vehicle alone (dimethyl sulfoxide, DMSO) served as a negative control, whereas matched T cells stimulated with phorbol myristate acetate (PMA) and ionomycin served as a positive control. Bar graphs (right) show the frequency of ex vivo CD4⁺ and CD8⁺ T cells from patients or healthy donor controls (n = 6) that secrete interferon-γ in response to stimulation with three different Cas9 peptide pools. The background frequency of interferon-γ–expressing T cells (unstimulated control group, DMSO alone) is subtracted from the values shown in the bar graph. Error bars represent SD.

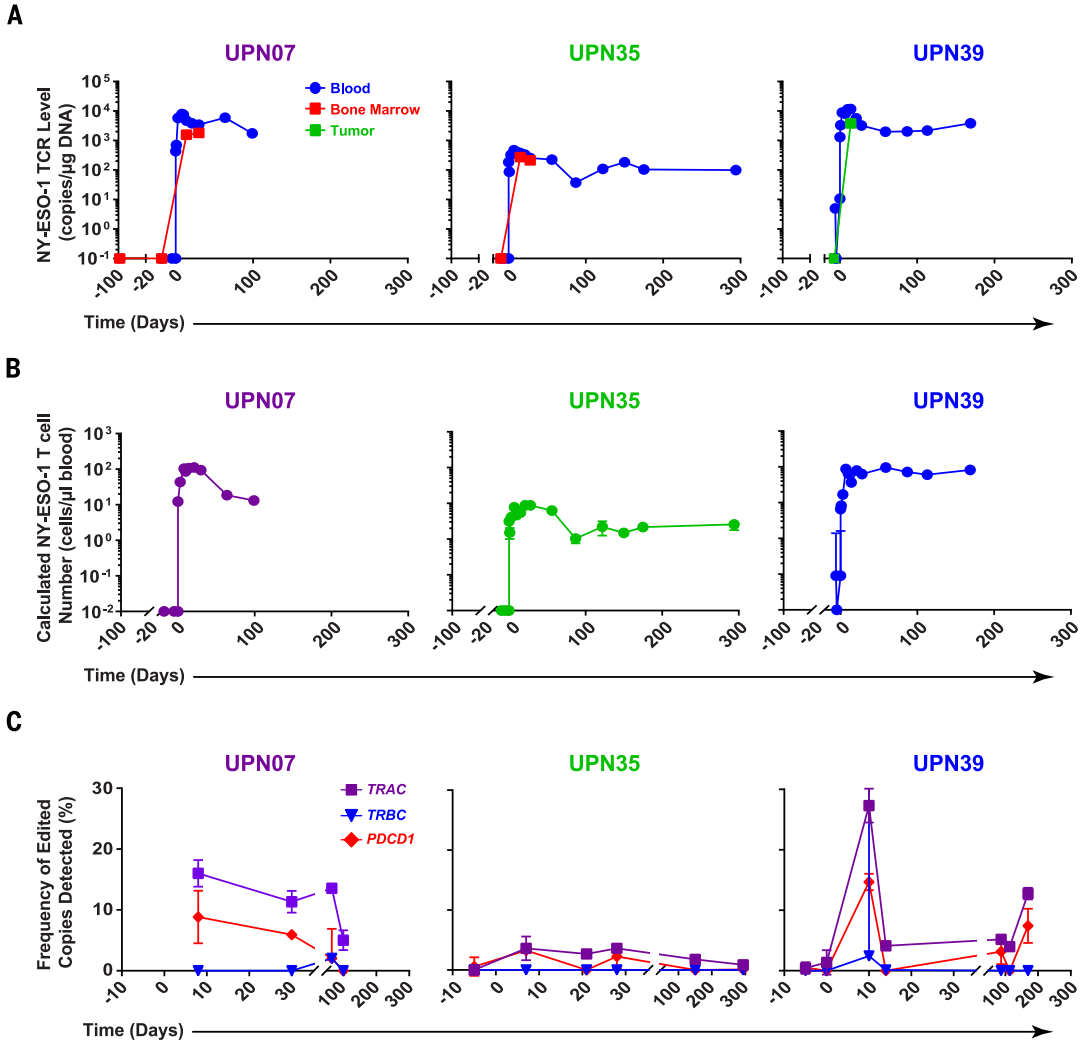


Fig. 3. Sustained in vivo expansion and persistence of CRISPR-Cas9–engineered T cells in patients. (**A**) The total number of vector copies per microgram of genomic DNA of the NY-ESO-1 TCR transgene in the peripheral blood (UPN07, UPN35, and UPN39), bone marrow (UPN07 and UPN35; multiple myeloma), and tumor (UPN39; sarcoma) is shown pre– and post–NYCE T cell infusion. (**B**) Calculated absolute numbers of NY-ESO-1 TCR–expressing T cells per microliter of whole blood from the time of infusion to various postinfusion time points in the study are shown. The limit of detection is about 2.5 cells per microliter of whole blood. (**C**) Frequencies of CRISPR-Cas9–edited T cells (*TRAC*, *TRBC*, and *PDCD1* knockout) before and after adoptive cell transfer are depicted. Error bars represent SD.

were infused with 1×10^8 cells/kg, and, owing to the considerable variation in TCR transduction efficiencies (table S2), the absolute number of infused engineered T cells ranged from 6.0×10^7 to 7.1×10^8 cells. Despite the variation in engineered cells, there were high peak levels and sustained persistence of the engineered cells in the blood of all three patients (Fig. 3A). The peak and steady-state levels of engineered cells were lowest in patient UPN35, who also had the lowest transduction efficiency (table S2). The persistence of the transduced cells is notably stable from 3 to 9 months after infusion, varying from 5 to 50 cells per microliter of blood (Fig. 3B). Using a subject-specific piecewise linear model, the decay half-lives of the transduced cells were 20.3, 121.8, and 293.5 days for UPN07, UPN35, and UPN39, respectively. The average decay half-life was 83.9 days (15 to 153 days, 95% confidence interval) for the three subjects, as estimated by a piecewise linear mixed-effects model that assumes cells decay linearly from day 14 postexpansion and random effects to allow varying level of expansion (or peak values) across subjects. The stable engraftment of our engineered T cells is notably different from previously reported trials with NY-ESO-1 TCR–engineered T cells, in which the half-life of the cells in blood was ~1 week (*11*, *32*, *33*). Biopsy specimens of bone marrow in the myeloma patients and tumor in the sarcoma patient demonstrated trafficking of the engineered T cells to the tumor in all three patients at levels approaching those in the blood compartment (Fig. 3A).

To determine the engraftment frequency of the CRISPR-Cas9 gene-edited cells, we initially used chip-based digital PCR. With this assay, engraftment of cells with editing at the *TRAC* and *PDCD1* loci was evident in all three patients (Fig. 3C). There was sustained persistence of *TRAC* and *PDCD1* edits in patients UPN39 and UPN07 at frequencies of 5 to 10% of circulating peripheral blood mononuclear cells (PBMCs), whereas *TRBC*-edited cells were lowest in frequency and only transiently detected. The low-level engraftment of *TRBC*-edited cells is likely related to the observation that this locus had the lowest level of editing efficiency in our preclinical studies (*25*) and in the harvested products (Fig. 1D).

Analysis of the fidelity of CRISPR-Cas9 genome editing

On- and off-target editing efficiency was assessed in the NYCE cells at the end of product manufacturing. Details of the analysis for UPN07 are shown as an example in Fig. 4, with detailed analysis of the other three manufactured products shown in table S3. The average on-target CRISPR-Cas9 editing efficiency for all engineered T cell products for each target is shown in Table 3. We used iGUIDE (*34*), a modification of the GUIDE sequencing (GUIDE-seq) method (*35*), to analyze the Cas9-mediated cleavage specificity. A complication of assays to assess repair by nonhomologous end joining (NHEJ) is that DNA double-strand breaks are formed spontaneously during cell division at high rates in the absence of added nucleases (*36*), which can

Table 2. List of adverse events in the study. "–" indicates no adverse event.

Adverse events category	Toxicity	All grades	Grade 1 or 2	Grade 3 or 4
Hematologic	Anemia	2	1	1
	Leukopenia	4	–	4
	Neutropenia	4	1	3
	Thrombocytopenia	6	3	3
	Lymphopenia	1	–	1
Infection	Upper respiratory	1	1	–
	Febrile neutropenia	2	–	2
Electrolyte	Hypercalcemia	1	1	–
	Hyperphosphatemia	1	1	–
	Hypoalbuminemia	1	1	–
	Hypocalcemia	3	2	1
	Hypokalemia	1	1	–
	Hypomagnesemia	1	1	–
	Hyponatremia	1	1	–
	Hypophosphatemia	1	–	1
Neurologic	Dysgeusia	1	1	–
	Headache	1	1	–
	Paresthesia	2	2	–
	Syncope	1	–	1
	Pain	3	3	–
Renal	Acute kidney injury	1	1	–
	Urinary obstruction	1	–	1
Respiratory	Aspiration	1	–	1
	Nasal congestion	1	1	–
	Cough	2	2	–
Gastrointestinal	Lower gastrointestinal bleed	1	1	–
	Vomiting	1	1	–
Other	Alopecia	1	1	–
	Phlebitis	1	1	–
	Lower-extremity edema	1	1	–
Total		50	30	20

increase the background in assays of off-target cleavage. The distribution of on- and off-target cleavage is expected to vary for the three sgRNAs that were used in the manufacturing process (fig. S1A). Of the three sgRNAs, there were more off-target mutations identified for *TRBC* than for the other loci (Fig. 4C and figs. S4 and S5). The sgRNA for *PDCD1* was the most specific, because very few off-target edits were identified in more than 7000 sites of cleavage and there were very few off-target reads identified at the *TRAC1* and *TRAC2* loci (Fig. 4C).

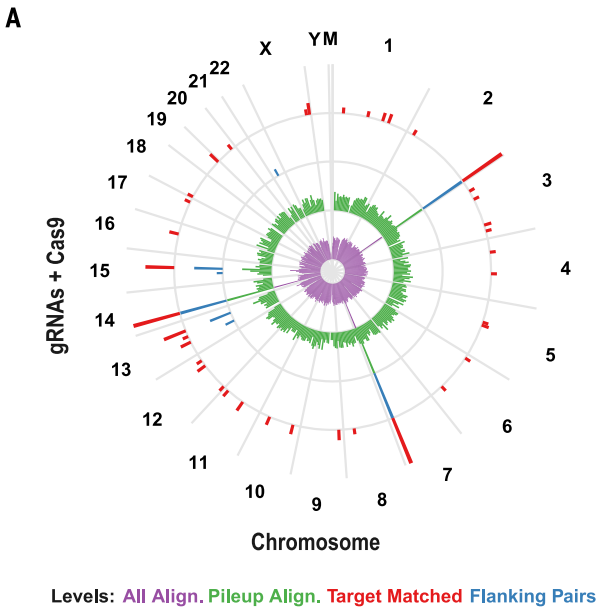
The genomic localization of identified DNA cleavage sites was as expected, given the chromosomal location of the three targeted genes on chromosomes 2, 7, and 14 (Fig. 4A). The distribution of the incorporation of the double-stranded oligodeoxynucleotide (dsODN) label around on-target sites, based on pileups within a window of 100 base pairs (bp), is shown in Fig. 4B and fig. S4. Although most mutations were on target, there were off-target mutations identified (Fig. 4C and fig. S5). For the *TRAC* sgRNA, there were low-abundance mutations within the transcriptional unit of *CLIC2* (chloride intracellular channel 2); however, disruption of *CLIC2* in T cells is not expected to have negative consequences because it is not reported to be expressed in T cells. For the *TRBC* sgRNA, off-target edits were identified in genes encoding a transcriptional regulator (ZNF609) and a long intergenic non–protein coding RNA (LINC00377) (table S3). In addition to the above post hoc investigations of multiplex editing specificity, all products were shown not to have cellular transformation by virtue of the absence of long-term growth before infusion (table S1).

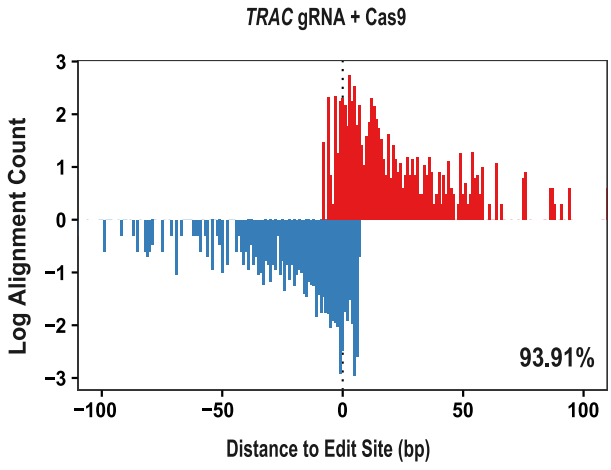
Detection of chromosomal translocations in CRISPR-Cas9–engineered T cells

In addition to the above detection of repair of double-strand DNA breaks by NHEJ, on-target mutagenesis by engineered nucleases can result in deletions, duplications, inversions, and translocations and can also lead to complex chromosomal rearrangements under some conditions (*37*). CRISPR-Cas9 has been used to intentionally create oncogenic chromosomal rearrangements (*38*). In preclinical studies with human T cells, simultaneous gene editing of *TRAC* and *CD52* using TALENs led to translocations that were detected at frequencies of 10^{-4} to 10^{-2} (*39*). In a subsequent clinical report using dual-gene editing with TALENs, chromosomal rearrangements were observed in 4% of infused cells (*40*). To study the safety and genotoxicity of multiplex CRISPR-Cas9 genome editing on three chromosomes, we used stringent release criteria of the manufactured cells and assays to detect translocations (fig. S6). We developed and qualified quantitative PCR (qPCR) assays to quantify the 12 potential translocations that could occur with the simultaneous editing of four loci: *TRAC*, *TRBC1*, *TRBC2*, and *PDCD1* (see materials and methods). We observed translocations in all manufactured products; however, the translocations were at the limit of detection for the assay in patient UPN39 (Fig. 5A). *TRBC1:TRBC2* was the most abundant rearrangement (Fig. 5A), resulting in a 9.3-kb deletion (supplementary materials). The deletion and translocations peaked on days 5 to 7 of manufacturing and then declined in frequency until cell harvest. The translocations and the *TRBC1:TRBC2* deletion were evident in the three patients between 10 days after infusion and 30 to 170 days after infusion (Fig. 5B). However, the rearrangements declined in frequency in vivo, suggesting that they conferred no evidence of a growth advantage over many generations of expansion in the patients on this trial (Fig. 3, A and B). At days 30, 150, and 170 in patients UPN07, UPN35, and UPN39, respectively, chromosomal translocations were at the limits of detection or not detected for all rearrangements except for the 9.3-kb deletion for *TRBC1:TRBC2*.

Single-cell RNA sequencing analysis reveals evolution of CRISPR-Cas9–engineered NYCE cells

We used single-cell RNA sequencing (scRNA-seq) to comprehensively characterize the transcriptomic phenotype of the NYCE T cells and their evolution over time in patient UPN39 (fig. S7). UPN39 was chosen because they had the highest level of cell engraftment and because this patient had evidence of tumor regression. CRISPR-Cas9–engineered T cells were infused to patient UPN39 and recovered after infusion from the blood on day 10 and at ~4 months (day 113) and were analyzed by scRNA-seq, as described in the materials and methods. For each sample (infusion product, day 10 and day 113), T cells were sorted on the basis of expression of CD4 or CD8 and processed using droplet-based 5′ scRNA-seq. From the gene expression libraries, PCR was used to further amplify cellular cDNA corresponding to the NY-ESO-1 TCR transgene, as well as *TRAC*, *TRBC*, and *PDCD1* target sequences, allowing us to genotype single cells as wild type or mutant. In the infusion product, cells





C

TRAC gRNA + Cas9

TGTGCTAGACATGAGGTCTANGG	Target	Abund.	Gene ID
. . . . . T. .	On	7,778	*TRA*~*
GA. . . A. . . . . A. . . . . A. .	Off	7	*CLIC2**
. TA. . . . . T. . TA. . . . TT. T	Off	6	*LINC00583*
C. G. . G. AT. . . . . A. GA.	Off	5	*C10orf67*
. T. AG. . . . . T. CA. C	Off	5	*LOC100131939*
. C. . . . TT. . T. A. A. . . . . CT.	Off	5	*ANKS1B**
. . . ATA. . G. . . . T. . . . G. G. T	Off	3	*IQCJ–SCHIP1**
. . . TT. . C. . . . A. C. . . . . G. .	Off	3	*ADCY10**
AA. . T. . . . . GA.	Off	2	*LDAH**
G. . . . . G. . . CAT. TCC.	Off	2	*MRPS27**

Fig. 4. Fidelity of CRISPR-Cas9 gene editing. (**A**) Genomic distribution of oligonucleotide (dsODN) incorporation sites, which mark locations of double-strand breaks. The ring indicates the human chromosomes aligned end to end, plus the mitochondrial chromosome (labeled M). The targeted cleavage sites are on chromosomes 2, 7, and 14. The frequency of cleavage and subsequent dsODN incorporation is shown on a log scale on each ring (pooled over 10-Mb windows). The purple innermost ring plots all alignments identified. The green ring shows pileups of three or more overlapping sequences, the blue ring shows alignments extending along either strand from a common dsODN incorporation site ("flanking pairs"), and the red ring shows reads with matches to the gRNA (allowing <6 mismatches) within 100 bp ("target matched"). (**B**) Distribution of inferred positions of cleavage and dsODN incorporation at an on-target locus. Incorporations in different strand orientations are shown on the positive (red) and negative (blue) *y* axis. The percentage in the bottom right corner is an estimate of the number of incorporations associated with the on-target site (based on pileups) captured within the allowed window of 100 bp. (**C**) Sequences of sites of cleavage and dsODN incorporation are shown, annotated by whether they are on target or off target ("Target"); the total number of unique alignments associated with the site ("Abund."); and an identifier indicating the nearest gene ("Gene ID"). An asterisk after the gene name indicates that the site is within the transcription unit of the specific gene, whereas "~" indicates that the gene appears on the allOnco cancer-associated gene list.

were identified that contained mutations in all three target sequences (Fig. 6, A and B). The most commonly mutated gene was *TRAC*. About 30% of cells had no mutations identified, whereas ~40% had one mutation, and ~20 and ~10% of the T cells in the manufactured product were double mutated and triple mutated, respectively, at the target sequences. Of the transgenic TCR$^+$ cells in the infusion product, monogenic mutations were less frequent than digenic and trigenic mutations (Fig. 6A). Single-cell genotyping of UPN39 cells at 10 days and 4 months after infusion showed a decline in the frequency of gene-edited T cells from the levels in the infusion product, and this decline occurred regardless of whether the cells were transduced with the NY-ESO-1 TCR (Fig. 6C). The frequency of gene-edited cells was quite stable between day 10 and 4 months postinfusion, and notably, about 40% of the peripheral blood–circulating T cells in this patient 4 months after infusion were mutated at any one of the targeted genes (Fig. 6, B and C, and table S4).

Of particular interest is the frequency and evolution of PD-1–deficient T cells owing to the previous mention that genetic disruption of *PDCD1* in CAR and TCR T cells enhances antitumor efficacy in preclinical models

(*19*, *21*–*24*). We found that ~25% of the T cells expressing the NY-ESO-1 TCR in the infusion product had mutations in the *PDCD1* locus (fig. S8). It is interesting that the frequency of cells with edits in the *PDCD1* locus decreased to ~5% of the cells expressing the transgenic TCR at 4 months postinfusion. This would be consistent with mouse studies of chronic infection in which PD-1–deficient T cells are less able to establish memory (*18*).

Figure 6D shows the distribution of engineered T cells expressing the NY-ESO-1 TCR transgene in the infusion product of patient UPN39, and again at 4 months in vivo as they evolve from the infused cells. In the heatmap (Fig. 6E), the most differentially expressed genes in the cells expressing the NY-ESO-1 transcript at the various time points are shown in table S5. Notably, UPN39 had increases in expression of genes associated with central memory (*IL7R* and *TCF7*) over time (Fig. 6, D and E, and table S4). This is in marked contrast to the recently published results with NY-ESO-1 T cells in the absence of genome editing, in which the infused transgenic T cells evolved to a terminally differentiated phenotype and displayed characteristics of T cell exhaustion in cancer patients (*12*).

Table 3. iGUIDE measurement of on-target editing efficiency for each gene by final product.

Manufactured NYCE T cell product (subject ID)	On-target editing efficiency (%)		
	PDCD1	*TRAC*	*TRBC*
UPN07	100.0	99.6	96.1
UPN27	99.6	99.1	96.8
UPN35	99.8	99.1	97.0
UPN39	98.2	96.7	93.5
Average ± SD	99.4 ± 0.8	98.6 ± 1.3	95.8 ± 1.6

Clinical observations

The clinical course of the three infused cancer patients is shown in Fig. 7 (and described in the materials and methods). No patient experienced cytokine release syndrome or overt side effects attributed to the cell infusion (table S5). The best clinical responses were stable disease in two patients. UPN39 had a mixed response, with a ~50% decrease in a large abdominal mass that was sustained for 4 months (Fig. 7D), although other lesions progressed. As of December 2019, all patients have progressed: Two are receiving other therapies, and UPN07 died from progressive myeloma.

Biopsies of bone marrow and tumor showed trafficking of the NYCE-engineered T cells to the sites of tumor in all three patients (Fig. 3A). It is interesting to note that even though the tumor biopsies revealed residual tumor, in both patients with myeloma, there was a reduction

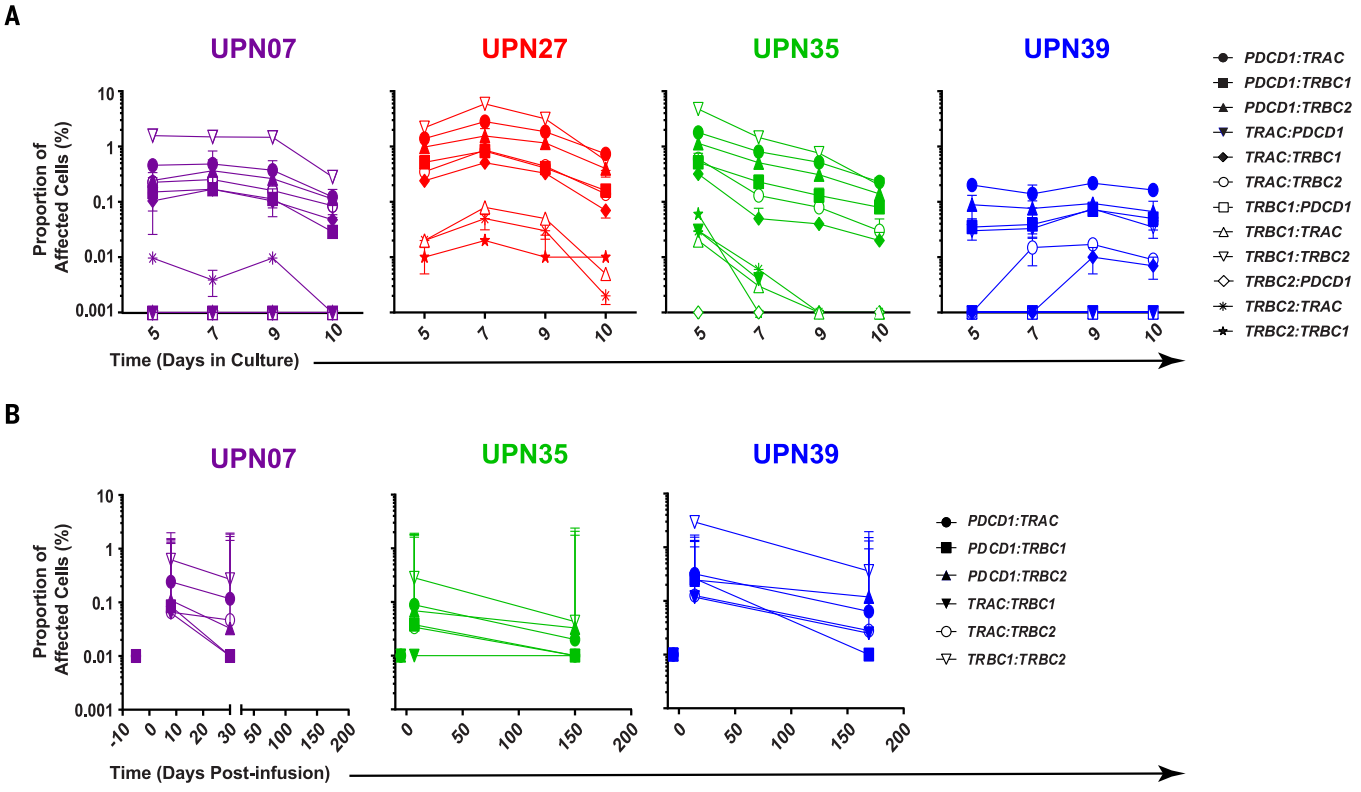


Fig. 5. Detection of chromosomal translocations in engineered T cells after CRISPR-Cas9 gene editing. (**A**) Evaluation of chromosomal translocations in NYCE T cell infusion products during the course of large-scale culture is shown. For the 12 monocentromeric translocation assays conducted, a positive reference sample that contains 1×10^3 copies of the synthetic template plasmid was evaluated as a control, and the percent difference between expected and observed marking was calculated. The absence of amplification from the 12 reactions that correspond to the different chromosomal translocations indicates assay specificity (see methods). (**B**) Longitudinal analysis of chromosomal translocations in vivo in three patients pre– and post–NYCE T cell product infusion is displayed. In (A) and (B), error bars represent SD. For graphical purposes, the proportions of affected cells were plotted on a log scale; a value of 0.001% indicates that translocations were not detected.

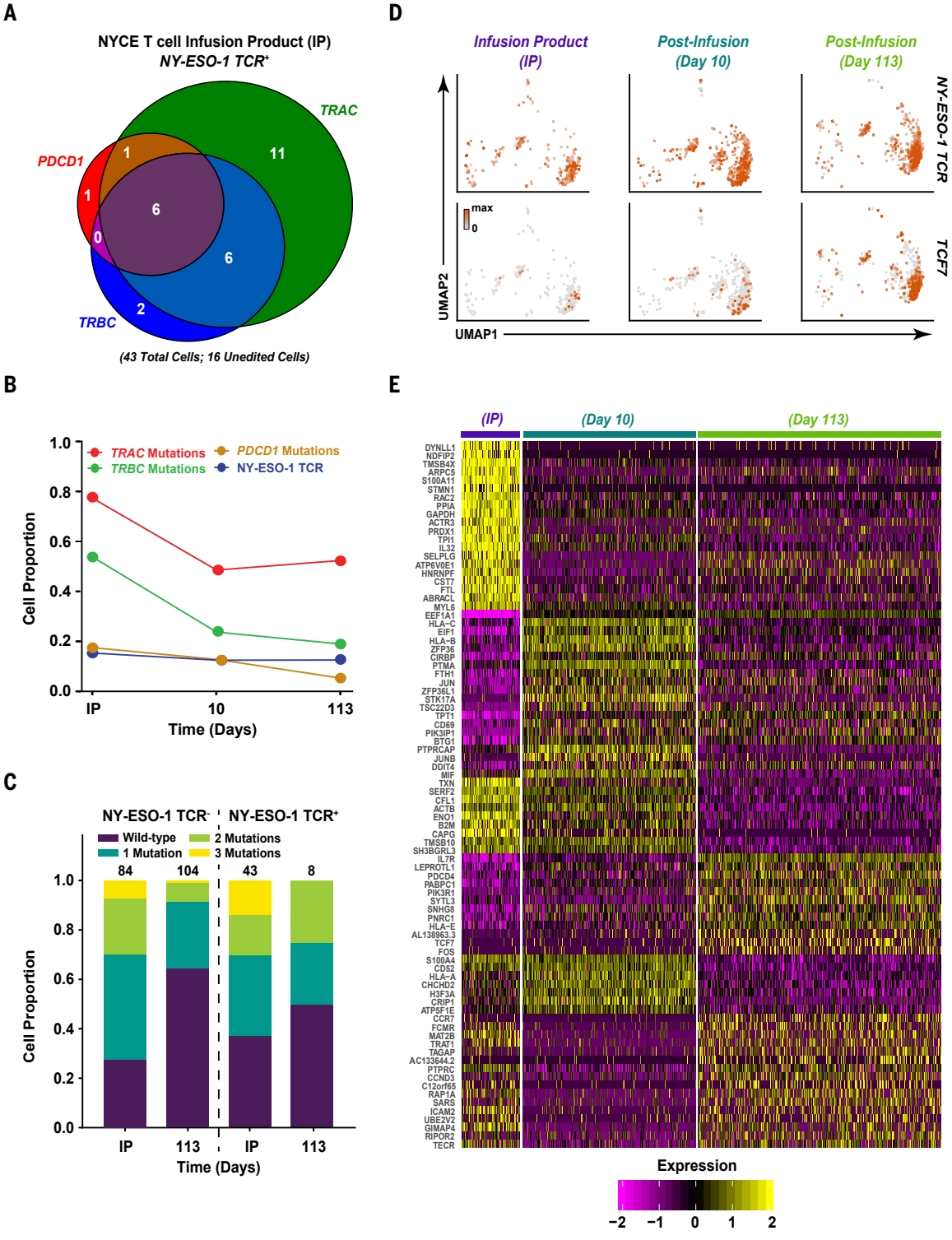


Fig. 6. Single-cell RNA sequencing of patient UPN39 CRISPR-Cas9–engineered NYCE T cells pre- and postinfusion. (**A**) Venn diagram showing relative numbers of NY-ESO-1 TCR–positive cells with *TRAC*, *TRBC*, and/or *PDCD1* mutations in the NYCE T cell infusion product (IP) (day 0). (**B**) Proportions of preinfusion (IP, day 0) and postinfusion (days 10 and 113) wild-type T cells with *TRAC*, *TRBC*, or *PDCD1* mutations or expressing the NY-ESO-1 TCR transgene. Numbers of cells belonging to each of these categories are listed below the graph. (**C**) Analysis of NY-ESO-1 TCR–positive (right) and NY-ESO-1 TCR–negative (left) cells without mutations (wild type) or with single, double, or triple mutations at day 0 (NYCE T cell infusion product) and day 113 post–NYCE T cell infusion. Numbers of analyzed cells for each time point are listed above the bars. (**D**) Uniform manifold approximation and projection (UMAP) plots of gene expression data. Analysis was performed on all T cells integrated across time points, but only NY-ESO-1 TCR–expressing cells, split by time point, are shown (top). The increase in *TCF7* expression is indicative of an acquired central memory phenotype (bottom, same cells). (**E**) Heatmap showing scaled expression of differentially expressed genes in NY-ESO-1 TCR–positive T cells across time points. Color scheme is based on scaled gene expression from −2 (purple) to 2 (yellow).

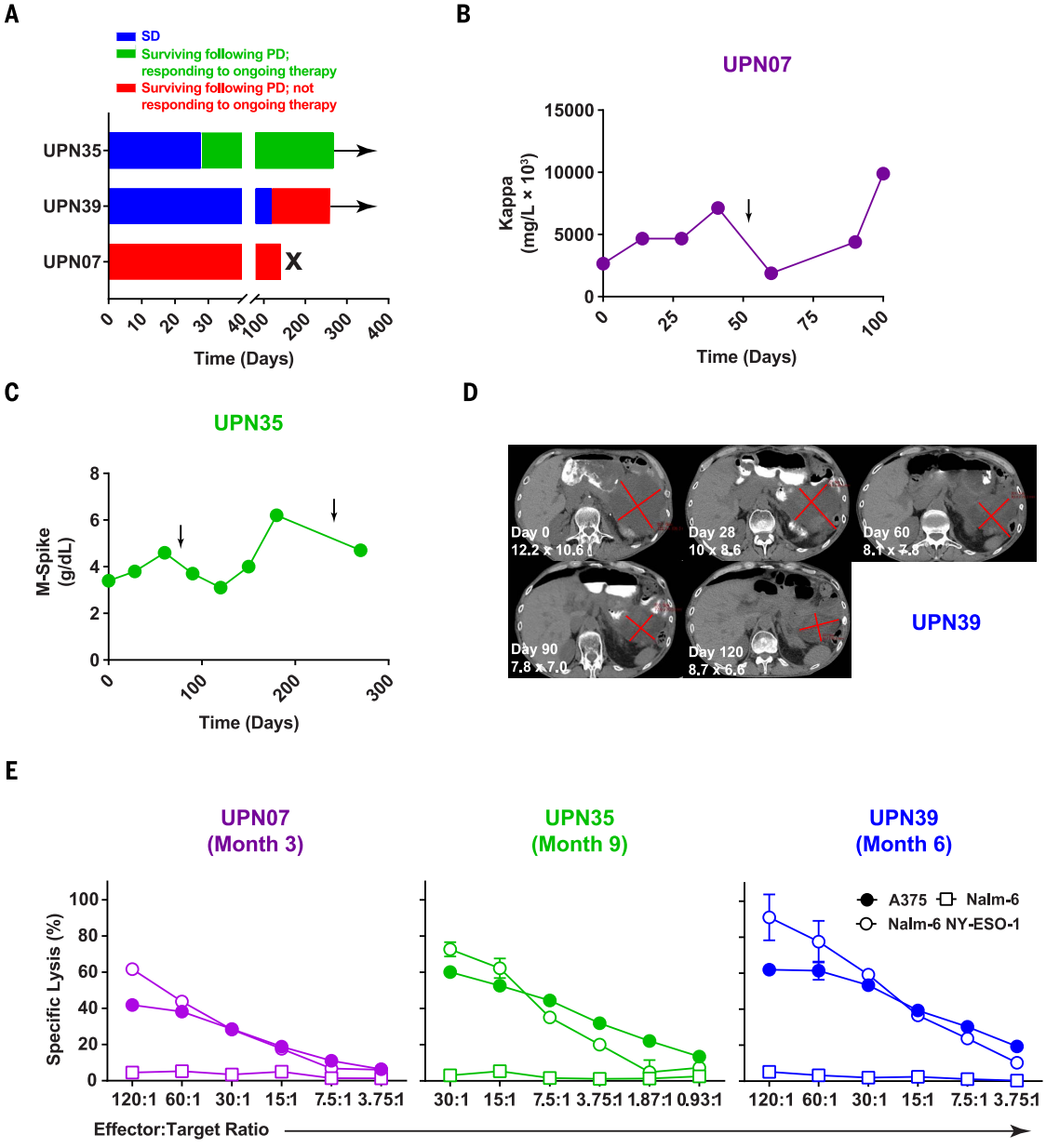


Fig. 7. Clinical responses and patient outcomes after infusion of CRISPR-Cas9–engineered NYCE T cells. (**A**) Swimmer's plot describing time on study for each patient, duration of follow-up off study (defined as survival beyond progression or initiation of other cancer therapy), and present status (differentially colored) is shown. Arrows indicate ongoing survival. SD, stable disease; PD, progressive disease. (**B**) Changes in kappa light chain levels (mg/liter × 10^3) in patient UPN07 after NYCE T cell product infusion are depicted. Vertical black arrow indicates initiation of a D-ACE salvage chemotherapy regimen (defined as intravenous infusion of cisplatin, etoposide, cytarabine, and dexamethasone). (**C**) Longitudinal M-spike levels (g/dl) in patient UPN35 post–NYCE T cell product administration are shown. Vertical black arrows indicate administration of combination therapy with elotuzumab, pomalidomide, and dexamethasone. (**D**) Computed tomography scans demonstrating tumor regression in patient UPN39 after administration of an autologous NYCE T cell infusion product. Radiologic studies were obtained before therapy and after adoptive transfer of NYCE T cells. Tumor is indicated by red X. (**E**) Cytolytic capacity of NY-ESO-1–specific $CD8^+$ T cells recovered at the indicated month after infusion and expanded from patients is shown. PBMC samples collected after NYCE T cell product infusion were expanded in vitro in the presence of NY-ESO-1 peptide and interleukin-2. The ability of expanded effector cells to recognize antigen and elicit cytotoxicity was tested in a 4-hour ^{51}Cr release assay incorporating Nalm-6 NY-ESO-1^+, parental Nalm-6 (NY-ESO-1^-), and A375 melanoma cells (NY-ESO-1^+). All target cell lines were HLA-A*02 positive. Assays were performed in triplicate, and error bars represent SD.

in the target antigens NY-ESO-1 and/or LAGE-1 (fig. S9). The reduction of target antigen was transient in patient UPN07 and persistent in patient UPN35. This result is consistent with an on-target effect of the infused cells, likely resulting in tumor editing (*41*).

To determine whether the NYCE cells retained antitumor activity after infusion, samples of blood obtained from patients 3 to 9 months after infusion were expanded in culture in the presence of NY-ESO-1 peptide and assessed for cytotoxicity against tumor cells (Fig. 7E and fig. S10). Antigen-specific cytotoxicity was observed in all three patients. It is interesting to note that the most potent antitumor cytotoxicity was observed in UPN39, because UPN39 was the only patient to have tumor regression after infusion of the CRISPR-Cas9–engineered T cells (Fig. 7D).

Discussion

Our phase 1 first-in-human pilot study demonstrates the initial safety and feasibility of multiplex CRISPR-Cas9 T cell human genome engineering in patients with advanced, refractory cancer. In one patient analyzed at depth, a frequency of 30% of digenic and trigenic editing was achieved in the infused cell population, and 20% of the TCR transgenic T cells in circulation 4 months later had persisting digenic and trigenic edits. We chose to redirect specificity of the T cells with a T cell receptor, rather than a CAR, to avoid the CAR-associated potential toxicities such as cytokine release syndrome (*31*). This provided a lower baseline toxicity profile, thus enhancing the ability to detect toxicity specifically associated with the CRISPR-Cas9–engineering process. We observed mild toxicity, and most of the adverse events were attributed to the lymphodepleting chemotherapy. We note that although the initial clinical results have acceptable safety, experience with more patients given infusions of CRISPR-engineered T cells with higher editing efficiencies, and longer observation after infusion, will be required to fully assess the safety of this approach.

Our large-scale product manufacturing process resulted in gene-editing efficiencies similar to those in our preclinical studies (*24*). A surprising finding was the high-level engraftment and long-term persistence of the infused CRISPR-Cas9–engineered T cells. In previous clinical studies testing adoptively transferred NY-ESO-1 transgenic T cells, the engrafted cells had an initial decay half-life of about 1 week (*10*–*12*). The explanation for the extended survival that we observed remains to be determined and could include the editing of the endogenous TCR, PD-1, and/or the choice of the TCR and vector design.

The use of scRNA-seq technology permitted the analysis of the transcriptome of the infused NY-ESO-1–specific T cells (i.e., CRISPR-Cas9–engineered T cells) at baseline and for up to 4 months in vivo. The results shown for UPN39 revealed that the infused cells evolved to a state consistent with central memory. These results are in contrast to a recent study in which the infused NY-ESO-1 T cells evolved to a state consistent with T cell exhaustion (*12*). A limitation of our in vivo single-cell analysis is that for purposes of feasibility, it is limited to the one patient who had the highest level of engraftment. Another limitation is that we were not able to compare the transcriptional state of the modified cells in the tumor microenvironment with circulating NYCE T cells.

Analysis of the manufacturing process in vitro demonstrated monochromosomal translocations and rearrangements, and some of these persisted in vivo. The translocations were not random in occurrence and occurred most frequently between *PDCD1:TRAC* and *TRBC1:TRBC2*. The frequency of translocations that we observed with trigenic editing is similar to that reported for digenic editing using TALEN-mediated gene editing in preclinical and clinical studies, in which rearrangements were detected in about 4% of cells (*39*, *40*). It is important to note that healthy individuals often harbor oncogenic translocations in B and T cells (*42*–*44*). T cells bearing translocations can persist for months to years without evidence of pathogenicity (*45*–*47*).

Antagonism of the PD-1:PD-L1 costimulatory pathway can result in organ-specific and systemic autoimmunity (*17*, *48*). PD-1 has been reported to function as a haploinsufficient tumor suppressor in mouse T cells (*49*). Our patients have had engraftment with PD-1–deficient T cells, and to date, there is no evidence of autoimmunity or T cell genotoxicity.

In conclusion, our phase 1 human pilot study has confirmed that multiplex CRISPR-Cas9 editing of the human genome is possible at clinical scale. We note that although the initial clinical results suggest that this treatment is safe, experience with more patients given infusions with higher editing efficiencies and longer observation after infusion will be required to fully assess the safety of this approach. The potential rejection of infused cells due to preexisting immune responses to Cas9 (*28*, *29*) does not appear to be a barrier to the application of this promising technology. Finally, it is important to note that our manufacturing was based on the reagents available in 2016, when our protocol had been reviewed by the National Institutes of Health (NIH) Recombinant DNA Advisory Committee and received approval. Our Investigational New Drug application was subsequently reviewed and accepted by the FDA. There has been rapid progress in the field since that time, with the development of reagents that should increase efficiencies and decrease off-target editing using CRISPR-based technology (*50*).

Materials and methods summary

Experimental design

The clinical protocol is listed at clinicaltrials.gov, trial NCT03399448. Protocol no. 1604-1524 "Phase 1 trial of autologous T cells engineered to express NY-ESO-1 TCR and CRISPR gene edited to eliminate endogenous TCR and PD-1 (NYCE T Cells)" was reviewed and approved by the U.S. National Institutes of Health Recombinant DNA Advisory Committee on 21 June 2016. See fig. S1B for clinical trial design. Patient demographics are shown in Table 1. A list of adverse events is depicted in Table 2.

Guide RNAs (gRNAs)

The genomic gRNA target sequences with protospacer adjacent motif (PAM) underlined were: *TRAC1* and *TRAC2*: 5′-TGTGCTAGACATGAGGTCTATGG-3′, *TRBC*: 5′-GGAGAATGACGAGTGGACCCAGG-3′, and *PDCD1*: 5′-GGCGCCCTGGCCAGTCGTCTGGG-3′. In vitro transcribed gRNA was prepared from linearized DNA (Aldevron) using Bulk T7 Megascript 5X (Ambion) and purified using RNeasy Maxi Kit (Qiagen).

Recombinant Cas9 protein

Cas9 recombinant protein derived from *S. pyogenes* was TrueCut Cas9 v2 (catalogue no. A36499, ThermoFisher). Cas9 RNP was made by incubating protein with gRNA at a molar ratio of 1:1 at 25°C for 10 min immediately before electroporation.

Lentiviral vector manufacturing

The 8F TCR recognizes the HLA-A*0201 SLLMWITQC epitope on NY-ESO-1 and LAGE-1. The 8F TCR was isolated from a T cell clone obtained from patient after vaccination with NY-ESO-1 peptide. The TCR sequences were cloned into a transfer plasmid that contains the EF-1α promoter, a cPPT sequence, a Rev response element and a woodchuck hepatitis virus posttranscriptional regulatory element (WPRE), as shown in fig. S1B. Plasmid DNA was manufactured at Puresyn, Inc. Lentiviral vector was produced at the University of Pennsylvania Center for Advanced Retinal and Ocular Therapeutics using transient transfection with four plasmids expressing the transfer vector, Rev, VSV-G, and gag-pol, in human embryonic kidney 293T cells.

REFERENCES AND NOTES

1. P. Rouet, F. Smih, M. Jasin, Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**, 8096–8106 (1994). doi: 10.1128/MCB.14.12.8096; pmid: 7969147
2. M. H. Porteus, A new class of medicines through DNA editing. *N. Engl. J. Med.* **380**, 947–959 (2019). doi: 10.1056/NEJMra1800729; pmid: 30855744
3. P. Tebas *et al.*, Gene editing of *CCR5* in autologous CD4 T cells of persons infected with HIV. *N. Engl. J. Med.* **370**, 901–910 (2014). doi: 10.1056/NEJMoa1300662; pmid: 24597865
4. L. Xu *et al.*, CRISPR-edited stem cells in a patient with HIV and acute lymphocytic leukemia. *N. Engl. J. Med.* **381**, 1240–1247 (2019). doi: 10.1056/NEJMoa1817426; pmid: 31509667
5. M. Jinek *et al.*, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012). doi: 10.1126/science.1225829; pmid: 22745249
6. L. Cong *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013). doi: 10.1126/science.1231143; pmid: 23287718
7. P. K. Mandal *et al.*, Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. *Cell Stem Cell* **15**, 643–652 (2014). doi: 10.1016/j.stem.2014.10.004; pmid: 25517468
8. J. F. Hultquist *et al.*, CRISPR-Cas9 genome engineering of primary CD4$^+$ T cells for the interrogation of HIV-host factor interactions. *Nat. Protoc.* **14**, 1–27 (2019). doi: 10.1038/s41596-018-0069-7; pmid: 30559373
9. R. O. Bak *et al.*, Multiplexed genetic engineering of human hematopoietic stem and progenitor cells using CRISPR/Cas9 and AAV6. *eLife* **6**, e27873 (2017). doi: 10.7554/eLife.27873; pmid: 28956530
10. P. F. Robbins *et al.*, Tumor regression in patients with metastatic synovial cell sarcoma and melanoma using genetically engineered lymphocytes reactive with NY-ESO-1. *J. Clin. Oncol.* **29**, 917–924 (2011). doi: 10.1200/JCO.2010.32.2537; pmid: 21282551

11. A. P. Rapoport *et al*., NY-ESO-1-specific TCR-engineered T cells mediate sustained antigen-specific antitumor effects in myeloma. *Nat. Med.* **21**, 914–921 (2015). doi: 10.1038/nm.3910; pmid: 26193344
12. T. S. Nowicki *et al*., A pilot trial of the combination of transgenic NY-ESO-1-reactive adoptive cellular therapy with dendritic cell vaccination with or without ipilimumab. *Clin. Cancer Res.* **25**, 2096–2108 (2019). doi: 10.1158/1078-0432.CCR-18-3496; pmid: 30573690
13. E. Provasi *et al*., Editing T cell specificity towards leukemia by zinc finger nucleases and lentiviral gene transfer. *Nat. Med.* **18**, 807–815 (2012). doi: 10.1038/nm.2128; pmid: 20400962
14. G. M. Bendle *et al*., Lethal graft-versus-host disease in mouse models of T cell receptor gene therapy. *Nat. Med.* **16**, 565–570, 1p, 570 (2010). doi: 10.1038/nm.2128; pmid: 20400962
15. K. Schober *et al*., Orthotopic replacement of T-cell receptor α- and β-chains with preservation of near-physiological T-cell function. *Nat. Biomed. Eng.* **3**, 974–984 (2019). doi: 10.1038/s41551-019-0409-0; pmid: 31182835
16. A. Schietinger, J. J. Delrow, R. S. Basom, J. N. Blattman, P. D. Greenberg, Rescued tolerant CD8 T cells are preprogrammed to reestablish the tolerant state. *Science* **335**, 723–727 (2012). doi: 10.1126/science.1214277; pmid: 22267581
17. H. Nishimura, M. Nose, H. Hiai, N. Minato, T. Honjo, Development of lupus-like autoimmune diseases by disruption of the PD-1 gene encoding an ITIM motif-carrying immunoreceptor. *Immunity* **11**, 141–151 (1999). doi: 10.1016/S1074-7613(00)80089-8; pmid: 10485649
18. P. M. Odorizzi, K. E. Pauken, M. A. Paley, A. Sharpe, E. J. Wherry, Genetic absence of PD-1 promotes accumulation of terminally differentiated exhausted $CD8^+$ T cells. *J. Exp. Med.* **212**, 1125–1137 (2015). doi: 10.1084/jem.20142237; pmid: 26034050
19. L. J. Rupp *et al*., CRISPR/Cas9-mediated PD-1 disruption enhances anti-tumor efficacy of human chimeric antigen receptor T cells. *Sci. Rep.* **7**, 737 (2017). doi: 10.1038/s41598-017-00462-8; pmid: 28389661
20. I. Serganova *et al*., Enhancement of PSMA-directed CAR adoptive immunotherapy by PD-1/PD-L1 blockade. *Mol. Ther. Oncolytics* **4**, 41–54 (2016). doi: 10.1016/j.omto.2016.11.005; pmid: 28345023
21. S. Su *et al*., CRISPR-Cas9 mediated efficient PD-1 disruption on human primary T cells from cancer patients. *Sci. Rep.* **6**, 20070 (2016). doi: 10.1038/srep20070; pmid: 26818188
22. L. Menger *et al*., TALEN-mediated inactivation of PD-1 in tumor-reactive lymphocytes promotes intratumoral T-cell persistence and rejection of established tumors. *Cancer Res.* **76**, 2087–2093 (2016). doi: 10.1158/0008-5472.CAN-15-3352; pmid: 27197251
23. L. Cherkassky *et al*., Human CAR T cells with cell-intrinsic PD-1 checkpoint blockade resist tumor-mediated inhibition. *J. Clin. Invest.* **126**, 3130–3144 (2016). doi: 10.1172/JCI83092; pmid: 27454297
24. J. Ren *et al*., Multiplex genome editing to generate universal CAR T cells resistant to PD1 inhibition. *Clin. Cancer Res.* **23**, 2255–2266 (2017). doi: 10.1158/1078-0432.CCR-16-1300; pmid: 27815355
25. J. Ren *et al*., A versatile system for rapid multiplex genome-edited CAR T cell generation. *Oncotarget* **8**, 17002–17011 (2017). doi: 10.18632/oncotarget.15218; pmid: 28199983
26. X. Liu *et al*., CRISPR-Cas9-mediated multiplex gene editing in CAR-T cells. *Cell Res.* **27**, 154–157 (2017). doi: 10.1038/cr.2016.142; pmid: 27910851
27. E. K. Moon *et al*., Blockade of programmed death 1 augments the ability of human T cells engineered to target NY-ESO-1 to control tumor growth after adoptive transfer. *Clin. Cancer Res.* **22**, 436–447 (2016). doi: 10.1158/1078-0432.CCR-15-1070; pmid: 26324743
28. C. T. Charlesworth *et al*., Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat. Med.* **25**, 249–254 (2019). doi: 10.1038/s41591-018-0326-x; pmid: 30692695
29. D. L. Wagner *et al*., High prevalence of *Streptococcus pyogenes* Cas9-reactive T cells within the adult human population. *Nat. Med.* **25**, 242–248 (2019). doi: 10.1038/s41591-018-0204-6; pmid: 30374197
30. V. L. Simhadri *et al*., Prevalence of pre-existing antibodies to CRISPR-associated nuclease Cas9 in the USA population. *Mol. Ther. Methods Clin. Dev.* **10**, 105–112 (2018). doi: 10.1016/j.omtm.2018.06.006; pmid: 30073181
31. D. T. Teachey *et al*., Identification of predictive biomarkers for cytokine release syndrome after chimeric antigen receptor T-cell therapy for acute lymphoblastic leukemia. *Cancer Discov.* **6**, 664–679 (2016). doi: 10.1158/2159-8290.CD-16-0040; pmid: 27076371
32. P. F. Robbins *et al*., A pilot trial using lymphocytes genetically engineered with an NY-ESO-1-reactive T-cell receptor: Long-term follow-up and correlates with response. *Clin. Cancer Res.* **21**, 1019–1027 (2015). doi: 10.1158/1078-0432.CCR-14-2708; pmid: 25538264
33. S. P. D'Angelo *et al*., Antitumor activity associated with prolonged persistence of adoptively transferred NY-ESO-1 ^{c259}T cells in synovial sarcoma. *Cancer Discov.* **8**, 944–957 (2018). doi: 10.1158/2159-8290.CD-17-1417; pmid: 29891538
34. C. L. Nobles *et al*., iGUIDE: An improved pipeline for analyzing CRISPR cleavage specificity. *Genome Biol.* **20**, 14 (2019). doi: 10.1186/s13059-019-1625-3; pmid: 30654827
35. S. Q. Tsai *et al*., GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015). doi: 10.1038/nbt.3117; pmid: 25513782
36. M. M. Vilenchik, A. G. Knudson, Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12871–12876 (2003). doi: 10.1073/pnas.2135498100; pmid: 14566050
37. M. Kosicki, K. Tomberg, A. Bradley, Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018). doi: 10.1038/nbt.4192; pmid: 30010673
38. D. Maddalo *et al*., In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* **516**, 423–427 (2014). doi: 10.1038/nature13902; pmid: 25337876
39. L. Poirot *et al*., Multiplex genome-edited T-cell manufacturing platform for "off-the-shelf" adoptive T-cell immunotherapies. *Cancer Res.* **75**, 3853–3864 (2015). doi: 10.1158/0008-5472.CAN-14-3321; pmid: 26183927
40. W. Qasim *et al*., Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells. *Sci. Transl. Med.* **9**, eaaj2013 (2017). doi: 10.1126/scitranslmed.aaj2013; pmid: 28123068
41. G. P. Dunn, A. T. Bruce, H. Ikeda, L. J. Old, R. D. Schreiber, Cancer immunoediting: From immunosurveillance to tumor escape. *Nat. Immunol.* **3**, 991–998 (2002). doi: 10.1038/ni1102-991; pmid: 12407406
42. J. Bäsecke, F. Griesinger, L. Trümper, G. Brittinger, Leukemia- and lymphoma-associated genetic aberrations in healthy individuals. *Ann. Hematol.* **81**, 64–75 (2002). doi: 10.1007/s00277-002-0427-x; pmid: 11907785
43. S. I. Ismail, R. G. Naffa, A. M. Yousef, M. T. Ghanim, Incidence of bcr-abl fusion transcripts in healthy individuals. *Mol. Med. Rep.* **9**, 1271–1276 (2014). doi: 10.3892/mmr.2014.1951; pmid: 24535287
44. R. Chiarle, Translocations in normal B cells and cancers: Insights from new technical approaches. *Adv. Immunol.* **117**, 39–71 (2013). doi: 10.1016/B978-0-12-410524-9.00002-5; pmid: 23611285
45. C. A. Michie, A. McLean, C. Alcock, P. C. Beverley, Lifespan of human lymphocyte subsets defined by CD45 isoforms. *Nature* **360**, 264–265 (1992). doi: 10.1038/360264a0; pmid: 1436108
46. A. R. Mclean, C. A. Michie, In vivo estimates of division and death rates of human T lymphocytes. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3707–3711 (1995). doi: 10.1073/pnas.92.9.3707; pmid: 7731969
47. K. George, L. J. Chappell, F. A. Cucinotta, Persistence of space radiation induced cytogenetic damage in the blood lymphocytes of astronauts. *Mutat. Res.* **701**, 75–79 (2010). doi: 10.1016/j.mrgentox.2010.02.007; pmid: 20176126
48. C. H. June, J. T. Warshauer, J. A. Bluestone, Is autoimmunity the Achilles' heel of cancer immunotherapy? *Nat. Med.* **23**, 540–547 (2017). doi: 10.1038/nm.4321; pmid: 28475571
49. T. Wartewig *et al*., PD-1 is a haploinsufficient suppressor of T cell lymphomagenesis. *Nature* **552**, 121–125 (2017). doi: 10.1038/nature24649; pmid: 29143824
50. C. A. Vakulskas *et al*., A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat. Med.* **24**, 1216–1224 (2018). doi: 10.1038/s41591-018-0137-0; pmid: 30082871

ACKNOWLEDGMENTS

We thank the Human Immunology Core at the University of Pennsylvania for providing leukocytes for research; the Clinical Cell and Vaccine Production Facility for GMP cell manufacturing; the Hospital of the University of Pennsylvania Apheresis Unit for peripheral blood mononuclear cell collections from which NYCE T cells were manufactured; regulatory assistance from E. Meagher, S. Emmanuel, E. Veloso, and the Office of Clinical Research, the Center for Advanced Retinal and Ocular Therapeutics Clinical Vector Core (CAROT CVC); RECIST Core for radiologic assessment of response; Data Safety Monitoring Board consisting of L. Schuchter, A. Rapoport, and M. Dhodapkar; J. Everett for data analysis with iGUIDE; C. Bartoszek, J. Finklestein, M. Gohil, A. Kim, N. Koterba, M. Mahir, B. Menchel, T. Mikheeva, F. Nazimuddin, H. Parakandi, R. Reynolds, M. Feldman, and T. Yoder for experimental support; W. Gladney for protocol development and management; and R. Chiarle at the Boston Children's Hospital for helpful discussions. **Funding:** This work was funded by the National Institutes of Health (grant 2R01CA120409 to Y.Z. and C.H.J.), Alliance for Cancer Gene Therapy Investigator's Award (J.A.F. and R.M.Y.), NCI P01 CA214278 (J.A.F., R.M.Y., S.F.L., and C.H.J.), NCI U54 CA24711 (J.A.F., M.M.D., R.M.Y., and C.H.J.), NIA U01 AG066100 (J.A.F., R.M.Y., and C.H.J.), NSF Engineering Research Center for Cell Manufacturing Technologies Seed Grant (J.A.F. and B.L.L.), Abramson Cancer Center Emerging Cancer Informatics Center of Excellence Award (J.A.F.), and sponsored research grants from the Parker Institute for Cancer Immunotherapy and Tmunity Therapeutics. **Author contributions:** Conceptualization: Y.Z., C.H.J., F.D.B.; Data curation: V.E.G.; Formal analysis: W.-T.H., J.A.F.; Funding acquisition: C.H.J.; Investigation: E.A.S., A.Ce., K.L.W., A.D.C., E.L., P.A.M., I.K., M.G., J.X., I.J., J.S., T.M., A.Ch., A.L.G., S.D., A.L., K.R.P., Y.Q., A.T.S., H.Y.C., B.M.C.; Methodology: L.T., A.F., J.S.-M., F.D.B.; Project administration: R.M.Y., A.C., G.P., J.K.J.; Software: C.L.N.; Supervision: J.A.F., S.F.L., E.O.H., J.J.M., R.M.Y., D.L.S.; Validation: V.E.G., F.C., L.T., M.M.D., S.F.L., J.S.-M.; Writing – original draft: E.A.S., J.A.F., S.F.L., C.H.J.; Writing – review and editing: B.L.L., D.L.S., H.Y.C., F.D.B. **Competing interests:** Y.Z. and C.H.J. are inventors on patent applications 15/516,052 and WO2016069282A1 submitted by the University of Pennsylvania that cover the use of gene modification in T cells for adoptive cell therapy. A.C., B.L.L., Y.Z., and C.H.J. are scientific founders of Tmunity and have equity in Tmunity. J.A.F. and M.M.D. have funding support from Tmunity. A.T.S. is a scientific founder of Immunai and has funding support from Arsenal Biosciences. S.F.L. has funding support from Tmunity, Cabaletta, and Novartis and is a consultant for Gilead/Kite. H.Y.C. is a co-founder of Accent Therapeutics and Boundless Bio and is a consultant for 10x Genomics, Arsenal Biosciences, and Spring Discovery. The other authors declare no competing interests. **Data and materials availability:** iGUIDE-seq data have been deposited with links to BioProject accession number PRJNA601142 in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/). scRNA-seq data have been deposited into the database of Genotypes and Phenotypes dbGaP (https://www.ncbi.nlm.nih.gov/gap/) under accession number phs001707. Reagents are available under an MTA with the University of Pennsylvania; address requests to C.H.J.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/eaba7365/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 to S6
References (*51*–*55*)

View/request a protocol for this paper from *Bio-protocol*.

3 January 2020; accepted 28 January 2020
Published online 6 February 2020
10.1126/science.aba7365

RESEARCH ARTICLE SUMMARY

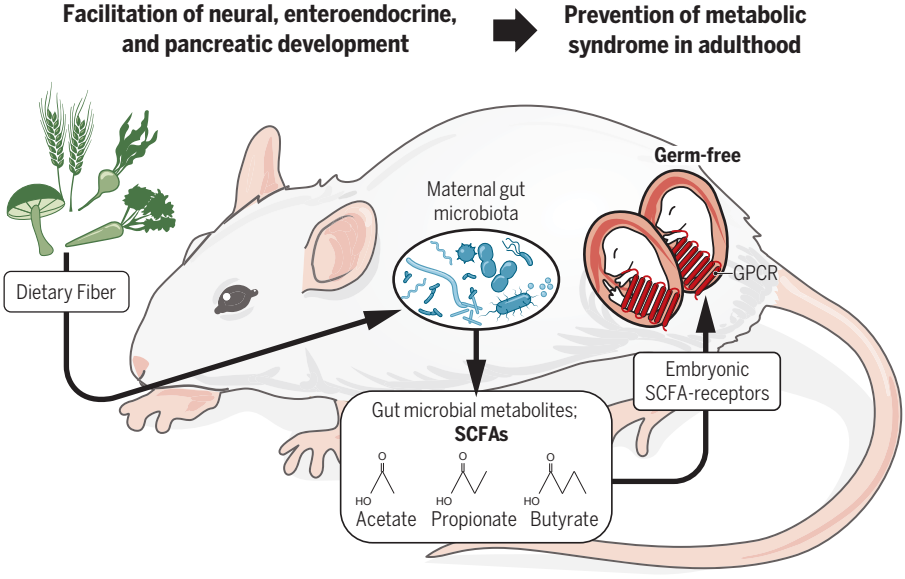
MICROBIOTA

Maternal gut microbiota in pregnancy influences offspring metabolic phenotype in mice

Ikuo Kimura*†, Junki Miyamoto*, Ryuji Ohue-Kitano, Keita Watanabe, Takahiro Yamada, Masayoshi Onuki, Ryo Aoki, Yosuke Isobe, Daiji Kashihara, Daisuke Inoue, Akihiko Inaba, Yuta Takamura, Satsuki Taira, Shunsuke Kumaki, Masaki Watanabe, Masato Ito, Fumiyuki Nakagawa, Junichiro Irie, Hiroki Kakuta, Masakazu Shinohara, Ken Iwatsuki, Gozoh Tsujimoto, Hiroaki Ohno, Makoto Arita, Hiroshi Itoh, Koji Hase†

INTRODUCTION: In recent decades, the rapid expansion of antibiotic use and intake of high-calorie, low-fiber diets have contributed to disturbances in the gut microbial community, predisposing humans to various diseases such as metabolic syndrome. Although the influence of microbiota on the postnatal environment has been well documented, much less is known regarding the impact of gut microbiota at the embryonic stage. Although accumulating evidence supports the notion of the developmental origins of health and disease (DOHaD), the underlying mechanisms remain obscure. In this study, we explored the impact of maternal gut microbiota on embryonic development and disease susceptibility late in life.

RATIONALE: Gut microbiota–derived metabolites represented by short-chain fatty acids (SCFAs; e.g., acetate, propionate, and butyrate) not only fuel host cells but also serve as signaling molecules between the gut microbiota and extraintestinal organs. GPR41 and GPR43 belong to the free fatty acids receptor (FFAR) family and are receptors for SCFAs. We have previously corroborated the biological importance of FFARs in energy metabolism through interactions with dietary ingredients, as well as gut microbiota–derived metabolites. Gut microbial SCFAs regulate host energy homeostasis via GPR41 and GPR43 in the sympathetic nervous system, adipose tissues, pancreas, and intestine. A recent study further showed that the gut microbiota of pregnant mice influence immune and brain functions of offspring. These findings raise the possibility that maternal SCFAs play a key role in the regulation of disease susceptibility during postnatal life in the context of the DOHaD theory.



During pregnancy, maternal gut microbiota influences offspring propensity for obesity via embryonic SCFA receptors. The maternal gut microbial SCFA-embryonic GPR41 and GPR43 axes facilitate the development of neural cells, GLP-1–expressing enteroendocrine cells, and pancreatic β cells to shape the development of energy metabolism in offspring, even as adults. GPCR, G protein–coupled receptor.

RESULTS: We found that maternal microbiota during pregnancy imparts resistance to obesity to their offspring. Pregnant mice were bred under specific pathogen–free (SPF) and germ-free (GF) conditions, after which newborns were raised by foster mothers under conventional conditions to align growth environments after birth. The offspring from GF mothers were highly susceptible to metabolic syndrome characterized by an exacerbation of obesity and glucose intolerance in association with reduced energy expenditure upon high-fat diet consumption during adulthood. A similar phenotype was also observed in offspring from mice fed a low-fiber (LFi) diet during pregnancy. Treatment of pregnant GF or LFi-fed mice with SCFA rendered adult offspring resistant to obesity. SCFA in the colonic lumen of pregnant mice reached the embryos via the maternal liver and bloodstream. Notably, the sympathetic nerves, intestinal epithelium, and pancreas of embryos highly expressed GPR41 and/or GPR43 to sense SCFAs originating from the maternal gut microbiota. Deficiency in embryonic GPR41 and GPR43 signaling compromised energy metabolism because of sympathetic dysfunction and hyperglycemia during the prenatal period. The SCFA-GPR41 and SCFA-GPR43 axes facilitate the development of neural cells, GLP-1–expressing enteroendocrine cells, and pancreatic β cells, thereby shaping embryonic energy metabolism. This developmental process contributes to maintaining postnatal energy homeostasis.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaw8429

CONCLUSION: We determined that, during pregnancy, the maternal gut microbiota confers resistance to obesity in offspring via the SCFA-GPR41 and SCFA-GPR43 axes. During pregnancy, SCFAs from the maternal gut microbiota were sensed by GPR41 and GPR43 in the sympathetic nerve, intestinal tract, and pancreas of the embryo, influencing prenatal development of the metabolic and neural systems. These findings indicate that the maternal gut environment during pregnancy is a key contributor to metabolic programming of offspring to prevent metabolic syndrome. Thus, the gut microbiota of pregnant mice provides an environmental cue that fine-tunes energy homeostasis in offspring to prevent the developmental origin of metabolic syndrome. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: ikimura@cc.tuat.ac.jp (I.K.); hase-kj@pha.keio.ac.jp (K.H.)
Cite this article as I. Kimura *et al.*, *Science* **367**, eaaw8429 (2020). DOI: 10.1126/science.aaw8429

on February 27, 2020

RESEARCH ARTICLE

MICROBIOTA

Maternal gut microbiota in pregnancy influences offspring metabolic phenotype in mice

Ikuo Kimura[1,2]*†, Junki Miyamoto[1,2]*, Ryuji Ohue-Kitano[1,2], Keita Watanabe[1], Takahiro Yamada[3], Masayoshi Onuki[3], Ryo Aoki[4,5], Yosuke Isobe[6], Daiji Kashihara[7], Daisuke Inoue[7], Akihiko Inaba[8], Yuta Takamura[9], Satsuki Taira[1], Shunsuke Kumaki[8], Masaki Watanabe[9], Masato Ito[3], Fumiyuki Nakagawa[10,11], Junichiro Irie[2,12], Hiroki Kakuta[9], Masakazu Shinohara[13], Ken Iwatsuki[8], Gozoh Tsujimoto[7], Hiroaki Ohno[2,14], Makoto Arita[6,15,16], Hiroshi Itoh[2,12], Koji Hase[3,17]†

Antibiotics and dietary habits can affect the gut microbial community, thus influencing disease susceptibility. Although the effect of microbiota on the postnatal environment has been well documented, much less is known regarding the impact of gut microbiota at the embryonic stage. Here we show that maternal microbiota shapes the metabolic system of offspring in mice. During pregnancy, short-chain fatty acids produced by the maternal microbiota dictate the differentiation of neural, intestinal, and pancreatic cells through embryonic GPR41 and GPR43. This developmental process helps maintain postnatal energy homeostasis, as evidenced by the fact that offspring from germ-free mothers are highly susceptible to metabolic syndrome, even when reared under conventional conditions. Thus, our findings elaborate on a link between the maternal gut environment and the developmental origin of metabolic syndrome.

The gut microbiota substantially contributes to energy extraction from indigestible polysaccharides, such as oligosaccharides and soluble dietary fibers, and influences host energy homeostasis during infancy and adulthood (*1*–*6*). Therefore, maintenance of the symbiotic microbial community is of paramount importance for host health. Changes in microbial composition may cause dysbiosis, leading to disease-prone phenotypes in the host. Dysbiosis has been implicated in a growing number of systemic disorders, including metabolic syndrome (*7*–*9*). Obesity is a major risk factor for metabolic syndrome, predisposing patients to cardiovascular disease and type 2 diabetes. In addition to genetic and epigenetic factors, changes in the gut microbiota have been implicated in the development of obesity, in combination with dietary factors. Gut microbiota–derived metabolites represented by short-chain fatty acids (SCFAs; e.g., acetate, propionate, and butyrate) (*9*–*14*) not only fuel host cells but also serve as signaling molecules between the gut microbiota and extraintestinal organs.

We have previously shown that SCFAs suppress insulin signaling in adipocytes and ultimately inhibit fat deposition via adipose GPR43 (*15*). Furthermore, GPR41 stimulation by propionate potently activates sympathetic neurons to regulate energy expenditure (*16*). A recent study further showed that feeding pregnant mice a high-fiber or acetate-supplemented diet decreases offspring susceptibility to allergic airway disease (*17*). These findings raise the possibility that maternal SCFAs play a key role in the regulation of disease susceptibility during postnatal life in the context of the developmental origins of health and disease theory (*18*). However, the underlying mechanisms and the biological importance of the maternal-embryonic cross-talk via microbial metabolites remain obscure.

In this study, we explored the impact of the maternal gut microbiota on energy homeostasis in offspring in a mouse model. We observed that the offspring of germ-free (GF) mothers are more prone to obesity and glucose intolerance than those of specific pathogen–free (SPF) mothers. Maternal microbiota–derived SCFAs translocated to the embryos to facilitate development of the sympathetic nervous system and regulation of insulin levels via GPR41 and GPR43 signaling. Thus, during pregnancy, the gut microbiota provides an environmental cue that fine-tunes energy homeostasis in offspring.

Offspring of GF mothers and obesity development

To investigate the impact of the maternal gut microbiota during pregnancy on offspring, pregnant mice were bred under SPF and GF conditions. On day 18.5 of gestation, pregnant GF mice received a fecal microbiota transplant from SPF mice of the corresponding strain to prevent overgrowth of unfavorable microbes. Newborn animals were raised by foster mothers under conventional conditions to align growth environments after birth. After weaning, the male mice were fed a high-fat diet (HFD) to induce obesity (Fig. 1A). Although the postnatal body weight of newborns from GF ICR mothers was less than that of offspring from SPF ICR mothers (fig. S1A), the offspring developed marked obesity upon HFD consumption during growth (Fig. 1A). In addition, the weight of perirenal or subcutaneous white adipose tissue (WAT) and the liver was significantly higher in offspring derived from GF mothers (GF offspring) than in those from SPF mothers (SPF offspring) at 16 weeks (Fig. 1A), which is in accordance with increases in WAT adipocyte size and hepatic triglycerides (TGs) (fig. S1, B and C). Concomitantly, plasma glucose, TGs, non-esterified fatty acids (NEFAs), and total cholesterol levels were significantly elevated in GF offspring (Fig. 1B). Body temperature and heart rate were significantly reduced (Fig. 1C), whereas plasma insulin levels and pancreatic islet sizes were significantly higher in GF offspring than in SPF offspring (fig. S1D). Moreover, the GF offspring exhibited elevated food intake (fig. S1E), with reduced plasma levels of the gut hormone peptide YY (PYY) and glucagon-like peptide-1 (GLP-1) (Fig. 1D) as well as reduced energy expenditure (Fig. 1E). These results indicate that the GF offspring exhibited an obese phenotype upon HFD feeding. In support of this interpretation, HFD-induced glucose intolerance and insulin resistance were significantly accelerated in GF offspring (Fig. 1F), indicating impaired insulin sensitivity. Notably, female GF offspring also exhibited similar phenotypes (fig. S2, A to H).

Mammalian neonates are initially exposed to the vaginal microbiota, which substantially contributes to the establishment of the gut

[1]Department of Applied Biological Science, Graduate School of Agriculture, Tokyo University of Agriculture and Technology, Fuchu-shi, Tokyo 183-8509, Japan. [2]AMED-CREST, Japan Agency for Medical Research and Development, Chiyoda-ku, Tokyo 100-0004, Japan. [3]Division of Biochemistry, Faculty of Pharmacy and Graduate School of Pharmaceutical Science, Keio University, Tokyo 105-8512, Japan. [4]Division of Gastroenterology and Hepatology, Department of Internal Medicine, Keio University School of Medicine, Tokyo 160-8582, Japan. [5]Institute of Health Sciences, Ezaki Glico Co., Ltd., Osaka 555-8502, Japan. [6]Laboratory for Metabolomics, RIKEN Center for Integrative Medical Sciences, Kanagawa 230-0045, Japan. [7]Department of Genomic Drug Discovery Science, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan. [8]Department of Nutritional Science and Food Safety, Tokyo University of Agriculture, Tokyo 156-8502, Japan. [9]Division of Pharmaceutical Sciences, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama 700-8530, Japan. [10]Department of Medicine, Shiga University of Medical Science, Shiga 520-2192, Japan. [11]Nishiwaki Laboratory, CMIC Pharma Science Co., Ltd., Hyogo 677-0032, Japan. [12]Department of Endocrinology, Metabolism and Nephrology, School of Medicine, Keio University, Tokyo 160-8582, Japan. [13]Division of Epidemiology, Kobe University Graduate School of Medicine, Kobe 650-0017, Japan. [14]Department of Bioorganic Medicinal Chemistry, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan. [15]Division of Physiological Chemistry and Metabolism, Keio University Faculty of Pharmacy, Tokyo, 105-0011, Japan. [16]Cellular and Molecular Epigenetics Laboratory, Graduate School of Medical Life Science, Yokohama City University, Kanagawa 230-0045, Japan. [17]International Research and Development Center for Mucosal Vaccines, The Institute of Medical Science, The University of Tokyo (IMSUT), Tokyo 108-8639, Japan.
*These authors contributed equally to this work.
†Corresponding author. Email: ikimura@cc.tuat.ac.jp (I.K.); hase-kj@pha.keio.ac.jp (K.H.)

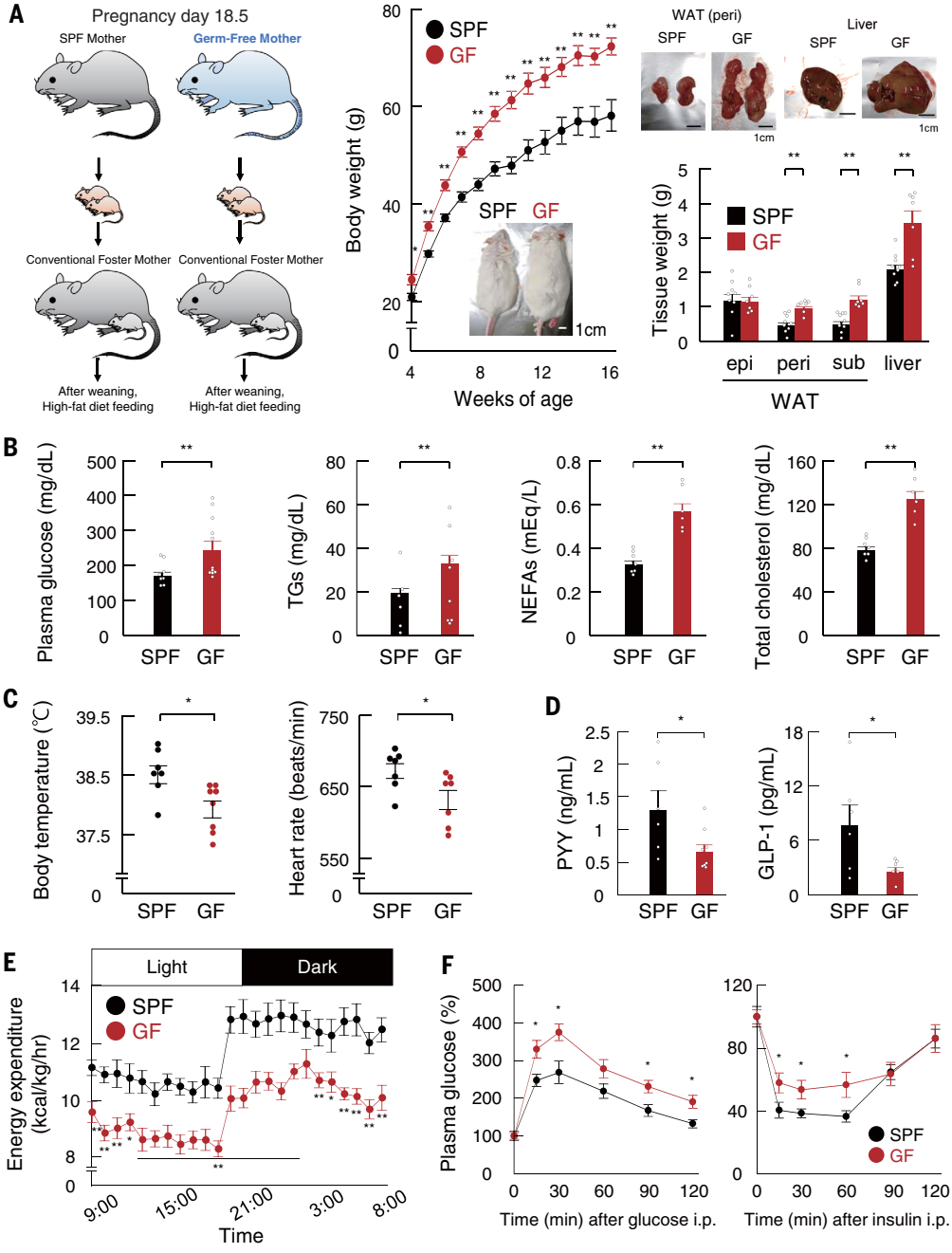


Fig. 1. Offspring from GF mothers exhibit severe obese phenotype when fed a HFD. (**A**) Experimental scheme (left). Body weight changes during the HFD trial (middle) (n = 14 animals per group) and tissue weights (right) (n = 7 to 9 tissues per group). epi, epididymal; peri, perirenal; sub, subcutaneous. (**B**) Plasma glucose, TGs, NEFAs, and total cholesterol levels (n = 6 to 11 plasma samples per group). mEq, milliequivalents. (**C**) Body temperature (left) (n = 7 or 8 animals per group) and heart rate (right) (n = 7 animals per group). (**D**) Gut hormone PYY (left) (n = 6 or 8 plasma samples per group) and GLP-1 (right) (n = 6 plasma samples per group) levels. (**E**) Energy expenditure (n = 7 or 8 animals per group). (**F**) Glucose tolerance test (left) and insulin tolerance test (right) (n = 8 animals per group). Male mice were analyzed at 16 weeks of age. Student's t test; $^{**}P < 0.01$ and $^{*}P < 0.05$. Data are presented as means ± SEM. Offspring from three or four litters per group were used. SPF, conventional offspring derived from ICR SPF mothers; GF, conventional offspring derived from ICR GF mothers.

microbial community in infants (*19*) and thus potentially influences the development of the host metabolic system. However, 16*S* ribosomal DNA (rDNA) amplicon sequencing showed that the relative abundance of bacterial families constituting the gut microbiota was similar in offspring from SPF and GF ICR mothers during adulthood, although GF offspring in infancy showed significant decreases in Streptococcaceae and Enterococcaceae compared with SPF offspring (fig. S3A). Principal coordinate analysis based on weighted UniFrac distances confirmed that there were no differences between the two groups during adulthood, but not during infancy (fig. S3B). To exclude the possible influence of the vaginal microbiota, newborns from SPF and GF mothers were also delivered by caesarean section. Consistent with vaginally delivered offspring, caesarean GF offspring exhibited severe obesity upon HFD feeding during growth (fig. S4A). WAT and liver weight, plasma metabolic parameters, and insulin levels were also significantly higher in the caesarean GF offspring at 16 weeks than in the caesarean SPF offspring (fig. S4, B to D). Furthermore, the former showed reduced energy expenditure (fig. S4E). Additionally, compositions of the gut microbiota were similar between caesarean-delivered offspring from the SPF and GF mothers during adulthood, whereas infant microbiota may be affected slightly by the delivery modes (fig. S5, A and B), consistent with the findings of a recent study (*19*). Collectively, these data show that the gut microbiota in adulthood is not a primary factor in the obesity-prone phenotype of the GF offspring.

Interaction between the gut microbiota and host genetics modulates the predisposition to obesity, whereas environment factors such as diet are indispensable for the regulation of host-microbe interaction (*20*–*22*). Although gut microbial compositions tended to be different between offspring from ICR and C57BL/6J mothers (fig. S6, A and B), they showed a similar obesity phenotype (Fig. 1, A to F and figs.

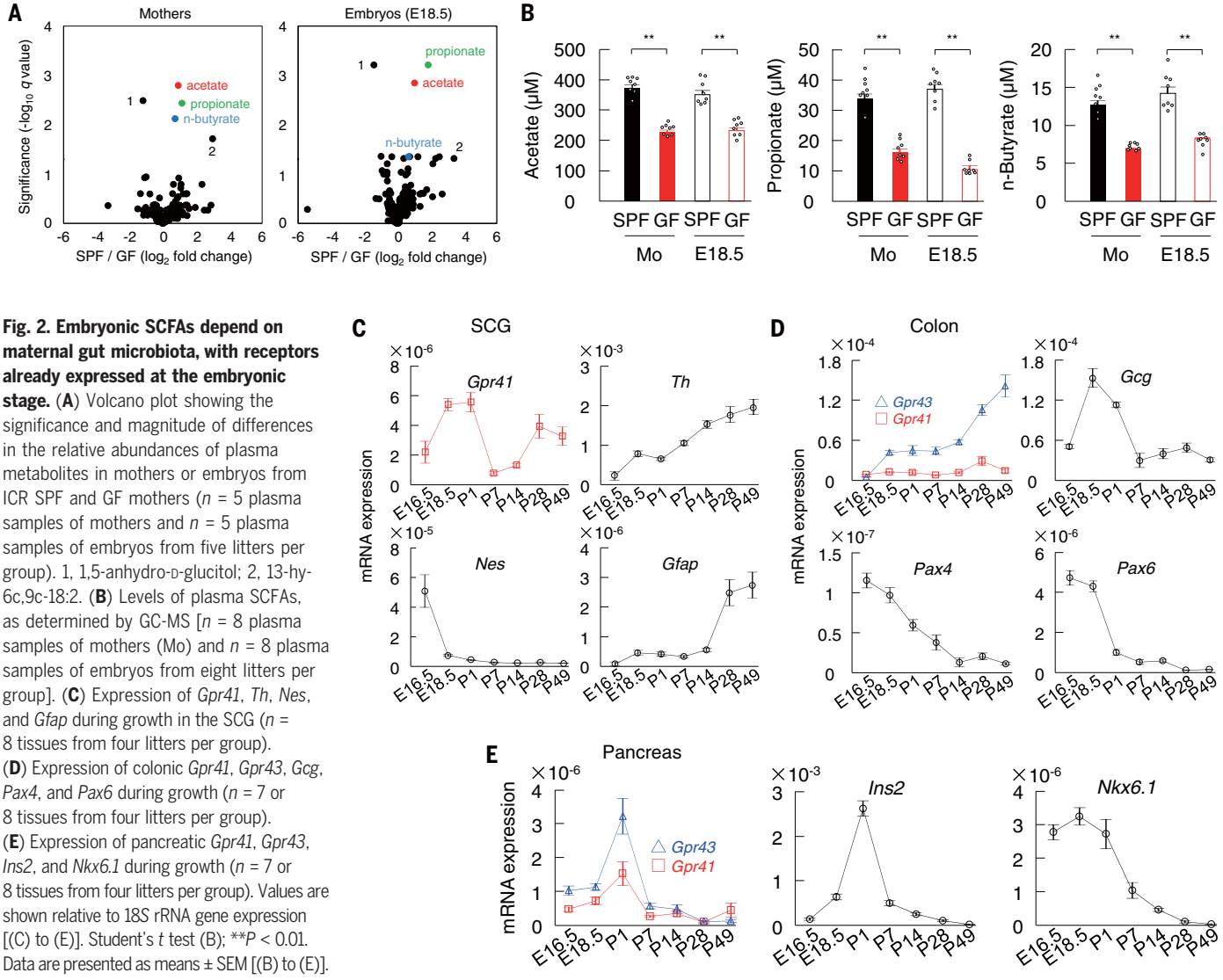


Fig. 2. Embryonic SCFAs depend on maternal gut microbiota, with receptors already expressed at the embryonic stage. (**A**) Volcano plot showing the significance and magnitude of differences in the relative abundances of plasma metabolites in mothers or embryos from ICR SPF and GF mothers (n = 5 plasma samples of mothers and n = 5 plasma samples of embryos from five litters per group). 1, 1,5-anhydro-D-glucitol; 2, 13-hy-6c,9c-18:2. (**B**) Levels of plasma SCFAs, as determined by GC-MS [n = 8 plasma samples of mothers (Mo) and n = 8 plasma samples of embryos from eight litters per group]. (**C**) Expression of *Gpr41*, *Th*, *Nes*, and *Gfap* during growth in the SCG (n = 8 tissues from four litters per group). (**D**) Expression of colonic *Gpr41*, *Gpr43*, *Gcg*, *Pax4*, and *Pax6* during growth (n = 7 or 8 tissues from four litters per group). (**E**) Expression of pancreatic *Gpr41*, *Gpr43*, *Ins2*, and *Nkx6.1* during growth (n = 7 or 8 tissues from four litters per group). Values are shown relative to 18*S* rRNA gene expression [(C) to (E)]. Student's t test (B); $^{**}P < 0.01$. Data are presented as means ± SEM [(B) to (E)].

S7, A to H). Furthermore, both male and female GF offspring from C57BL/6J mothers developed severe obesity and metabolic disorders upon HFD feeding (figs. S7, A to H, and S8, A to H). Thus, to a greater or lesser extent, metabolic disorders in GF offspring were commonly observed regardless of strain and sex.

Sensing maternal SCFAs in the embryo

We could not detect bacteria in the amniotic fluid of pregnant SPF ICR mice at our animal facility (fig. S9, A and B). We hypothesized that metabolite signals (*23*–*25*), derived from the maternal gut microbiota, may translocate to the fetus and influence the development of the metabolic system. Thus, we profiled the plasma levels of hydrophilic and lipophilic metabolites in SPF and GF ICR mothers and their embryos during pregnancy. The levels of 5 metabolites in the mothers and 12 metabolites in the embryos were significantly different between the SPF and GF groups (fig. S10); among them, only 5 metabolites showed similar changes in the mothers and embryos in response to breeding conditions (Fig. 2A). In particular, the plasma levels of SCFAs (acetate, propionate, and butyrate) were significantly lower in GF mothers and embryos than in their SPF counterparts (Fig. 2B). Given that plasma SCFA levels were constant in both mothers and embryos of the SPF group during pregnancy (fig. S11), the maternal gut microbiota appears to constitutively supply SCFAs to embryos via the bloodstream.

Regulation of host energy metabolism through activation of GPR41 and GPR43 by gut microbial SCFAs has been documented (*15*, *16*, *26*, *27*). We detected *Gpr41* mRNA in the sympathetic ganglia of embryos (Fig. 2C and fig. S12, A and B) with biphasic expression in the embryonic and adult stages (Fig. 2C). Such an expression pattern was not observed for nestin (*Nes*; an undifferentiated neural marker), tyrosine hydroxylase (*Th*; a sympathetic neuronal marker), or glial fibrillary acidic protein (*Gfap*; a glial marker) (Fig. 2C). Meanwhile, *Gpr43* mRNA was detected in the intestinal tract from embryonic day 15.5 (E15.5) onward (fig. S12C), although its expression was restricted to enteroendocrine cells in the adult stage (fig. S12D), as previously reported (*28*). Biphasic increases in *Gpr43* expression were observed in the colon in the embryonic and adult stages, with expression peaking later than that of *Pax4* and *Pax6* (regulators of intestinal enteroendocrine cell differentiation) and at a similar time point to that of *Gcg* (intestinal enteroendocrine cell maker, GLP-1) (Fig. 2D). These SCFA receptors are also expressed in the pancreas and regulate insulin secretion in adult mice (*27*, *29*). Levels of pancreatic *Gpr43* mRNA, but not *Gpr41*, transiently increased during the perinatal-postnatal period, with expression peaking later than that of *Nkx6.1* (early β cell differentiation–related factor) and at a similar time point to that of *Ins2* (pancreatic β cell maker) (Fig. 2E).

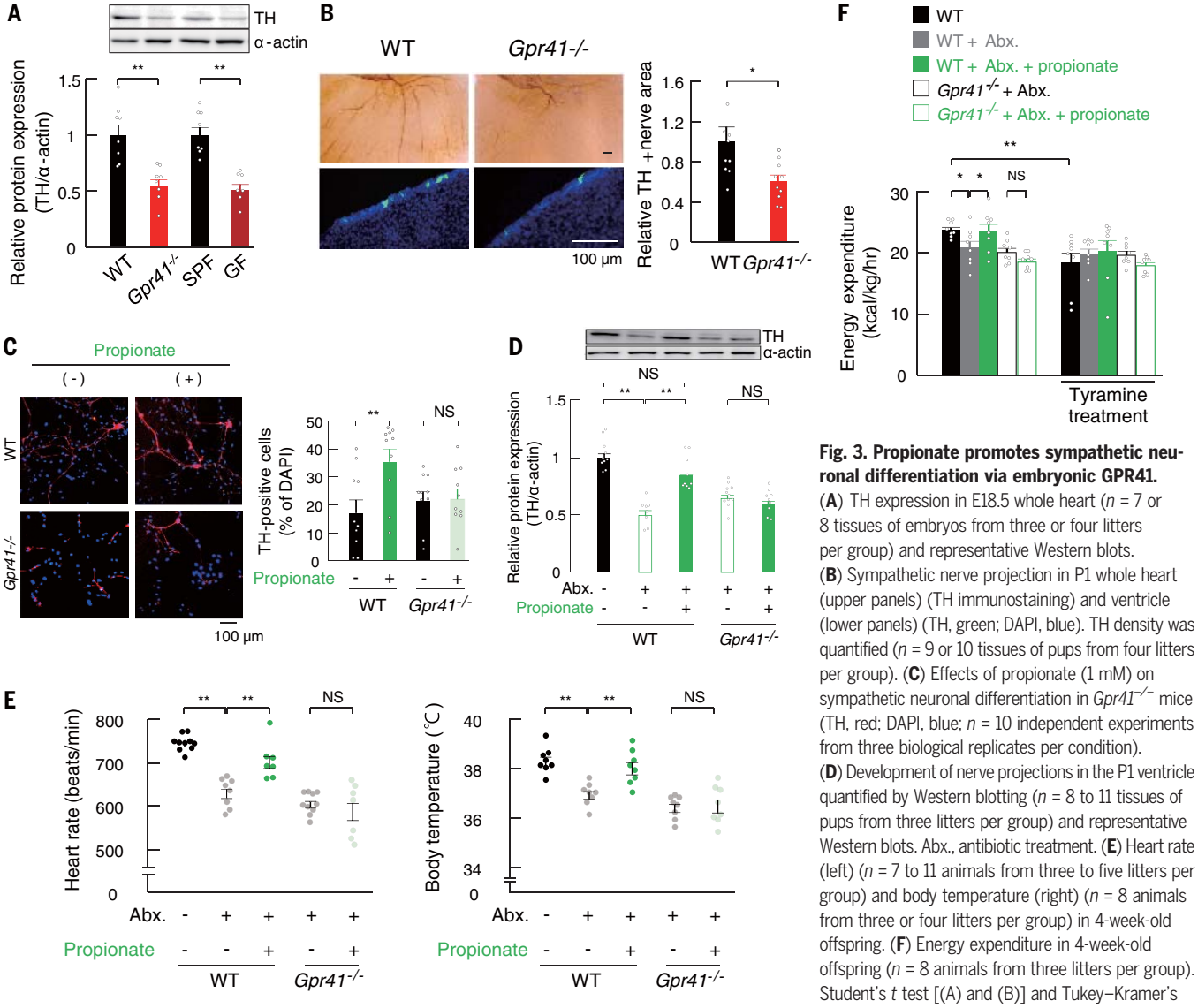


Fig. 3. Propionate promotes sympathetic neuronal differentiation via embryonic GPR41. (**A**) TH expression in E18.5 whole heart (n = 7 or 8 tissues of embryos from three or four litters per group) and representative Western blots. (**B**) Sympathetic nerve projection in P1 whole heart (upper panels) (TH immunostaining) and ventricle (lower panels) (TH, green; DAPI, blue). TH density was quantified (n = 9 or 10 tissues of pups from four litters per group). (**C**) Effects of propionate (1 mM) on sympathetic neuronal differentiation in *Gpr41*$^{-/-}$ mice (TH, red; DAPI, blue; n = 10 independent experiments from three biological replicates per condition). (**D**) Development of nerve projections in the P1 ventricle quantified by Western blotting (n = 8 to 11 tissues of pups from three litters per group) and representative Western blots. Abx., antibiotic treatment. (**E**) Heart rate (left) (n = 7 to 11 animals from three to five litters per group) and body temperature (right) (n = 8 animals from three or four litters per group) in 4-week-old offspring. (**F**) Energy expenditure in 4-week-old offspring (n = 8 animals from three litters per group). Student's t test [(A) and (B)] and Tukey–Kramer's test [(C) to (F)]; $^{**}P < 0.01$ and $^{*}P < 0.05$. NS, not significant. Data are presented as means ± SEM.

Notably, *Gpr43* mRNA expression in both the colon and pancreas of GF embryos was significantly lower than that in SPF embryos, although *Gpr41* mRNA expression in the sympathetic ganglia of embryos was similar among the two groups (fig. S12, E to G). These findings imply that the embryonic metabolic tissues, such as the sympathetic nervous system, intestinal tract, and pancreas, may sense maternal gut microbe–derived SCFAs by expressing GPR41 and GPR43.

Sympathetic development via GPR41

We further investigated the functions of GPR41 in the sympathetic nervous system and of GPR43 in the intestine and pancreas during embryonic development. We found that sympathetic nerve projections to the heart were significantly reduced in *Gpr41*$^{-/-}$ C57BL/6J embryos compared with wild-type (WT) embryos, and such an abnormality was also apparent in GF embryos with the same background (Fig. 3A). Moreover, sympathetic nerve projections to the heart were significantly reduced in *Gpr41*$^{-/-}$ pups on postnatal day 1 (P1), even though these pups were delivered from mothers maintained under SPF conditions (Fig. 3B). Gut microbial compositions were similar for WT, *Gpr41*$^{-/-}$, and *Gpr43*$^{-/-}$ pregnant C57BL/6J mice (fig. S13, A and B). Plasma SCFA levels were also comparable among the three groups (fig. S13C).

We further sought to investigate the contribution of GPR41 signaling to sympathetic neuronal differentiation. In a primary culture of embryonic superior cervical ganglion (SCG)–derived neural cells, each of the tested SCFAs, especially propionate, significantly promoted sympathetic neuronal differentiation (fig. S14A). This finding is consistent with the notion that the most potent agonist of GPR41 is propionate, followed by butyrate and acetate, with median effective concentration (EC_{50}) values of approximately 10, 40, and 2000 μM, respectively (*30*). Propionate-induced sympathetic neuronal differentiation was abolished in sympathetic neural cells from *Gpr41*$^{-/-}$ embryos (Fig. 3C). Furthermore, $G_{i/o}$-mediated (but not $G_{i/o}\alpha$-mediated) GPR41 propionate signaling increased sympathetic neurite length via Gβγ-mediated mitogen-activated protein kinase activation (fig. S14, B and C). These findings strongly suggest that propionate-mediated GPR41 activation promotes sympathetic neuronal differentiation. To test this notion, we treated WT pregnant mice with a cocktail of antibiotics to eliminate intestinal microbiota. Sympathetic nerve projections to the heart were significantly attenuated in pups from antibiotic-treated mice on P1. Notably, this abnormality

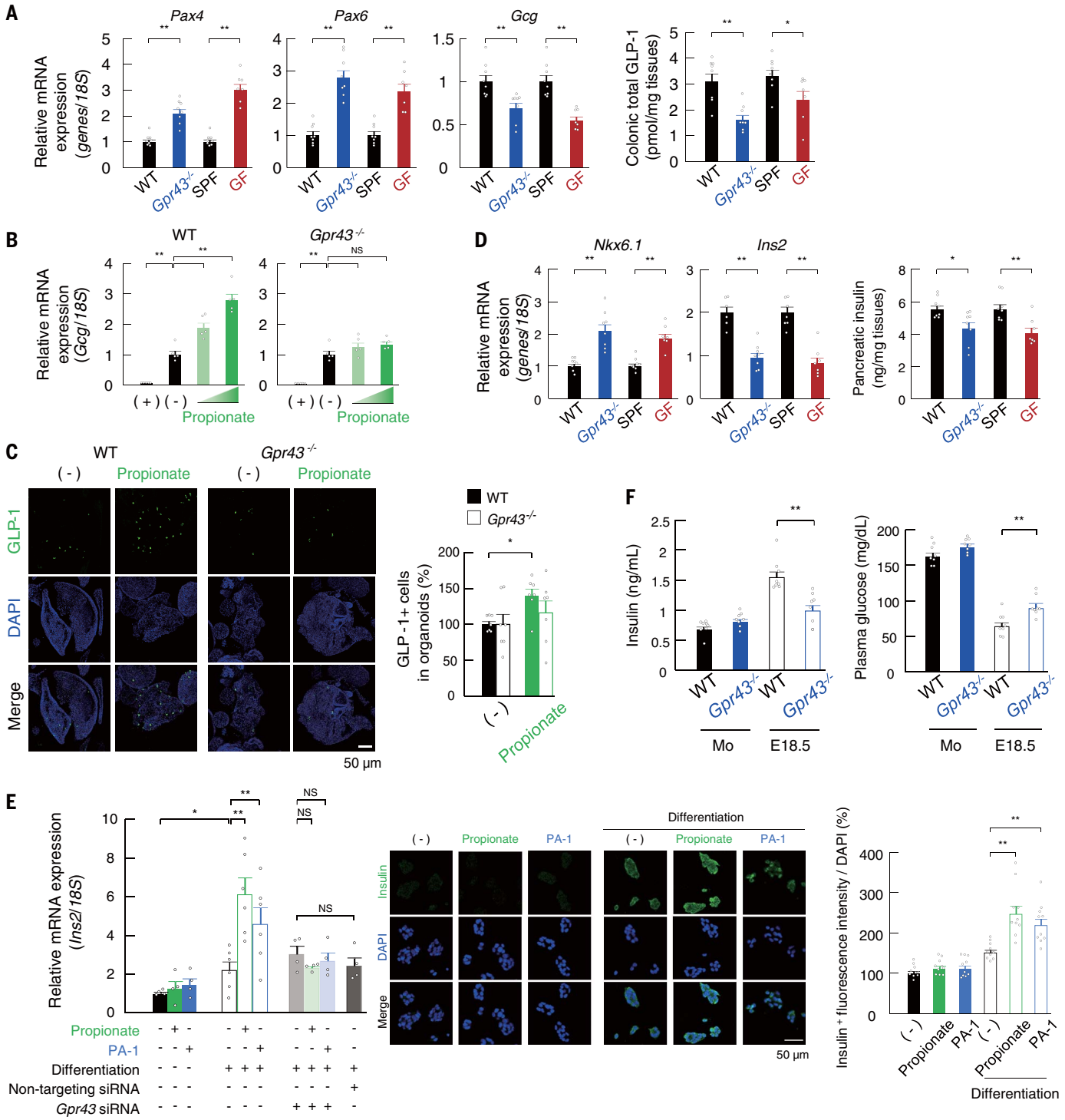


Fig. 4. Propionate promotes enteroendocrine and pancreatic β cell differentiation via embryonic GPR43. (**A**) *Pax4*, *Pax6*, and *Gcg* mRNA expression and GLP-1 protein levels in the colon (E18.5) (n = 7 or 8 tissues of embryos from three litters per group). (**B** and **C**) Expression of *Gcg* (B) and distribution of GLP-1–positive cells (C) after 24-hour treatment of embryonic organoids (E15.5) in the presence (+) or absence (-) of Wnt with GPR43 ligands [1 and 10 mM (B); 100 μM propionate (C)]. GLP-1, green; DAPI, blue; n = 5 to 7 independent experiments from two biological replicates per condition. (**D**) Expression of *Nkx6.1* and *Ins2* mRNA and insulin protein levels in the mouse pancreas (E18.5) (n = 7 or 8 tissues of embryos from three litters per group). (**E**) Expression of *Ins2* (left) and localization (middle; immunostaining) of insulin, and insulin positive cell count (right) in the presence or absence of GPR43 ligands (10 μM PA-1, 1 mM propionate) after 72-hour treatment (insulin, green; DAPI, blue; n = 4 to 10 independent experiments from two or three biological replicates per condition). For induction of β cell differentiation, cells were cultured with betacellulin and activin A. (**F**) Plasma insulin (left) and plasma glucose (right) levels in mothers and embryos of E18.5 (n = 8 plasma samples and n = 8 plasma samples of embryos from eight litters per group). GF and SPF mice with C57BL/6J background were analyzed [(A), (D), and (F)]. Student's t test [(A), (D), and (F)] and Tukey–Kramer's test [(B) and (E)]; $^{**}P < 0.01$ and $^{*}P < 0.05$. NS, not significant. All data are presented as means ± SEM.

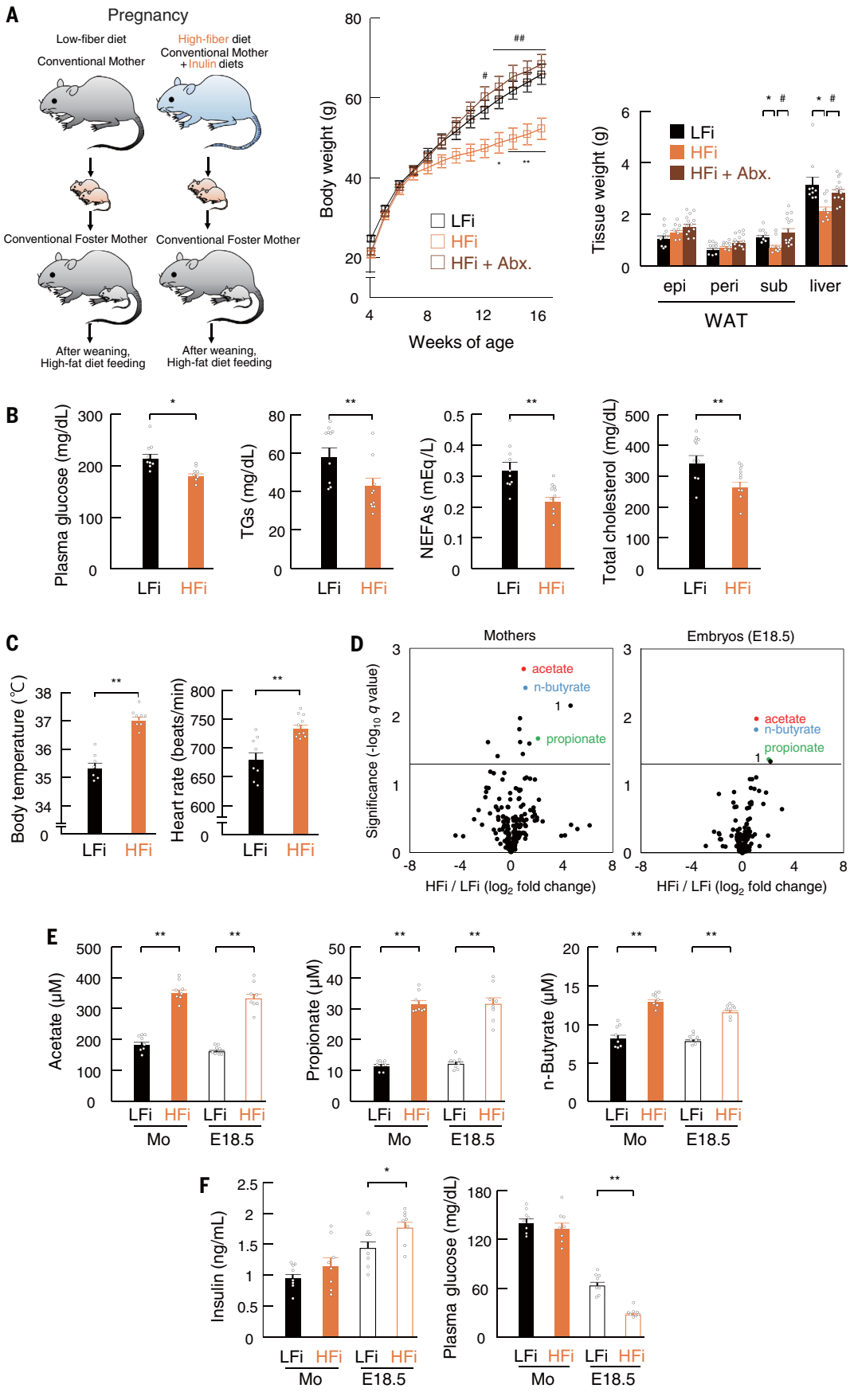


Fig. 5. Dietary fiber supplementation of pregnant mothers results in offspring resistance to obesity. (**A**) Experimental scheme (left). Body weight changes during HFD trial (middle) (n = 8 to 13 animals from four litters per group). Tissue weight (right) (n = 8 to 13 tissues from four litters per group). (**B**) Plasma glucose, TGs, NEFAs, and total cholesterol levels (n = 8 to 10 plasma samples from four litters per group). (**C**) Body temperature (left) (n = 7 or 8 animals from three litters per group) and heart rate (right) (n = 8 or 10 animals from three litters per group). (**D**) Volcano plot showing the significance and magnitude of differences in the relative abundances of plasma metabolites in the mothers or embryos from LFi and HFi mothers (n = 5 plasma samples and n = 5 plasma samples of embryos from five litters). 1, galactose. (**E**) Plasma SCFAs determined by GC-MS (n = 8 plasma samples of mothers and n = 8 plasma samples of embryos from eight litters per group). (**F**) Plasma insulin (left) and plasma glucose (right) levels in the mothers and embryos of E18.5 (n = 8 plasma samples of mothers and n = 8 plasma samples of embryos from eight litters per group). Male mice were analyzed at 16 weeks of age [(A) to (C)]. Student's t test [(B), (C), (E), and (F)] and Tukey–Kramer's test (A); $^{**}P < 0.01$ and $^{*}P < 0.05$ (LFi versus HFi); $\#\#P < 0.01$ and $\#P < 0.05$ (HFi versus HFi + Abx). Data are presented as means ± SEM [(A) to (C), (E), and (F)]. LFi, offspring derived from ICR LFi mothers; HFi, offspring derived from ICR HFi mothers [(A) to (C)].

was ameliorated by administration of propionate during pregnancy (Fig. 3D). Heart rate and body temperature after weaning, an index of sympathetic nerve projection, exhibited a similar trend, as did oxygen consumption (Fig. 3, E and F). Treatment with tyramine markedly decreased oxygen consumption in offspring from untreated or propionate- or antibiotic-treated WT mothers, whereas these effects were weakened in offspring from antibiotic-treated WT mothers or untreated *Gpr41*$^{-/-}$ mothers (Fig. 3F). On the basis of these observations, we reasoned that propionate from the maternal gut microbiota facilitates sympathetic nerve development via GPR41. Deprivation of propionate resulted in sympathetic dysfunction, including a reduction in body temperature and heart rate fluctuations, as observed in the GF offspring (Fig. 1C).

Embryonic insulin regulation via GPR43

GPR43 is expressed in adult enteroendocrine L cells and promotes the secretion of gut hormones upon activation (*28*). Because *Gpr43* was also detected in the embryonic colon (fig. S12C), we investigated the effect of GPR43 on enteroendocrine cell differentiation at the embryonic stage. *Pax4* and *Pax6* were significantly up-regulated in the colon of *Gpr43*$^{-/-}$ C57BL/6J embryos in comparison with WT embryos (Fig. 4A and fig. S15A). Such changes were also observed in GF embryos with the same background. By contrast, *Gcg* as well as GLP-1 were down-regulated in *Gpr43*$^{-/-}$ and GF mice (Fig. 4A and fig. S15A). This result suggests a retardation of enteroendocrine cell differentiation in *Gpr43*$^{-/-}$ and GF mice. To directly assess the role of the SCFA-GPR43 axis in enteroendocrine cell differentiation, we employed intestinal organoids, which recapitulate cell differentiation in vivo by reducing canonical Wnt signaling (fig. S15B). Propionate, which is a more potent (EC_{50}: ~30 μM) GPR43 ligand than acetate or butyrate (EC_{50}: ~50 or 100 μM, respectively) (*31*), significantly promoted differentiation of GLP-1^{+} enteroendocrine cells in intestinal organoids from WT embryos (Fig. 4, B and C). In sharp contrast, this effect was abrogated in organoids from *Gpr43*$^{-/-}$ embryos (Fig. 4, B and C). Notably, the effect of propionate on enteroendocrine cell differentiation was not observed in organoids from adult WT mice (fig. S15C), illustrating that propionate-mediated GPR43 signaling is a prerequisite for the development of enteroendocrine cells during the prenatal period.

Because GPR43 regulates insulin secretion in pancreatic β cells (*27*, *29*) and this receptor was expressed in the pancreas during the perinatal-postnatal period (Fig. 2E), we subsequently investigated the effect of GPR43 on β cell differentiation during embryonic development. *Nkx6.1* was significantly up-regulated in the pancreas of embryos from *Gpr43*$^{-/-}$ and GF mice compared with those from WT and SPF mice, respectively, whereas *Ins2* expression and insulin level were down-regulated in *Gpr43*$^{-/-}$ and GF mice (Fig. 4D and fig. S15D). Moreover, during the induction of differentiation of the rat pancreatic tumor cell line AR42J into β cells, expression levels of *Gpr43* mRNA, but not *Gpr41* mRNA, were significantly elevated in line with those of *Ins2* (fig. S15E). Notably, propionate and the synthetic GPR43 agonist phenylacetamide-1 (PA-1) promoted differentiation into insulin^{+} β cells; however, this effect was compromised by RNA interference–mediated knockdown of *Gpr43* (Fig. 4E and fig. S15F). Thus, propionate-mediated activation of GPR43 facilitates differentiation into pancreatic β cells as well as GLP-1–positive enteroendocrine cells. GF ICR as well as GF C57BL/6J embryos also demonstrated abnormalities in these differentiation and functional markers of sympathetic neuron, enteroendocrine, and pancreatic β cells (fig. S15, A, D, and G).

These findings raise the possibility that SCFA-GPR43 signaling may play a vital role in embryonic insulin secretion. Plasma insulin levels in *Gpr43*$^{-/-}$ embryos were markedly lower than in WT embryos, although there were no differences between WT and *Gpr43*$^{-/-}$ mothers (Fig. 4F). Likewise, reduced insulin levels were also observed in GF embryos (fig. S16, A and B). Correspondingly, plasma glucose levels in *Gpr43*$^{-/-}$ (Fig. 4F) and GF embryos (fig. S16, A and B) were significantly higher than those in their control groups. Given that dysregulation of fetal glucose levels renders offspring susceptible to metabolic syndromes, such as obesity and type 2 diabetes (*32*), we speculate that the lack of SCFA-GPR43 signaling during the prenatal period causes metabolic syndrome in adulthood, most likely by compromising energy homeostasis in the embryos. Such retardation of the differentiation of the sympathetic nerve, intestinal tract, and pancreas was also evident in *Gpr41*$^{-/-}$*Gpr43*$^{-/-}$ double-mutant C57BL/6J mice (fig. S17, A to E).

Dietary fiber intake during pregnancy

To provide further evidence of the importance of SCFAs in the developmental origin of obesity resistance, we performed a dietary intervention study in which pregnant ICR mice were fed a high-fiber (HFi) or low-fiber (LFi) diet under conventional conditions, after which the susceptibility of their offspring to obesity was examined (Fig. 5A). Although the body weight of postpartum offspring from HFi-fed mothers (HFi offspring) was significantly higher than that of offspring from LFi-fed mothers (LFi offspring) (fig. S18A), HFi intake suppressed the HFD-induced body weight gain from 13 weeks of age onward, in accordance with reduced subcutaneous WAT and liver weights (Fig. 5A). However, the effect of a HFi diet was abrogated when antibiotics were administered to pregnant mice to eradicate the gut microbiota (Fig. 5A and fig. S18A), indicating that microbial fermentation of dietary fiber contributes to obesity suppression. Plasma metabolic parameters were also improved in HFi offspring compared with LFi offspring (Fig. 5B and fig. S18, B and C). Likewise, HFi offspring were resistant to HFD-induced glucose intolerance and insulin resistance (fig. S18D), in association with improved energy expenditure (fig. S18E). Furthermore, sympathetic dysfunction, such as reduction in body temperature and heart rate fluctuations, in LFi offspring was ameliorated in HFi offspring (Fig. 5C). Compositions of the gut microbiota were comparable between LFi offspring and HFi offspring during adulthood but were significantly different between the two groups during infancy (fig. S19, A and B). Metabolome profiling of maternal and embryonic plasma samples revealed that 11 and 4 metabolites were significantly altered between the LFi- and HFi-fed groups (fig. S20), with 4 metabolites commonly increased in the HFi-fed mothers and their embryos (Fig. 5D). Among these 4 metabolites, SCFAs were the only common factor in the SPF versus GF and LFi versus HFi comparisons. SCFA levels were significantly higher in the embryos of HFi-fed mice (HFi embryos) than in those of LFi-fed mice (LFi embryos) (Fig. 5E). We also observed that plasma insulin levels were significantly higher in HFi embryos than in LFi embryos (Fig. 5F), and thereby plasma glucose levels were significantly decreased in HFi embryos (Fig. 5F). Thus, SCFAs generated by the maternal gut microbiota through the fermentation of dietary fiber are provided to embryos via maternal circulation, improving fetal glucose homeostasis and imparting resistance to obesity in the offspring.

SCFA supplementation during pregnancy

Plasma propionate levels in HFi embryos were likely sufficient to activate the GPR41 and/or GPR43 receptors, considering that the determined values were superior to their EC_{50} values (Fig. 5E). Positron emission tomography (PET) imaging showed that [^{11}C]-labeled propionate in the colonic lumen reached the embryos via the maternal liver and bloodstream within 40 min after infusion (fig. S21, A to D). Thus, to rigorously examine the role of propionate in obesity resistance of the offspring, we fed pregnant ICR mice a LFi diet supplemented with propionate (Fig. 6A). The intake of this diet raised plasma levels of propionate in both mothers and embryos (fig. S22A). Treatment with propionate suppressed the HFD-induced increases in body weight, perirenal or subcutaneous WAT mass, and liver weight of adult offspring (Fig. 6A). Plasma metabolic parameters were also improved in the offspring of propionate-treated mothers (Pro offspring)

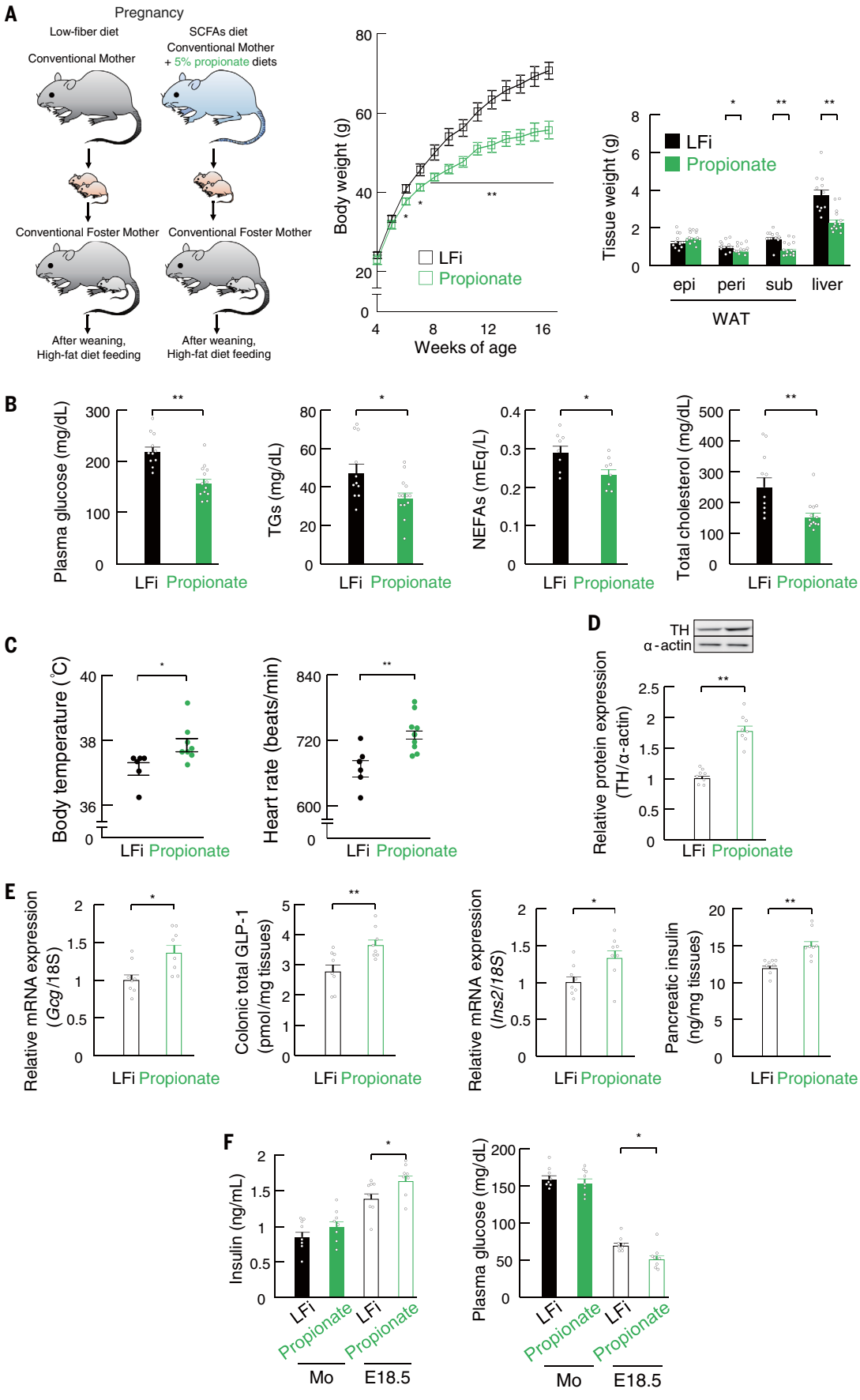


Fig. 6. Propionate supplementation of pregnant mothers results in offspring resistance to obesity. (**A**) Experimental scheme (left). Body weight changes during HFD trial (middle) (n = 11 to 13 animals from four litters per group). Tissue weight (right) (n = 11 to 13 tissues from four litters per group). (**B**) Plasma glucose, TGs, NEFAs, and total cholesterol levels (n = 8 to 13 plasma samples from four litters per group). (**C**) Body temperature (left) (n = 6 or 8 animals from four litters per group) and heart rate (right) (n = 6 or 9 animals from four litters per group). (**D**) TH expression in E18.5 whole heart (n = 8 tissues of embryos from three litters per group) and representative Western blots. (**E**) *Gcg* mRNA expression and GLP-1 protein levels in the colon (E18.5) (left) (n = 8 tissues of embryos from three litters per group). Expression of *Ins2* mRNA and insulin protein levels in the pancreas (E18.5) (right) (n = 8 tissues of embryos from three litters per group). (**F**) Plasma insulin (left) (n = 8 plasma samples of mothers and n = 8 plasma samples of embryos from eight litters per group) and plasma glucose (right) (n = 8 plasma samples of mothers and n = 8 plasma samples of embryos from eight litters per group) levels in mothers and embryos (E18.5). Male mice were analyzed at 16 weeks of age [(A) to (C)]. Student's t test [(A) to (F)]; $^{**}P < 0.01$ and $^{*}P < 0.05$. Data are presented as means ± SEM. LFi, offspring derived from ICR LFi mothers; propionate, offspring derived from ICR propionate-supplemented mothers [(A) to (C)].

compared with control LFi offspring (Fig. 6B and fig. S22, B and C). HFD-induced glucose intolerance and insulin resistance were also significantly ameliorated in Pro offspring (fig. S22D), and energy expenditure was improved in Pro offspring (fig. S22E). Furthermore, sympathetic dysfunction in the LFi offspring was rescued in Pro offspring (Fig. 6C). Compositions of the gut microbiota were similar between LFi offspring and Pro offspring during both infancy and adulthood (fig. S23, A and B). Additionally, maternal intervention with propionate reversed the retardation of sympathetic nerve projections to the heart as well as the retardation of GLP-1^{+} enteroendocrine cell and pancreatic β cell differentiation in embryos of LFi-fed mothers (Fig. 6, D and E); it also enhanced plasma insulin levels in the embryos, restoring them to levels comparable to those of HFi embryos (Fig. 6F). Consequently, an increase in plasma glucose levels in control embryos was efficiently suppressed in embryos of mothers given propionate (Fig. 6F). Consistent with the offspring from HFi-fed mothers treated with antibiotics, offspring from HFi-fed GF ICR mothers recapitulated the obesity-prone phenotype, which was rescued by administration of propionate during pregnancy (fig. S24, A to H). Together, these observations define the importance of maternal propionate, which renders offspring resistant to obesity.

Discussion

In this study, we determined that maternal gut microbiota during pregnancy confers resistance to obesity in offspring via the SCFA-GPR41 and SCFA-GPR43 axes. During pregnancy, SCFAs from the maternal gut microbiota are sensed by GPR41 and GPR43 in the sympathetic nerve, intestinal tract, and pancreas of the embryo. SCFAs are known to exert pleiotropic effects through several mechanisms (*10*, *11*, *13*, *15*, *16*, *24*, *33*–*36*) such as histone deacetylase (HDAC) inhibition by butyrate [median inhibitory concentration (IC_{50}): ~90 to 170 μM] (*37*, *38*) and activation of GPR109A by butyrate (EC_{50}: ~700 μM) (*39*) and Olfr78 by acetate (EC_{50}: ~2300 μM) and propionate (EC_{50}: ~1000 μM) (*33*), in addition to GPR41 and GPR43. Considering the low concentrations of SCFAs (acetate: ~400 μM, propionate: ~50 μM, and butyrate: ~10 μM) in the embryonic circulating plasma, SCFAs were unlikely to interact with Olfr78 and GPR109A. Meanwhile, the concentrations of SCFAs were sufficient to activate GPR41 and GPR43 (*30*, *31*). Activation of embryonic GPR41 and GPR43 by SCFAs promoted sympathetic neuronal, enteroendocrine, and pancreatic β cell differentiation, which were essential for maintaining energy homeostasis (e.g., thermogenesis and heart rate) via the sympathetic nervous system and fetal glucose homeostasis. Furthermore, given that the expression of *Gpr43* tended to be down-regulated in the pancreas and colon of GF offspring, SCFAs such as butyrate and propionate [which also serves as an HDAC inhibitor (*40*)] may regulate embryonic development by regulating *Gpr41* and *Gpr43* gene expression through epigenetic modifications.

Several metabolites other than SCFAs showed similar changes in the plasma of mothers and embryos between SPF and GF conditions. However, in HFi-fed mothers and their embryos, only SCFAs were commonly elevated compared with those in the LFi-fed counterparts. Treatment of LFi-fed mothers with propionate repaired the defects in the differentiation of intestinal enteroendocrine cells and sympathetic neurons in the embryos, and the effects of propionate were abolished under *Gpr41* and *Gpr43* deficiencies. These observations underscore the contribution of the SCFA-GPR41 and SCFA-GPR43 axes in prenatal development of the metabolic and neural systems. Additionally, although offspring from GF-ICR and GF-C57BL/6J mothers developed obesity, the severity of the obese phenotype was slightly different between the two groups. Such variation may be attributed to gut microbial composition, which is substantially affected by host genetic backgrounds (*20*, *22*). Consistent with this notion, we found that plasma SCFA levels differed between ICR and C57BL/6J mothers (Fig. 2B and fig. S13C).

Whereas compositions of the gut microbiota during adulthood were comparable in the SPF and GF offspring, compositions in infancy were different between the two groups. A similar trend was also observed in infant HFi offspring compared with LFi offspring, raising the possibility that altered infant microbiota partially contributes to development of the obesity phenotype of GF and LFi offspring. Nevertheless, propionate administration to LFi mothers during pregnancy ameliorated the obesity-prone phenotype of the offspring without affecting the infant microbiota (fig. S23). Furthermore, propionate treatment also improved hyperglycemia in GF embryos and lower energy expenditure in LFi and GF offspring. Although GPR41 deficiency also caused lower energy expenditure, propionate administration during pregnancy failed to prevent this phenotype. On the basis of these observations, we reason that maternal microbiota–derived SCFAs, particularly propionate, play a vital role in preventing the development of metabolic disorder in offspring. Meanwhile, it should be noted that GF offspring exhibited the abnormality in energy expenditure in both light and dark cycles, whereas LFi offspring displayed the abnormality only in the dark cycle. Therefore, we cannot formally exclude the possibility that the presence of intestinal microbiota and/or their products, other than SCFAs, may contribute to the enhancement of energy expenditure in the light cycle.

Embryonic insulin regulation was impaired in embryos from GF mothers, and insulin levels were significantly elevated in the adult stage. Excessive insulin levels in adult GF offspring are most likely attributed to metabolic adaptations in response to low birth weight and the retardation of pancreatic β cell differentiation, eventually increasing susceptibility to obesity upon HFD feeding. This abnormality is reminiscent of catch-up growth, whereby children born small for gestational age (SGA) face a risk of excessive body weight gain and metabolic syndrome later in life (*18*, *41*), as evidenced by multiple birth cohort studies (*42*–*44*). Although the etiology of SGA births remains to be clarified, maternal factors, including malnutrition, smoking, and alcohol consumption, have been implicated (*45*). We further propose that deprivation of SCFAs as a result of disturbances in intestinal microbiota may be another causative factor for SGA births, leading to catch-up growth and susceptibility to obesity. Our study provides evidence for the crucial contribution of the maternal gut environment during pregnancy to the metabolic programming of offspring to prevent metabolic syndrome. This finding opens new research avenues into preemptive therapies for metabolic disorders by targeting the maternal gut microbiota.

Materials and methods

Animal study

All animal diets unless otherwise indicated were provided by Research Diets. GF ICR (Sankyo Labo Service) and C57BL/6J (Clea Japan) mice were housed in vinyl isolators under a 12-hour light-dark cycle and given regular chow (CMF, Oriental Yeast). For fecal microbiota transplantation, pregnant GF mice (day 18.5) received oral gavage of fecal suspension in phosphate-buffered saline (PBS) from strain-matched pregnant SPF mice. In caesarean section experiments, progesterone (2 mg per mouse) was subcutaneously injected into pregnant GF or SPF IQI mice (Clea, Japan) on days 17.5 and 18.5 to avoid preterm delivery. On day 19.5, newborns were delivered by caesarean section and reared by foster IQI mothers under conventional conditions for 4 weeks. Then, 4-week-old male mice were fed a HFD (D12492) for 12 weeks and housed individually (table S1). In separate experiments, conventional ICR mice (SLC, Inc.) were fed AIN-93G formula–based LFi or 10% inulin–containing HFi diets during pregnancy (table S2). To eliminate commensal bacteria, HFi-fed ICR mice were treated with 1 mg/ml neomycin (Nacalai Tesque) in drinking water during pregnancy. Offspring were reared by CMF-fed conventional foster mothers for 4 weeks; weaned mice were fed a HFD for 12 weeks and housed individually. In SCFA-supplementation experiments, conventional ICR mice were given an AIN-93G–based diet supplemented with 5% (w/w) propionate for

the pregnancy period (table S3). Offspring were reared by CMF-fed conventional foster mothers for 4 weeks. Over the course of the experiments, newborns [litter size aligned to 10 ± 2 (ICR) or 6 ± 2 (C57BL/6J) mice] from the GF and SPF mothers were reared by strain-matched foster mothers under conventional conditions for 4 weeks. Then, more than three groups of littermates from each mother were analyzed in individual experiments.

Heart rates were determined in conscious mice using a tail-cuff system (Softron and Muromachi Kikai), and body temperature was determined in conscious mice using Lifechip and a pocket reader (Destron Fearing), as described elsewhere (*16*).

C57BL/6J-background WT, *Gpr41*$^{-/-}$, and *Gpr43*$^{-/-}$ mice and *Gpr41*$^{-/-}$*Gpr43*$^{-/-}$ double-mutant pregnant mice were given CMF. *Gpr41*$^{-/-}$ and *Gpr43*$^{-/-}$ mice were generated as described previously (*15*, *16*); *Gpr41*$^{-/-}$*Gpr43*$^{-/-}$ double-mutant mice were generated by the CRISPR-Cas9 system (fig. S17, A and B). Two guide RNAs were identified using the publicly available optimized CRISPR design tool (http://crispr.mit.edu). Transgene insertion was established by homology-directed DNA repair recombination, and mutations were analyzed using genotyping primers. Pregnant WT and *Gpr41*$^{-/-}$ mice were treated with propionate (150 mM) and antibiotics (1 mg/ml neomycin) in drinking water. After birth, pups were analyzed on P1.

Plasma samples were obtained from embryos (6 to 12 per litter) delivered by WT and mutant mice at E18.5 and subjected to metabolome analysis. Plasma glucose concentrations were measured in individual embryos at E18.5. The SCG, heart, colon, and pancreas were collected from WT or mutant mice at E16.5, E18.5, P1, P7, P14, P28, or P49.

All experimental procedures involving mice were performed according to protocols approved by the Institutional Animal Care and Use Committee of Keio University School of Medicine [permission no. 09036-(12)]; the Committee on the Ethics of Animal Experiments of the Tokyo University of Agriculture and Technology (permit no. 28–87); and the Animal Care and Use Committee, Okayama University (permission no. OKU-2018346). Mice were treated with lethal anesthesia with somnopentyl, and all efforts were made to minimize animal suffering.

Histology

Livers, pancreas, and whole embryos were embedded in OCT compound (Sakura Finetek), and adipose tissues were embedded in paraffin. These samples were sectioned into 7-μm-thick slices that were stained with oil red O (Sigma-Aldrich) or hematoxylin and eosin for microscopic examination. Sections and cells were fixed in 4% paraformaldehyde and immunostained using primary antibodies raised against tyrosine hydroxylase (TH) (Millipore) to detect sympathetic neurons, nestin (BD Biosciences) to detect undifferentiated neural cells, GLP-1 (Abcam) to identify enteroendocrine cells, and insulin (Sigma-Aldrich) to identify differentiated pancreatic exocrine cells. This was followed by signal development using secondary antibodies conjugated with a fluorescent marker. Nuclei were counterstained with 4′,6-diamidino-2-phenylindole (DAPI) (Roche).

In situ hybridization

Mouse embryos (transverse and sagittal sections) and adult tissues were frozen in OCT compound, after which 18-μm sections were cut using a cryostat and stored at −80°C until hybridization. Mouse antisense *Gpr41* and *Gpr43* RNA probes were transcribed using T7 RNA polymerase and uridine 5′-α-[^{35}S]-thiotriphosphate (Perkin Elmer). Tissue sections were examined as described previously (*15*, *16*) and counterstained with hematoxylin and eosin and TH antibodies (Millipore) to visualize SCGs.

Biochemical analyses

Plasma samples were obtained from mice fasted for 5 hours. Plasma glucose levels were determined using OneTouch Ultra (LifeScan, Inc.). Determinations of plasma TGs, free fatty acids, and total cholesterol levels were performed using commercial kits (TG, LabAssay Triglyceride; free fatty acids, LabAssay NEFA; and total cholesterol, LabAssay Cholesterol; Wako Chemicals). Levels of plasma PYY [Mouse/Rat PYY enzyme-linked immunosorbent assay (ELISA) kit; Wako Chemicals], GLP-1 [GLP-1 (Active) ELISA kit; Shibayagi], and insulin [Insulin ELISA kit (RTU); Shibayagi] were determined using ELISA kits, following the respective manufacturer's instructions. Levels of total GLP-1 in colonic lysates were determined using a commercial kit (Multi Species GLP-1 Total ELISA; Millipore). For GLP-1 determinations, plasma and tissue samples were treated with dipeptidyl peptidase IV inhibitor (Merck Millipore) to prevent the degradation of active GLP-1.

Glucose- and insulin-tolerance tests

For the glucose-tolerance test, mice fasted for 24 hours were intraperitoneally (i.p.) administered 1.5 mg of glucose (Wako Chemicals) per gram of body weight. For the insulin-tolerance test, mice fasted for 3 hours were administered human insulin (3 mU/g, i.p.; Sigma-Aldrich). Plasma glucose concentration was monitored before injection and 15, 30, 60, 90, and 120 min after injection using OneTouch Ultra.

PET and CT imaging

Ten-week-old female Sprague-Dawley rats (Charles River) were housed under a 12-hour light-dark cycle and provided an AIN-93G diet during pregnancy. For PET imaging, rats on day 16.5 and 17.5 of pregnancy were used. Synthesis of [^{11}C]-propionate was performed by modifying [^{11}C]-acetate preparation as previously described (*46*). Briefly, [^{11}C]-CO_2 was bubbled into a mixture of 0.1 M ethyl magnesium bromide and tetrahydrofuran (Wako Chemicals) at −10°C, and then the mixture was purified using solid-phase extraction columns (OnGuard II Ag and OnGuard II H, DIONEX). The resultant solution was converted to sodium salt with $NaHCO_3$aq (Maylon). [^{11}C]-propionate was confirmed by high-performance liquid chromatography [Partisil-10-SAX (4.6 mm by 250 mm) and 20 mM phosphate buffer (pH 4) containing KH_2PO_3 and phosphoric acid (for adjusting pH) as an eluent at 0.7 ml/min and at 40°C] at 96.5% radiochemical purity and 5.77% radiochemical yield. [^{11}C]-Propionate was diluted with physiological saline to 20 megabecquerels (250 to 300 μl per rat) and administered to the colonic lumen via the anus using a feeding needle for rats under isoflurane anesthesia. PET and computed tomography (CT) scans were performed using ClairvivoPET (Shimadzu Co., Ltd.) and Aquilion TSX-01A (Tokyo Medical Systems) instruments, respectively. After PET and CT imaging, rats were laparotomized to collect blood from the inferior vena cava under isoflurane anesthesia. Subsequently, portions of the liver, kidney, spleen, left hind paw muscle, and fetus with placenta were dissected. Each tissue was washed twice with physiological saline. After sufficiently wiping off physiological saline, the radioactivity of each organ was measured using a γ-counter. The distribution of [^{11}C]-propionate in each organ and in embryos was measured using an AccuFLEX γ7001 instrument (ARC-7001; Hitachi Aloka Medical). Thereafter, the weight of each tissue was measured. The obtained radioactivity value was calculated using the following equation

$$\text{Value of attenuation correction} = \text{radioactivity value}/(2^{\{(T1\ -\ T2)/20.4\}})$$

where $T1$, $T2$, and 20.4 (in minutes) indicate the time when each tissue was measured, [^{11}C] propionic acid administration time, and ^{11}C half-life, respectively. After attenuation correction, each radioactivity value was divided by the corresponding weight to calculate the radioactivity value per fixed weight. The pharmacokinetics of the tracer was evaluated as counts per minute per gram of each tissue relative to that in the liver taken from the same pregnant rat.

SCFA determinations

Plasma SCFAs were determined as previously described (*47*). The SCFA-containing ether layers were collected and pooled for gas chromatography–mass spectrometry (GC-MS) analysis using a GCMS-QP2010 Ultra instrument (Shimadzu). The concentration of each

SCFA was determined using external standard calibration over an appropriate concentration range.

Commensal bacteria composition

Fecal, cecal, placental, colonic, and embryonic DNA were extracted from frozen samples using the FastDNA SPIN Kit for Feces (MP Biomedicals) or Gentra Puregene Mouse Tail Kit (Qiagen) according to the manufacturer's instructions. Bacteria (*Bacteroides vulgatus* JCM5826T) were provided by the Japan Collection of Microorganisms of RIKEN BRC and used as standards specifically for the DNA-based determination of bacterial counts. Bacterial DNA was isolated using the MonoFas Bacterial Genomic Kit IV (GLC Science) according to manufacturer's instructions. Quantitative polymerase chain reaction (PCR) analysis was performed using SYBR Premix Ex Taq II (TaKaRa Bio) and a StepOne Real-Time PCR System (Applied Biosystems). Standard curves for quantification consisted of 10-fold serial dilutions in the range of 10^8 to 10^0 copies of the target 16*S* ribosomal RNA (rRNA) gene. Universal bacterial primer sequences were as follows: 5′-CRAACAGGATTAGAACCCT-3′ (forward) and 5′-GGTAAGGTTCCTCGCGTAT-3′ (reverse) (*47*).

For bacterial culture, amniotic fluids and cecal contents from E16.5 pregnant ICR mice were cultured in MRS (BD Biosciences) and BL medium (Eiken Chemical) at 37°C for 24 hours under anaerobic conditions. The samples were then suspended at 0.1 mg/ml in sterile PBS and diluted in 10-fold series.

For 16*S* rDNA amplicon sequencing, sequencing of the 16*S* rRNA variable regions 3 and 4 was performed as described previously (*47*). The libraries were sequenced using the MiSeq system (Illumina) with 2 × 300 base pair protocols. Microbial diversity and composition analyses were performed using QIIME with the Greengenes reference database clustered at 97% identity.

Indirect calorimetry

Energy expenditure was calculated as the product of the calorific value of oxygen (3.815 + 1.232 × the respiratory exchange ratio) and oxygen consumption (VO_2) as previously described (*15*). Tyramine (100 mg/kg), a catecholamine releasing agent, was administrated at a volume per i.p. injection as described (*16*).

RNA isolation and quantitative reverse transcription PCR

Total RNA was extracted using the RNeasy Mini Kit (Qiagen) and ISOGEN (Nippon Gene). cDNA was reverse transcribed using isolated RNA samples as templates and Moloney murine leukemia virus reverse transcriptase (Invitrogen). Expression analyses were performed using SYBR Premix Ex Taq II and the StepOnePlus Real-Time PCR System. Real-time PCR cycling conditions were as follows: 95°C for 30 s, followed by 40 cycles of 95°C for 5 s, 58°C for 30 s, and 72°C for 1 min. In addition, dissociation was examined at 95°C for 15 s, followed by 1 cycle of 60°C for 1 min and 95°C for 15 s. The 18*S* rRNA gene was used as an internal control. Each sample was tested in duplicate for the average Ct value. Relative mRNA expression was calculated after normalization to the 18*S* rRNA reference gene using the $2^{-\Delta\Delta Ct}$ method. Primer sequences are listed in tables S4 and S5.

Sympathetic neural cell culture

SCGs were dissected from P1 mice, trypsinized in 0.05% trypsin in Hanks' balanced salt solution for 20 min at 37°C, and dissociated by trituration. Dissociated cultures were plated onto dishes coated with poly-L-lysine (20 μg/ml; Sigma-Aldrich) in DF medium (Gibco; Thermo Fisher Scientific, Waltham, MA) containing 1% penicillin-streptomycin solution (Gibco). Cells were cultured in conditioned medium containing nerve growth factor (10 ng/ml; Upstate Biotech) and 10% fetal bovine serum (FBS) as described elsewhere (*16*). Neuronal differentiation was quantified by counting TH-positive cells; cells with outgrowths longer than the cell body diameter were scored as positive for neurites. ImageJ (NIH, Bethesda, MD) was used to determine neurite outgrowth (*48*).

Western blotting and cAMP determination

Hearts were homogenized in 0.1 M sodium phosphate (pH 7.4) and centrifuged at 10,000*g* for 20 min at 4°C. Supernatants were analyzed by Western blotting as previously described (*15*). Proteins were detected by Western blotting using anti-TH (Millipore) and anti-α-actin antibodies (Sigma-Aldrich) for normalization. For cyclic adenosine monophosphate (cAMP) determinations, cells were lysed in 0.1 N HCl. After acetylation, cAMP levels were determined in duplicate using enzyme immunoassay kits (Cayman) as previously described (*16*).

Pancreatic cell line culture

AR42J cells (rat pancreatic exocrine cells) were purchased from the American Type Culture Collection (ATCC) and cultured in F-12K (Thermo Fisher Scientific) containing 20% FBS and maintained at 37°C under 5% CO_2. For cell differentiation, the cells were plated in 24-well plates (1×10^5 cells per well) and cultured with betacellulin (2 nM; Wako Chemicals) and activin A (1 nM; Wako Chemicals), in the presence or absence of propionate (1 mM; Sigma-Aldrich) or PA-1 (10 μM) for 72 hours. Then, total RNA was isolated using ISOGEN. For small interfering RNA (siRNA) knockdown experiments, AR42J cells were transfected with 30 nM siRNA (*Gpr43*; target sequence: GGCCAUUGCACCAUCGUCA; GE Healthcare) using Lipofectamine 2000 transfection reagent (Invitrogen) according to the manufacturers' instructions.

Organoid culture

Colonic crypts were isolated from 7-week-old mice or embryos on day 15.5 and cultured on Matrigel, as previously described (*49*). Culture medium contained advanced DMEM/F12 (Life Technologies) supplemented with gentamicin/amphotericin B solution (Life Technologies), 10 mM HEPES (Nacalai Tesque), 1% (v/v) N2 (Life Technologies), 1% (v/v) B27 (Life Technologies), 1 μM *N*-acetylcysteine (Sigma-Aldrich), 2 mM L-alanyl-L-glutamine (Nacalai Tesque), 500 nM A-83-01 (Wako Chemicals), 10 nM [Leu^{15}]-gastrin 1 (Sigma-Aldrich), 1 mM nicotinamide (Wako Chemicals), 10 μM SB202190 (Wako Chemicals), 2.5 μM CHIR99021 (StemRD), and 2.5 μM thiazovivin (StemRD) supplemented with the growth factors EGF (epidermal growth factor; 50 ng/ml; Peprotech), noggin (100 ng/ml; Sigma-Aldrich), and R-spondin (1 μg/ml; Sigma-Aldrich). The medium was incubated in the presence or absence of Wnt3a for 24 hours. Organoid cultures were then treated with propionate (1 or 10 mM) for 24 hours, and total RNA was isolated using the RNeasy Mini Kit.

Metabolite analysis

Liquid chromatography–tandem MS (LC-MS/MS)–based lipidomics analysis was performed as described elsewhere (*50*). Briefly, samples were subjected to solid-phase extraction on a Sep-Pak C18 cartridge (Waters) with deuterium-labeled internal standards (arachidonic acid-d_8, leukotriene B4-d_4, 15-hydroxyeicosatetraenoic acid-d_8, and prostaglandin E2-d_4). Lipidomics analyses were performed using a Waters UPLC system with a linear ion-trap quadrupole mass spectrometer (QTRAP 5500; AB SCIEX) equipped with an Acquity UPLC BEH C18 column (1.0 mm by 150 mm by 1.7 μm; Waters). MS/MS analyses were conducted in negative-ion mode, and lipophilic metabolites were identified and quantified by multiple-reaction monitoring. Hydrophilic metabolites from biological samples were extracted and derivatized as described previously (*51*). GC-MS analysis was performed using a GCMS-QP2010 Ultra instrument (Shimadzu) with a fused silica capillary column (CP-SIL 8 CB low bleed/MS; 0.25 mm by 30 m by 0.25 μm; Agilent Technologies) as described previously (*51*). Acquired data were exported in the CSV file format and analyzed using in-house analytical software (AI output).

Statistical analysis

All values are presented as means ± SEM. Statistical significance of differences between groups was determined using two-tailed unpaired Student's *t* test (two groups) or two-tailed one-way analysis of variance followed by Tukey-Kramer's post hoc test and Dunnett's post hoc test (three or more groups). *P* values

< 0.05 were considered statistically significant. False discovery rates (q value) of the metabolomics and 16*S* rDNA amplicon sequencing data were estimated with the Benjamini-Hochberg procedure. The Smirnov-Grubbs' test was used for evaluating outliers. The similarity of microbiomes was tested using PERMANOVA (permutational multivariate analysis of variance).

REFERENCES AND NOTES

1. P. J. Turnbaugh *et al*., An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006). doi: 10.1038/nature05414; pmid: 17183312
2. V. K. Ridaura *et al*., Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**, 1241214 (2013). doi: 10.1126/science.1241214; pmid: 24009397
3. M. T. Khan, M. Nieuwdorp, F. Bäckhed, Microbial modulation of insulin sensitivity. *Cell Metab.* **20**, 753–760 (2014). doi: 10.1016/j.cmet.2014.07.006; pmid: 25176147
4. L. M. Cox *et al*., Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell* **158**, 705–721 (2014). doi: 10.1016/j.cell.2014.05.052; pmid: 25126780
5. A. D. Kostic *et al*., The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe* **17**, 260–273 (2015). doi: 10.1016/j.chom.2015.01.001; pmid: 25662751
6. J. L. Sonnenburg, F. Bäckhed, Diet-microbiota interactions as moderators of human metabolism. *Nature* **535**, 56–64 (2016). doi: 10.1038/nature18846; pmid: 27383980
7. I. Cho *et al*., Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature* **488**, 621–626 (2012). doi: 10.1038/nature11400; pmid: 22914093
8. P. Kovatcheva-Datchary *et al*., Dietary fiber-induced improvement in glucose metabolism is associated with increased abundance of prevotella. *Cell Metab.* **22**, 971–982 (2015). doi: 10.1016/j.cmet.2015.10.001; pmid: 26552345
9. H. J. Flint, E. A. Bayer, M. T. Rincon, R. Lamed, B. A. White, Polysaccharide utilization by gut bacteria: Potential for new insights from genomic analysis. *Nat. Rev. Microbiol.* **6**, 121–131 (2008). doi: 10.1038/nrmicro1817; pmid: 18180751
10. S. Fukuda *et al*., Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469**, 543–547 (2011). doi: 10.1038/nature09646; pmid: 21270894
11. Y. Furusawa *et al*., Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**, 446–450 (2013). doi: 10.1038/nature12721; pmid: 24226770
12. G. Frost *et al*., The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. *Nat. Commun.* **5**, 3611 (2014). doi: 10.1038/ncomms4611; pmid: 24781306
13. W. J. Lee, K. Hase, Gut microbiota-generated metabolites in animal health and disease. *Nat. Chem. Biol.* **10**, 416–424 (2014). doi: 10.1038/nchembio.1535; pmid: 24838170
14. E. E. Canfora, J. W. Jocken, E. E. Blaak, Short-chain fatty acids in control of body weight and insulin sensitivity. *Nat. Rev. Endocrinol.* **11**, 577–591 (2015). doi: 10.1038/nrendo.2015.128; pmid: 26260141
15. I. Kimura *et al*., The gut microbiota suppresses insulin-mediated fat accumulation via the short-chain fatty acid receptor GPR43. *Nat. Commun.* **4**, 1829 (2013). doi: 10.1038/ncomms2852; pmid: 23652017
16. I. Kimura *et al*., Short-chain fatty acids and ketones directly regulate sympathetic nervous system via G protein-coupled receptor 41 (GPR41). *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8030–8035 (2011). doi: 10.1073/pnas.1016088108; pmid: 21518883
17. A. N. Thorburn *et al*., Evidence that asthma is a developmental origin disease influenced by maternal diet and bacterial metabolites. *Nat. Commun.* **6**, 7320 (2015). doi: 10.1038/ncomms8320; pmid: 26102221
18. P. D. Wadhwa, C. Buss, S. Entringer, J. M. Swanson, Developmental origins of health and disease: Brief history of the approach and current focus on epigenetic mechanisms. *Semin. Reprod. Med.* **27**, 358–368 (2009). doi: 10.1055/s-0029-1237424; pmid: 19711246
19. M. G. Dominguez-Bello *et al*., Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. *Nat. Med.* **22**, 250–253 (2016). doi: 10.1038/nm.4039; pmid: 26828196
20. S. Ussar *et al*., Interactions between gut microbiota, host genetics and diet modulate the predisposition to obesity and metabolic syndrome. *Cell Metab.* **22**, 516–530 (2015). doi: 10.1016/j.cmet.2015.07.007; pmid: 26299453
21. R. N. Carmody *et al*., Diet dominates host genotype in shaping the murine gut microbiota. *Cell Host Microbe* **17**, 72–84 (2015). doi: 10.1016/j.chom.2014.11.010; pmid: 25532804
22. S. Fujisaka *et al*., Diet, genetics, and the gut microbiome drive dynamic changes in plasma metabolites. *Cell Rep.* **22**, 3072–3086 (2018). doi: 10.1016/j.celrep.2018.02.060; pmid: 29539432
23. R. A. Koeth *et al*., Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* **19**, 576–585 (2013). doi: 10.1038/nm.3145; pmid: 23563705
24. E. Y. Hsiao *et al*., Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* **155**, 1451–1463 (2013). doi: 10.1016/j.cell.2013.11.024; pmid: 24315484
25. S. Yoshimoto *et al*., Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. *Nature* **499**, 97–101 (2013). doi: 10.1038/nature12347; pmid: 23803760
26. F. De Vadder *et al*., Microbiota-generated metabolites promote metabolic benefits via gut-brain neural circuits. *Cell* **156**, 84–96 (2014). doi: 10.1016/j.cell.2013.12.016; pmid: 24412651
27. J. C. McNelis *et al*., GPR43 potentiates β-cell function in obesity. *Diabetes* **64**, 3203–3217 (2015). doi: 10.2337/db14-1938; pmid: 26023106
28. G. Tolhurst *et al*., Short-chain fatty acids stimulate glucagon-like peptide-1 secretion via the G-protein-coupled receptor FFAR2. *Diabetes* **61**, 364–371 (2012). doi: 10.2337/db11-1019; pmid: 22190648
29. C. Tang *et al*., Loss of FFA2 and FFA3 increases insulin secretion and improves glucose tolerance in type 2 diabetes. *Nat. Med.* **21**, 173–177 (2015). doi: 10.1038/nm.3779; pmid: 25581519
30. E. Le Poul *et al*., Functional characterization of human receptors for short chain fatty acids and their role in polymorphonuclear cell activation. *J. Biol. Chem.* **278**, 25481–25489 (2003). doi: 10.1074/jbc.M301403200; pmid: 12711604
31. A. J. Brown *et al*., The Orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *J. Biol. Chem.* **278**, 11312–11319 (2003). doi: 10.1074/jbc.M211609200; pmid: 12496283
32. A. Jawerbaum, V. White, Animal models in diabetes and pregnancy. *Endocr. Rev.* **31**, 680–701 (2010). doi: 10.1210/er.2009-0038; pmid: 20534704
33. J. L. Pluznick *et al*., Olfactory receptor responding to gut microbiota-derived signals plays a role in renin secretion and blood pressure regulation. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4410–4415 (2013). doi: 10.1073/pnas.1215927110; pmid: 23401498
34. I. Kimura, A. Ichimura, R. Ohue-Kitano, M. Igarashi, Free fatty acid receptors in health and disease. *Physiol. Rev.* **100**, 171–210 (2020). doi: 10.1152/physrev.00041.2018; pmid: 31487233
35. N. Singh *et al*., Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. *Immunity* **40**, 128–139 (2014). doi: 10.1016/j.immuni.2013.12.007; pmid: 24412617
36. K. M. Maslowski *et al*., Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461**, 1282–1286 (2009). doi: 10.1038/nature08530; pmid: 19865172
37. K. Huber *et al*., Inhibitors of histone deacetylases: Correlation between isoform specificity and reactivation of HIV type 1 (HIV-1) from latently infected cells. *J. Biol. Chem.* **286**, 22211–22218 (2011). doi: 10.1074/jbc.M110.180224; pmid: 21531716
38. M. Waldecker, T. Kautenburger, H. Daumann, C. Busch, D. Schrenk, Inhibition of histone-deacetylase activity by short-chain fatty acids and some polyphenol metabolites formed in the colon. *J. Nutr. Biochem.* **19**, 587–593 (2008). doi: 10.1016/j.jnutbio.2007.08.002; pmid: 18061431
39. A. K. Taggart *et al*., (D)-beta-Hydroxybutyrate inhibits adipocyte lipolysis via the nicotinic acid receptor PUMA-G. *J. Biol. Chem.* **280**, 26649–26652 (2005). doi: 10.1074/jbc.C500213200; pmid: 15929991
40. J. A. Sanford *et al*., Inhibition of HDAC8 and HDAC9 by microbial short-chain fatty acids breaks immune tolerance of the epidermis to TLR ligands. *Sci. Immunol.* **1**, eaah4609 (2016). doi: 10.1126/sciimmunol.aah4609; pmid: 28783689
41. S. D. Parlee, O. A. MacDougald, Maternal nutrition and risk of obesity in offspring: The Trojan horse of developmental plasticity. *Biochim. Biophys. Acta* **1842**, 495–506 (2014). doi: 10.1016/j.bbadis.2013.07.007; pmid: 23871838
42. R. C. Whitaker, J. A. Wright, M. S. Pepe, K. D. Seidel, W. H. Dietz, Predicting obesity in young adulthood from childhood and parental obesity. *N. Engl. J. Med.* **337**, 869–873 (1997). doi: 10.1056/NEJM199709253371301; pmid: 9302300
43. N. J. Binkin, R. Yip, L. Fleshood, F. L. Trowbridge, Birth weight and childhood growth. *Pediatrics* **82**, 828–834 (1988). pmid: 3186371
44. T. Harder, E. Rodekamp, K. Schellong, J. W. Dudenhausen, A. Plagemann, Birth weight and subsequent risk of type 2 diabetes: A meta-analysis. *Am. J. Epidemiol.* **165**, 849–857 (2007). doi: 10.1093/aje/kwk071; pmid: 17215379
45. H. K. Nam, K. H. Lee, Small for gestational age and obesity: Epidemiology and general risks. *Ann. Pediatr. Endocrinol. Metab.* **23**, 9–13 (2018). doi: 10.6065/apem.2018.23.1.9; pmid: 29609444
46. F. Oberdorfer, A. Theobald, C. Prenant, A practical method for the preparation of ^{11}C-acetate. *Appl. Radiat. Isot.* **46**, 317–321 (1995). doi: 10.1016/0969-8043(95)00008-2
47. J. Miyamoto *et al*., Ketone body receptor GPR43 regulates lipid metabolism under ketogenic conditions. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 23813–23821 (2019). doi: 10.1073/pnas.1912573116; pmid: 31685604
48. M. Kasubuchi *et al*., Membrane progesterone receptor beta (mPRβ/Paqr8) promotes progesterone-dependent neurite outgrowth in PC12 neuronal cells via non-G protein-coupled receptor (GPCR) signaling. *Sci. Rep.* **7**, 5168 (2017). doi: 10.1038/s41598-017-05423-9; pmid: 28701790
49. T. Sato *et al*., Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011). doi: 10.1038/nature09637; pmid: 21113151
50. M. Arita, Mediator lipidomics in acute inflammation and resolution. *J. Biochem.* **152**, 313–319 (2012). doi: 10.1093/jb/mvs092; pmid: 22923733
51. K. Tanaka *et al*., Compensatory glutamine metabolism promotes glioblastoma resistance to mTOR inhibitor treatment. *J. Clin. Invest.* **125**, 1591–1602 (2015). doi: 10.1172/JCI78239; pmid: 25798620
52. I. Kimura *et al*., Maternal gut microbiota in pregnancy dictates offspring metabolic phenotype, v6, Dryad (2020); https://doi.org/10.5061/dryad.5hqbzkh29.

ACKNOWLEDGMENTS

We thank K. Igarashi for metabolome analysis; A. Nagata for SCFA analysis; M. Arita and S. Kasuga for in vitro assay; and T. Sasaki, M. Akehi, and H. Hirano for PET imaging. **Funding:** This work was supported by research grants from the JSPS KAKENHI (JP17H05344 to I.K., JP15H05897 and JP15H05898 to M.A., and JP18H04680 and JP17KT0055 to K.H.), AMED (JP18gm1010007 to I.K. and JP18gm1010004h0103 to K.H.), the Lotte Foundation (to I.K.), the Institute for Fermentation Osaka (to I.K.), the Takeda Science Foundation (to H.K.), the Asahi Grass Foundation (to K.H.), and the Yakult Science Foundation (to K.H.). **Author contributions:** I.K., J.M., R.O.-K., K.W., T.Y., M.O., R.A., Y.I., D.K., D.I., A.I., Y.T., S.T., S.K., M.W., M.I., F.N., H.K., M.S., K.I., and K.H. performed the experiments. I.K., J.M., R.O.-K., and K.H. wrote the manuscript. I.K., J.M., R.O.-K., J.I., G.T., H.O., M.A., H.I., and K.H. interpreted the data. I.K. and K.H. supervised the project. All authors read and approved the final manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The source data underlying Figs. 1 to 6, figs. S1 to S24, and metabolome analysis have been deposited into the Dryad repository (*52*). The raw data for 16S rDNA amplicon sequencing have been deposited at the DNA Data Bank of Japan (DDBJ) under the accession nos. DRA007699 (fig. S3, A and B), DRA007700 (fig. S5, A and B), DRA009267 (fig. S6, A and B), DRA007701 (fig. S13, A and B), DRA009265 (fig. S19, A and B), and DRA009266 (fig. S23, A and B). All other data generated or analyzed during this study are included in this article and its supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/eaaw8429/suppl/DC1
Figs. S1 to S24
Tables S1 to S5

View/request a protocol for this paper from *Bio-protocol*.

30 January 2019; resubmitted 25 November 2019
Accepted 13 January 2020
10.1126/science.aaw8429

RESEARCH ARTICLE

PLANT SCIENCE

Pectin homogalacturonan nanofilament expansion drives morphogenesis in plant epidermal cells

Kalina T. Haas[1,2]*†, Raymond Wightman[3], Elliot M. Meyerowitz[4], Alexis Peaucelle[5]*

The process by which plant cells expand and gain shape has presented a challenge for researchers. Current models propose that these processes are driven by turgor pressure acting on the cell wall. Using nanoimaging, we show that the cell wall contains pectin nanofilaments that possess an intrinsic expansion capacity. Additionally, we use growth models containing such structures to show that a complex plant cell shape can derive from chemically induced local and polarized expansion of the pectin nanofilaments without turgor-driven growth. Thus, the plant cell wall, outside of the cell itself, is an active participant in shaping plant cells. Extracellular matrix function may similarly guide cell shape in other kingdoms, including Animalia.

The plant cell wall, a component of the plant extracellular matrix, responds to extra- and intracellular cues to affect cell shape, size, and division (*1*–*3*). It is an intricate composite of polysaccharides, such as cellulose, hemicellulose, and pectins. The cell wall is considered biphasic: Crystalline cellulose microfibrils tethered by hemicellulose are submerged in a gel-like matrix of pectins and proteins. How its components associate into a coherent, strong, and dynamic material remains unknown. Cellulose microfibrils, the main load-bearing components that drive growth anisotropy, are deposited directly in the cell wall by plasma membrane–localized cellulose synthase complexes that are guided by cortical microtubules (*4*–*7*). Pectins constitute a structurally diverse family of polysaccharides with the defining feature of 1,4-linked α-D-galactosyluronic acid (GalpA). We focus on homogalacturonan (HG) polysaccharides that contain exclusively linear chains of GalpA. HG also exists as HG glycan domains in heteroglycans containing more complex and branched pectins (rhamnogalacturonan type I and II) and in glycoconjugates, such as the proteoglycan APAP1 (*8*, *9*). HGs are synthesized in methylated form and undergo a demethylesterification only after insertion into the cell wall (*2*). The demethylated HG and the resulting negatively charged acidic carboxyl groups it contains are present in rapidly expanding cells, and demethylation correlates with wall elasticity change (*2*, *3*). However, the relationship between HG methylesterifiction and growth remains elusive. In the primary cell wall, matrix polysaccharides may form a continuous cross-linked network with covalent bonds to structural glycoproteins. Nuclear magnetic resonance indicates numerous intermolecular links between wall polysaccharides—notably, direct association of cellulose and pectins—though their supramolecular complexes are yet to be determined (*10*).

In the currently accepted model, plant cell growth is driven by turgor pressure (*11*) causing strain in the wall and leading to irreversible deformation, or creep (*12*). On the molecular level, this is thought to be mediated by the reorganizing of the cellulose microfibrils (wall remodeling) (*11*). Anisotropic expansion is caused by ordered arrays of cellulose microfibrils that restrict growth parallel to their orientation (*13*); however, some evidence suggests that microfibril alignment is not sufficient for anisotropic expansion (*14*).

In this work, we study growth and morphogenesis of epidermal cells, the pavement cells, in the *Arabidopsis* cotyledon (Fig. 1A). These cells possess an undulatory pattern in their lateral (anticlinal) walls, called lobes, with mathematical order that can be perceptualized by data sonification (fig. S1 and audio S1 and S2). Lobes are initiated by ROP guanosine triphosphatases concentrating microtubules at the future convex position (*15*, *16*), associated with increased cell wall thickness, radial microfibril distribution, and HG demethylesterification (*16*, *17*). Turgor-induced buckling of anticlinal walls, caused by tension in periclinal walls and their local reinforcement linked to pectin demethylation, has been proposed to explain pavement cell shape (*18*).

The prevailing view of pectin tertiary structure in the wall is as an amorphous collection of polymers. However, there is a rich literature that proposes a crystalline structure for HG (*19*, *20*, and references therein). Prior in vitro x-ray diffraction studies have identified HG tertiary structures as helices with three galacturonic acid residues per helical turn. These helices arrange uniaxially in fibrous quaternary structures packed in hexagonal and rectangular lattices for methylated and demethylated HG, respectively (*19*, *20*). Yet little is known about HG higher-order structure in intact tissues.

Here, we present the in muro nanostructure of the homoglycan polymeric form of HG using super-resolution three-dimensional direct stochastic optical reconstruction microscopy (3D-dSTORM) (*21*–*25*) and cryo–scanning electron microscopy (cryo-SEM). We show that, in the cotyledon anticlinal walls, HG assembles into discrete nanofilaments rather than a continuously interlinked network. Although the fine structure of these nanofilaments is not discernable with our techniques, we propose that it may be a quaternary structure similar to that observed by x-ray diffraction. This hypothesis led us to formulate an intrinsic cell wall expansion "expanding beam" model of pavement cell morphogenesis. In this model, local HG demethylesterification leads to nanofilament radial swelling, which is caused by conversion between quaternary structures with different packaging. We further test this hypothesis by showing that demethylesterification of HG alone is sufficient to induce tissue expansion. Finally, we formalize the model as a 3D nonlinear finite element method (FEM) model that predicts tissue topology, local cell wall thickness, tension, and growth.

Quaternary structure of polymeric HGs as nanofilaments

The 3D-dSTORM nanoscopy provides insight into biological structures at the nanometer scale (*21*–*25*). Combining 3D-dSTORM and immunolabeling using antibodies against highly methylesterified (LM20) and low- or non-esterified (2F4) HG on 4-μm-thick tissue sections, we obtained ~40- to 50-nm lateral and ~80-nm axial resolution and depth reconstruction of ~800 nm (*26*, *27*) (figs. S2 and S3). Both antibodies bind in the cell wall near the plasma membrane but rarely in the middle lamella, which suggests limited epitope accessibility to an antibody or lack of such epitopes in this location (Fig. 1B). 3D-dSTORM revealed that, in the anticlinal walls, HG forms aligned filaments perpendicular to the cotyledon surface, hereafter referred to as HG nanofilaments (Fig. 1C, figs. S4 and S5A, and movie S1). Their estimated width is ~40 nm (Fig. 1D

[1]Medical Research Council Cancer Unit, University of Cambridge, Hutchison/MRC Research Centre, Hills Road, Cambridge CB2 0XZ, UK. [2]Laboratoire Matière et Systèmes Complexes, Université Paris Diderot and CNRS UMR7057, 10 rue Alice Domon et Léonie Duquet, 75013 Paris, France. [3]Microscopy Core Facility, Sainsbury Laboratory, University of Cambridge, 47 Bateman Street, Cambridge CB2 1LR, UK. [4]Howard Hughes Medical Institute and Division of Biology and Biological Engineering 156-29, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA. [5]Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, 78000 Versailles, France.
†Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, 78000 Versailles, France.
***Corresponding author. Email: alexis.peaucelle@inrae.fr (A.P.); kalina.haas@inrae.fr (K.T.H.)**

and fig. S5B), which is the limit of our reported lateral resolution. In the periclinal walls, in contrast, we did not detect a filamentous pattern, suggesting distinctive HG organization in different walls of the same cells (Fig. 1E, fig. S5C, and movies S2 to S4). Different photoswitching properties of fluorescent probes used in this study, Alexa647 and CF568, make the nanofilament organization more apparent when using Alexa647 (*28*) (fig. S3C). Swapping the fluorophore confirmed that HG in anticlinal walls organizes into parallel and often-interspaced nanofilaments, which are either predominantly methylated or demethylated (fig. S5D and movies S5 and S6). These observations allowed us to hypothesize that HG nanofilaments are the quaternary structures predicted by x-ray diffraction studies. However, 3D-dSTORM resolution was insufficient to discern any difference in LM20- and 2F4-tagged nanofilaments.

HG methylation affects cell wall filament diameter

To gain better insight into the HG nanofilament architecture, we used cryo-fracture SEM, which, through the use of hydrated and ultra-frozen samples, permits examination of the native tissue structure. The cryogenically prepared tissue is fractured, allowing the observation of otherwise-hidden anticlinal walls (Fig. 1F). Similarly to 3D-dSTORM, cryo-SEM revealed organized filaments in anticlinal but not periclinal walls (Fig. 1G and fig. S6A), with median (± median absolute deviation) full width at half-maximum (FWHM in nanometers) of 26 ± 8, 20 ± 4, and 21 ± 7 and a median filament interspacing of 58 ± 18, 42 ± 10, and 46 ± 12 in convex and concave sides of lobed walls and in straight walls, respectively (Fig. 1H and fig. S6, B and C). To test the potential effect of HG differential methylation on the cryo-SEM structures, we used plants with inducible overexpression of PECTIN METHYLESTERASE 5 (PME5oe) to promote HG demethylation and inducible overexpression of PECTIN METHYLESTERASE INHIBITOR 3 (PMEI3oe) to inhibit demethylation (*2*, *29*, *30*). We observed filaments with a median width of 26 ± 9 and 19 ± 7 nm and median interspacing of 49 ± 17 and 39 ± 15 nm, respectively, for PME5oe and PMEI3oe (Fig. 1, G and H). The cross sections of filaments in HG-demethylated walls (PME5oe) were ~1.4 times the length of those in HG-methylated walls (PMEI3oe), consistent with in vitro differential packaging of methylated and demethylated HG in distinct quaternary structures (*19*, *20*). Although the cryo-SEM lacks chemical information, the filament position, size, and reaction to PME activity suggest that they contain HG. These data and the close association of cellulose with pectin supported by nuclear magnetic resonance allow us to speculate that cellulose and HG may coexist in the same higher-order structures (*10*).

Demethylation-mediated nanofilament inflation underlies anisotropic growth

We propose that lobes in the anticlinal wall could emerge by spatially varying HG demethylation along and across the wall causing local radial swelling of the HG nanofilaments, thus causing local wall expansion. To formalize this hypothesis we refer to the in vitro quaternary structure of HGs (*19*), where uniaxially oriented helices of methylated HG pack in a hexagonal net with a side length of a = 0.837 nm, whereas demethylated and calcium-bound HG pack in a rectangular lattice with unit cell dimensions of b = 1.23 nm and c = 0.99 nm. We hypothesize that, in vivo, HG demethylation converts between these quaternary structures, leading to the expansion ratio of 1.42 in average lateral dimension and a shortening in the axial dimension of −0.013 nm per helical repeating unit (materials and methods). The ~1.42-fold expansion is compatible with our cryo-SEM measurements of HG nanofilaments' widths going predominantly from methylated (PMEI3oe) to demethylated (PME5oe) forms (Fig. 1H).

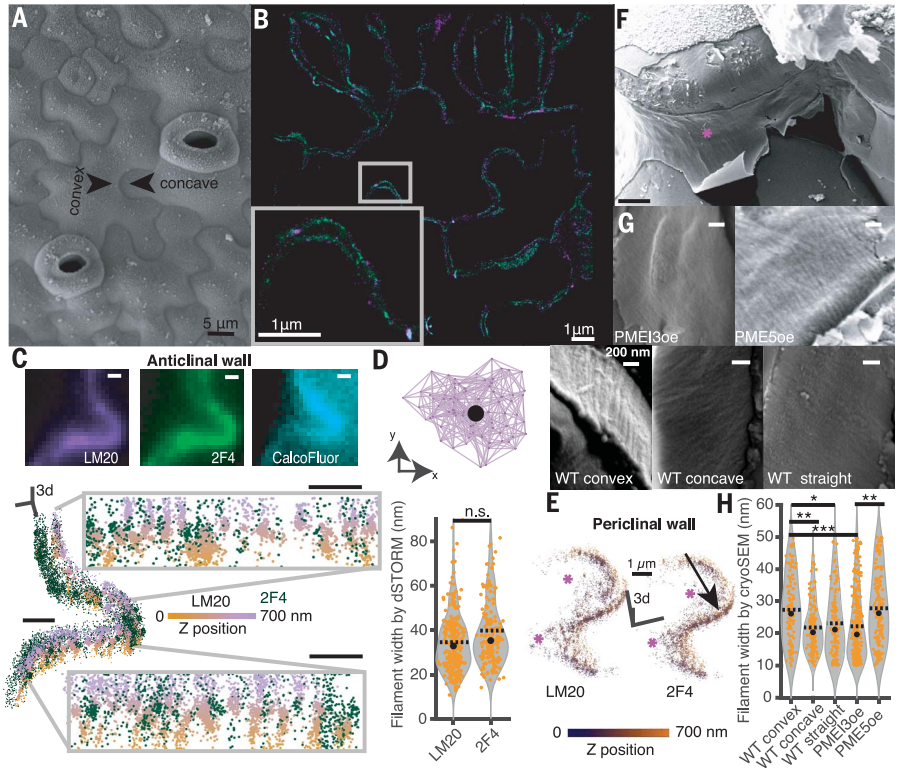


Fig. 1. The 3D-dSTORM nanoscopy and cryo-SEM reveal HG nanofilaments. (**A**) Surface view of cotyledon epidermal pavement cells using cryo-SEM. (**B**) High-magnification view of an epidermal region observed by dSTORM; methylated (violet) and demethylated (green) HG. (**C**) Diffraction-limited image of the lobed region of the anticlinal wall (top), and the same region imaged with 3D-dSTORM shown as a 3D scatterplot of the coordinates of localized emitters (bottom); demethylated (green) and methylated (orange-violet colormap encoding the Z position) HG nanofilaments. Gray insets represent two wall segments in the orthogonal view. Scale bars, 500 nm. (**D**) Lateral view of segmented nanofilament represented as a bidirectional graph (top); the filament width is estimated as the median pairwise distance between the centroid (black dot) and each point in the graph. The nanofilament width is also shown (bottom), representing, respectively, 159 and 96 methylated and demethylated filaments. n.s., not significant. (**E**) The 3D-dSTORM imaging of lobed wall segment showing the junction between the anticlinal wall (black arrow) and periclinal walls (violet stars). The orange-blue colormap encodes the Z position. (**F**) Representative gross-scale cryo-SEM picture with the fracture exposing convex anticlinal wall (violet star). Scale bar, 1 µm. (**G**) Fine-scale pictures of the convex anticlinal walls in PMEI3oe and PME5oe (top row) and WT convex, concave, and straight anticlinal walls (bottom row). Scale bars, 200 nm. (**H**) The filament width distribution observed by cryo-SEM, representing 109, 137, and 104 filaments measured in, respectively, WT convex, concave, and straight walls. For PME5oe and PMEI3oe, the convex and concave wall data were pooled together; 139 and 322 filaments were measured in PME5oe and PMEI3oe. *P < 0.05; **P < 0.001; ***P < 0.0001. P values were obtained with multiple group comparison Kruskal-Wallis test and Bonferroni correction. In (D) and (H), the black dot and dashed line represent median and mean, respectively.

FEM implementation of the model

To test the above hypothesis, we developed a 3D nonlinear FEM model that simulates lobe growth and morphogenesis by local swelling of oriented HG nanofilaments (*31*, *32*). The two-cell junction is composed of the bottom and top periclinal walls connected by a perpendicular anticlinal wall (fig. S7). At each growth iteration, exocytosis deposits an amount of methylated HG that is proportional to the area of the element's face directly in contact with the cytoplasm. The change of the element's edge relaxation length is proportional to the HG demethylation rate obtained from 3D-dSTORM data and follows the local HG anisotropic organization (Fig. 2 and figs. S8 and S9). HG exocytosis rate was the only non-experimental parameter and was initially set to 0.05. The primary outputs of the model are the shape prediction—the lobes in anticlinal walls and the topography of the periclinal walls—and the local cell wall thickness. Starting from a straight wall after ~200 iterations, we observe maturing lobes in the anticlinal wall, which correspond in shape to those seen 2 to 3 days after germination in plant tissue (Fig. 3, A to C).

One outcome of this model is variability in the thickness along the length of the anticlinal wall (Fig. 3, C and D) that does not exceed 250 nm. This is less than that measured in dSTORM images (~250 to 500 nm) and notably less than the >1-μm measurements based on transmission electron microscopy (TEM) data that form the basis of other models (*17*). These (dSTORM and TEM) experimental methods rely on sections and pretreatments that may relax and distort tissues. Cryo-SEM, with cryo-fracture, is expected to preserve the native constrained state of the walls, and imaging of the wall in wild-type (WT) and overexpressor lines demonstrates wall thicknesses (<250 nm) that are within the range of those in the model (Fig. 3E).

Next, to model the effect of HG methylation on lobe formation, we induced in silico PME and PMEI overexpression starting at iteration 25, which corresponds to 1-day post-germination induction in the cotyledon. Simulation of PMEI induction agreed with observations leading to growth impairment (Fig. 3, A to C, and fig. S10A). To achieve size concordance for PME induction between the model and experiment, we had to decrease the exocytosis rate by 60%, which suggests the existence of a regulatory loop between wall composition and exocytosis. As exocytosis of new wall material is expected to affect wall thickness, we compared outputs of the model (Fig. 3D) with cryo-fractured walls (Fig. 3E). Thickness variability along the anticlinal wall was more pronounced for WT (comparing thicknesses for model, Fig. 3D, with cryo-fractures lobed versus straight, Fig. 3E) than

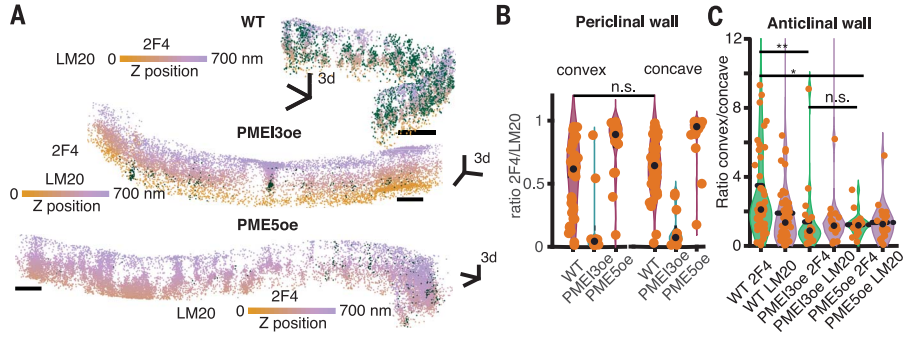


Fig. 2. HG methylation asymmetry affects lobe formation. (**A**) Representative lobed wall segments imaged with 3D-dSTORM in WT, PME5oe, and PMEI3oe cotyledons. The orange-violet colormap encodes the Z position of 2F4, LM20, and 2F4, and green marks LM20, 2F4, and LM20, from the top to bottom panels, respectively. Scale bars, 500 nm. (**B** and **C**) HG methylation asymmetry between convex and concave regions of a lobe in the periclinal walls (B) and in the lobed anticlinal walls (C) for WT, PME5oe, and PMEI3oe and straight walls for WT plants. **P* < 0.05; ***P* < 0.001; ****P* < 0.0001. *P* values were obtained with multiple group comparison Kruskal-Wallis test and Bonferroni correction. The number of analyzed regions was 100, 39, and 23 in (B) and 53, 22, and 25 in (C) for WT, PME5oe, and PMEI3oe cotyledons, respectively.

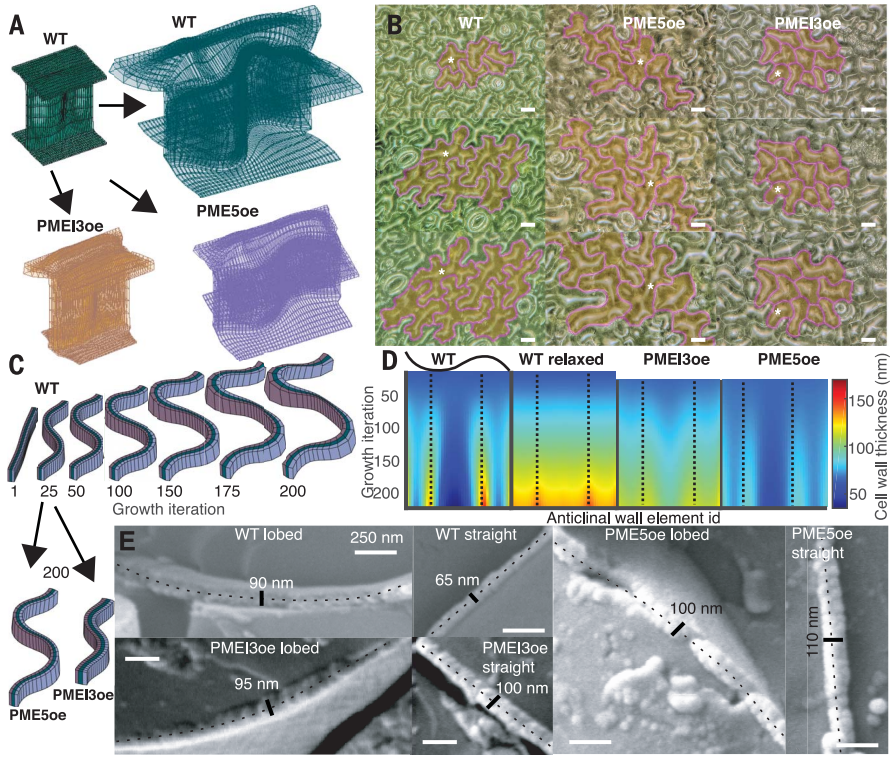


Fig. 3. The expanding beam model of pavement cell shape. (**A**) FEM implementation of the model in WT (green), PME5oe (violet), and PMEI3oe (orange) cotyledons. (**B**) Digital microscope images of a time course of WT, PME5oe, and PMEI3oe cotyledons at 1, 2, and 3 days after germination. Tracked cells are highlighted in yellow with magenta contours, and white stars indicate selected tracked cells. Scale bars, 10 μm. (**C**) Single element–high cross sections of the anticlinal wall at selected iteration steps showing the development of lobes and the cell wall thickness. Black arrows point from the 25th iteration (the start of PME5oe and PMEI3oe induction) to the 200th iteration. (**D**) A spatiotemporal heat map of cell wall thickness output from the computational model, obtained using a single element–high middle segment of the anticlinal wall. Each color represents the wall thickness at a particular local portion or segment of the wall used in the model, termed an anticlinal wall element (*x* axis), for any given growth iteration of the model (*y* axis). The wall begins with a uniform thickness (iteration 0 at the top of the WT and WT relaxed images) that increases in variability between elements with increasing numbers of iterations. Dashed lines represent the center of the lobes. The WT relaxed corresponds to the stress-free incompatible growth state before elastic transformation is applied (fig. S7B). (**E**) Cryo-SEM images of cryo-fractured lobed and straight walls in WT, PME5oe, and PMEI3oe cotyledons. In each image, the fractured anticlinal wall is labeled with a dotted line and the measured thickness. Scale bars, 250 nm.

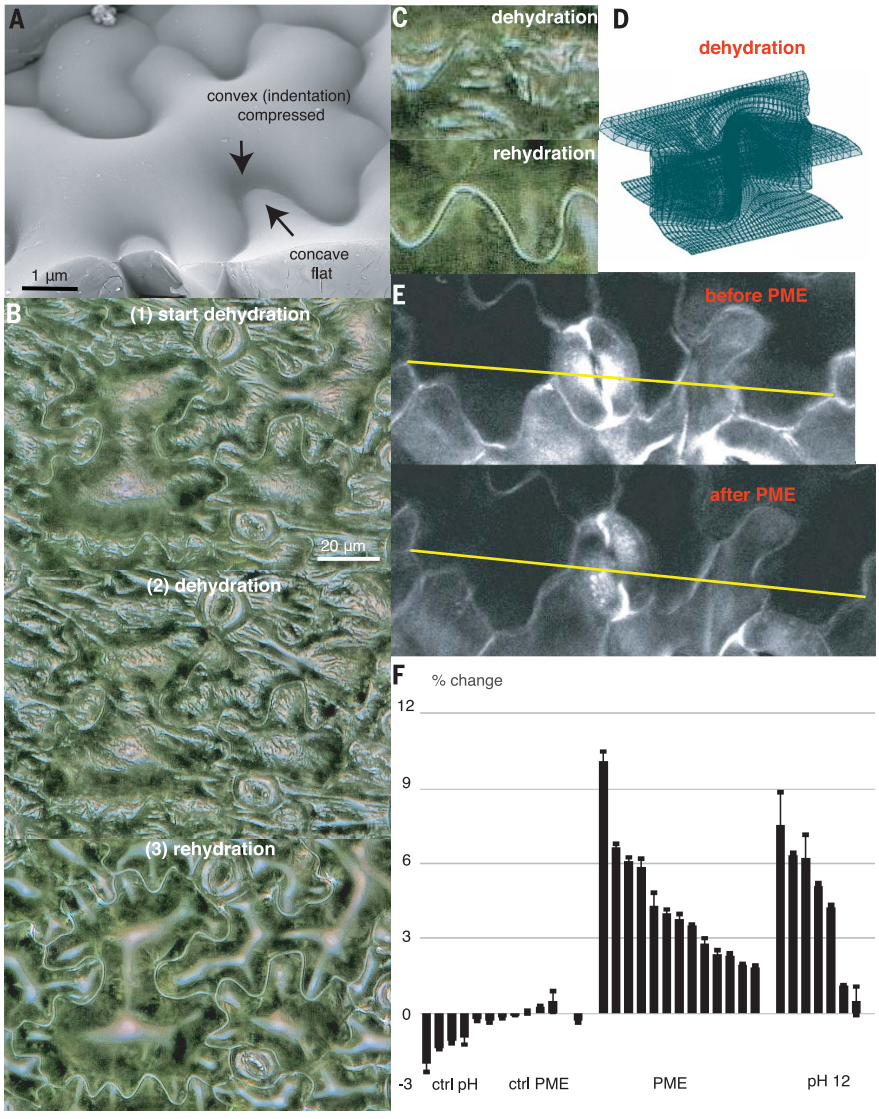


Fig. 4. Model tests by dehydration. (**A**) The topography of a pavement cell by cryo-SEM. (**B**) Digital microscope images of a WT cotyledon at the onset of (top), during (middle), and after (bottom) dehydration. (**C**) Enlargement of undulating walls from (B). (**D**) FEM model of a cell wall upon dehydration. (**E**) Confocal images of a WT cotyledon before (top) and after (bottom) PME enzyme treatment. (**F**) Tissue expansion quantification upon HG demethylation with PME enzyme or with a pH 12 buffer measured as the percent change in the distance between two reference points [yellow line in (E)]. Bars (mean ± SD) show different measurements from different cotyledon regions, ordered according to the amount of expansion. ctrl, control.

for the PME5oe and PMEI3oe lines, where anticlinal wall thickness was found to be more homogeneous between lobed and straight portions (Fig. 3, D and E).

Testing model predictions by reversible dehydration

In contrast to models that correlate tissue tension and growth, our model predicts a limited amount of local tension and compression in the anticlinal walls. If this is similar to the situation in vivo, then lobes should not change their shape in cotyledons when turgor pressure is reduced. On the other hand, the model does predict that the periclinal walls are under tension in the lobed regions and under compression in indentations. This agrees with the observed shape of periclinal walls over lobes and indentations and with the microtubules aligned along the direction of maximal tensional stress in the periclinal walls (*33*) (Fig. 4A). To test for the presence of wall tension, we followed cell shape changes after partial and reversible cotyledon dehydration (Fig. 4, B and C, and movie S7). Drying is characterized by wrinkling and inversion of the periclinal wall, which indicates that the turgor was reduced enough to return wall regions to their resting length. We observed no change in the outline of the anticlinal walls, despite the drying being inhomogeneous throughout the tissue. Therefore, alleviating turgor in anticlinal walls does not lead to observable strain, in contrast to the lobed periclinal walls, which flatten and invert upon dehydration, consistent with the computational model (Fig. 4D).

HG demethylation alone leads to tissue expansion

Finally, we tested whether demethylation of HG could lead to cell wall expansion in the absence of turgor pressure. Plasmolyzed cotyledons were maintained at a high calcium concentration, and HG demethylation was achieved either by adding PME enzyme or by incubating the tissue at high pH (fig. S11). Assuming that polymeric HG constitutes 10% of the wall and, upon demethylation, it expands by a factor of 1.42 in lateral dimension (Fig. 5A), for a tissue length of Δ there is an expected expansion of $(10\% \times \Delta \times 1.42) - (10\% \times \Delta) = \sim 4.2\%$ if all HG is demethylated. As presented in Fig. 4, E and F, and fig. S10B, both treatments led to tissue expansion within this range; however, this may be further influenced by contaminating cell wall remodeling enzymes.

Discussion

Here, we observe that, in the anticlinal walls of pavement cells, HG organizes into nanofilaments—either methylated or demethylated—that are oriented perpendicular to the cotyledon surface. We propose that these nanofilaments are the quaternary structures of HG previously observed in vitro by Walkinshaw and Arnott (*19*, *20*), which are composed of individual three-screw symmetry helical chains arranged on a hexagonal (methylated HG) or a rectangular (demethylated HG) lattice (Fig. 5A). On the basis of our observations and the prior x-ray characterization, we propose an expanding beam model for lobe formation. In this framework, the undulating form of anticlinal walls is initiated by demethylation-induced and spatially varying radial swelling and axial contraction of HG nanofilaments, which causes local and oriented wall expansion in the direction perpendicular to the dominant nanofilament orientation (Fig. 5, B and C). The HG methylation influences the width of filaments detected with cryo-SEM, supporting the hypothesis that nanofilament swelling is caused by the conversion between two quaternary forms of HG. Our model indicates that local methylation asymmetry alone may not be sufficient to explain lobe formation, suggesting that HG organization in nanofilaments is necessary for lobe morphogenesis (*30*).

It remains to be determined if cellulose and other more complex associations of heteroglycans and glycoconjugates reside in the

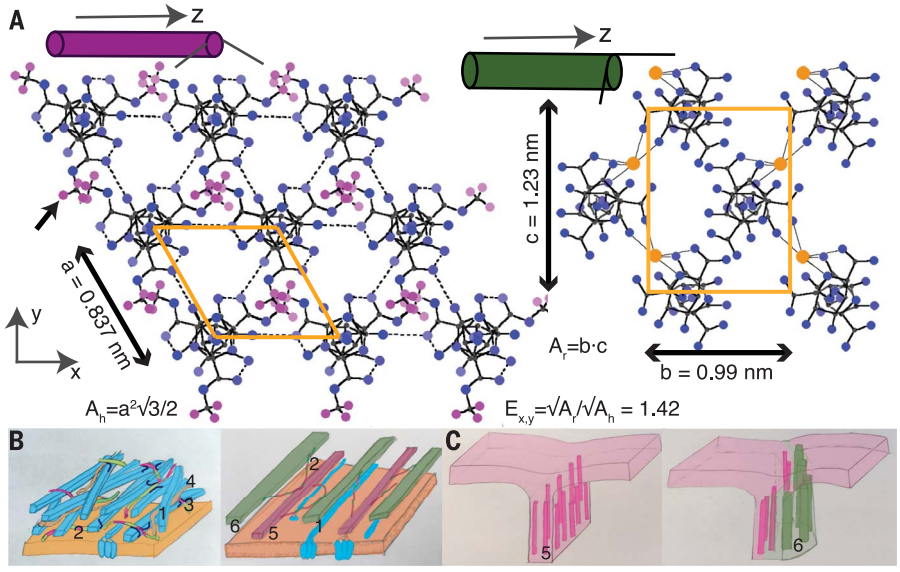


Fig. 5. Cell wall model including HG nanofilaments. (**A**) HG nanofilaments as quaternary structures composed of helical chains uniaxially arranged on a hexagonal or a rectangular net for (left) methylated and (right) demethylated HG with unit cell areas (orange polygons) A_h and A_r, respectively. $E_{x,y}$, expansion in a lateral dimension; blue circles, oxygen; small gray circles, carbon; magenta circles, methyl hydrogen (black arrow); large orange circles, calcium ions. Color scheme is based on the monochrome diagram of Walkinshaw and Arnott (*19*, *20*). (**B**) The cell wall models (left) without and (right) with HG nanofilaments. (1) Cellulose microfibrils (blue) embedded in pectin matrix containing (2) HGs methylated and demethylated (violet and green segments) and (3) and (4) rhamnogalacturonans type II and I (red). Updated cell wall model contains (5) methylated (violet) and (6) demethylated (green) HG nanofilaments. (**C**) The expanding beam model. (Left) Young wall composed of (5) methylated HG. (Right) Local HG nanofilament demethylation causes asymmetric cell wall expansion and lobe formation (green).

filaments. The HG nanofilaments appear to be absent in the periclinal walls, which suggests that HG forms an amorphous mesh, as depicted in current cell wall models (Fig. 5B) (*6*). The nanofilament orientation is compatible with the springlike behavior of the epidermis, because filaments would stiffen the tissue along their axis and permit reversible lateral extension (*34*).

In our model, wall thickening and cell expansion are inevitably linked, even though thickening is not always observed in rapidly expanding cells (*4*). Additionally, the model focuses on the most recently deposited material, assuming that previous layers adapt to expansion, and thus clearly demonstrates the importance of cell wall remodeling in permitting cell expansion.

Although the basis for the proposal of the expanding beam model for *Arabidopsis* pavement cells is an observation of HG nanofilaments, similar self-expansion models may well apply to other types of extracellular matrix polymers in which organized structures could be changed after synthesis (*35*). For example, the carrageenan of red algae (*36*) and the polarized, calcium-dependent alginates in brown algae (*37*) have the requisite attributes. Thus, biochemical processes similar to those that act upon pectins could underlie growth in organisms that have cell walls but do not have HG. Our observations could inspire the development of smart materials with the regulated capacity to expand. Localized swelling of the extracellular matrix in nonwalled organisms, such as animals, may even alter morphogenetic or metastatic processes.

REFERENCES AND NOTES

1. O. Hamant *et al.*, *Science* **322**, 1650–1655 (2008).
2. A. Peaucelle, R. Wightman, H. Höfte, *Curr. Biol.* **25**, 1746–1752 (2015).
3. M. Bosch, A. Y. Cheung, P. K. Hepler, *Plant Physiol.* **138**, 1334–1346 (2005).
4. F. Nicol *et al.*, *EMBO J.* **17**, 5563–5576 (1998).
5. A. Bichet, T. Desnos, S. Turner, O. Grandjean, H. Höfte, *Plant J.* **25**, 137–148 (2001).
6. T. Zhang, D. Vavylonis, D. M. Durachko, D. J. Cosgrove, *Nat. Plants* **3**, 17056 (2017).
7. A. R. Paredez, C. R. Somerville, D. W. Ehrhardt, *Science* **312**, 1491–1495 (2006).
8. L. Tan *et al.*, *Plant Cell* **25**, 270–287 (2013).
9. M. A. Atmodjo, Z. Hao, D. Mohnen, *Annu. Rev. Plant Biol.* **64**, 747–779 (2013).
10. T. Wang, M. Hong, *J. Exp. Bot.* **67**, 503–514 (2016).
11. A. Peaucelle, in *Concepts in Cell Biology - History and Evolution*, vol. 23 of *Plant Cell Monographs*, V. P. Sahi, F. Baluška, Eds. (Springer, 2018), pp. 139–154.
12. D. J. Cosgrove, *J. Exp. Bot.* **67**, 463–476 (2016).
13. T. I. Baskin, *Annu. Rev. Cell Dev. Biol.* **21**, 203–222 (2005).
14. A. M. D. Wiedemeier *et al.*, *Development* **129**, 4821–4830 (2002).
15. Y. Fu, T. Xu, L. Zhu, M. Wen, Z. Yang, *Curr. Biol.* **19**, 1827–1832 (2009).
16. A. Sapala, A. Runions, R. S. Smith, *Curr. Opin. Plant Biol.* **47**, 1–8 (2019).
17. M. Majda *et al.*, *Dev. Cell* **43**, 290–304.e4 (2017).
18. B. Altartouri *et al.*, *Plant Physiol.* **181**, 127–141 (2019).
19. M. D. Walkinshaw, S. Arnott, *J. Mol. Biol.* **153**, 1075–1085 (1981).
20. M. D. Walkinshaw, S. Arnott, *J. Mol. Biol.* **153**, 1055–1073 (1981).
21. B. Huang, W. Wang, M. Bates, X. Zhuang, *Science* **319**, 810–813 (2008).
22. E. Betzig *et al.*, *Science* **313**, 1642–1645 (2006).
23. M. J. Rust, M. Bates, X. Zhuang, *Nat. Methods* **3**, 793–795 (2006).
24. M. Bates, B. Huang, G. T. Dempsey, X. Zhuang, *Science* **317**, 1749–1753 (2007).
25. M. Heilemann *et al.*, *Angew. Chem. Int. Ed.* **47**, 6172–6176 (2008).
26. Y. Verhertbruggen, S. E. Marcus, A. Haeger, J. J. Ordaz-Ortiz, J. P. Knox, *Carbohydr. Res.* **344**, 1858–1862 (2009).
27. F. Liners, J.-J. Letesson, C. Didembourg, P. Van Cutsem, *Plant Physiol.* **91**, 1419–1424 (1989).
28. G. T. Dempsey, J. C. Vaughan, K. H. Chen, M. Bates, X. Zhuang, *Nat. Methods* **8**, 1027–1036 (2011).
29. A. Peaucelle *et al.*, *Curr. Biol.* **21**, 1720–1726 (2011).
30. Y. Rui *et al.*, *Plant Physiol.* **180**, 66–77 (2019).
31. N.-H. Kim, *Introduction to Nonlinear Finite Element Analysis* (Springer, 2015).
32. E. K. Rodriguez, A. Hoger, A. D. McCulloch, *J. Biomech.* **27**, 455–467 (1994).
33. A. Sampathkumar *et al.*, *eLife* **3**, e01967 (2014).
34. P. Sotiriou, E. Giannoutsou, E. Panteris, B. Galatis, P. Apostolakos, *Plant Biol.* **20**, 223–237 (2018).
35. Z. A. Popper *et al.*, *Annu. Rev. Plant Biol.* **62**, 567–590 (2011).
36. E. Ficko-Blean, C. Hervé, G. Michel, *Perspect. Phycol.* **2**, 51–64 (2015).
37. M. Terauchi, C. Nagasato, A. Inoue, T. Ito, T. Motomura, *Planta* **244**, 361–377 (2016).
38. K. T. Haas, A. Peaucelle, Pectin homogalacturonan nanofilament expansion drives morphogenesis in plant epidermal cells, version 1, Mendeley Data (2020); http://doi.org/10.17632/xb3f5jrbwv.1.
39. K. T. Haas, Grafeo V2, version V2.1.0, Zenodo (2020); http://doi.org/10.5281/zenodo.3663608.
40. K. T. Haas, SEMfilaments V1, version V.1, Zenodo (2019); http://doi.org/10.5281/zenodo.3581097.
41. K. T. Haas, A. Peaucelle, Shape Spectra Analysis, version V1.1.0, Zenodo (2020); http://doi.org/10.5281/zenodo.3665413.
42. K. T. Haas, The expanding beam model V1.1.0, version V1.1.0, Zenodo (2020); http://doi.org/10.5281/zenodo.3664101.

ACKNOWLEDGMENTS

The dSTORM was performed at the Medical Research Council Laboratory of Molecular Biology, Cambridge, and we thank N. Barry and J. Howe for their support. We thank A. Gurzadyan for the critical evaluation of our FEM model and H. Höfte for help with the fundraising and discussion of the results. **Funding:** A.P. has received support from the French National Research Agency (ANR) GoodVibration ANR-17-CE13-0007 and the European Union in the framework of the Marie-Curie FP7 COFUND People Program, through the award of an AgreenSkills+ fellowship (under grant agreement 201310). The Institut Jean-Pierre Bourgin (IJPB) benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007). The Microscopy Facility at the Sainsbury Laboratory is supported by the Gatsby Charitable Foundation, and work at the Sainsbury Laboratory was supported by the Gatsby Charitable Foundation (through fellowship GAT3395/DAA). The Meyerowitz Laboratory has support from the Howard Hughes Medical Institute. This work has benefited from the support of IJPB's Plant Observatory technological platforms. **Author contributions:** K.T.H.: software and visualization; K.T.H. and A.P.: investigation, methodology, and formal analysis; R.W. carried out the cryo-SEM; A.P., E.M.M., K.T.H., and R.W.: writing, original draft, and writing, review and editing; A.P. with the help of other authors: conceptualization. **Competing interests:** E.M.M. is a member of the Management Board of the Sainsbury Laboratory at Cambridge University. The authors declare no other competing interests. **Data and materials availability:** All data, codes, and materials used in this study are available. Please contact A.P. for the material request. All data are available in the manuscript, the supplementary materials, or at Mendeley (*38*). All codes used in this study are available on Zenodo (*39*–*42*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/1003/suppl/DC1
Materials and Methods
Figs. S1 to S11
References (*43*, *44*)
MDAR Reproducibility Checklist
Movies S1 to S7
Audio S1 and S2

View/request a protocol for this paper from *Bio-protocol*.

15 September 2019; accepted 24 January 2020
10.1126/science.aaz5103

IMMUNOLOGY

Structure of the secretory immunoglobulin A core

Nikit Kumar, Christopher P. Arthur, Claudio Ciferri*, Marissa L. Matsumoto*

Secretory immunoglobulin A (sIgA) represents the immune system's first line of defense against mucosal pathogens. IgAs are transported across the epithelium, as dimers and higher-order polymers, by the polymeric immunoglobulin receptor (pIgR). Upon reaching the luminal side, sIgAs mediate host protection and pathogen neutralization. In recent years, an increasing amount of attention has been given to IgA as a novel therapeutic antibody. However, despite extensive studies, sIgA structures have remained elusive. Here, we determine the atomic resolution structures of dimeric, tetrameric, and pentameric IgA-Fc linked by the joining chain (JC) and in complex with the secretory component of the pIgR. We suggest a mechanism in which the JC templates IgA oligomerization and imparts asymmetry for pIgR binding and transcytosis. This framework will inform the design of future IgA-based therapeutics.

Secreted polymeric immunoglobulins A and M (pIgA and pIgM, respectively) play vital roles in protecting the ~400 m^2 of human mucosa from invasion by pathogens. IgA and IgM contain 18-residue tailpiece extensions on their heavy chains that bestow polymer-forming capabilities (*1*). IgM can independently oligomerize to form hexamers, whereas IgA polymerization requires the 137-residue joining chain (JC), a protein with no known structural homologs (*2*). The predominant IgA oligomer is a dimer, though a smaller fraction of higher-order polymers up to pentamers have been described (*1*). At the core of pIgA is a JC-clasped, tail-to-tail dimer stabilized by two disulfides between the JC and one tailpiece of each monomer (*3*). Additional monomers are linked to this precursor via tailpiece-to-tailpiece–mediated disulfides (*3*, *4*). Although less abundant than the dimer, higher-order polymers display better neutralizing capabilities, specifically for low-affinity antigens (*5*).

The pIgA that is synthesized by mucosal plasma cells undergoes transcytosis, crossing the epithelium to reach the external secretions and perform its protective function. To initiate this process, pIgA binds the ectodomain of basolaterally expressed pIg receptor (pIgR), an interaction that requires the JC and is stabilized through a disulfide bond between the receptor and one Fc (*6*). Once bound, the receptor and pIgA undergo transcytosis through the epithelial cell to the mucosa. Upon transcytosis, proteolytic cleavage by unidentified protease(s) releases the secretory component (SC) of the receptor, which remains covalently attached to pIgA to form secretory IgA (sIgA) (*7*, *8*). In this mature form, sIgA performs its antimicrobial, neutralization, and protective functions (*8*).

Department of Structural Biology, Genentech, Inc., South San Francisco, CA, USA 94080.
*Corresponding author. Email: ciferri.claudio@gene.com (C.C.); matsumoto.marissa@gene.com (M.L.M.)

In recent years, extensive efforts have been made to develop vaccines that induce immunity via the mucosal route by eliciting sIgA responses (*1*). In the field of biotechnology, a major focus has been to engineer antibodies to target disease-relevant tissues, and therapeutic IgAs may allow delivery to mucosal tissues inaccessible to traditional IgG-based therapeutics (*1*). Thus, a detailed structural characterization of sIgA is critical for the development of future therapeutics and vaccines.

In this study, we use cryo–electron microscopy (cryo-EM) to determine the architecture of sIgA-Fc dimers, tetramers, and pentamers at atomic resolution. We propose a mechanism of JC-templated IgA oligomerization and describe a model for pIgR recognition. These findings will inform the design of unexplored IgA-based therapeutics with the potential for tissue-specific targeting and mucosal delivery.

Reconstitution of sIgA complexes

Previous cryo-EM studies of pIgA indicated that these molecules are refractory to high-resolution structure determination due to the inherent flexibility of the hinge regions as well as the tailpieces interconnecting monomers (*9*). We therefore reasoned that truncation

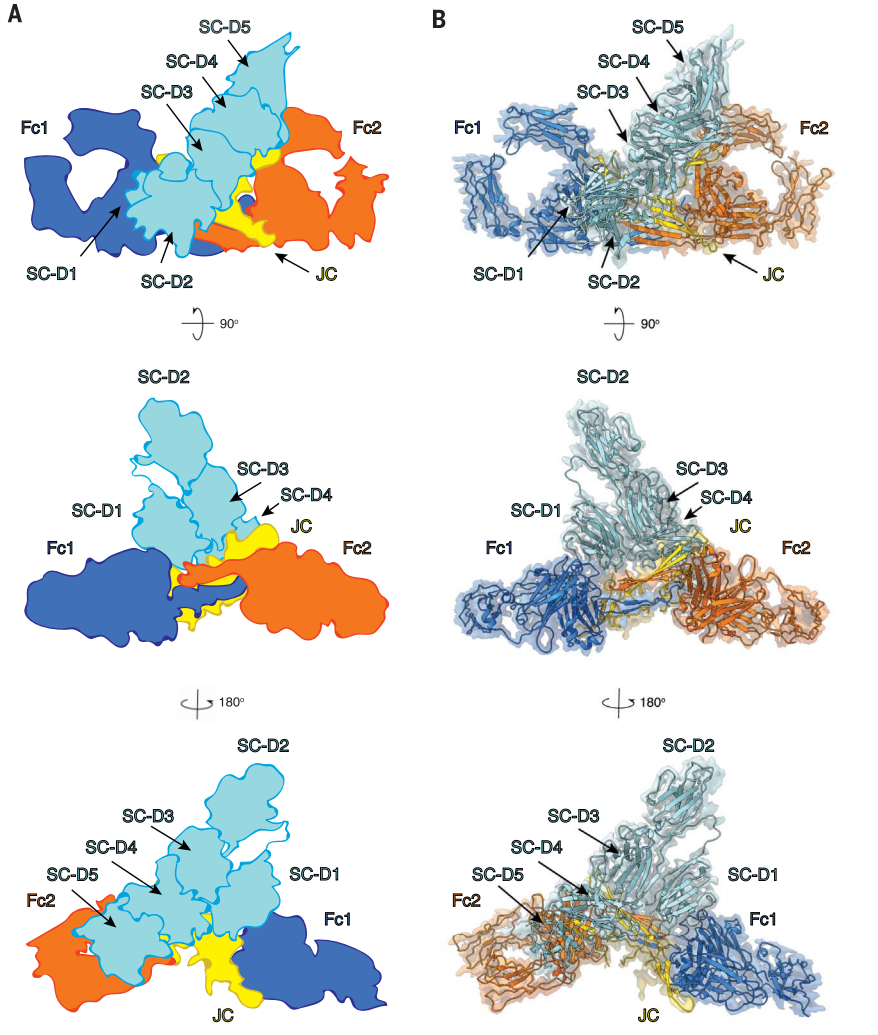


Fig. 1. Cryo-EM structure of dimeric sIgA1. (**A**) Top, back, and front view schematics of subunit arrangements in the dimer. (**B**) Cryo-EM reconstruction of dimeric sIgA1. Transparent maps overlaid with the model are shown.

of IgA to the Fc core and the addition of the SC would result in the formation of a more rigid molecule amenable to high-resolution structure determination. In humans, there are two IgA isotypes, IgA1 and IgA2. Both isotypes can form dimers and higher-order oligomers, though there is a propensity for recombinant IgA1 to form dimers and for IgA2m2 to form higher-order oligomers (*9*). Therefore, we co-expressed human IgA1 or IgA2m2 Fc with JC to form dimers or tetramers and pentamers, respectively (figs. S1 and S2), and subsequently assembled them with the SC. Single-particle cryo-EM was used to determine the structures of the resulting dimeric, tetrameric (two classes), and pentameric sIgAs at 2.9-, 3.0-, 2.9-, and 3.0-Å resolution [Fourier shell correlation (FSC) = 0.143], respectively (figs. S1 to S3 and table S1). Two classes were obtained for the tetramer; the Fcs were slightly bent toward each other in class 1 and were more planar in class 2 (fig. S2). The difference was subtle and did not affect the pIgA/SC interaction interface. The local resolution was highest at the central core, which contains the key interaction interfaces (fig. S4).

Overall architecture of sIgA

sIgA1 adopted a tail-to-tail planar dimer with the two Fcs positioned at an ~110° angle and held tightly by the JC, which functioned as a clasp (Fig. 1). The SC made extensive contacts with both Fcs and the JC, binding diagonally across the ~50° gap between the two monomers. The structure of the core sIgA1 dimer was relatively unchanged in the sIgA2m2 tetramer and pentamer, consistent with the 94% sequence identity between the Fcs of the two isotypes (fig. S5). Therefore higher-order polymers are assembled through incorporation of two or three additional Fcs in-plane to the original dimer (Fig. 2). The pentamer adopted an asymmetric pentagon containing an ~50° gap between Fc1 and Fc2 (Fig. 2), reminiscent of the JC-containing IgM pentamer (*10*). The SC bound at the gap, cross-bridging two Fcs in a manner similar to that of the IgM-binding apoptosis inhibitor of macrophage protein (*10*), suggesting a common binding mode.

Structure of the JC

We used the 2.9-Å map of the sIgA1 dimer (fig. S6) to de novo build a model containing all but six residues of the 137–amino acid JC (residues 5 to 94 and 97 to 137), as well as the tailpieces of each Fc. The JC was almost entirely composed of β-sheets and loops (Fig. 3, A and B), consistent with secondary structure predictions and circular dichroism studies (*11*). Previous biochemical studies identified intramolecular disulfides between JC_{C13} and JC_{C101}, JC_{C72} and JC_{C92}, and JC_{C109} and JC_{C134}, and intermolecular disulfides between JC_{C15} and $Fc2_{C471}$ and JC_{C69} and $Fc1_{C471}$ (*3*). Our structure confirmed these disulfides (Fig. 3C) and additionally revealed a much larger network of noncovalent interactions. The JC intricately joined the two Fcs of the dimer core though its three β-hairpins and four-stranded, twisted β-sandwich. It lay at an angle bridging the top of Fc2 (defined as the SC-binding face) through β-hairpins 1 and 2, to the bottom of Fc1 through β-hairpin 3, creating an asymmetric dimer (Fig. 3A). The JC β-hairpins made predominantly hydrophobic interactions with the two Fcs (Fig. 3D). β-hairpins 2 and 3 interacted with the same region of Fc2 and Fc1, respectively, with their β-strands packing against Fc_{M433} and Fc_{F443} and their loops contacting Fc_{L258} (Fig. 3D). This same Fc-interaction site is also exploited by the human Fcα receptor I (FcαRI) and the *Staphylococcus aureus* SSL7 toxin (*12*, *13*), suggesting a hot spot for Fcα recognition (fig. S7). The dimer was further stabilized through interaction of the Fc tailpieces with the JC, where Fc1 and

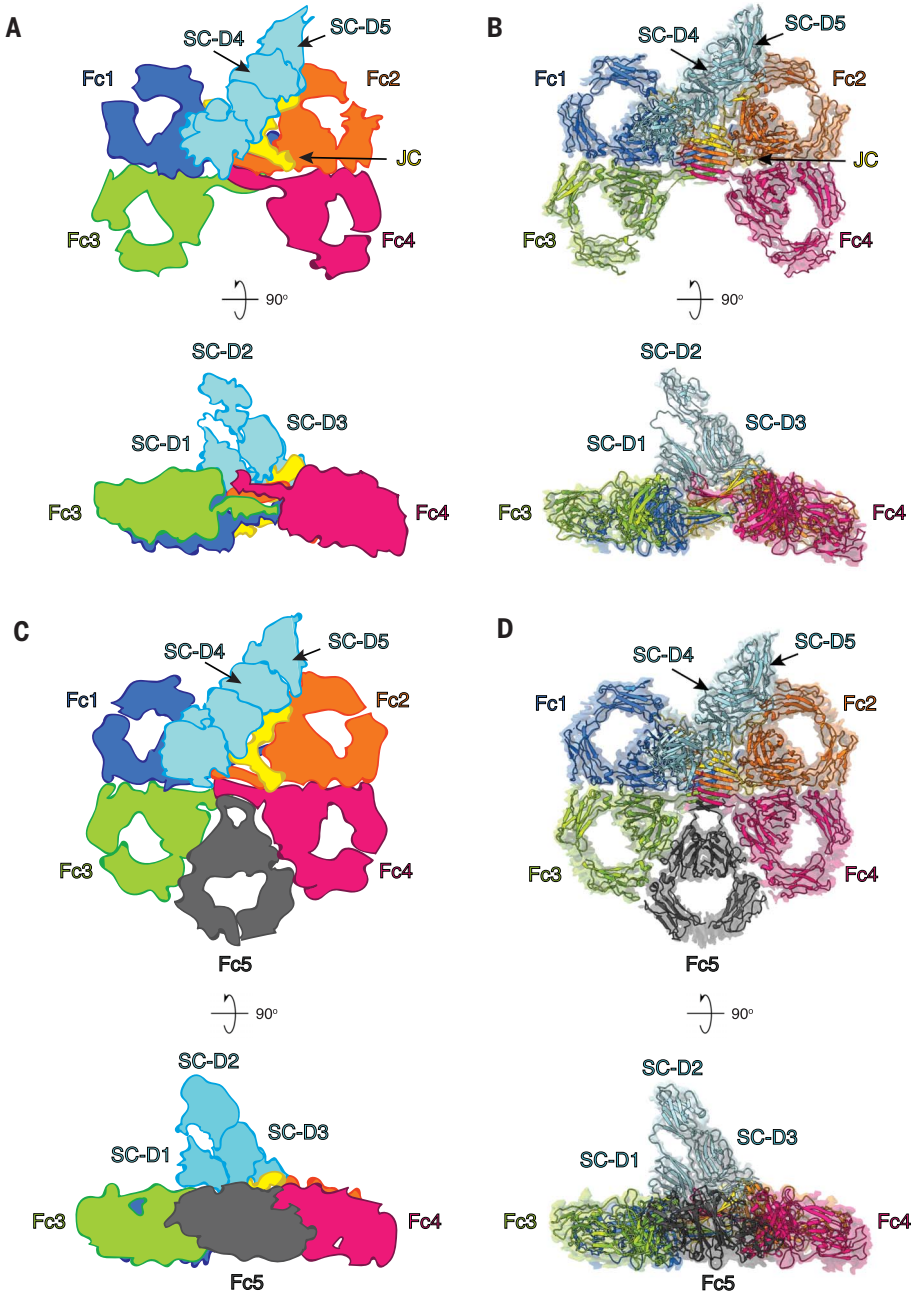


Fig. 2. Cryo-EM structures of tetrameric and pentameric sIgA2m2. (**A**) Top and back view schematics of subunit arrangements in the tetramer. (**B**) Cryo-EM reconstruction of tetrameric sIgA2m2. Transparent maps overlaid with the model are shown. (**C**) Same as in (A) but for pentameric sIgA2m2. (**D**) Same as in (B) but for pentameric sIgA2m2.

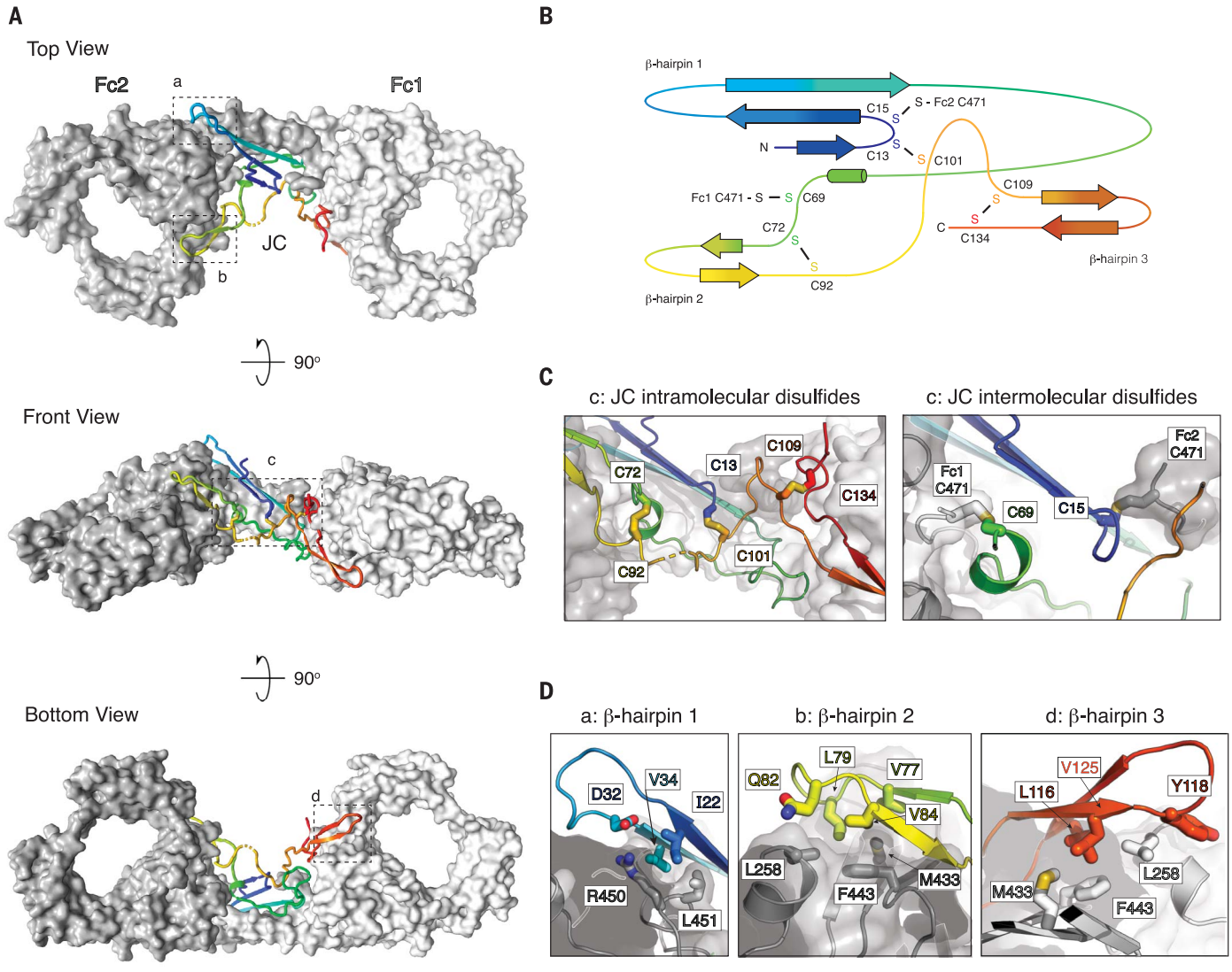


Fig. 3. Structure of the JC. (**A**) Top, front, and bottom views of dimeric IgA1 with the JC colored rainbow from the N terminus (blue; model starts at residue 5) to the C terminus (red). The SC is omitted for clarity. (**B**) Topology diagram of the JC with both intramolecular and intermolecular disulfide bonds shown. (**C**) Magnified views of box c in (A) showing JC intramolecular (left) and intermolecular (right) disulfide bonds. (**D**) Magnified views of β-hairpins 1, 2, and 3 from boxes a, b, and d, respectively, in (A) and the interactions they make with the Fcs.

Fc2 contributed two parallel β-strands each, extending the bottom and top sheets, respectively, of the twisted JC β-sandwich (Fig. 4A). Despite its small size, the JC created an extensive, ~3600-Å^2 interface with the Fcs in the dimer, underscoring its apropos title of "joining" chain.

Mechanism of IgA oligomerization

The JC-stabilized dimer core acted as a building block for larger polymers, consistent with previous studies showing dissociation of tetrameric IgA into a JC-containing dimer and two monomers upon mild reduction (*4*). In the tetramer, an additional set of two parallel β-strands, originating from the Fc3 and Fc4 tailpieces, continued the extension of the twisted β-sandwich, with Fc3 adding to the bottom sheet and Fc4 to the top (Fig. 4B). However, in the pentamer, the β-sheet extension continued, with Fc5 adding one tailpiece as a parallel strand to the top sheet and the other tailpiece as a parallel strand to the bottom sheet to cap the β-sandwich (Fig. 4C). Fc–Fc contacts occurred at the Cα2–Cα3 interface in both the tetramer and pentamer using residues similarly involved in JC β-hairpin and FcαRI binding (Fig. 4D and fig. S7). The arrangement of Fc tailpieces into the JC β-sandwich resulted in a striking repetitive pattern of equivalent residues packed at the core of pIgA, running parallel within each β-sheet and antiparallel within the β-sandwich (Fig. 4, E to G). In this arrangement, polar residues were solvent exposed, whereas hydrophobic residues were buried at the β-sandwich interface (Fig. 4, E to G). Thus, we propose a model for IgA oligomerization in which the JC acts as a template for the incorporation of IgA monomers via their tailpieces. This creates a molecular zipper of hydrophobic side chains, which stabilizes the polymer (Fig. 4G).

Structure of the SC and interaction with pIgA

The SC comprises five Ig-like domains, of which D1 is both necessary and sufficient for pIgA binding, whereas D2 to D5 provide affinity enhancement (*14*, *15*). Previous structural analysis of the SC showed that in the absence of pIgA, the SC adopts a closed conformation stabilized by the interaction among D1, D4, and D5 (fig. S8A) (*15*). Our structures revealed that the SC underwent a large conformational change and interacted through domains D1 and D5 with both Fcs of the IgA dimer core and the JC (Fig. 5A and fig. S8A). The asymmetry imparted by the JC allowed one-to-one binding of the asymmetric SC to the

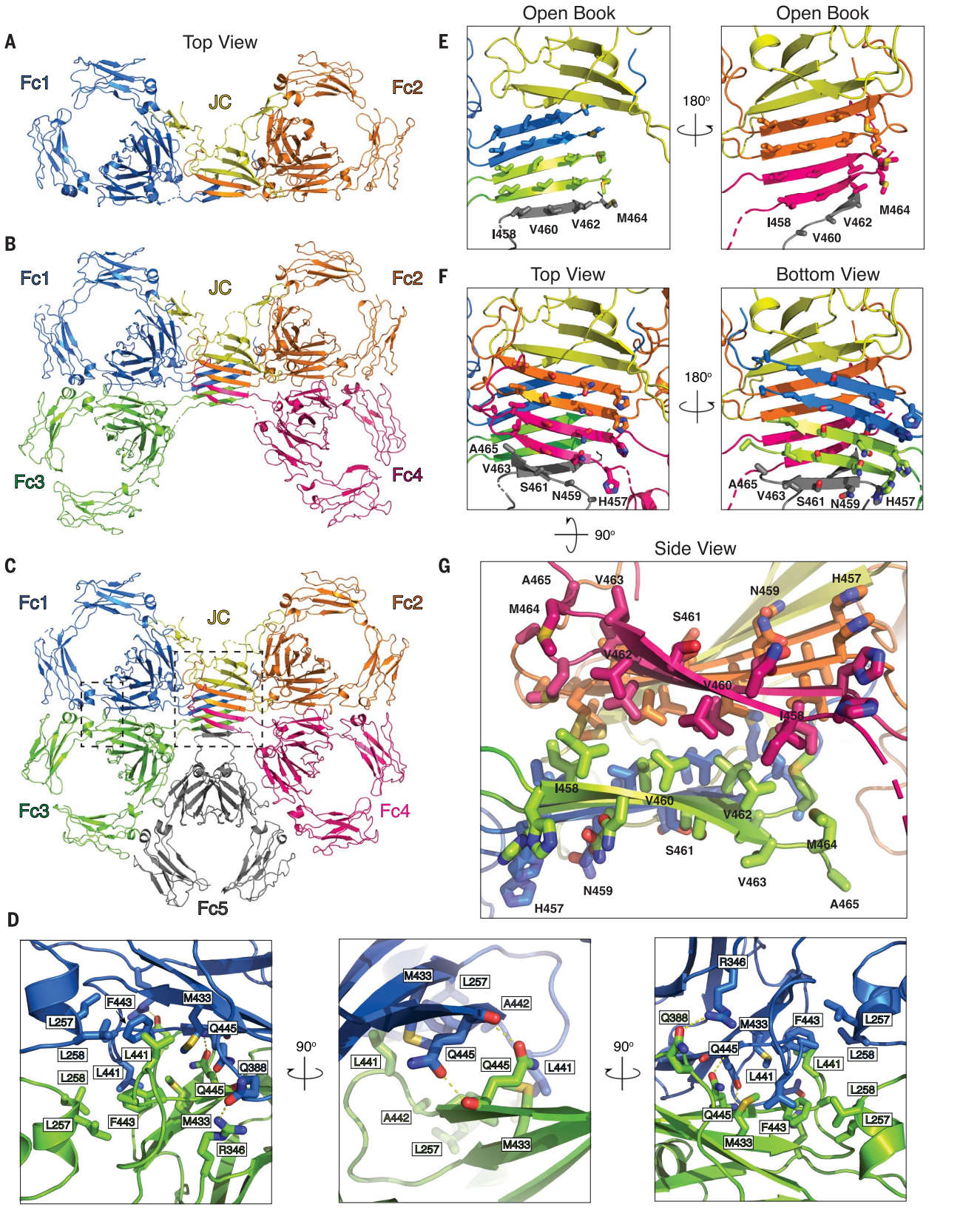


Fig. 4. Mechanism of IgA oligomerization. Top view of (**A**) dimeric IgA1, (**B**) tetrameric IgA2m2, and (**C**) pentameric IgA2m2. The SC is omitted for clarity. (**D**) Magnification of the left boxed region in (C) highlighting Fc–Fc contacts at the Cα2–Cα3 interface. (**E**) Magnification of the central boxed region in (C) highlighting the JC β-sandwich extension shown as an open book preparation of the Fc tailpieces with repeating internal residues labeled and side chains shown as sticks. (**F**) Magnification of the central boxed region in (C) showing repeating external residues of the β-sandwich. (**G**) Cross-section of the extended β-sandwich of the pentamer with Fc5 omitted for clarity.

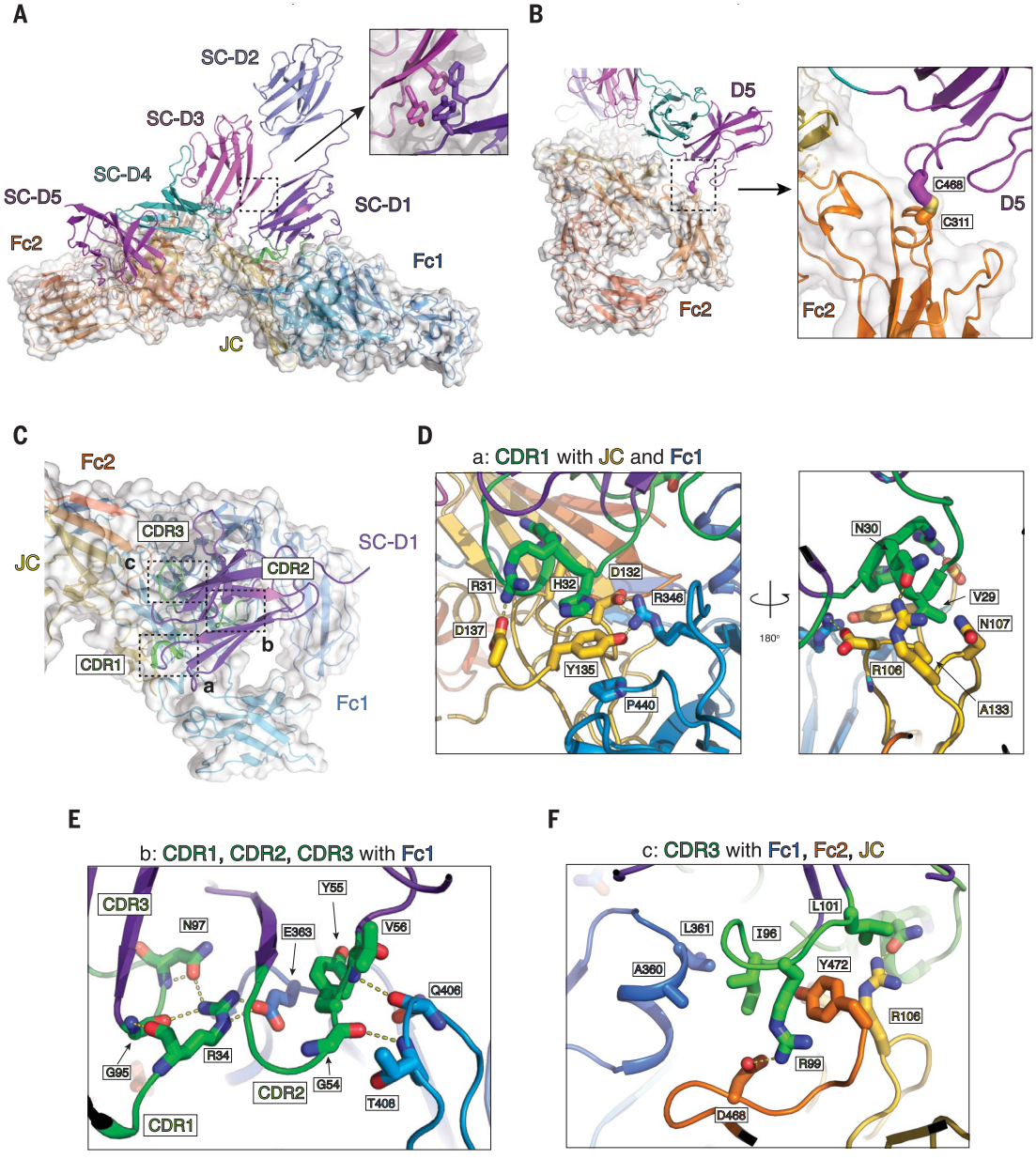


Fig. 5. Recognition of pIgA by the SC. (**A**) sIgA1 dimer with IgA shown as a transparent surface with cartoon representations underneath. Domains D1 to D5 of the SC are color coded as indicated with D1 CDR1, CDR2, and CDR3 highlighted in green. Magnification of the aromatic-rich D1–D3 interface shown in the inset. (**B**) Interaction of SC D5 and Fc2 with inset magnifying the $D5_{C468}$–$Fc2_{C311}$ disulfide bond. (**C**) Relative positioning of D1 CDR1 (box a), CDR2 (box b), and CDR3 (box c) and their interactions with pIgA. Domains D2 to D5 are omitted for clarity. (**D**) Magnification of box a from (C) to show D1 CDR1 interactions with Fc1 and JC. Side chains are shown as sticks and polar interactions as dashed yellow lines. (**E**) Magnification of box b from (C) showing interactions among D1 CDR1, CDR2, CDR3, and Fc1. (**F**) Magnification of box c from (C) showing interactions among D1 CDR3, Fc1, Fc2, and JC.

dimer. The interaction with D1 was noncovalent and involved all three complementarity determining region (CDR)–like loops, whereas the interaction with D5 was mediated by a single disulfide bond (Fig. 5). Although D2 to D4 did not make direct contacts with pIgA, they provided the correct spacing to allow the interactions of D1 and D5 with the dimer. D2 to D5 were arranged in a head-to-tail manner, whereas a sharp 180° turn in the D1–D2 linker allowed D1 to position its three CDRs to bind pIgA (Fig. 5A). The described kinked SC conformation was stabilized via an aromatic-rich D1–D3 interface (Fig. 5A). Approximately one-third of the ~1055 $Å^2$ of D1 buried surface area resulted from interactions with the JC C terminus, consistent with the requirement of the JC for pIgR binding and transcytosis (*16*).

Discussion

This work presents the first atomic-level resolution structures of pIgA bound to the elusive JC and sheds light on the mechanism involved in higher-order polymer formation. It also presents for the first time the architecture of sIgA and describes at the atomic level the molecular basis for recognition of pIgA by the SC of the pIgR. Our structures show that upon pIgA binding, the SC undergoes a large conformational change. In the closed SC conformation, the majority of the Fc1-binding D1 residues are at the D1–D4–D5 interface and at least partially buried, whereas the JC and Fc2

tailpiece-binding residues are solvent accessible (Fig. 6, A and B). Expanding on the hypothesis of Stadtmueller *et al.* (*15*), we propose a model in which the closed pIgR conformation solvent-accessible residues make initial contact with the JC and Fc2 tailpiece. This leads to a conformational change in the receptor that breaks the D1–D4–D5 interface and exposes the remaining contact residues in D1 to bind Fc1 (Fig. 6C). The open pIgR conformation is stabilized through a new D1–D3 interface that creates optimal spacing between D1 and D5 to enable secondary site binding to Fc2. Reduction of the $D5_{C468}$–$D5_{C502}$ intramolecular disulfide frees $D5_{C468}$ to form the intermolecular disulfide with $Fc2A_{C311}$, locking pIgR into a bent, pIgA-bound conformation (fig. S8). When pIgR is proteolytically cleaved upon transcytosis, SC remains covalently attached to pIgA, leading to the release of sIgA at the mucosa.

Both the SC and pIgA are modified with N-linked glycans, which not only facilitate proper protein folding, but also mediate the attachment and neutralization of pathogens (*1*, *8*). Our structures revealed ordered N-linked glycans on JC_{N49}, Fc_{N337}, Fc_{N459}, SC_{N65}, SC_{N72}, SC_{N168}, SC_{N403}, SC_{N451}, and SC_{N481}, which are not in contact with the protein and are placed away from any SC–pIgA interaction interfaces (fig. S9). These glycosylation sites could be relevant in facilitating host and pathogen lectin binding.

Extensive mutational analysis has previously been performed on both pIgA and the pIgR, assessing both binding and transcytosis. As seen in our structures, the C-terminal 25 residues of the JC are critical for SC binding, whereas JC_{C15} and JC_{C69} are important for efficient transcytosis, likely due to stabilization of the pIgA assembly through their covalent bonds to the Fc tailpieces (*16*). Mutagenesis of the human SC D1 residues N30, R31, H32, and R34 in the highly homologous rabbit SC either significantly reduced or abolished SC binding to human pIgA (*17*), consistent with the identification of these as key interaction residues in our structures. Additionally, swapping of the rabbit SC D1 CDR2 or CDR3 loops with the corresponding structurally similar yet sequence diverse CDRs from D2 also abolished binding (*17*), consistent with both CDRs making important contacts to pIgA in our structures. Mutation of Fc_{C311}, $Fc_{P440-F443}$, or deletion of the Fc tailpieces results in a significant impairment in transcytosis (*18*), a functional consequence highlighting the important roles these regions play in SC binding, as observed in our structures. Taken together, these mutagenesis studies underscore the functional importance of the interactions described in our structures.

The coreceptor Mac-1 (CD11b/CD18) is required to bind and activate FcαRI signaling

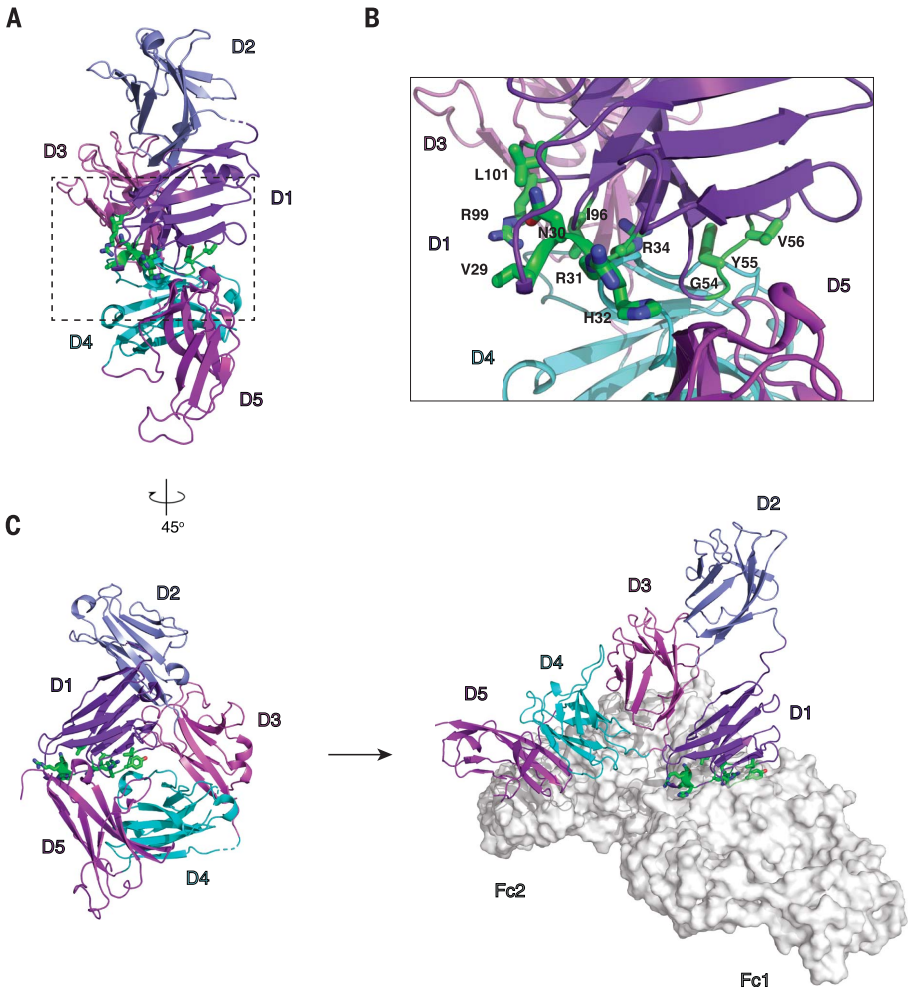


Fig. 6. Model for pIgA recognition by pIgR. (**A**) Structure of the free SC (PDB 5D4K) in a closed conformation stabilized by a D1–D4–D5 interface. CDR residues of D1 involved in pIgA binding are colored green and their side chains are shown as sticks. (**B**) Magnification of the box in (A) highlighting pIgA contact residues of D1 that are solvent accessible or buried. (**C**) Model of the conformational change in pIgR that converts it from the closed, unbound state (left, free SC) to the open, pIgA-bound state (right, dimer structure aligned to D1 of free SC).

by sIgA but not dimeric IgA, possibly due to steric hindrance by SC (*12*, *19*). Our sIgA dimer structure showed that JC binding to the Fcs overlapped with two of the four FcαRI-binding sites (fig. S7). Similarly, in the tetramer, only two sites were available because the Fc–Fc-mediated contacts also occurred at the Cα2–Cα3 interface. We saw no evidence of additional steric hindrance between the SC and the two accessible FcαRI-binding sites in the dimer or tetramer. In the pentamer, no FcαRI-binding sites were accessible, suggesting that pentameric IgA cannot bind the receptor. Because the oligomeric state of sIgA used in the Mac-1 studies is unknown (*19*), it is difficult to interpret the requirement for Mac-1 in FcαRI signaling.

Although many recombinant IgGs have been approved for therapeutic use in fields such as oncology, immunology, and ophthalmology, an IgA-based therapeutic has yet to be approved, likely in part due to the short serum half-life of this isotype (*9*). The structural characterization of sIgA provided here opens the possibility for engineering half-life extension properties into pIgA while maintaining its fundamental mucosal tissue–targeting properties. Furthermore, using the structural information as a framework, recombinant sIgA could be designed for potential oral delivery of a therapeutic antibody because the SC is known to provide stability and protection to pIgA from proteolytic degradation (*8*). Our studies also raise the question of whether the JC-containing pentameric IgM is formed through a similar oligomerization mechanism and if pIgR binding is analogous. High sequence conservation between IgA and IgM tailpieces (fig. S5B) suggests a shared mechanism of polymerization. However, the presence of an additional constant domain, Cμ4, in IgM may require a modified mode of pIgR recognition. Additional structural studies will be needed to address

these fundamentally important questions in mucosal immunity.

REFERENCES AND NOTES

1. P. de Sousa-Pereira, J. M. Woof, *Antibodies* **8**, 57 (2019).
2. M. S. Halpern, M. E. Koshland, *Nature* **228**, 1276–1278 (1970).
3. A. Bastian, H. Kratzin, K. Eckart, N. Hilschmann, *Biol. Chem. Hoppe Seyler* **373**, 1255–1264 (1992).
4. R. M. Chapuis, M. E. Koshland, *Biochemistry* **14**, 1320–1326 (1975).
5. S. Saito *et al*., *PLOS Pathog.* **15**, e1007427 (2019).
6. E. Fallgren-Gebauer *et al*., *Adv. Exp. Med. Biol.* **371**, 625–628 (1995).
7. T. B. Tomasi Jr., E. M. Tan, A. Solomon, R. A. Prendergast, *J. Exp. Med.* **121**, 101–124 (1965).
8. C. S. Kaetzel, *Immunol. Rev.* **206**, 83–99 (2005).
9. T. N. Lombana *et al*., *MAbs* **11**, 1122–1138 (2019).
10. E. Hiramoto *et al*., *Sci. Adv.* **4**, eaau1199 (2018).
11. J. Zikan *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5905–5909 (1985).
12. A. B. Herr, E. R. Ballister, P. J. Bjorkman, *Nature* **423**, 614–620 (2003).
13. P. A. Ramsland *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15051–15056 (2007).
14. S. Frutiger, G. J. Hughes, W. C. Hanly, M. Kingzette, J.-C. Jaton, *J. Biol. Chem.* **261**, 16673–16681 (1986).
15. B. M. Stadtmueller *et al*., *eLife* **5**, e10640 (2016).
16. F. E. Johansen, R. Braathen, P. Brandtzaeg, *J. Immunol.* **167**, 5185–5192 (2001).
17. R. S. Coyne, M. Siebrecht, M. C. Peitsch, J. E. Casanova, *J. Biol. Chem.* **269**, 31620–31625 (1994).
18. M. J. Lewis, R. J. Pleass, M. R. Batten, J. D. Atkin, J. M. Woof, *J. Immunol.* **175**, 6694–6701 (2005).
19. A. B. van Spriel, J. H. W. Leusen, H. Vilé, J. G. J. Van De Winkel, *J. Immunol.* **169**, 3831–3836 (2002).

ACKNOWLEDGMENTS

We thank N. Lombana for ordering constructs, Research Materials Group for protein expression, F. Farahi and A. Gill for small-scale purification, A. Estevez and M. Kschonsak for data collection and discussions, and A. Rohou and B. Barad for assistance with data processing. **Funding:** This work was funded by Genentech, Inc. **Author contributions:** N.K.: conceptualization; data curation; formal analysis; investigation; validation; visualization; writing original draft; writing, review, and editing; C.P.A.: data curation; investigation; supervision; C.C.: supervision; writing original draft; writing, review, and editing; M.L.M: conceptualization; investigation; formal analysis; project administration; supervision; validation; visualization; writing original draft; writing, review, and editing; **Competing interests:** All authors are employees of Genentech, Inc., a member of the Roche Group, and may hold stock and options. C.C. and M.L.M. are inventors on unpublished patent applications that relate to IgA antibodies and IgG-IgA fusion molecules and methods of making and using them. **Data and materials availability:** All maps and coordinate models have been deposited to the Worldwide Protein Data Bank: sIgA1 dimer (EMD-20749, PDB ID 6UE7); sIgA2m2 tetramer class 1 (EMD-20750, PDB ID 6UE8); sIgA2m2 tetramer class 2 (EMD-20751, PDB ID 6UE9); sIgA2m2 pentamer (EMD-20752, PDB ID 6UEA). Materials used in this study are available under a material transfer agreement from Genentech, Inc.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/1008/suppl/DC1
Materials and Methods
Figs. S1 to S9
Table S1
References (*20*–*34*)

20 September 2019; accepted 24 January 2020
Published online 6 February 2020
10.1126/science.aaz5807

REPORT

IMMUNOLOGY

Structural insights into immunoglobulin M

Yaxin Li[1,2]*, Guopeng Wang[3]*, Ningning Li[2,3], Yuxin Wang[1,2], Qinyu Zhu[1,2], Huarui Chu[1,2], Wenjun Wu[4,5], Ying Tan[4,5], Feng Yu[4,5,6], Xiao-Dong Su[1], Ning Gao[2,3], Junyu Xiao[1,2]†

Immunoglobulin M (IgM) plays a pivotal role in both humoral and mucosal immunity. Its assembly and transport depend on the joining chain (J-chain) and the polymeric immunoglobulin receptor (pIgR), but the underlying molecular mechanisms of these processes are unclear. We report a cryo–electron microscopy structure of the Fc region of human IgM in complex with the J-chain and pIgR ectodomain. The IgM-Fc pentamer is formed asymmetrically, resembling a hexagon with a missing triangle. The tailpieces of IgM-Fc pack into an amyloid-like structure to stabilize the pentamer. The J-chain caps the tailpiece assembly and bridges the interaction between IgM-Fc and the polymeric immunoglobulin receptor, which undergoes a large conformational change to engage the IgM-J complex. These results provide a structural basis for the function of IgM.

Immunoglobulin M (IgM) is the first class of antibody produced after B cell activation. Secretory IgM, together with IgA, plays a critical role in the mucosal immune system. IgM forms oligomers, and the presence of multivalent antigen-binding sites in IgM oligomers is a key factor in their ability to agglutinate pathogens. The heavy chain of IgM contains a C-terminal extension known as the tailpiece that is essential for oligomerization. In the presence of the joining chain (J-chain), a 15-kDa protein that has no homology to other proteins, IgM forms a pentamer, in which five IgM monomers are linked by disulfide bonds between each other and with the J-chain (*1*–*3*). The J-chain also facilitates the dimerization of IgA, which contains a similar tailpiece. The overall structural organization of the IgM pentamer is not completely understood, nor is the function of the J-chain in regulating the assembly processes of these polymeric immunoglobulins.

Furthermore, to function at the mucosal surface, IgM secreted by the plasma cells must be transcytosed through the mucosal epithelial cells. This process critically depends on the polymeric immunoglobulin receptor (pIgR) (*4*, *5*). pIgR is a type I transmembrane protein that contains five extracellular immunoglobulin-like domains (D1 to D5). It specifically binds to J-chain–containing secretory IgM and IgA at the basolateral surface of epithelial cells and escorts them to the apical side. There, the ectodomain is released by proteolysis and secreted together with IgM and IgA. The free ectodomain is often referred to as the secretory component (SC). The molecular mechanism of how pIgR/SC facilitates the secretion of IgM and IgA also remains elusive.

To gain insight into the assembly and secretion of IgM, we reconstituted a tripartite complex containing the Fc region of human IgM (Fcμ), J-chain, and SC (fig. S1), which was then analyzed using cryo–electron microscopy (cryo-EM) (fig. S2). The final constructed model reveals that an IgM Cμ3-Cμ4-tailpiece pentamer and a J-chain molecule form a near-planar structure, and a triangular SC docks perpendicularly to the Fcμ-J plane (Fig. 1). The center region of the structure, including the IgM-Cμ4 domains and tailpieces, the J-chain, and the D1 domain of pIgR/SC, displayed better than 3 Å resolution, with most of the side chains clearly visualizable (figs. S2 and S3).

Although the structures of individual mouse IgM-Cμ2, IgM-Cμ3, and IgM-Cμ4 domains have been previously characterized (*6*), the structure of an entire Fcμ has yet to be elucidated. In our structure, IgM-Cμ4 forms a dimer within each Fcμ monomer (fig. S4A). The IgM-Cμ4 dimer here highly resembles the dimers formed by IgA-Cα3, IgG-Cγ3, and IgE-Cε4, but is remarkably distinct from the "parallel" dimer observed in the mouse IgM-Cμ4 crystal structure (fig. S4). The IgM-Cμ4 dimer here buries 2540 Å^2 of surface area, larger than the 1900-Å^2 surface concealed by the mouse IgM-Cμ4 dimer. The physiological importance of these different IgM-Cμ4 dimers remains unclear. The IgM-Cμ3 domains are not involved in direct contacts within each Fcμ monomer (fig. S4A). In some determined structures of IgE, the IgE-Cε2 dimer bends acutely and packs against IgE-Cε3 and IgE-Cε4 (fig. S4E). On the basis of homology modeling, the IgM-Cμ2 dimer was thought to bend similarly (*7*). However, the densities corresponding to the IgM-Cμ2 dimers, albeit weak, suggest that they do not adopt a stably bent conformation (fig. S4F).

IgM forms a pentamer in the presence of the J-chain. Earlier EM studies depicted a stellate appearance of the IgM pentamer with a five-fold symmetry (*8*). Although this model is widely documented in textbooks, a recent negative-stain EM study shows that the IgM pentamer is an asymmetric pentagon with a 50° gap (*9*). A similar gap is present in our structure and is occupied by the J-chain (Fig. 2A). Nonetheless, the gap in our structure has a 61° angle, and the five Fcμ units are arranged in an almost perfect hexagonal symmetry. We speculate that if the J-chain was not present, a sixth Fcμ unit could be readily accommodated to form a hexamer. This is consistent with a recent electron tomography study showing that the IgM pentamer is equivalent to the hexamer except for the J-chain (*10*). IgM-Cμ3 and IgM-Cμ4 as well as the tailpieces all contribute to pentamer formation (Fig. 2A). Cys^{414} residues from two neighboring IgM-Cμ3 domains are adjacent to one another and likely form interchain disulfide bonds, consistent with earlier analyses (*11*). The FG loops mediate the interaction between two neighboring IgM-Cμ4 domains. The seven residues Tyr^{562} to Met^{568} in each tailpiece form a β strand, and the 10 strands are arranged into two five-stranded parallel β sheets that pack onto one another in an antiparallel fashion. This is reminiscent of the cross-β fibers seen in amyloid proteins and peptides (*12*) and provides the most prominent interactions to stabilize the IgM pentamer. The Tyr^{562}, Val^{564}, Leu^{566}, and Met^{568} side chains face inward and mediate hydrophobic interactions between the two sheets (Fig. 2B and fig. S3A). Mutation of each of these residues either abolished or significantly reduced the oligomerization of mouse IgM (*13*). Val^{567} residues also stack onto one another to mediate packing interactions between adjacent strands, and mutation of the corresponding Ile^{567} in mouse IgM disrupted its oligomerization (*13*). Notably, Val^{567} residues are exposed to the solvent and form extended hydrophobic surface patches (Fig. 2B). The highly conserved Asn^{563} and Ser^{565} residues conform to the N-linked glycosylation consensus motif and facilitate the attachment of glycans on Asn^{563} (Fig. 2B and fig. S3A), which is likely necessary to prevent IgM from forming aggregations. Cys^{575}, the penultimate Cys, mediates the formation of disulfide bonds between adjacent Fcμ monomers, which is a prerequisite for IgM oligomerization (*13*, *14*). Indeed, the Cys^{575} residues of Fcμ5A and Fcμ4B are adjacent to one another and likely form a disulfide bond (Fig. 3A and fig. S3C). By contrast, Cys^{575} of Fcμ5B and Cys^{575} of Fcμ1A are linked to the J-chain.

[1]State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, China. [2]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. [3]State Key Laboratory of Membrane Biology, School of Life Sciences, Peking University, Beijing, China. [4]Renal Division, Department of Medicine, Peking University First Hospital, Beijing, China. [5]Institute of Nephrology, Peking University, Beijing, China. [6]Department of Nephrology, Peking University International Hospital, Beijing, China.
*These authors contributed equally to this work.
†Corresponding author. Email: junyuxiao@pku.edu.cn

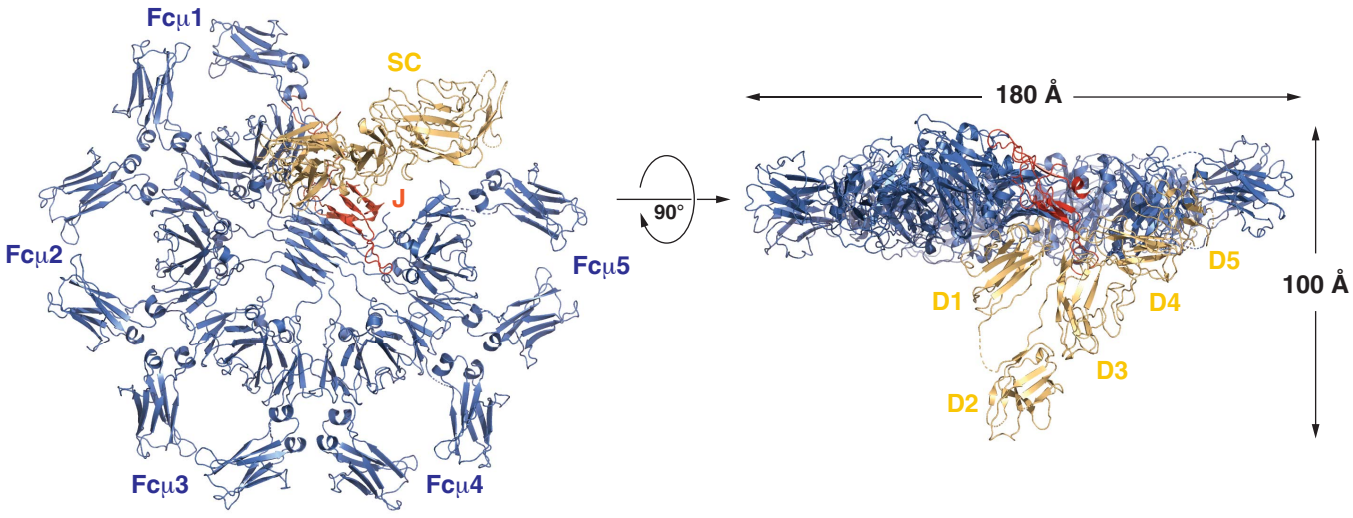


Fig. 1. The cryo-EM structure of the Fcμ pentamer in complex with the J-chain and SC. The five Fcμ monomers are shown in blue and labeled Fcμ1 to Fcμ5. The J-chain and SC are shown in red and gold, respectively.

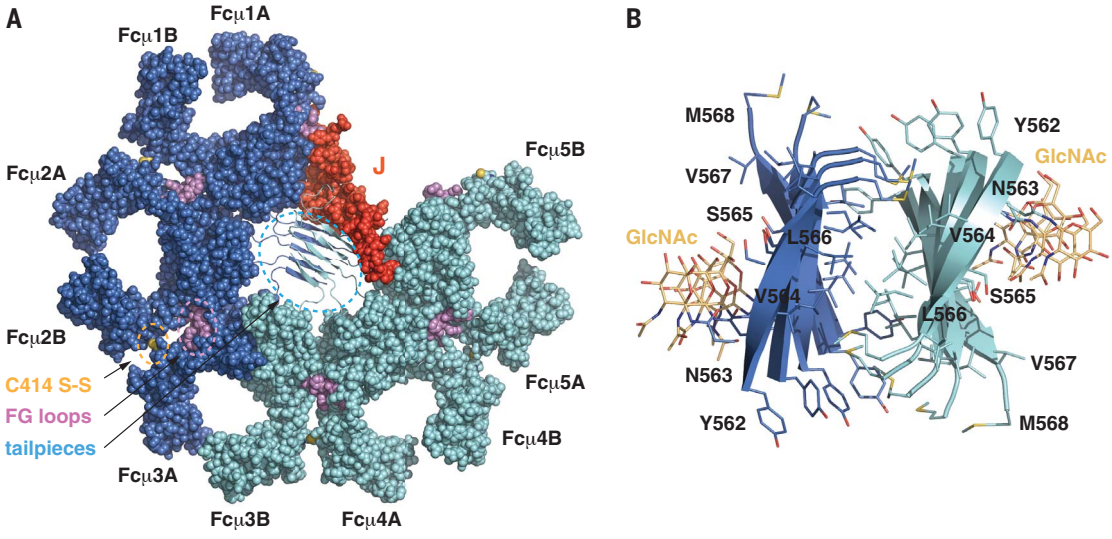


Fig. 2. The Fcμ pentamer. (**A**) The Fcμ pentamer is shown as a space-filling model except for the tailpieces, which are shown as ribbon diagrams. Five Fcμ chains are shown in blue and five are shown in cyan. The Cys^{414} disulfide bonds and the FG loops are highlighted. (**B**) Detailed view of the tailpiece assembly. The *N*-acetylglucosamine (GlcNAc) molecules attached to Asn^{563} are shown as orange sticks. Amino acid abbreviations in this or later figures: A, Ala; C, Cys; E, Glu; F, Phe; H, His; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; Y, Tyr.

The J-chain was identified almost 50 years ago as an integral subunit of secretory IgA and IgM (*15*, *16*). Guided by the disulfide bond assignments from earlier biochemical studies (*17*), we were able to build an atomic model for the majority of this protein (Fig. 3A and fig. S3B). It interacts with both Fcμ1 and Fcμ5 to seal the Fcμ pentamer. The N-terminal wing comprises four β strands (β1 to β4) and a short helix. The β3 and β4 strands pack on the tailpieces of Fcμ5B and Fcμ1A, respectively, to cap the fiber-like assembly of the IgM tailpieces. J-chain residue Cys^{14} at the beginning of β2 forms a disulfide bond with Cys^{575} of Fcμ5B, and J-chain residue Cys^{68} in the short helix forms a bond with Cys^{575} of Fcμ1A (Fig. 3A and fig. S3, D and E). The β2-β3 loop interacts with the base of Fcμ5B. The C-terminal wing contains a long hairpin-like structure, which reaches up to the Cμ3-Cμ4 junction of Fcμ1A and makes extensive hydrophobic contacts (Fig. 3B and fig. S3, F to H). J-chain residues Val^{113} and Pro^{114} interact with Met^{489}, Pro^{494}, and Val^{537} of Fcμ1A, whereas J-chain residues Leu^{115}, Val^{124}, and Thr^{126} interact with Phe^{358}, Leu^{359}, Phe^{485}, and Val^{547} of Fcμ1A. J-chain residues Ala^{127}, Pro^{130}, and Tyr^{134} form a pocket to accommodate Pro^{544} of Fcμ1A. Furthermore, this C-terminal hairpin also interacts with pIgR/SC to connect IgM with the receptor.

pIgR/SC binds selectively to IgA or IgM that contains the J-chain (*18*). The D1 domain is the major binding site, and the three loops that are analogous to the complementarity-determining regions (CDRs) of immunoglobulin variable domains are all involved (Fig. 4A) (*19*–*22*). In the ligand-free state of pIgR/SC, D1 to D5 are arranged in the form of an isosceles triangle (*22*), with the D2-D3 and D4-D5 sides having similar lengths (Fig. 4B). In the Fcμ-J-SC complex, the D2-D3 side remains unaltered, whereas D1 and D4-D5 undergo drastic conformational

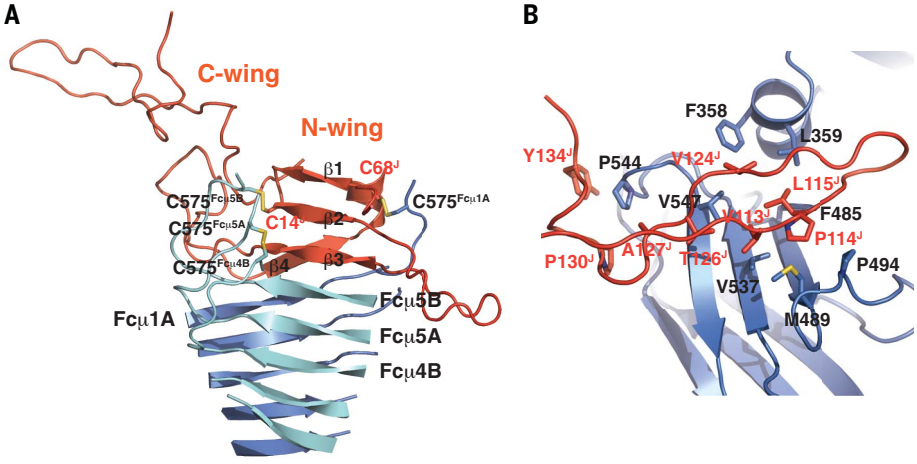


Fig. 3. Structure of the J-chain and its interactions with Fcμ. (**A**) The J-chain has a two-winged structure and interacts with the tailpieces of the Fcμ pentamer. (**B**) The C-terminal hairpin of the J-chain interacts with Fcμ1A.

changes, leading to the formation of a very different triangular shape (Fig. 4C). D4-D5 rotates 120° en bloc to become co-linear with D2-D3, and the newly formed D3-D4 interface buries a large surface area (1486 Å^2). D1 is rotated 84° and packs onto D2-D3. This large conformational change is consistent with a previous double electron-electron resonance (DEER) spectroscopy study (*22*). In particular, the distance between the Cα atoms of D1-Thr^{67} and D5-Gln^{491} is 78 Å in this new conformation, in good agreement with DEER measurements showing that ligand binding induces a separation of these two residues beyond 70 Å. The three CDRs function as a central hub to mediate a network of interactions with Fcμ1, Fcμ5, and the J-chain (Fig. 4, A, D, and E, and fig. S3, I to N). In CDR1, pIgR/SC residue Val^{29} interacts with J-chain residue Ala^{132} in the J-chain C-terminal hairpin (Fig. 4D). pIgR/SC residue Asn^{30} forms a hydrogen bond with

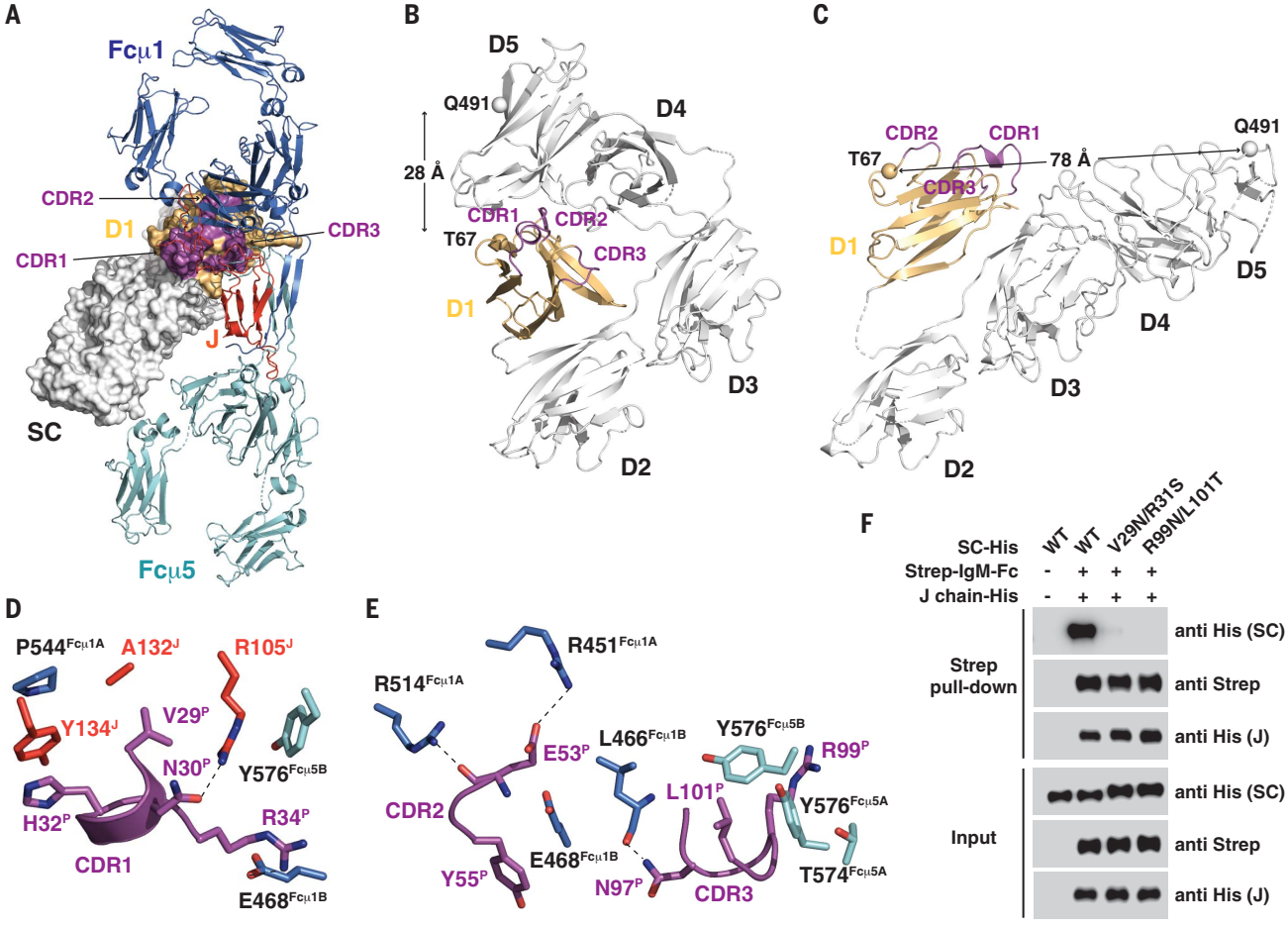


Fig. 4. Conformational change of pIgR/SC and its interaction with the Fcμ-J complex. (**A**) The three CDRs (magenta) in pIgR/SC-D1 (gold) interact with Fcμ1 (blue), Fcμ5 (cyan), and the J-chain (red). pIgR/SC is shown as a surface representation. Its D2 to D5 domains are shown in white. (**B**) The structure of apo pIgR/SC (PDB ID 5D4K). D1 is shown in gold, with the three CDR loops highlighted in magenta. The Cα atoms of D1-Thr^{67} and D5-Gln^{491} are shown as spheres, and the distance between them is indicated. (**C**) The structure of pIgR/SC in the Fcμ-J-SC complex. (**D**) Detailed view of the interactions at the D1-CDR1 region. Polar interactions are indicated by dashed lines. (**E**) Detailed view of the interactions at the CDR2 and CDR3 region. (**F**) SC mutants display reduced interactions with Fcμ-J.

J-chain residue Arg^{105}, which in turn packs onto Tyr^{576} of Fcμ5B. pIgR/SC residue His^{32} packs against J-chain residue Tyr^{134}, which interacts with Pro^{544} of Fcμ1A as described above. pIgR/SC residue Arg^{34} forms a salt bridge with Glu^{468} of Fcμ1B. In CDR2 and CDR3, pIgR/SC residue Glu^{53} interacts with both Arg^{451} and Arg^{514} of Fcμ1A (Fig. 4E). It has been reported that rabbit and rodent pIgR can transport IgA but not IgM (*23*, *24*). Rabbit pIgR lacks a residue entirely at this position, whereas rodents feature an Asn (*4*, *21*). Arg^{451} and Arg^{514} are also not conserved in mouse IgM (fig. S5A). pIgR/SC residue Tyr^{55} packs against Glu^{468} of Fcμ1B. pIgR/SC residue Asn^{97} interacts with the main-chain carbonyl group of Leu^{466} of Fcμ1B. pIgR/SC residue Arg^{99} is sandwiched between Tyr^{576} of Fcμ5B and Thr^{574} of Fcμ5A. pIgR/SC residue Leu^{101} interacts with Tyr^{576} of Fcμ5A and Tyr^{576} of Fcμ5B. Two SC mutants, V29N/R31S and R99N/L101T, which are designed to introduce bulky N-linked glycans in the CDR1 and CDR3 regions, respectively, display greatly reduced interactions with the Fcμ-J complex (Fig. 4F), confirming the functional relevance of the molecular interactions described above.

Two other IgM receptors exist in mammals in addition to pIgR: FcαμR and FcμR/Toso/Faim3 (*25*). They each contain a domain that is homologous to the D1 domain of pIgR. Like pIgR, FcαμR binds both IgA and IgM. Its D1-like domain also shows high sequence similarity to pIgR-D1 (fig. S5B), and residues corresponding to pIgR/SC residues Val^{29}, Asn^{30}, His^{32}, Arg^{34}, Tyr^{55}, and Leu^{101} that are involved in binding to the Fcμ-J complex in pIgR are all present. Thus, FcαμR may interact with IgM in a manner similar to pIgR. FcμR/Toso/Faim3, on the other hand, binds only to IgM. Furthermore, it can interact with both pentameric and hexameric IgM with similar affinities (*26*). This is in contrast to pIgR, which selectively binds to the IgM pentamer that contains the J-chain. Indeed, the D1-like domain of FcμR/Toso/Faim3 is more divergent than pIgR-D1 and lacks most of the critical Fcμ-J–interacting residues (fig. S5C). Thus, FcμR likely binds IgM in a different fashion.

Our high-resolution cryo-EM structure of the Fcμ-J-SC complex provides a framework for further understanding the functions of IgM, and also sheds light on the interaction between IgM and other receptors. As a result of the stronger binding of IgM to its targets and its more potent activity to induce complement-dependent cytotoxicity, IgM can be potentially exploited for therapeutic applications. Our results pave the way for structure-based engineering of these molecules.

REFERENCES AND NOTES

1. I. N. Norderhaug, F. E. Johansen, H. Schjerven, P. Brandtzaeg, *Crit. Rev. Immunol.* **19**, 481–508 (1999).
2. F. E. Johansen, R. Braathen, P. Brandtzaeg, *Scand. J. Immunol.* **52**, 240–248 (2000).
3. J. M. Woof, J. Mestecky, *Immunol. Rev.* **206**, 64–82 (2005).
4. C. S. Kaetzel, *Immunol. Rev.* **206**, 83–99 (2005).
5. H. Turula, C. E. Wobus, *Viruses* **10**, 237 (2018).
6. R. Müller *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10183–10188 (2013).
7. D. M. Czajkowsky, Z. Shao, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14960–14965 (2009).
8. A. Feinstein, E. A. Munn, *Nature* **224**, 1307–1309 (1969).
9. E. Hiramoto *et al.*, *Sci. Adv.* **4**, eaau1199 (2018).
10. T. H. Sharp *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 11900–11905 (2019).
11. F. W. Putnam, G. Florent, C. Paul, T. Shinoda, A. Shimizu, *Science* **182**, 287–291 (1973).
12. D. Eisenberg, M. Jucker, *Cell* **148**, 1188–1203 (2012).
13. D. Pasalic *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E8575–E8584 (2017).
14. R. Sitia *et al.*, *Cell* **60**, 781–790 (1990).
15. M. S. Halpern, M. E. Koshland, *Nature* **228**, 1276–1278 (1970).
16. J. Mestecky, J. Zikan, W. T. Butler, *Science* **171**, 1163–1165 (1971).
17. S. Frutiger, G. J. Hughes, N. Paquet, R. Lüthy, J. C. Jaton, *Biochemistry* **31**, 12643–12647 (1992).
18. P. Brandtzaeg, H. Prydz, *Nature* **311**, 71–73 (1984).
19. S. Frutiger, G. J. Hughes, W. C. Hanly, M. Kingzette, J. C. Jaton, *J. Biol. Chem.* **261**, 16673–16681 (1986).
20. R. S. Coyne, M. Siebrecht, M. C. Peitsch, J. E. Casanova, *J. Biol. Chem.* **269**, 31620–31625 (1994).
21. A. E. Hamburger, A. P. West Jr., P. J. Bjorkman, *Structure* **12**, 1925–1935 (2004).
22. B. M. Stadtmueller *et al.*, *eLife* **5**, e10640 (2016).
23. B. J. Underdown, I. Switzer, G. D. Jackson, *J. Immunol.* **149**, 487–491 (1992).
24. M. Røe, I. N. Norderhaug, P. Brandtzaeg, F. E. Johansen, *J. Immunol.* **162**, 6046–6052 (1999).
25. S. Akula, L. Hellman, *Curr. Top. Microbiol. Immunol.* **408**, 1–23 (2017).
26. H. Kubagawa *et al.*, *Curr. Top. Microbiol. Immunol.* **408**, 25–45 (2017).

ACKNOWLEDGMENTS

We thank the Core Facilities at the School of Life Sciences, Peking University for help with negative-staining EM; the Cryo-EM Platform of Peking University for help with data collection; the High-Performance Computing Platform of Peking University for help with computation; the National Center for Protein Sciences at Peking University for assistance with Amersham Imager; and G. Li for help with cDNA library. **Funding:** Supported by the National Key Research and Development Program of China (2017YFA0505200 to J.X.; 2016YFC0906000 to J.X. and X.-D.S.; 2019YFA0508904 to N.G.), the National Science Foundation of China (31570735, 31822014 to J.X.; 31725007, 31630087 to N.G.; 21727806 to X.-D.S.; 31922036 to N.L.), the Qidong-SLS Innovation Fund to J.X. and N.G.; and the Clinical Medicine Plus X Project of Peking University to J.X. and Y.T. **Author contributions:** Y.L. performed protein purification; Y.L. and G.W. prepared the cryo-EM sample and collected data; G.W. and N.L. processed the cryo-EM data; J.X. and N.G. built the structural model; Y.L. performed the pull-down experiments; J.X. wrote the manuscript, with inputs from all authors. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** The cryo-EM map and atomic coordinates of the Fcμ-J-SC complex have been deposited in the EMDB and PDB with accession codes EMD-0782 and 6KXS, respectively. All constructs used for protein expression in this study are available upon request.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/1014/suppl/DC1
Materials and Methods
Figs. S1 to S5
Table S1
References (*27*–*36*)

18 September 2019; accepted 24 January 2020
Published online 6 February 2020
10.1126/science.aaz5425

OPTICS

Ultrafast control of vortex microlasers

Can Huang[1], Chen Zhang[1], Shumin Xiao[1,2], Yuhan Wang[1], Yubin Fan[1], Yilin Liu[1], Nan Zhang[1], Geyang Qu[1], Hongjun Ji[1], Jiecai Han[2], Li Ge[3,4]*, Yuri Kivshar[5]*, Qinghai Song[1,6]*

The development of classical and quantum information–processing technology calls for on-chip integrated sources of structured light. Although integrated vortex microlasers have been previously demonstrated, they remain static and possess relatively high lasing thresholds, making them unsuitable for high-speed optical communication and computing. We introduce perovskite-based vortex microlasers and demonstrate their application to ultrafast all-optical switching at room temperature. By exploiting both mode symmetry and far-field properties, we reveal that the vortex beam lasing can be switched to linearly polarized beam lasing, or vice versa, with switching times of 1 to 1.5 picoseconds and energy consumption that is orders of magnitude lower than in previously demonstrated all-optical switching. Our results provide an approach that breaks the long-standing trade-off between low energy consumption and high-speed nanophotonics, introducing vortex microlasers that are switchable at terahertz frequencies.

Because of their mutual orthogonality, vortex beams with different topological charges have been proposed as an effective approach to revolutionize classical and quantum communications (*1*–*8*). Vortex beams with well-defined topological charges have been developed using external phase elements, such as spiral phase plates, computer generated holograms, and metasurfaces (*5*–*10*). Recently, driven by the demand for compact displays and high-density integration, on-chip vortex microlasers have attracted much attention (*5*–*7*, *11*). Compact vortex microlasers are usually created by transforming conventional optical cavities into spiral waveguides (*6*) or micropillar chains (*7*) and modulating them with additional asymmetric scatterers (*5*). Although the reported performance in both directional output and generation efficiency of orbital angular momentum (OAM) beams is notable, the quality (*Q*) factors of such vortex microlasers are strongly reduced by scattering losses and, thus, their energy consumption is large (*5*–*7*, *12*). Additionally, on-chip integrated vortex microlasers are either static (*5*, *6*) or controllable in emission chirality via the circularly polarized optical pump (*7*), making them unsuitable for ultrafast optical networks (*8*). Finally, because of the rundown time of an optical resonance, there appears to be a trade-off between low energy consumption and high modulation speed in nanophotonics, which restricts their application in modern optical computing and optical communications.

In this work, we solve these problems by employing bounded states in the continuum (BICs), which represent special solutions of the wave equation where the wave function exhibits localization in a radiation band (*13*, *14*). In optical systems, BICs appear through interference between localized resonances and radiation modes, and they have been observed in the form of quasi-BICs in many systems, ranging from isolated nanoparticles to periodic structures (*14*–*19*). In addition to ultrahigh *Q* factors and low-threshold lasing, it has been predicted that the BIC modes can possess vortex behaviors with different topological charges, which are important for vector beams (*20*–*22*). These findings make BICs very attractive for application in active photonics. We employ the specific characteristics of the topologically protected optical BICs and demonstrate the ultrafast control of perovskite-based vortex microlasers at room temperature.

Our metasurface is created using a 220-nm lead bromide perovskite ($MAPbBr_3$) film, patterned with a square array of circular holes (Fig. 1A). The whole structure is placed on a glass substrate (n_{sub} = 1.5, where n is the refractive index) and coated with polymethyl methacrylate (PMMA) (n_{pmma} = 1.49). The radius of the air holes is R = 105 nm, and the lattice spacing is p = 280 nm. We calculate the resonances of the transverse magnetic (TM) polarized field within the perovskite metasurface (Fig. 1B). Mode 1 has an appreciable *Q* factor within the gain spectral range of $MAPbBr_3$ perovskites (Fig. 1C). By changing

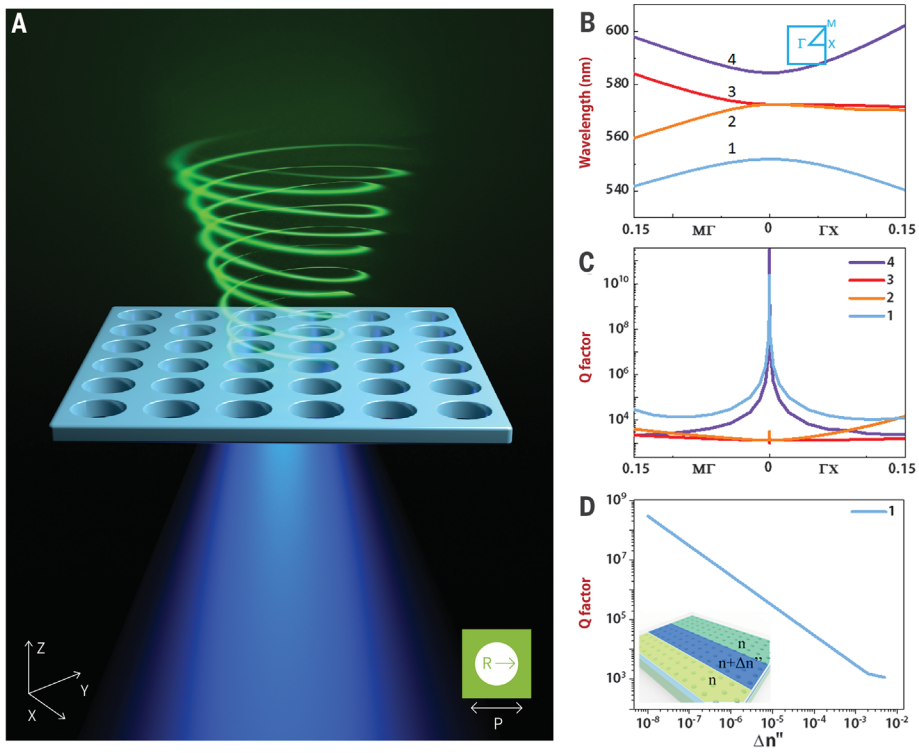


Fig. 1. Design and control of the quasi-BIC modes. (**A**) Schematic of the designed perovskite metasurface. The metasurface is pumped by a blue laser light, producing a green vortex beam in the vertical direction. (**B**) Dispersion relation around 550 nm for laser resonances in both the ΓX and ΓM directions. The inset shows the first Brillouin zone of the square lattice. (**C**) Calculated *Q* factors of four resonances. (**D**) Reduction of the *Q* factor for the quasi-BIC mode, with a growth of the imaginary part of the refractive index, $\Delta n''$.

[1]State Key Laboratory on Tunable Laser Technology, Ministry of Industry and Information Technology Key Laboratory of Micro-Nano Optoelectronic Information System, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China. [2]National Key Laboratory of Science and Technology on Advanced Composites in Special Environments, Harbin Institute of Technology, Harbin 150080, China. [3]The Graduate Center, CUNY, New York, NY 10016, USA. [4]Department of Physics and Astronomy, College of Staten Island, CUNY, Staten Island, NY 10314, USA. [5]Nonlinear Physics Centre, Research School of Physics, Australian National University, Canberra, ACT 2601, Australia. [6]Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan 030006, China.
*Corresponding author. Email: qinghai.song@hit.edu.cn (Q.S.); ysk@internode.on.net (Y.K.); li.ge@csi.cuny.edu (L.G.)

the radius of the hole, we preserve the Q factor of mode 1 throughout the operational range. This mode represents the topology-protected BIC with an embedded polarization vortex, which has been discussed previously (*20*, *23*–*26*).

The perovskite metasurfaces were fabricated from $MAPbBr_3$ with a combined process of electron-beam lithography and reactive ion etching (*27*). Figure 2A shows the top-view scanning electron microscope (SEM) image of the sample. To attain the designed BIC vortex lasers, we optically pump the perovskite metasurface with a frequency-doubled Ti: sapphire laser at room temperature. Figure 2B shows the evolution of the emission spectrum at different pumping densities. A broad, spontaneous emission peak of lead halide centered at 520 nm is observed at low pumping density. With an increase of pumping power, a single peak appears at 552 nm (the resonant wavelength of mode 1 in the normal direction) and quickly dominates the emission spectrum at higher pumping fluence. Figure 2C shows the output laser intensity as a function of the pumping density, and it confirms the onset of lasing with the characteristic S-shaped curve.

The vortex characteristics of the emission are studied from the far-field angular distributions by the back focal plane imaging technique (*27*) (Fig. 2). The intensity of the perovskite laser emission is spatially distributed in a donut shape with a dark zone at the center (Fig. 2D). The dominant bright ring corresponds to the far-field angle of $\theta_{FF} = 2°$. The dark center is caused by a topological singularity at the beam axis. Figure 2D shows the self-interference pattern of a donut beam (*27*), where a fork-shaped interference pattern can be seen. In the profiles of the donut beam behind a linear polarizer (Fig. 2, E to H), two lobes are observed, and their direction follows the axis of the linear polarizer, demonstrating the radial polarization. These experimental results for the beam profiles, self-interference patterns, and polarization states confirm the onset of the vector vortex lasing at the BIC mode (*28*). The BIC vortex microlasers are robust to global changes. These observations have been reproduced in more than 10 samples (*27*), and they have proven to be robust to a change of the excitation power. The donut-shaped laser beams and fork-shaped interference patterns are well preserved from the laser threshold to gain saturation (Fig. 2D).

Vortex emission is typically produced by real-space chiral structures (*9*), which are absent in our experiment. Instead, in our system, the topology that protects the BICs manifests itself as the rotation of polarization along the beam axis in the far field (*28*, *29*). Acting alone, it affects an OAM in the cross-polarized transmission of circularly polarized beams (*29*). Together with the transverse spin angular momentum introduced in real samples, we observe the emergence of the effective Pancharatnam-Berry phase (*27*).

Notably, the laser emission at the symmetry-protected BICs can be all-optically controlled. Although the BIC lasers are robust to a global change, they are extremely sensitive to symmetry-breaking perturbations (*14*, *21*) and, thus, are easily controllable. In passive systems, such control is realized via a deformation of nanostructures (*26*) or the Kerr nonlinearity, but these options are not suitable for post-fabrication control or require strong optical

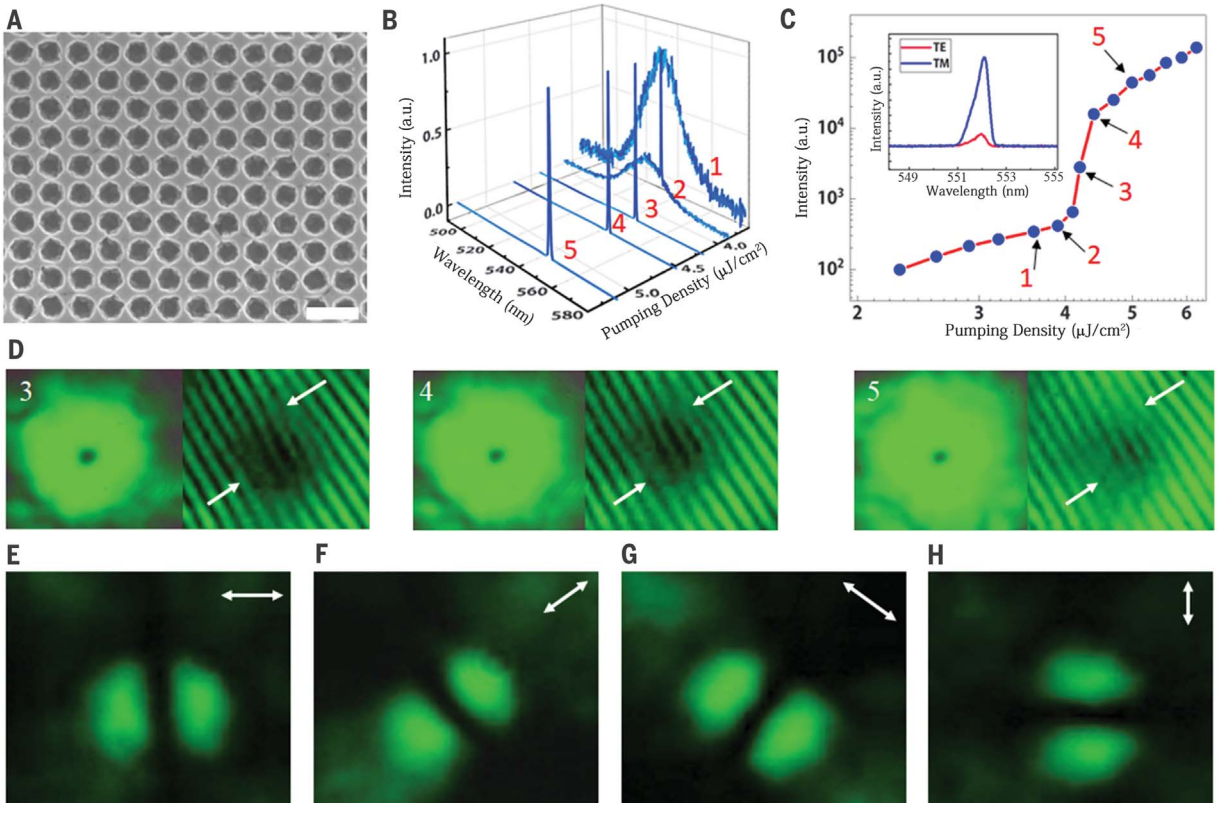


Fig. 2. Demonstration of vortex perovskite microlasers. (**A**) SEM image of the fabricated perovskite metasurface, following the design of Fig. 1. Scale bar, 500 nm. (**B**) Evolution of the normalized emission spectrum with pumping density. a.u., arbitrary units. (**C**) Integrated output intensity as a function of pumping density. The S-shaped curve shows a laser threshold at ~4.2 μJ/cm^2. Inset shows the laser polarization within the plane of the device. TE, transverse electric polarization with E in-plane. (**D**) Far-field patterns and corresponding self-interference at different pumping densities. Numbers 3 to 5 denote the same spectra plotted in (B) and (C). (**E** to **H**) Measured intensity distribution of the vortex laser beam after a linear polarizer with polarization orientations along 0°, 45°, 90°, and 135°. The scale and color bars are the same across the panels 3 to 5 in (D). White arrows denote the direction of the linear polarizer.

excitation. The laser systems can solve this problem in a simpler way. The exceptional gain corresponds to the imaginary part of the refractive index (n''), which is a new parameter to control the symmetry. One example is illustrated in Fig. 1D and (*27*), where a regular change of n'' of a laser system in a selected region can break the fourfold rotational symmetry of the system. As a result, the resonance at the Γ-point degrades from BIC to quasi-BIC, and the corresponding Q factor is reduced by orders of magnitude. In this sense, the pumping geometry can provide a more flexible approach to control the performance of the vortex BICs lasers.

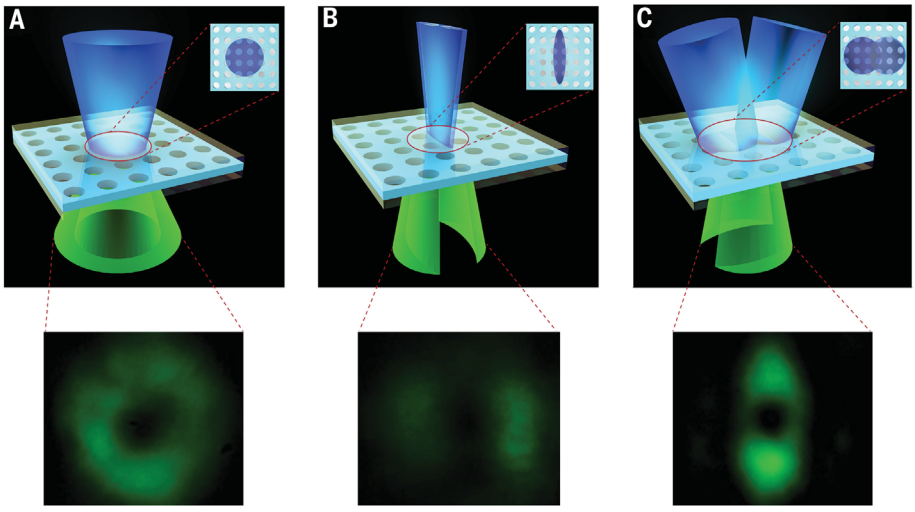


Fig. 3. Optically controlled far-field vortex lasing. (**A** to **C**) Schematics of the experiments (top) and experimentally measured far-field patterns (bottom). The emission profile switches to two lobes when the pumping laser beam is changed to an ellipse (B) or two overlapped circular beams (C).

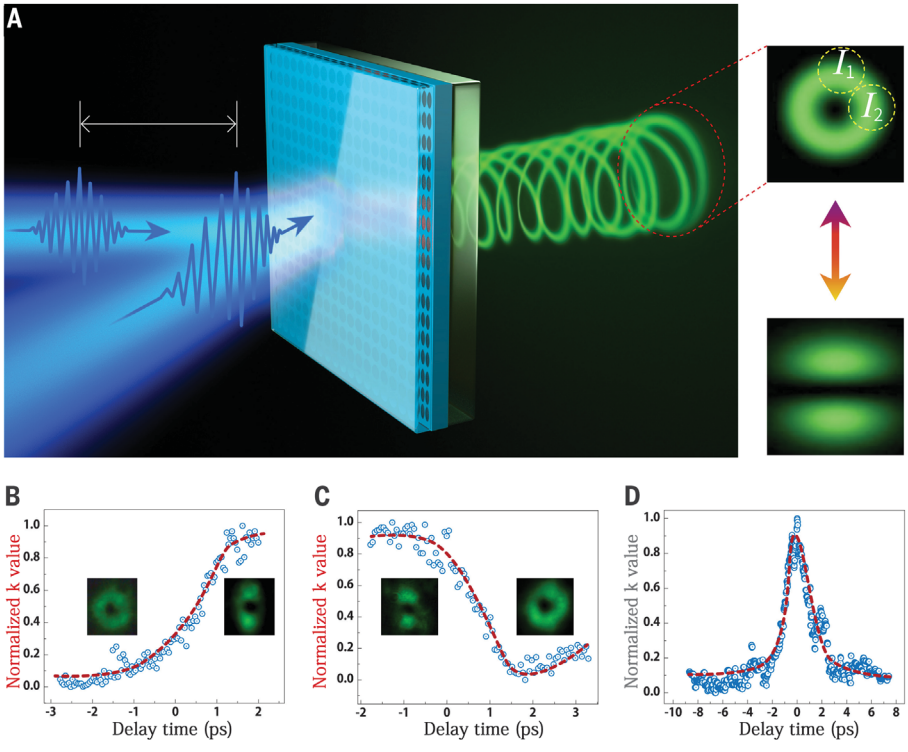


Fig. 4. Ultrafast control of the quasi-BIC microlasers. (**A**) Schematic of the two-beam pumping experiment. Two beams are spatially detuned with a distance of $d < 2R$, shifted temporally with a delay time, τ. The insets show the far-field emission patterns from the perovskite metasurface under both symmetric and asymmetric excitations. (**B**) Transition from a BIC microlaser to a linearly polarized laser. I_1 and I_2 are the intensities at the marked regions in the top inset in (A). Insets in (B) show the corresponding beam profiles. (**C**) Reverse process to that shown in (B). (**D**) Transition from a donut beam to two-lobe beam, and back, within a few picoseconds. Red curves are guiding lines for the calculation of the transition time.

To demonstrate this kind of all-optical control of the vortex microlasers, the perovskite metasurface is pumped with a circular laser beam, the symmetry is protected, and thus a donut-shaped beam is generated (Fig. 3A). Once the pumping region is transferred from a circle to an ellipse (Fig. 3B), the symmetry protection breaks, and two linearly diffracted beams are produced (*30*). Similar symmetry breaking can also be realized with a two-beam configuration. As depicted in Fig. 3C, one circular beam with a density above a threshold ($1.2P_{th}$) is applied onto the perovskite metasurface. The second circular beam with $0.8P_{th}$ is pumped onto the same sample with a 10-μm lateral shift. The second beam does not produce laser emission, but it could pump an overlapping region far above the lasing threshold. In microlasers, a small change in the pumping density above a threshold can induce an appreciable variation in the gain coefficient. Consequently, the symmetry at BICs breaks, and two linearly polarized beams without OAM are generated along the radial direction (*27*).

In addition to the spatial deviation, the two-beam configuration also allows a time delay, τ (Fig. 4A), which can characterize the temporal behavior of the transition process. To accurately characterize the switching time, we introduce the parameter $K = (I_1–I_2)/I_2$ and study its dependence on τ. Here, I_1 and I_2 correspond to the intensities in the regions marked as 1 and 2, respectively, in Fig. 4A. When the nanostructure is pumped only by the first beam, the output is a uniform donut, and the normalized ratio is $K \sim 0$. Once the second beam overlaps temporally with the first beam, the optical symmetry breaks. Correspondingly, the intensity at region 2 decreases, almost vanishing (*27*) and giving the ratio of $K \sim 1$ (Fig. 4B). The switching time from a vortex lasing to a regular linearly polarized lasing is only ~1.5 ps. By changing the asymmetric pumping beam to a circular beam, the two linearly polarized beams can be switched back to the vortex lasers with a similar transition time of ~1.5 ps (Fig. 4C).

The switching on and off at the quasi-BIC mode is intrinsically different from the formation of short laser pulses. The latter is usually restricted by the build-up time and cannot repeat with high speed. To illustrate this difference, we apply the third pump beam to recover the symmetry. As depicted in Fig. 4D, the donut-lobes-donut states can be combined within a single process. Thus, the transition time is not limited. Because of a strong relationship between the output beam and the symmetry, the emission beam has only two

states: a donut and two lobes (*27*). Considering the transition time, such a binary transition is reliable for applications to all-optical switching.

The transition time of the BIC lasers is orders of magnitude faster than that previously reported for directly modulated microlasers (*27*). It is also faster than the lifetime of these BIC lasers. Such an improvement is attributed to the far-field characteristics of BICs. In principle, the BICs are formed by destructive interference at the radiation channels. The transition from BIC vortex lasers to linear lasers represents a redistribution of the laser emission instead of a direct switching of the lasing mode. Thus, we do not need to wait for the rundown of an initial laser mode, and the trade-off between high-Q values ($Q \propto \omega t$, relates to a low threshold) and high-speed operation ($\propto 1/t$) can be broken. Nonlinear all-optical switching can have a similar or even shorter transition time. But here, the exceptional gain coefficient makes the energy consumption (122 W for peak power) orders of magnitude lower. This approach also allows the removal of the fundamental limitation for microlasers, and the energy consumption can be further reduced by fully exploiting the very high Q factor at the BICs (*18*).

The high sensitivity to symmetry-breaking perturbations and the far-field characteristics of the BIC mode that are associated with exceptional optical gain make these BIC lasers all-optically controllable, with ultralow energy consumption and, simultaneously, ultrahigh speed. Therefore, breaking the traditional trade-off between low energy and high speed with the BIC lasers provides a route to develop high-speed classical and quantum communication systems.

REFERENCES AND NOTES

1. C. W. Qiu, Y. Yang, *Science* **357**, 645 (2017).
2. L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, J. P. Woerdman, *Phys. Rev. A* **45**, 8185–8189 (1992).
3. A. M. Yao, M. J. Padgett, *Adv. Opt. Photonics* **3**, 161–204 (2011).
4. S. K. Özdemir, S. Rotter, F. Nori, L. Yang, *Nat. Mater.* **18**, 783–798 (2019).
5. P. Miao *et al*., *Science* **353**, 464–467 (2016).
6. D. Stellinga *et al*., *ACS Nano* **12**, 2389–2394 (2018).
7. N. Carlon Zambon *et al*., *Nat. Photonics* **13**, 283–288 (2019).
8. J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, W. H. P. Pernice, *Nature* **569**, 208–214 (2019).
9. R. C. Devlin, A. Ambrosio, N. A. Rubin, J. P. B. Mueller, F. Capasso, *Science* **358**, 896–901 (2017).
10. X. Cai *et al*., *Science* **338**, 363–366 (2012).
11. D. Naidoo *et al*., *Nat. Photonics* **10**, 327–332 (2016).
12. L. Feng, Z. J. Wong, R. M. Ma, Y. Wang, X. Zhang, *Science* **346**, 972–975 (2014).
13. J. von Neumann, E. Wigner, *Phys. Z.* **30**, 465 (1929).
14. C. W. Hsu, B. Zhen, A. D. Stone, J. D. Joannopoulos, M. Soljačić, *Nat. Rev. Mater.* **1**, 16048 (2016).
15. M. V. Rybin *et al*., *Phys. Rev. Lett.* **119**, 243901 (2017).
16. A. Kodigala *et al*., *Nature* **541**, 196–199 (2017).
17. K. Koshelev *et al*., *Science* **367**, 288–292 (2020).
18. C. W. Hsu *et al*., *Nature* **499**, 188–191 (2013).
19. J. Jin *et al*., *Nature* **574**, 501–504 (2019).
20. B. Zhen, C. W. Hsu, L. Lu, A. D. Stone, M. Soljačić, *Phys. Rev. Lett.* **113**, 257401 (2014).
21. H. M. Doeleman, F. Monticone, W. den Hollander, A. Alù, A. F. Koenderink, *Nat. Photonics* **12**, 397–401 (2018).
22. Y. Zhang *et al*., *Phys. Rev. Lett.* **120**, 186103 (2018).
23. S. Fan, J. D. Joannopoulos, *Phys. Rev. B* **65**, 235112 (2002).
24. K. Hirose *et al*., *Nat. Photonics* **8**, 406–411 (2014).
25. Y. Yang, C. Peng, Y. Liang, Z. Li, S. Noda, *Phys. Rev. Lett.* **113**, 037401 (2014).
26. W. Liu *et al*., *Phys. Rev. Lett.* **123**, 116104 (2019).
27. Materials and methods are available as supplementary materials.
28. B. Bahari *et al*., arXiv:1707.00181 [physics.optics] (16 July 2017).
29. B. Wang *et al*., arXiv:1909.12618 [physics.optics] (27 September 2019).
30. S. T. Ha *et al*., *Nat. Nanotechnol.* **13**, 1042–1047 (2018).

ACKNOWLEDGMENTS

The authors thank B. Kanté and K. Koshelev for useful discussions and suggestions. **Funding:** This research was supported by the National Key Research and Development Program of China (grant no. SQ2018YFB220027), the Shenzhen Fundamental Research Fund (grant no. JCYJ20180507184613841), the Australian Research Council (grant no. DP200101168), and the National Science Foundation (grant no. PHY-1847240). The authors also acknowledge support from the Shenzhen Engineering Laboratory on Organic-Inorganic Perovskite Devices. **Author contributions:** Q.S. conceived the concept and supervised this research. C.H. and Y.F. performed numerical simulations and optical characterization. C.Z., C.H., Y.W., H.J., and J.H. fabricated the samples. Q.S., Y.K., S.X., C.H., and L.G. discussed the results and wrote the manuscript. All authors contributed to editing and preparing the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/1018/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S15
References (*31*–*36*)

5 December 2019; accepted 29 January 2020
10.1126/science.aba4597

ORGANIC CHEMISTRY

Aminoalkyl radicals as halogen-atom transfer agents for activation of alkyl and aryl halides

Timothée Constantin[1], Margherita Zanini[1], Alessio Regni[1], Nadeem S. Sheikh[2], Fabio Juliá[1]*, Daniele Leonori[1]*

Organic halides are important building blocks in synthesis, but their use in (photo)redox chemistry is limited by their low reduction potentials. Halogen-atom transfer remains the most reliable approach to exploit these substrates in radical processes despite its requirement for hazardous reagents and initiators such as tributyltin hydride. In this study, we demonstrate that α-aminoalkyl radicals, easily accessible from simple amines, promote the homolytic activation of carbon-halogen bonds with a reactivity profile mirroring that of classical tin radicals. This strategy conveniently engages alkyl and aryl halides in a wide range of redox transformations to construct sp^3-sp^3, sp^3-sp^2, and sp^2-sp^2 carbon-carbon bonds under mild conditions with high chemoselectivity.

Carbon radicals are versatile synthetic intermediates central to the preparation of high-value compounds (*1*, *2*). The advent of visible-light photoredox catalysis (*3*) has offered a broadly applicable radical generation strategy, transforming a variety of redox-active precursors into open-shell intermediates by single-electron transfer (SET) and fragmentation (*4*–*6*). However, photoredox activation has thus far rarely extended to organic halides, one of the largest classes of building blocks available to organic chemists. The current synthetic gap is especially evident in the case of unactivated alkyl halides, for which only dehalogenation and intramolecular cyclization of iodides have been reported (*7*–*10*). The difficulties in engaging these feedstocks in redox chemistry arise from their highly negative reduction potentials [E_{red} < –2 V versus saturated calomel electrode (SCE) for unactivated alkyl and aryl iodides], which in turn necessitate the use of strongly reducing systems (*11*, *12*) (Fig. 1A). Furthermore, the mechanisms involved in photoredox reactions are often uncertain (*9*), displaying large redox mismatches (>1 V) for SET activation and thus thwarting the exploitation of the carbon radicals accessed in this manner.

This lack of synthetic applicability stands in stark contrast to the fundamental role alkyl and aryl halides have played in the development of radical chemistry. Methods based on tin or silicon reagents and trialkylborane–O_2 systems have proven to be highly reliable in accessing carbon radicals from organic halides, generating the open-shell intermediate by homolytic carbon-halogen bond cleavage via halogen-atom transfer (XAT) (*13*–*15*). However, the toxic, hazardous nature of these reagents and initiators is problematic and has been one of the main drivers toward the identification of alternative precursors and chemical strategies for carbon radical generation. Nevertheless, silicon radicals have been recently used in metallaphotoredox catalysis to overcome sluggish carbon-halogen oxidative additions with transition metals (*16*, *17*).

We questioned whether α-aminoalkyl radicals could serve as a distinct class of halogen-abstracting reagents (Fig. 1B). Our idea for this reactivity stemmed from the fact that although classical XAT processes benefit from the formation of strong halogen-tin or halogen-silicon bonds, it is the high degree of charge transfer in the transition state that facilitates halogen-atom abstraction by these nucleophilic radicals (*18*). We therefore reasoned that strongly nucleophilic α-aminoalkyl radicals might benefit from related kinetic polar effects and manifest the same reactivity. Such radicals can be easily generated from simple amines, a class of abundant and inexpensive reagents that would offer ample opportunity for fine steric and electronic tuning.

Here we report the successful realization of this concept and its implementation as part of a mild and general strategy for the engagement of unactivated alkyl and aryl halides in redox chemistry (Fig. 1C). Because α-aminoalkyl radicals display a reactivity profile similar to that of tin radicals, their capacity to abstract iodine and bromine atoms has enabled the development of deuteration, cross-electrophile coupling, Heck-type olefination, and aromatic C–H alkylation protocols.

We initiated our study by evaluating the iodine-atom transfer reaction from cyclohexyl iodide **2** to the α-aminoalkyl radical **I-a**, derived from triethylamine (Et_3N, **1a**) (Fig. 2A).

[1]Department of Chemistry, University of Manchester, Manchester M13 9PL, UK. [2]Department of Chemistry, College of Science, King Faisal University, Al-Ahsa 31982, Saudi Arabia. *Corresponding author. Email: fabio.juliahernandez@manchester.ac.uk (F.J.); daniele.leonori@manchester.ac.uk (D.L.)

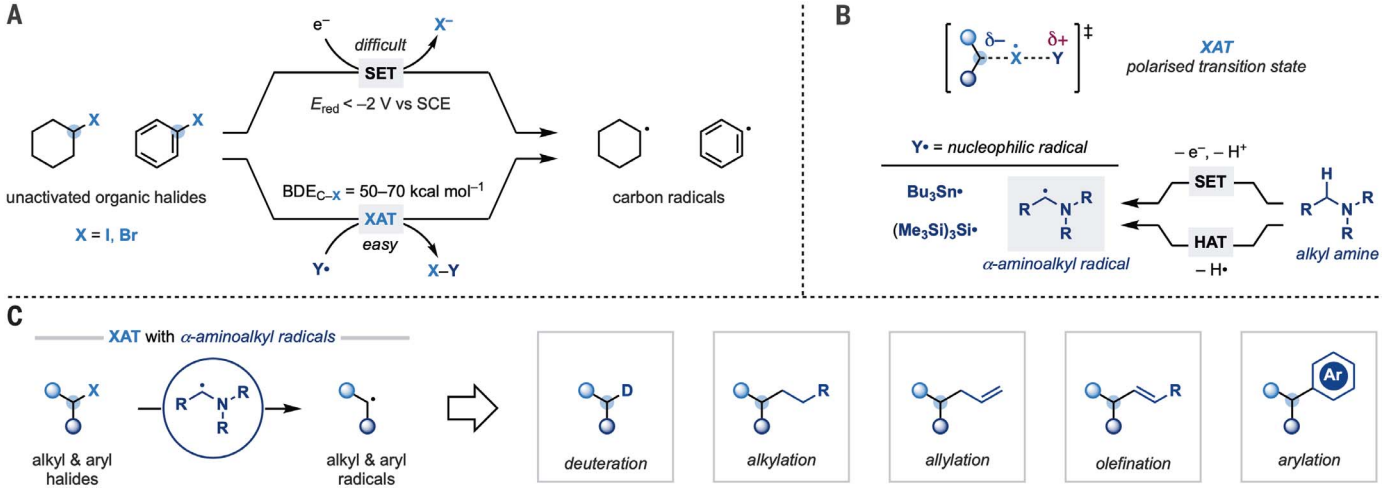
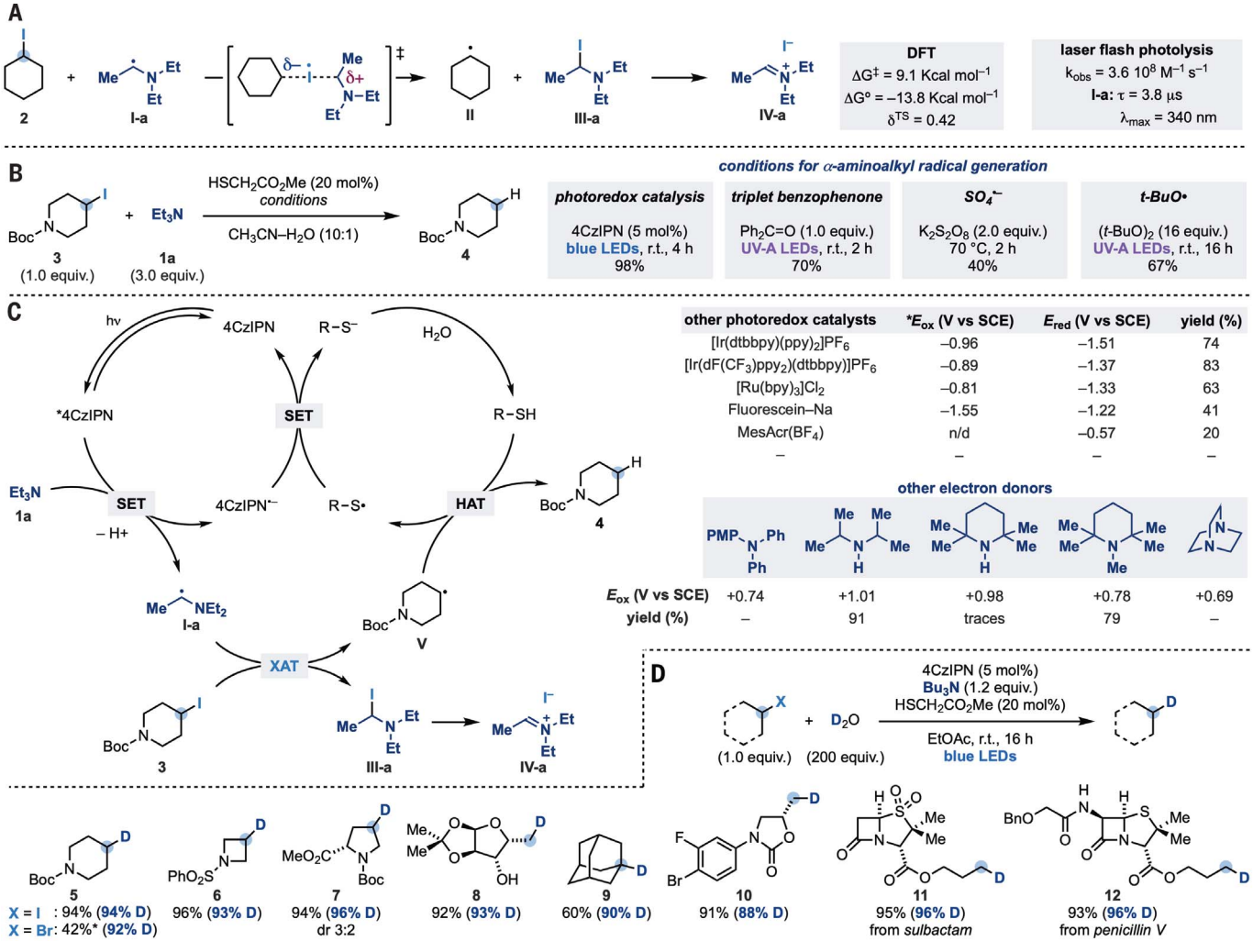


Fig. 1. Homolysis of carbon-halogen bonds by α-aminoalkyl radicals. (**A**) Activation modes for the generation of carbon radicals from alkyl and aryl halides. e⁻, electron. (**B**) Nucleophilic α-aminoalkyl radicals abstract halogen atoms (X) through polarized transition states, in analogy to tin and silicon radicals. Me, methyl; Bu, butyl; R, alkyl, aryl. (**C**) Outline of the transformations possible using alkyl and aryl halides activated via α-aminoalkyl radical-mediated XAT. Ar, aryl.

Density functional theory calculations predicted this XAT to be kinetically feasible, involving a polarized transition state with a notable charge-transfer character (δ^{TS} = 0.42), which supports the anticipated interplay of polar effects. Although the XAT is only slightly exothermic (*19*), the fast and irreversible dissociation of the resulting α-iodoamine **III-a** into the iminium iodide **IV-a** provides the thermodynamic driving force to the process. To gather direct experimental evidence, we generated and monitored **I-a** using laser flash photolysis (*20*, *21*) and observed a noticeable reactivity toward **2**. Data analysis provided a fast rate constant (k_{XAT} = 3.6 10^8 M^{-1} s^{-1}) that is only one order of magnitude slower than reported rates for I-abstraction by Bu_3Sn• and $(Me_3Si)_3Si$• (~10^9 M^{-1} s^{-1}) (*22*), showing promise for implementation in synthetic radical chemistry.

To explore the applicability of this strategy in radical reactions, we chose the dehalogenation of 4-iodo-*N*-Boc-piperidine **3**, using Et_3N as the XAT-agent precursor and methyl thioglycolate–H_2O as the H-atom donor (Fig. 2B). At the outset, we were particularly interested to evaluate whether the photochemical or thermal modes for α-aminoalkyl radical generation could be recruited for XAT reactivity. We therefore began by testing four known systems based on amine SET oxidation (Et_3N: E_{ox} = +0.77 V versus SCE) followed by deprotonation [i.e., photoredox catalysis (*23*), triplet benzophenone (*24*), and $SO_4^{\bullet-}$ (*25*)] or direct H-atom transfer (HAT) [Et_3N: α-N-C–H bond dissociation energy (BDE) = 91 kcal mol^{-1}] using *t*-BuO• (*26*). The desired product **4** was obtained in all cases in excellent to good yields, exemplifying the variety of conditions for α-aminoalkyl radical generation and ensuing XAT.

The proposed mechanism under photoredox conditions is depicted in Fig. 2C. Upon blue light irradiation, the excited organic photocatalyst 4CzIPN ($*E_{red}$ = +1.35 V versus SCE) oxidizes **1a**, which, after subsequent deprotonation, furnishes the key α-aminoalkyl radical **I-a**. This species undergoes XAT with **3**, and the resulting alkyl radical **V** provides the



other photoredox catalysts	$*E_{ox}$ (V vs SCE)	E_{red} (V vs SCE)	yield (%)
$[Ir(dtbbpy)(ppy)_2]PF_6$	–0.96	–1.51	74
$[Ir(dF(CF_3)ppy)_2(dtbbpy)]PF_6$	–0.89	–1.37	83
$[Ru(bpy)_3]Cl_2$	–0.81	–1.33	63
Fluorescein–Na	–1.55	–1.22	41
$MesAcr(BF_4)$	n/d	–0.57	20
–	–	–	–

other electron donors					
E_{ox} (V vs SCE)	+0.74	+1.01	+0.98	+0.78	+0.69
yield (%)	–	91	traces	79	–

Fig. 2. Mechanistic analysis and application to dehalogenation and deuteration reactions. (**A**) Computational [B3LYP-D3/def2-TZVP] and laser flash photolysis studies on a model XAT reaction with an alkyl iodide. Et, ethyl; DFT, density functional theory; $\Delta G^{\ddagger}$, Gibbs energy of activation; $\Delta G°$, Gibbs free energy; λ_{max}, wavelength of maximum absorption. (**B**) Evaluation of photochemical and thermal strategies for α-aminoalkyl radical generation and their use in the dehalogenation of alkyl iodide **3**. Boc, *tert*-butoxycarbonyl; LEDs, light-emitting diodes; r.t., room temperature; Ph, phenyl; UV, ultraviolet; *t*-Bu, *tert*-butyl. (**C**) Proposed mechanism for the photoredox-based dehalogenation of alkyl iodide **3**. Mechanistic studies support the intermediacy of an α-aminoalkyl radical in the activation of the C–I bond. h, Planck's constant; ν, photon frequency; dtbbpy, 4,4′-di-*tert*-butyl-2,2′-dipyridyl; ppy, 2-phenylpyridyl; dF, difluoro; bpy, 2,2′-bipyridine; Mes, mesityl; Acr, acridinium; n/d, not determined. (**D**) Application of the XAT methodology in deuteration of alkyl halides. All yields are isolated. Deuteration was determined by gas chromatography–mass spectrometry/quantitative ^{13}C nuclear magnetic resonance spectroscopy. *Tribenzylamine **1b** was used as the amine. Ac, acetyl; Bn, benzyl; dr, diastereomeric ratio.

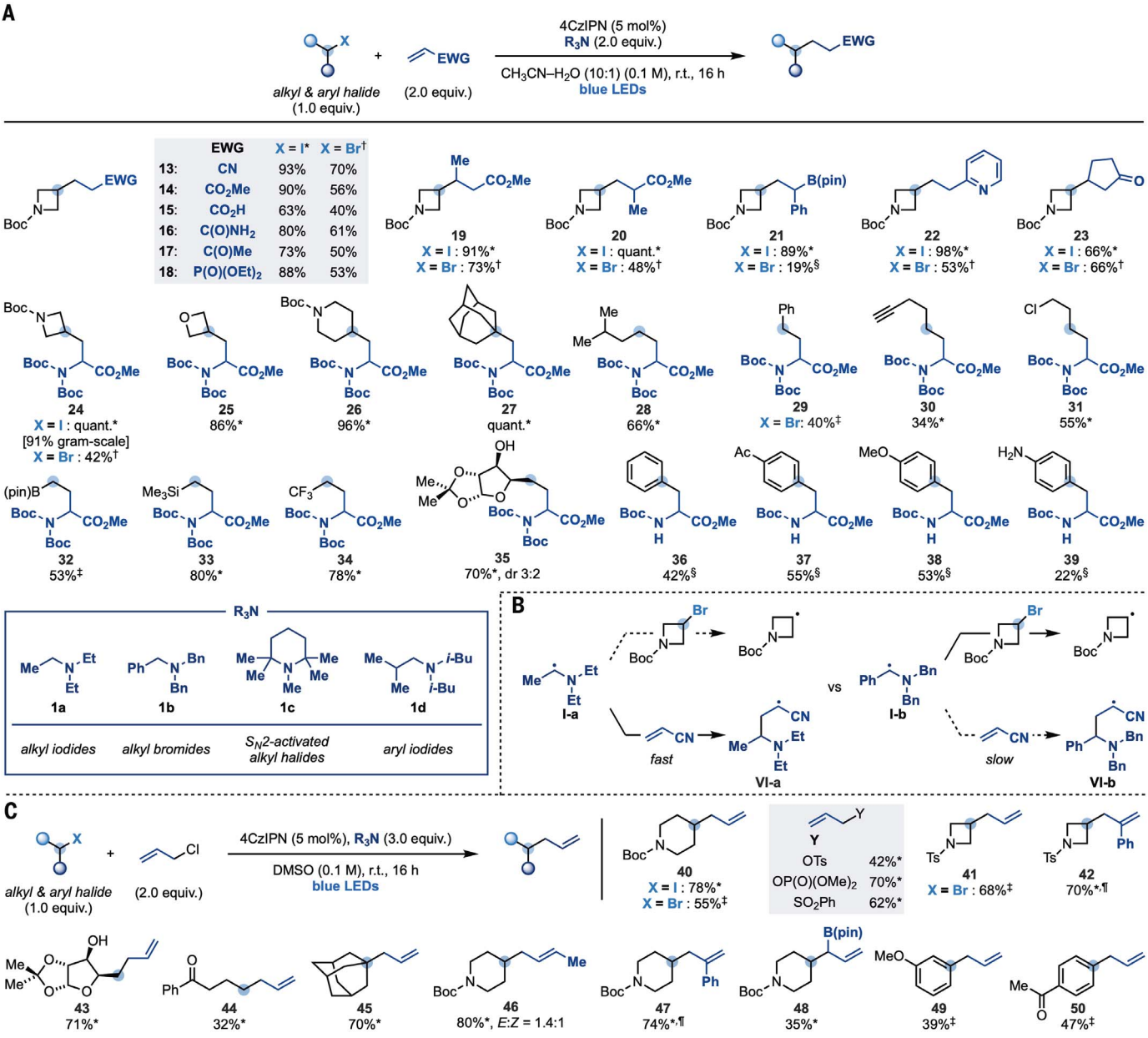


Fig. 3. Application to hydroalkylation and allylation. (**A**) Scope for the alkylation of alkyl iodides, alkyl bromides, and aryl iodides. *__1a__ was used as the amine. †__1b__ was used as the amine. ‡__1c__ was used as the amine. §__1d__ was used as the amine. EWG, electron withdrawing group; quant., quantitative; pin, pinacolato; Ac, acetyl; *i*-Bu, *iso*-butyl. (**B**) Tailoring XAT reactivity by modifying the α-aminoalkyl radical structure. (**C**) Scope for the allylation of alkyl iodides, alkyl bromides, and aryl iodides. All yields are isolated. ¶The corresponding allyl sulfone was used. Ts, tosyl.

product **4** by favorable HAT from methyl thioglycolate (S–H BDE = 87 kcal mol^{-1}). Lastly, SET between the thiyl radical and 4CzIPN$^{\bullet-}$, followed by protonation with H_2O, regenerates the thiol along with the ground-state photocatalyst. The choice of 4CzIPN and Et_3N is relevant to our mechanistic hypothesis because neither the excited nor the reduced state of the photocatalyst [$^*E_{ox}$ = −1.04 V; E_{red} = −1.21 V versus SCE (*27*)] or **I-a** [E_{ox} = −1.12 V versus SCE (*27*)] is strong enough to promote direct SET reduction of **3** (E_{red} = −2.35 V versus SCE). This means that the carbon radical generation is now dissected by the redox requirements of the system, and therefore the reductive ability of the photocatalyst is not crucial to the outcome of the reaction. Indeed, this process can be achieved with a diverse range of photocatalysts, including those of limited reductive power (e.g., Fukuzumi's acridinium; E_{red} = −0.57 V versus SCE). The replacement of Et_3N with other common electron donors [e.g., Ph_2N(PMP), sodium ascorbate, or Hantzsch ester] suppressed the reactivity, despite all compounds effectively quenching the excited photocatalyst (*19*). Moreover, other alkyl amines were tested, but only those able to generate an α-aminoalkyl radical promoted the desired reactivity (*19*). These results suggest that alkyl iodide activation via a reductive-quenching photoredox cycle is not operative and that the amine plays a fundamental role in the C–I bond cleavage that goes beyond its capacity to act as an electron donor.

The high yields obtained with the photoredox system, along with the use of H_2O as a stoichiometric H-atom source, prompted exploration of dehalogenation-deuteration reactions using D_2O (Fig. 2D). After optimization, we

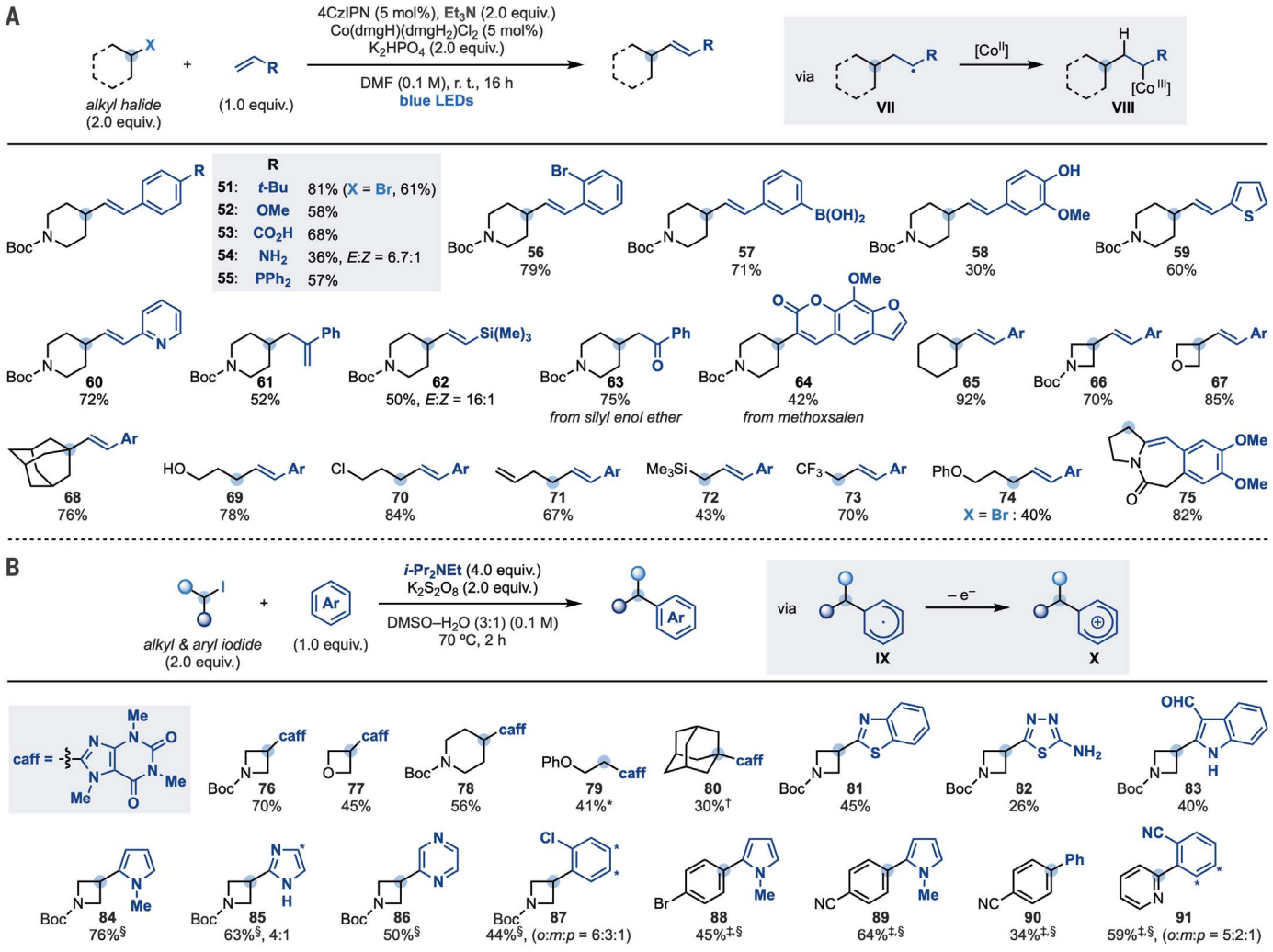


Fig. 4. Application to olefinations and arylations. (**A**) Scope for olefination of alkyl iodides and alkyl bromides. dmg, dimethylglyoximate; DMF, dimethylformamide. (**B**) Scope for the C–H alkylation and arylation of aromatics. All yields are isolated. *__1c__ was used as the amine. †Me_3N was used as the amine. ‡Bu_3N was used as the amine. §The reaction was run with 50 equiv of the arene. DMSO, dimethyl sulfoxide. Asterisks in structures indicate the position of the minor constitutional isomer.

achieved efficient deuteration of primary, secondary, and tertiary alkyl iodides in nearly quantitative yields (**5** to **12**). The mild reaction conditions tolerated multiple functional groups, showcasing the strong chemoselectivity of this XAT approach. Activation of alkyl bromides is still a challenging task in radical chemistry and is considered unfeasible using trialkylborane–O_2 systems (*28*). Our α-aminoalkyl radical–based XAT strategy is applicable to bromides, albeit in lower conversion compared with the results obtained for iodides.

The XAT strategy oxidatively generates carbon radicals from organic halides, representing an umpolung approach relative to the natural redox requirement for SET activation of these building blocks. We posited that the generated radicals could therefore be used in similar mechanistic scenarios to those involving carboxylic acids or potassium trifluoroborates, allowing their modular application in net reductive processes such as cross-electrophile couplings (*29*, *30*).

We explored this premise by developing Giese-type hydroalkylation of electron-poor olefins. Although these transformations have been performed with the aid of nickel catalysis, they typically require the use of stoichiometric metal reductants (e.g., Mn^0, Zn^0) or silane H-donors (*31*, *32*). In our case, because α-aminoalkyl radicals have been used as substrates in Giese additions (*33*), the success of this strategy hinged on their capacity to undergo XAT preferentially over their known reaction with the olefin. Exploration began with 3-iodo-*N*-Boc-azetidine in the presence of Et_3N and 4CzIPN under blue light irradiation (Fig. 3A; see fig. S10 for a proposed mechanism). A diverse range of electron-poor olefins were efficiently converted to the corresponding products in high to excellent yields (**13** to **23**). A variety of functionalities—including polar groups such as free carboxylic acid, primary amide, pyridine, and boronic ester—were readily accommodated. When the same reactions were attempted using 3-bromo-*N*-Boc-azetidine, no desired product was obtained and a substantial amount of the adduct arising from direct addition of **I-a** to the olefin acceptor was identified (Fig. 3B). In this case, XAT is slower owing to the stronger nature of the C–Br bond, thus rendering the direct Giese reaction of **I-a** with the acceptor competitive [observed rate constant k_{obs} ~ 10^7 M^{-1} s^{-1} (*21*)]. We therefore reasoned that the modulation of the electronic and steric properties of the α-aminoalkyl radical could be used to tune its reactivity. Indeed, we used tribenzylamine (**1b**) to restore XAT as the favored pathway for reactions of unactivated alkyl bromides in these hydroalkylations. Because the stabilized α-aminoalkyl

radical **I-b** was essentially unreactive toward electron-poor olefins [calculated rate constant $k_{calc} \sim 10^{-1}\ M^{-1}\ s^{-1}$ (*21*)], bromine abstraction became possible, providing the desired products in good yields.

We next explored the alkyl iodide scope by using Boc-protected dehydroalanine as an olefin acceptor, thus providing convenient access to unnatural amino acids (**24** to **35**). In this case, a variety of organyl groups bearing common functionalities (e.g., free alcohol, alkyl chloride, silane, and terminal alkyne) were compatible, reflecting the mildness of the reaction conditions. This protocol has also been carried out at gram scale without erosion in yield. The ability to generate primary alkyl radicals complements approaches that use oxalates and trifluoroborates, which are known to experience sluggish fragmentations (*34*, *35*). When alkyl halides activated toward S_N2 (second-order nucleophilic displacement) attack by Et_3N (e.g., **29** and **32**) were employed, the desired products were, unsurprisingly, obtained in low yields. This hurdle was addressed by adjusting the steric properties of the XAT reagent: Efficient couplings were achieved with the use of the bulkier amine 1,2,2,6,6-pentamethylpiperidine (PMP; **1c**). We have also been able to extend this methodology to unactivated aryl iodides by using the more hindered but less stabilized α-aminoalkyl radical derived from triisobutylamine (**1d**). These conditions enabled direct access to aryl radicals by sp^2 C–I bond cleavage and were applied to the one-pot transformation of tosylated serine into phenylalanine derivatives (**36** to **39**). Overall, these results illustrate how the large structural diversity of available tertiary amines facilitates the rational tailoring of the α-aminoalkyl radical reactivity to address different challenges in carbon-halogen bond activation.

The XAT strategy for cross-electrophile coupling is not restricted to electron-poor olefins. We also achieved efficient allylation of alkyl and aryl halides by using simple allyl chlorides and other pseudohalides (**40** to **50**) (Fig. 3C; see fig. S12 for a proposed mechanism). This approach bypasses the conventional conversion of one of the two coupling partners into a Grignard or organozinc reagent (*36*) and therefore tolerates functionalities, such as free alcohol and ketone, that are often troublesome with organometallics.

To further demonstrate the versatility of this activation mode, we sought to adapt it to target the use of alkyl halides in Heck-type olefinations, a long-standing challenge in conventional palladium catalysis, owing to undesired β-hydride elimination (*37*–*39*). Specifically, we questioned whether, after addition of alkyl radicals to suitable olefins (**VII**), a cobaloxime cocatalyst might trigger a dehydrogenation reaction (*40*), thus leading to sp^3-sp^2 C–C bond formation (via **VIII**) without the need for precious metals (see fig. S14 for a proposed mechanism). As shown in Fig. 4A, we found this dual XAT–[Co] protocol feasible, thus allowing direct olefination of primary, secondary, and tertiary alkyl iodides and bromides exclusively as the *E* isomers (**51** to **74**, with the exception of **54** and **62**). The broad functional group compatibility was demonstrated with the successful engagement of substrates containing phenol, aniline, and benzoic acid moieties, as well as aryl bromide, boronic acid, and phosphine groups that could limit application under transition metal catalysis. The olefination was also very effective in intramolecular settings, as showcased by the construction of tricyclic **75** in good yield. Couplings with aryl iodides were attempted but generally resulted in low yields.

In a final effort to establish the generality of this XAT strategy, we turned our attention to the direct aromatic C–H alkylation via radical intermediates (Fig. 4B; see fig. S15 for a proposed mechanism). Recently, the use of zinc alkylsulfinates has provided a powerful and effective solution to this synthetic challenge (*41*, *42*). Because these reagents are often prepared from the corresponding halides, a methodology that directly uses these building blocks would obviate multistep synthesis of any reactive intermediate. In this case, however, a photoredox system for α-aminoalkyl radical generation is difficult to implement, owing to the mechanistic requirement of a second oxidation after radical addition to the arene to allow rearomatization (**IX**→**X**). The broad set of reactivity modes for α-aminoalkyl radical generation enabled identification of simple thermal, net oxidative conditions for the direct alkylation of caffeine with alkyl iodides, without the need for light or catalysts (**76** to **80**). This manifold for aromatic C–H alkylation was compatible with the installation of primary, secondary, and tertiary alkyl groups and could be extended to other heteroarenes commonly found in bioactive molecules, such as indoles and azoles as well as benzenoids (*43*) (**81** to **87**). Furthermore, we demonstrated that aryl iodide activation and subsequent sp^2-sp^2 coupling (*44*) is also possible, as shown by the successful preparation of **88** to **91**.

The results presented here demonstrate that alkyl and aryl halides can be converted to carbon radicals by XAT using α-aminoalkyl radicals. We believe that the broad scope, functional group tolerance, and modularity of this approach for carbon-halogen bond activation will likely be of great utility to chemists in both academia and industry.

REFERENCES AND NOTES

1. J. M. Smith, S. J. Harwood, P. S. Baran, *Acc. Chem. Res.* **51**, 1807–1817 (2018).
2. S. P. Pitre, N. A. Weires, L. E. Overman, *J. Am. Chem. Soc.* **141**, 2800–2813 (2019).
3. C. K. Prier, D. A. Rankic, D. W. C. MacMillan, *Chem. Rev.* **113**, 5322–5363 (2013).
4. J. Xuan, Z.-G. Zhang, W.-J. Xiao, *Angew. Chem. Int. Ed.* **54**, 15632–15641 (2015).
5. J. K. Matsui, S. B. Lang, D. R. Heitz, G. A. Molander, *ACS Catal.* **7**, 2563–2575 (2017).
6. A. Togni *et al*., *Angew. Chem. Int. Ed.* 10.1002/anie.201911660 (2019).
7. J. D. Nguyen, E. M. D'Amato, J. M. R. Narayanam, C. R. J. Stephenson, *Nat. Chem.* **4**, 854–859 (2012).
8. H. Kim, C. Lee, *Angew. Chem. Int. Ed.* **51**, 12303–12306 (2012).
9. Y. Shen, J. Cornella, F. Juliá-Hernández, R. Martin, *ACS Catal.* **7**, 409–412 (2017).
10. D. Alpers, M. Gallhof, J. Witt, F. Hoffmann, M. Brasholz, *Angew. Chem. Int. Ed.* **56**, 1402–1406 (2017).
11. J. L. Kuo, C. Lorenc, J. M. Abuyuan, J. R. Norton, *J. Am. Chem. Soc.* **140**, 4512–4516 (2018).
12. G. Nocera *et al*., *J. Am. Chem. Soc.* **140**, 9751–9757 (2018).
13. W. P. Neumann, *Synthesis* **1987**, 665–683 (1987).
14. C. Chatgilialoglu, C. Ferreri, Y. Landais, V. I. Timokhin, *Chem. Rev.* **118**, 6516–6572 (2018).
15. H. Yorimitsu, K. Oshima, in *Radicals in Organic Synthesis*, P. Renaud, M. P. Sibi, Eds. (Wiley, 2008), pp. 11–27.
16. C. Le, T. Q. Chen, T. Liang, P. Zhang, D. W. C. MacMillan, *Science* **360**, 1010–1014 (2018).
17. P. Zhang, C. C. Le, D. W. C. MacMillan, *J. Am. Chem. Soc.* **138**, 8084–8087 (2016).
18. W. H. Tamblyn, E. A. Vogler, J. K. Kochi, *J. Org. Chem.* **45**, 3912–3915 (1980).
19. See the supplementary materials for more information.
20. J. C. Scaiano, in *Reactive Intermediate Chemistry*, R. A. Moss, M. S. Platz, M. Jones, Eds. (Wiley, 2005), pp. 847–871.
21. J. Lalevée, B. Graff, X. Allonas, J. P. Fouassier, *J. Phys. Chem. A* **111**, 6991–6998 (2007).
22. K. U. Ingold, J. Lusztyk, J. C. Scaiano, *J. Am. Chem. Soc.* **106**, 343–348 (1984).
23. K. Nakajima, Y. Miyake, Y. Nishibayashi, *Acc. Chem. Res.* **49**, 1946–1956 (2016).
24. S. G. Cohen, A. Parola, G. H. Parsons, *Chem. Rev.* **73**, 141–161 (1973).
25. J. K. Laha, K. S. S. Tummalapalli, A. Nair, N. Patel, *J. Org. Chem.* **80**, 11351–11359 (2015).
26. D. D. M. Wayner, J. J. Dannenberg, D. Griller, *Chem. Phys. Lett.* **131**, 189–191 (1986).
27. J. Luo, J. Zhang, *ACS Catal.* **6**, 873–877 (2016).
28. V. Soulard, G. Villa, D. P. Vollmar, P. Renaud, *J. Am. Chem. Soc.* **140**, 155–158 (2018).
29. D. A. Everson, D. J. Weix, *J. Org. Chem.* **79**, 4793–4798 (2014).
30. E. Richmond, J. Moran, *Synthesis* **50**, 499–513 (2018).
31. K. M. M. Huihui, R. Shrestha, D. J. Weix, *Org. Lett.* **19**, 340–343 (2017).
32. F. Zhou, J. Zhu, Y. Zhang, S. Zhu, *Angew. Chem. Int. Ed.* **57**, 4058–4062 (2018).
33. R. A. Aycock, C. J. Pratt, N. T. Jui, *ACS Catal.* **8**, 9115–9119 (2018).
34. C. C. Nawrat, C. R. Jamison, Y. Slutskyy, D. W. C. MacMillan, L. E. Overman, *J. Am. Chem. Soc.* **137**, 11270–11273 (2015).
35. C. Lévêque, L. Chenneberg, V. Corcé, C. Ollivier, L. Fensterbank, *Chem. Commun.* **52**, 9877–9880 (2016).
36. H. Yorimitsu, K. Oshima, *Angew. Chem. Int. Ed.* **44**, 4435–4439 (2005).
37. M. R. Netherton, G. C. Fu, in *Palladium in Organic Synthesis*, J. Tsuji, Ed. (Springer, 2005), pp. 85–108.
38. G.-Z. Wang, R. Shang, W.-M. Cheng, Y. Fu, *J. Am. Chem. Soc.* **139**, 18307–18312 (2017).
39. D. Kurandina, M. Parasram, V. Gevorgyan, *Angew. Chem. Int. Ed.* **56**, 14212–14216 (2017).
40. X. Sun, J. Chen, T. Ritter, *Nat. Chem.* **10**, 1229–1233 (2018).
41. Y. Fujiwara *et al*., *Nature* **492**, 95–99 (2012).
42. R. Gianatassio *et al*., *Angew. Chem. Int. Ed.* **53**, 9851–9855 (2014).
43. Z. Jiao, L. H. Lim, H. Hirao, J. S. Zhou, *Angew. Chem. Int. Ed.* **57**, 6294–6298 (2018).
44. I. Ghosh, T. Ghosh, J. I. Bardagi, B. König, *Science* **346**, 725–728 (2014).

ACKNOWLEDGMENTS

We thank D. Heyes for help with laser flash photolysis studies. **Funding:** D.L. thanks EPSRC for a Fellowship (EP/P004997/1) and the European Research Council for a research grant (758427). **Author contributions:** F.J. and D.L. designed the project and wrote the manuscript. T.C., M.Z., A.R., and F.J. performed all experiments. N.S.S. performed the computational studies. All authors analyzed the results. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/1021/suppl/DC1
Materials and Methods
Figs. S1 to S24
Tables S1 to S16
NMR Spectra
References (*45*–*101*)

16 November 2019; accepted 31 January 2020
10.1126/science.aba2419

FLUID DYNAMICS

Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations

Maziar Raissi[1,2]*†, Alireza Yazdani[1], George Em Karniadakis[1]†

For centuries, flow visualization has been the art of making fluid motion visible in physical and biological systems. Although such flow patterns can be, in principle, described by the Navier-Stokes equations, extracting the velocity and pressure fields directly from the images is challenging. We addressed this problem by developing hidden fluid mechanics (HFM), a physics-informed deep-learning framework capable of encoding the Navier-Stokes equations into the neural networks while being agnostic to the geometry or the initial and boundary conditions. We demonstrate HFM for several physical and biomedical problems by extracting quantitative information for which direct measurements may not be possible. HFM is robust to low resolution and substantial noise in the observation data, which is important for potential applications.

Quantifying the flow dynamics in geophysical, living, and engineering systems requires detailed knowledge of velocity and pressure fields and has been the centerpiece of experimental and theoretical fluid mechanics for centuries. Available experimental techniques include point measurements (such as a hot wire anemometer or pitot tube) as well as smoke or dye visualization for qualitative characterization of entire flow fields. There are also quantitative flow visualizations such as particle image velocimetry and magnetic resonance imaging (*1*–*3*), but these are limited to small domains and laboratory settings. Furthermore, although experimental measurements of external flows (such as flow past a bluff object) are obtained relatively easily, albeit in small subdomains, the quantification of velocity fields for internal flows (such as blood flow in the vasculature) is very difficult or impractical. Despite substantial advances in experimental fluid mechanics, the use of measurements to reliably infer fluid velocity and pressure or stress fields is not a straightforward task. From the theoretical standpoint, the governing equations of fluid mechanics have been derived from conservation laws (conservation of mass, momentum, and energy), leading to partial differential equations (PDEs) such as the well-known Navier-Stokes (NS) equations (*4*). Although accurate solutions of such equations in a "forward" setting are now available by using direct numerical simulations or other approximate forms, the computational cost is prohibitively high for realistic conditions and "inverse" problems. Here, we address the question of leveraging the underlying laws of physics to extract quantitative information from available flow visualizations of passive scalars, such as the transport of dye or smoke in physical systems and contrast agents in biological systems. Data assimilation techniques have been used mostly in geophysics but rely on discrete point measurements rather than images of flow visualizations. We developed an alternative approach, which we call hidden fluid mechanics (HFM), that simultaneously exploits the information available in snapshots of flow visualizations and the NS equations, combined in the context of physics-informed deep learning (*5*) by using automatic differentiation. In mathematics, statistics, and computer science—in particular, in machine learning and inverse problems—regularization is the process of adding information in order to prevent overfitting or to solve an ill-posed problem. The prior knowledge of the NS equations introduces important structure that effectively regularizes the minimization procedure in the training of neural networks.

[1]Division of Applied Mathematics, Brown University, Providence, RI 02906, USA. [2]NVIDIA Corporation, Santa Clara, CA 95051, USA.
*Present address: Department of Applied Mathematics, University of Colorado, Boulder, CO 80309, USA.
†Corresponding author. Email: maziar.raissi@colorado.edu (M.R.); george_karniadakis@brown.edu (G.E.K.)

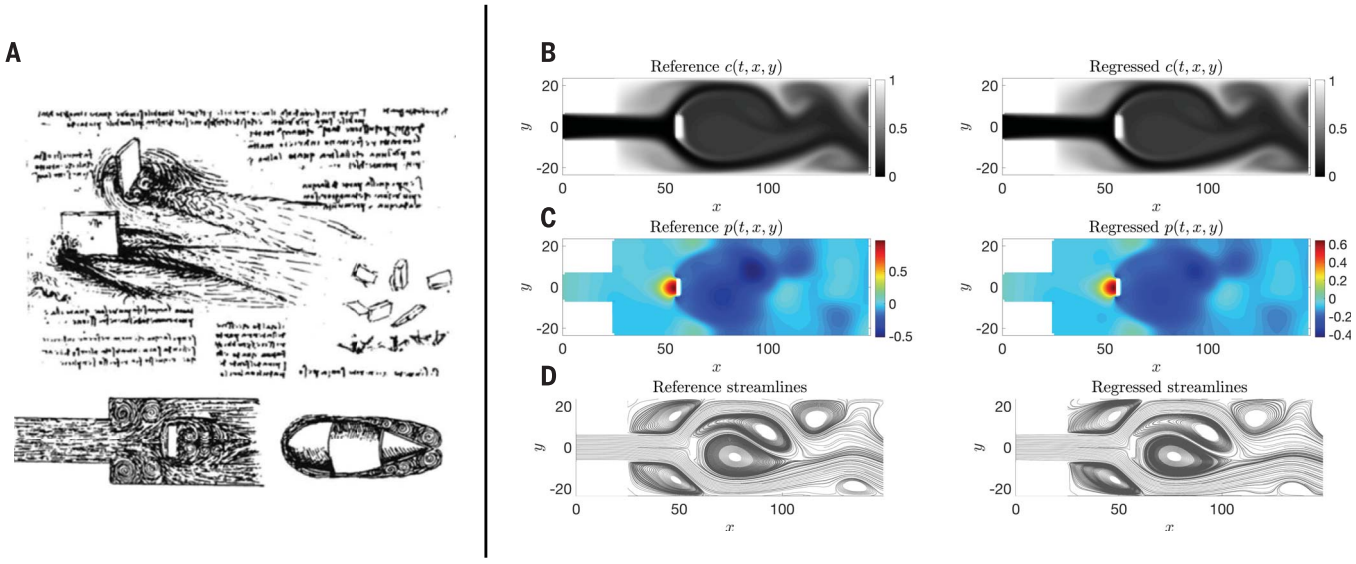


Fig. 1. Quantifying flow visualizations. (**A**) Leonardo da Vinci's scientific artistry led him to draw accurate patterns of eddies and vortices for various flow problems. [Reprinted from figure 1.4 and figure 1.5 of (*17*) with permission from Elsevier.] (**B** to **D**) We used HFM to quantify the velocity and pressure fields in a geometry similar to the drawing in the lower left corner of (A); our input was a point cloud of data on the concentration field $c(t, x, y)$ shown in (B), left. In (B) to (D), left, we plot the reference concentration, pressure, and streamlines, while at right, we plot the corresponding regressed quantities of interest produced by the algorithm. [(C) and (D)] Hidden states of the system—pressure $p(t, x, y)$ and velocity fields—obtained by using HFM based on the data on the concentration field, randomly scattered in time and space. (D) A comparison of the reference (left) and regressed (right) instantaneous streamlines. The streamlines are computed by using the velocity fields.

For example, using several snapshots of concentration fields (inspired by the drawings of da Vinci in Fig. 1A), we obtained quantitatively the velocity and pressure fields (Fig. 1, B to D).

We considered the transport of a passive scalar $c(t, x, y, z)$ by a velocity field $\boldsymbol{u}(t, x, y, z) = [u(t, x, y, z), v(t, x, y, z), w(t, x, y, z)]$, which satisfies the incompressible NS equations. The passive scalar is advected by the flow and diffused but has no dynamical effect on the fluid motion itself. Smoke and dye are two typical examples of passive scalars. In this work, we assumed that the only observables are noisy data $\{t^n, x^n, y^n, z^n, c^n\}_{n=1}^N$ on the concentration $c(t, x, y, z)$ of the passive scalar (Fig. 2B). This set of time-space coordinates represents a single point cloud of scattered data consisting of N data points (t^n, x^n, y^n, z^n) and their corresponding labels c^n, the measured concentration value at the point (t^n, x^n, y^n, z^n). Here, the superscript n denotes the nth data point and runs from 1 to N. Given such data, scattered in space and time, we were interested in inferring the latent (hidden) quantities of interest $u(t, x, y, z)$, $v(t, x, y, z)$, and $w(t, x, y, z)$ and pressure $p(t, x, y, z)$. We aimed to develop a flexible framework that could deal with data acquired in arbitrarily complex domains such as flow around vehicles or blood flow in brain or aortic aneurysms. We approximated the function $(t, x, y, z) \mapsto (c, u, v, w, p)$ by means of a physics-uninformed deep neural network, which was followed by a physics-informed deep neural network $(t, x, y, z) \mapsto (e_1, e_2, e_3, e_4, e_5)$, in which the coupled dynamics of the passive scalar and the NS equations were encoded in the outputs e_1, e_2, e_3, e_4, and e_5 by using automatic differentiation (Fig. 3C and fig. S1). Here, e_1 is the residual of the transport equation modeling the dynamics of the passive scalar, and e_2, e_3, and e_4 represent the momentum equations in x, y, and z directions, respectively. Moreover, e_5 corresponds to the residual of the continuity equation. In the following, we minimize the norms of these residuals to satisfy the corresponding equations that describe the underlying laws of fluid mechanics. The shared parameters of the physics-uninformed neural networks for c, u, v, w, and p and the physics-informed ones e_1, e_2, e_3, e_4, and e_5 can be learned by minimizing the following mean squared error loss function

$$MSE = \frac{1}{N}\sum_{n=1}^{N}\left|c(t^n, x^n, y^n, z^n) - c^n\right|^2 + \sum_{i=1}^{5}\frac{1}{M}\sum_{m=1}^{M}\left|e_i(t^m, x^m, y^m, z^m)\right|^2 \quad (1)$$

where the first term corresponds to the training data $\{t^n, x^n, y^n, z^n, c^n\}_{n=1}^N$ on the concentration of the passive scalar, whereas the last term enforces the structure imposed by the NS and transport equations at a finite set of residual points $\{t^m, x^m, y^m, z^m\}_{m=1}^M$ whose number and locations can be different from the actual training data. The number and locations of these points at which we penalize the equations are in our full control, whereas the data on the concentration of the passive scalar are available at the measurement points. Mini-batch gradient descent algorithms and their modern variants such as the Adam

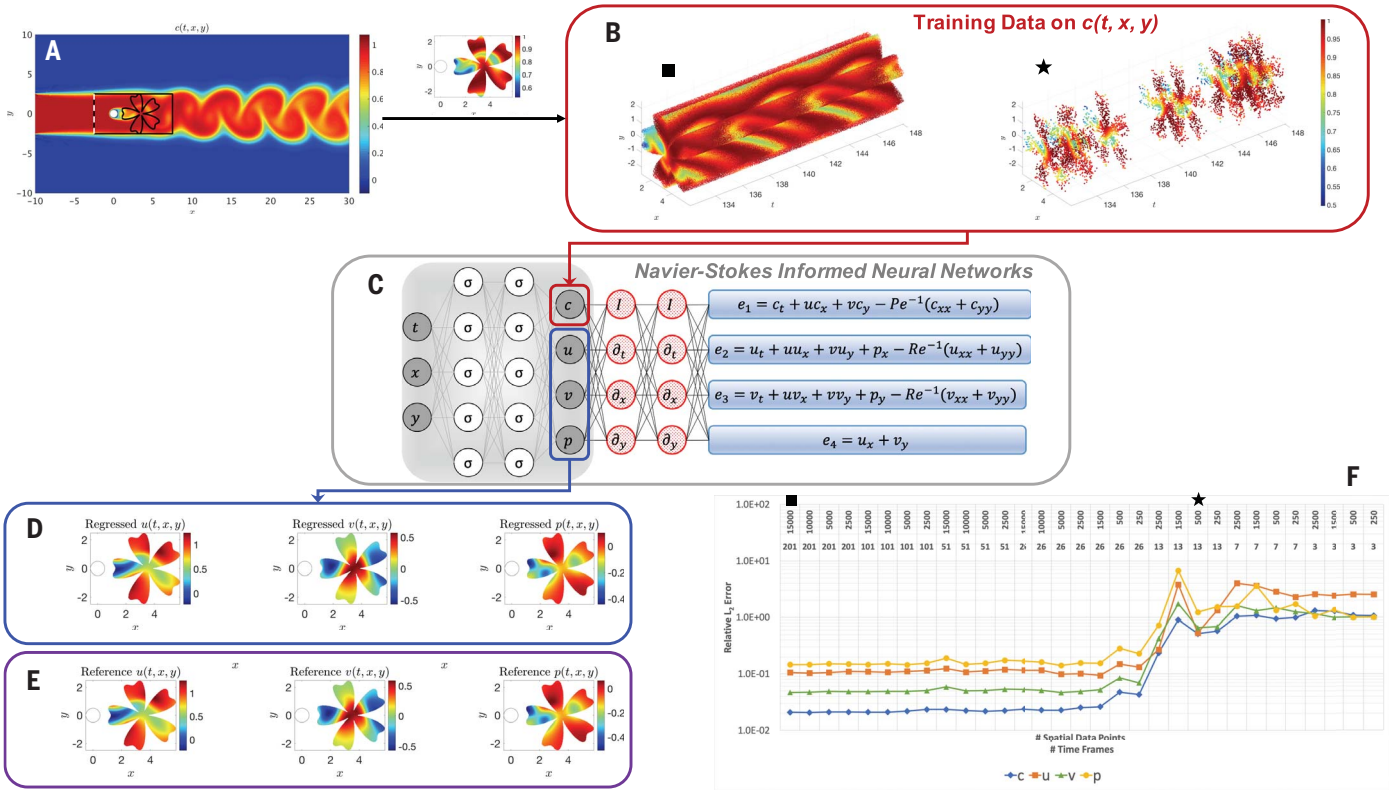


Fig. 2. Arbitrary training domain in the wake of a cylinder. (**A**) Domain where the training data for concentration and reference data for the velocity and pressure are generated by using direct numerical simulation. (**B**) Training data on concentration $c(t, x, y)$ in an arbitrary domain in the shape of a flower located in the wake of the cylinder. The solid black square corresponds to a very refined point cloud of data, whereas the solid black star corresponds to a low-resolution point cloud. (**C**) A physics-uninformed neural network (left) takes the input variables t, x, and y and outputs c, u, v, and p. By applying automatic differentiation on the output variables, we encode the transport and NS equations in the physics-informed neural networks e_i, i = 1, ..., 4 (right). (**D**) Velocity and pressure fields regressed by means of HFM. (**E**) Reference velocity and pressure fields obtained by cutting out the arbitrary domain in (A), used for testing the performance of HFM. (**F**) Relative L_2 errors estimated for various spatiotemporal resolutions of observations for c. On the top line, we list the spatial resolution for each case, and on the line below, we list the corresponding temporal resolution over 2.5 vortex shedding cycles.

optimizer enable us to penalize the equations at virtually "infinitely" many points. Moreover, in tables S2 and S3, we demonstrate that in addition to the velocity and pressure fields, it is possible to discover other unknown parameters of the flow, such as the Reynolds and Péclet numbers, directly from the data on concentration of the passive scalar.

We first considered external flows, starting with the prototypical problem of a two-dimensional (2D) flow past a circular cylinder at Reynolds number Re = 100 and Péclet number for the passive scalar Pe = 100 (Fig. 2). We have performed a direct numerical simulation [using the spectral element method (*6*)] to generate the training data and also the reference velocity and pressure fields in order to investigate the accuracy of HFM. The passive scalar is injected at the inlet on the left boundary. As illustrated in Fig. 2, the shape and extent of the boundaries of the training domains that we chose for our analysis could be arbitrary. However, there are two important factors that need to be considered when choosing the training domain. First, the concentration field of the passive scalar must be present within the training domain so that its information can be used to infer other flow variables. Second, to avoid the need for specifying appropriate boundary conditions for velocities, there must be sufficient gradients of concentration normal to the boundaries ($\partial c/\partial n \neq 0$) in order for the method to be able to infer a single solution for the velocity field. In regions of the training domain in which the concentration profile alone does not carry sufficient information to guarantee a single velocity or pressure field, one could provide the algorithm with additional information such as data on the velocities or pressure (for example, the no-slip boundary condition imposed for velocity on the wall boundaries).

However, in all of the examples considered in this work except for one (figs. S8 and S9), we have relied solely on the information encapsulated within the data on the concentration of the passive scalar. As shown in Fig. 2, good quantitative agreement can be achieved between the predictions of the algorithm and the reference data within a completely arbitrary training domain downstream of the cylinder. The input to the algorithm is essentially a point cloud of data on the concentration of the passive scalar, scattered in space and time (Fig. 2B). We performed a systematic study (Fig. 2F and fig. S5) with respect to the spatiotemporal resolution of the training data for the concentration profile of the passive scalar and verified that the proposed algorithm is very robust with respect to the spatiotemporal resolution of the point cloud of data. Specifically, the algorithm breaks down if for training there are fewer than five time snapshots per vortex shedding cycle or fewer than 250 points in the spatial domain. More details for this case—including quantifying the fluid forces on the cylinder, a benchmark problem for 3D external flow past a circular cylinder, and specifically an analysis of HFM robustness to significant noise levels in the concentration data—are given in figs. S3 to S14. Moreover, we demonstrate the effectiveness of HFM in using information from streaklines (fig. S11 and table S4) and show its robustness irrespective of the boundary

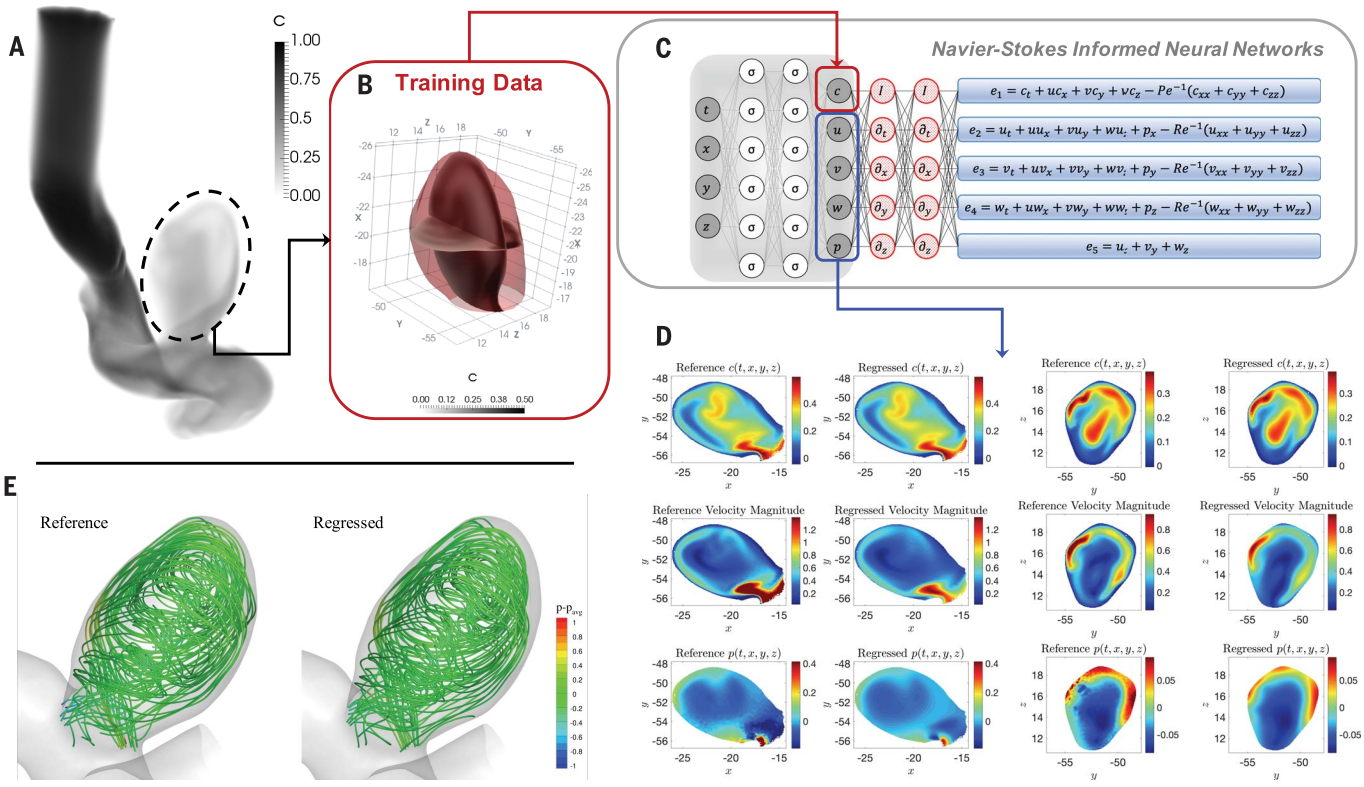


Fig. 3. Inferring quantitative hemodynamics in a 3D intracranial aneurysm. (**A**) Domain (right internal carotid artery with an aneurysm) where the training data for concentration and reference data for the velocity and pressure are generated by using a direct numerical simulation. (**B**) The training domain containing only the ICA sac is shown, in which two perpendicular planes have been used to interpolate the reference data and the predicted outputs for plotting 2D contours. (**C**) Schematic of the NS-informed neural networks, which take $c(t, x, y, z)$ data as the input and infer the velocity and pressure fields. (**D**) Contours of instantaneous reference and regressed fields plotted on two perpendicular planes for concentration c, velocity magnitude, and pressure p in each row. The first two columns show the results interpolated on a plane perpendicular to the z axis, and the next two columns are plotted for a plane perpendicular to the y axis. (**E**) Flow streamlines are computed from the reference and regressed velocity fields colored by the pressure field. The range of contour levels is the same for all fields for better comparison [movies S1 and S2, which correspond to (A) and (E)].

conditions imposed at the cylinder wall (c = 0 or $\partial c/\partial n$ = 0) to generate the training data.

Next, we focused on internal flows, first demonstrating the effectiveness of HFM for a 2D channel with a stenosis for which we aimed to infer the wall shear stresses (figs. S15 to S18). We studied a realistic biomedical application of HFM for 3D physiologic blood flow in a patient-specific intracranial aneurysm (ICA) (Fig. 3). The aneurysm was located in the cavernous segment of the right internal carotid artery at the level of the eye and beneath the brain (*7*). Exact concentration fields were generated numerically by using patient-specific boundary conditions, a physiologic flow waveform at the inlet along with a uniform concentration for the passive scalar. Because of the characteristics of our algorithm, we could focus only on the regions where the velocity and pressure fields were needed, reducing substantially the size of data and the cost of training. We first cropped the aneurysm sac out from the rest of geometry and then used only the passive scalar data within the ICA sac (Fig. 3B) for training, whereas no information was used for the boundary conditions. The reference and regressed concentration, velocity, and pressure fields within the ICA sac at a sample time instant were then projected on two separate planes perpendicular to the y and z axes (Fig. 3D). We observed very good agreement between the reference and regressed fields, given the complexity of the flow field. More details on this case, including the predictions for wall shear stress components on the wall of the ICA sac, are given in figs. S19 to S21.

The algorithm we developed is agnostic to the geometry, initial, and boundary conditions, hence providing flexibility in choosing the domain of interest for data acquisition as well as subsequent training and predictions. Moreover, the current methodology allows us to construct computationally efficient and fully differentiable surrogates for velocity and pressure fields that can be further used to estimate other quantities of interest, such as shear stresses and vorticity fields. Transport of scalar fields in fluid flow has been studied in many applications, such as aerodynamics, biofluid mechanics, and nonreactive flow mixing, to name a few. The use of smoke in wind tunnels or dye in water tunnels for flow visualization and quantification has long been practiced in experimental fluid mechanics (*8*). Moreover, recent techniques in planar laser-induced fluorescence imaging combined with particle image velocimetry have been developed to assess the relationships between scalar and velocity-vorticity fields (*9*, *10*). The use of scalar transport in conjunction with advanced imaging modalities to quantify blood flow in the vascular networks is now a common practice. For example, coronary computed tomography (CT) angiography is typically performed on multidetector CT systems after the injection of a nondiffusible iodine contrast agent, which allows coronary artery visualization and the detection of coronary stenoses (*11*). Another example is the quantification of cerebral blood flow, either in the prognostic assessment of strokes in patients by using a contrast agent and perfusion CT (*12*) or in cognitive neuroscience with the use of functional magnetic resonance imaging that only relies on the blood-oxygen-level–dependent contrast to measure brain activity (*13*). As demonstrated in this work, a direct implication of the current method is quantifying the hemodynamics in the vasculature. This could potentially have a substantial impact on the clinical diagnosis (especially with the noninvasive methods) of vascular diseases associated with important pathologies such as heart attack and stroke. Blood flow shear stresses acting on the vascular wall are crucial in the prognosis of a vascular disease, and their quantification is clinically important (*14*, *15*). Using the proposed method, it is possible to estimate the wall shear stresses at no extra cost. This will simplify the complexities of the state-of-the-art methods in which extracting the exact boundaries of the vessels from the clinical images is required (*16*).

Our framework is general and can be extended to other disciplines; for example, in electromagnetics, when given data on the electric field and knowing the Maxwell's equations, we can infer the magnetic field. We have also verified the robustness of HFM to low resolution and substantial noise in the observed concentration fields (Fig. 2 and figs. S5 and S6), which suggests that HFM may find applications in engineering and biomedicine.

REFERENCES AND NOTES

1. J. Westerweel, D. Dabiri, M. Gharib, *Exp. Fluids* **23**, 20–28 (1997).
2. F. Pereira, J. Lu, E. Castano-Graff, M. Gharib, *Exp. Fluids* **42**, 589–599 (2007).
3. M. Shattuck, R. Behringer, G. Johnson, J. Georgiadis, *J. Fluid Mech.* **332**, 215–245 (1997).
4. G. K. Batchelor, *An Introduction to Fluid Dynamics* (Cambridge Univ. Press, 2000).
5. M. Raissi, P. Perdikaris, G. E. Karniadakis, *J. Comput. Phys.* **378**, 686–707 (2019).
6. G. E. Karniadakis, S. Sherwin, *Spectral/hp Element Methods for Computational Fluid Dynamics* (Oxford Univ. Press, ed. 2, 2013).
7. H. Baek, G. E. Karniadakis, *J. Comput. Phys.* **231**, 629–652 (2012).
8. J. B. Barlow, W. H. Rae Jr., A. Pope, *INCAS Bull.* **7**, 133 (2015).
9. M. Koochesfahani, P. Dimotakis, *AIAA J.* **23**, 1700–1707 (1985).
10. J. Crimaldi, *Exp. Fluids* **44**, 851–863 (2008).
11. S. Voros *et al*., *JACC Cardiovasc. Imaging* **4**, 537–548 (2011).
12. M. Wintermark *et al*., *Ann. Neurol.* **51**, 417–432 (2002).
13. A. R. Aron, T. E. Behrens, S. Smith, M. J. Frank, R. A. Poldrack, *J. Neurosci.* **27**, 3743–3752 (2007).
14. A. M. Shaaban, A. J. Duerinckx, *AJR Am. J. Roentgenol.* **174**, 1657–1665 (2000).
15. C. K. Zarins *et al*., *Circ. Res.* **53**, 502–514 (1983).
16. L. Boussel *et al*., *Magn. Reson. Med.* **61**, 409–417 (2009).
17. Y. Nakayama, *Introduction to Fluid Mechanics* (Butterworth-Heinemann, 2018).
18. M. Raissi, maziarraissi/HFM: Hidden fluid mechanics (version v1.0). Zenodo (2019); https://doi.org/10.5281/zenodo.3566161.
19. A. Yazdani, alirezayazdani1/HFM: HFM - Synthetic data generators (version v1.0). Zenodo (2019); https://doi.org/10.5281/zenodo.3567215.

ACKNOWLEDGMENTS

We thank the anonymous referees for their very helpful suggestions. **Funding:** This work received support from the Defense Advanced Research Projects Agency (DARPA) Enabling Quantification of Uncertainty in Physical Systems (EQUiPS) grant N66001-15-2-4055 and DARPA– Artificial Intelligence Research Associate (AIRA) grant HR00111990025, the Air Force Office of Scientific Research (AFOSR) grant FA9550-17-1-0013, the NIH grant U01HL142518, and the U.S. Department of Energy grant DE-AC05-76RL01830. **Author contributions:** M.R., A.Y., and G.E.K. selected the problems; M.R. conceptualized the methodology and developed the software; M.R. and A.Y. wrote the manuscript, curated the data, and visualized the results; M.R., A.Y., and G.E.K. reviewed and edited the manuscript; and G.E.K. acquired the funds and supervised the project. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data and codes used in this manuscript are in the supplementary materials and are publicly available on GitHub (*18*, *19*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/1026/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S21
Tables S1 to S4
References (*20*–*37*)
Movies S1 and S2

22 December 2018; accepted 21 January 2020
Published online 30 January 2020
10.1126/science.aaw4741

BATTERIES

Kinetic pathways of ionic transport in fast-charging lithium titanate

Wei Zhang[1]*†, Dong-Hwa Seo[2]*‡, Tina Chen[2,3]*, Lijun Wu[4], Mehmet Topsakal[5], Yimei Zhu[4], Deyu Lu[5], Gerbrand Ceder[2,3]§, Feng Wang[1]§

Fast-charging batteries typically use electrodes capable of accommodating lithium continuously by means of solid-solution transformation because they have few kinetic barriers apart from ionic diffusion. One exception is lithium titanate ($Li_4Ti_5O_{12}$), an anode exhibiting extraordinary rate capability apparently inconsistent with its two-phase reaction and slow Li diffusion in both phases. Through real-time tracking of Li^+ migration using operando electron energy-loss spectroscopy, we reveal that facile transport in $Li_{4+x}Ti_5O_{12}$ is enabled by kinetic pathways comprising distorted Li polyhedra in metastable intermediates along two-phase boundaries. Our work demonstrates that high-rate capability may be enabled by accessing the energy landscape above the ground state, which may have fundamentally different kinetic mechanisms from the ground-state macroscopic phases. This insight should present new opportunities in searching for high-rate electrode materials.

Ionic transport in solids provides the basis of operation for electrochemical energy conversion and storage devices, such as lithium (Li)–ion batteries (LIBs), which function by storing and releasing Li^+ ions in electrode materials. During these processes, Li^+-ion transport is often coupled with phase transformations in the operating electrodes (*1*, *2*). For fast-charging applications, electrode materials capable of accommodating Li continuously through solid-solution transformation are preferentially used because they have few kinetic barriers apart from Li^+-ion diffusion in the solid state (*3*, *4*). An exception is lithium titanate (LTO), an appealing anode capable of fast charging without the issue of Li plating identified in graphite (*5*). LTO accommodates Li through a two-phase process, during which the initial disordered spinel phase ($Li_4Ti_5O_{12}$; space group $Fd\bar{3}m$) transforms directly into a rock-salt phase ($Li_7Ti_5O_{12}$; $Fm\bar{3}m$) with negligible volume change (i.e., zero strain) (*6*–*8*). Microscopically, Li insertion into the octahedral 16c sites is accompanied by Li^+-ion migration from the tetrahedral 8a to the 16c sites. However, because Li^+-ion mobility is poor in the end-members ($Li_4Ti_5O_{12}$ and $Li_7Ti_5O_{12}$), a model in which these two phases coexist macroscopically conflicts with the high Li^+ -ion mobility observed at intermediate concentrations (*9*–*11*).

This puzzling behavior has recently been attributed to the existence of an intermediate state ($Li_{4+x}Ti_5O_{12}$; $0 \le x \le 3$) with Li^+ ions simultaneously occupying face-sharing 8a and 16c sites, in the form of either a homogenous solid solution or a mixture of phase-separated nanometer-sized domains (*10*, *12*). In situ x-ray absorption spectroscopy studies provided evidence of the metastable $Li_{4+x}Ti_5O_{12}$ state, which emerges upon Li insertion even at low rates (*13*). Computational studies have predicted that face-sharing 8a and 16c local motifs are stabilized at the $Li_4Ti_5O_{12}/Li_7Ti_5O_{12}$ phase boundaries owing to the presence of Li occupying the 16d sites (*12*). However, because of the nonequilibrium nature of the ionic transport (*12*, *13*), the kinetic pathways and underlying mechanisms enabling facile ionic transport in LTO remain unresolved.

With available characterization techniques, it has been challenging to determine the atomic configuration of the metastable intermediates ($Li_{4+x}Ti_5O_{12}$) and the associated Li^+-ion transport pathways (*6*, *13*, *14*). Li K-edge electron energy-loss spectroscopy (Li-EELS), more specifically the energy-loss near-edge structure, shows promise for probing the site occupancy of Li in lithiated electrodes because of its high sensitivity to the local environment surrounding Li (*15*). Compared with other Li-sensitive techniques, such as nuclear magnetic resonance (NMR) spectroscopy (*10*) and neutron diffraction (*16*), EELS in the transmission electron microscope (TEM) has the intrinsic advantages of high spatial and temporal resolution (*17*). However, because the low-lying Li K edge (~60 eV) is in immediate proximity to the strong plasmon excitations, plural inelastic scattering may arise in the presence of thick membranes and liquid electrolyte in electrochemical cells, complicating operando Li-EELS measurements (*18*).

We developed an ionic liquid electrolyte (ILE)–based electrochemical cell for operation inside a TEM, with a configuration resembling that of a real battery, enabling operando Li-EELS probing of Li occupancy and transport in LTO upon galvanostatic (dis)charging at varying rates. Through combined operando Li-EELS and first-principles studies, we identified representative metastable $Li_{4+x}Ti_5O_{12}$ configurations, consisting of distorted Li polyhedra at the reaction front, which provide distinct Li^+-ion migration pathways with substantially lower activation energy than those in the end-members and which dominate the kinetics of Li^+-ion transport in LTO.

Figure 1A shows the design of the electrochemical cell for operando Li-EELS measurements, adapted from a TEM grid–based cell (*19*). The cell uses ILE containing 1.0 M lithium bis (trifluoromethanesulfonyl)imide in 1-butyl-1-methylpyrrolidinium bis(trifluoromethylsulfonyl) imide, a nonflammable electrolyte that has been increasingly used for batteries (*20*). Because of its low vapor pressure, ILE is compatible with the high-vacuum environment in the TEM column, thereby avoiding the thick membranes generally required in liquid cells. By controlling the ILE thickness (to 10 nm or less) and collection angle (below 1.0 mrad), plural plasmon excitation was largely suppressed (fig. S1), enabling the collection of high-quality, low-energy lying Li-EELS spectra (*21*). The electrochemical functionality of the cell was tested by galvanostatic cycling of LTO electrodes, with rates spanning from 0.8 to 8 C [see the method of estimating C-rate in the Operando TEM-EELS experiments of (*22*)]. The electrochemical performance, with flat voltage plateaus at ~1.55 V and sharp redox peaks in the cyclic voltammetry curves (Fig. 1B and fig. S2), is comparable to that in regular LIB cells. Such an ILE-based electrochemical cell was used in operando Li-EELS experiments to track Li^+-ion migration in LTO nanoparticles with well-defined structure and morphology (fig. S3). Li-EELS in the pre-edge region provides key information about the occupancy and migration of Li^+ ions among different sites (e.g., 8a in $Li_4Ti_5O_{12}$, 16c in $Li_7Ti_5O_{12}$, and other polyhedral sites associated with $Li_{4+x}Ti_5O_{12}$), as illustrated in Fig. 1C.

Figure 2 presents the time-resolved Li-EELS spectra obtained from a few selected nanoparticles (Fig. 2A and fig. S4) during the first cycle at a rate equivalent to 2 C. EELS spectra of the Ti L- and O K-edges were also obtained before and after (dis)charge (fig. S5), confirming active Ti redox [section 2 in (*22*)]. As shown in Fig. 2C, the main peak in the Li-EELS spectra remained at the same energy position

[1]Sustainable Energy Technologies Department, Brookhaven National Laboratory, Upton, NY 11973, USA. [2]Department of Materials Science and Engineering, University of California, Berkeley, Berkeley, CA 94720, USA. [3]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA [4]Department of Condensed Matter Physics and Materials Science, Brookhaven National Laboratory, Upton, NY 11973, USA. [5]Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, NY 11973, USA.
*These authors contributed equally to this work.
†Present address: Key Laboratory of Advanced Energy Materials Chemistry (Ministry of Education), College of Chemistry, Nankai University, Tianjin 300071, P. R. China. ‡Present address: Department of Energy Engineering, School of Energy and Chemical Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea.
§Corresponding author. Email: gceder@berkeley.edu (G.C.); fwang@bnl.gov (F.W.)

(~61.5 eV) during cycling. Nonetheless, subtle but clear changes occurred within the pre-edge region, as shown in the intensity map of the spectra in Fig. 2, C and D. In the spectrum of $Li_4Ti_5O_{12}$, a broad peak in the pre-edge region (called pre-peak hereafter) appeared at ~58.9 eV (labeled by "S") (see also fig. S6), which mainly comes from the inelastic scattering of $Li_{(8a)}$ sites (*15*). The pre-peak from $Li_{(16c)}$ sites in $Li_7Ti_5O_{12}$ appears at the same energy, despite the different local Li environment in the two end-members (fig. S6).

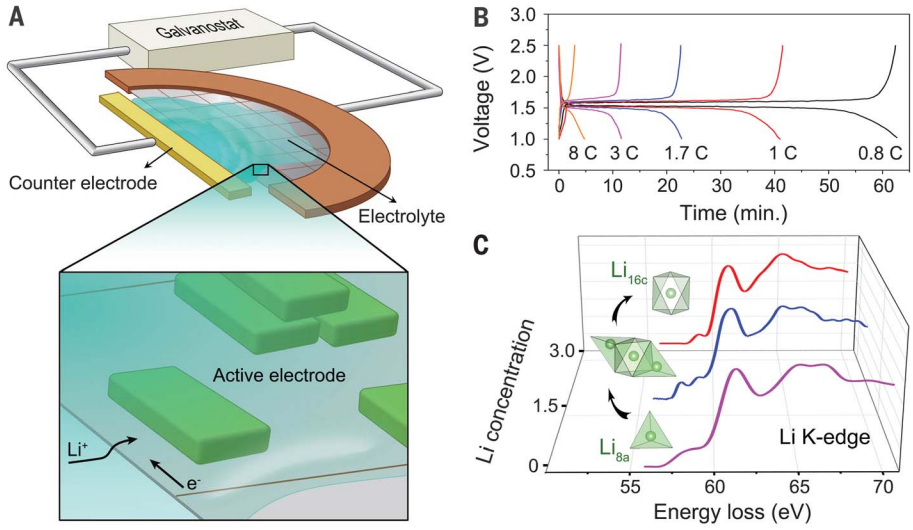


Fig. 1. Design of an electrochemically functional cell for operando characterization of battery materials inside a TEM. (**A**) Schematic of the ILE-based electrochemical cell, with a configuration similar to that of regular batteries with the active electrode (e.g., LTO nanoparticles marked by green color; inset) loaded on a carbon film (gray) as a working electrode, matched with Li metal (yellow) immersed in ILE (light green). (**B**) Voltage profiles of LTO nanoparticles in the ILE-based electrochemical cell during galvanostatic discharge and charge at different rates (fig. S2 shows comparison with cycling tests in regular coin-type cells). (**C**) Representative Li-EELS spectra obtained from LTO nanoparticles at different lithiated states, revealing the migration of Li^+ ions from the initial tetrahedral 8a sites in $Li_4Ti_5O_{12}$ to polyhedral sites in an intermediate to the final octahedral 16c sites in $Li_7Ti_5O_{12}$ (as illustrated in the inset).

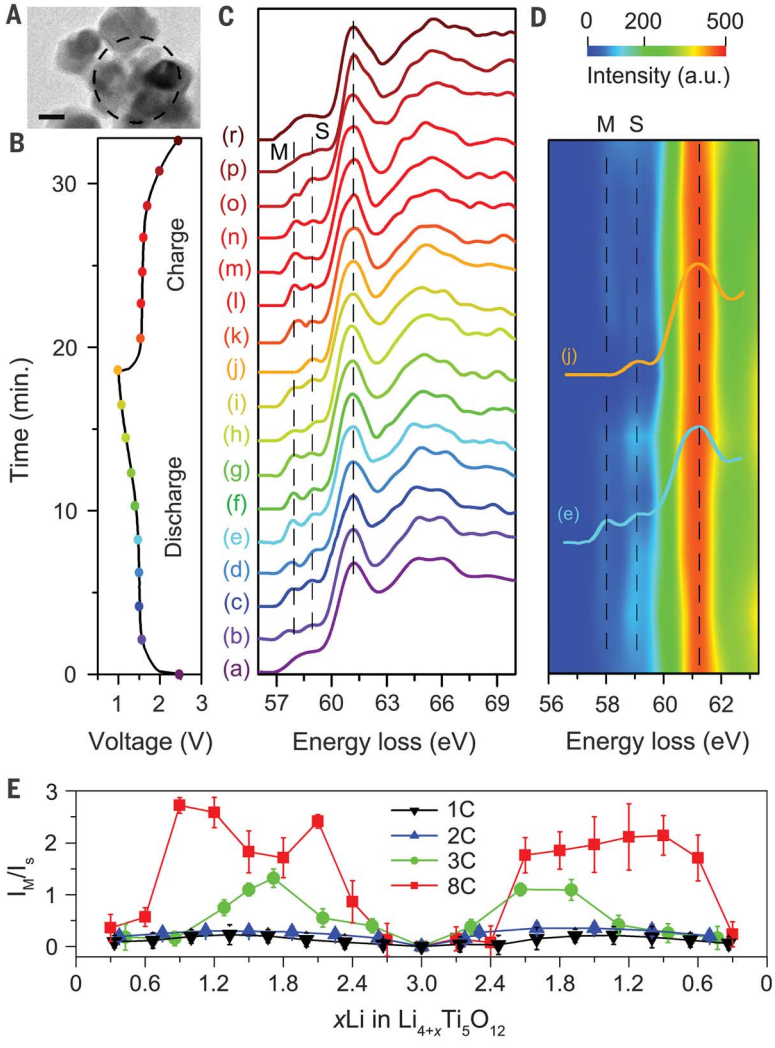


Fig. 2. Real-time probing of Li^+-ion transport in LTO during discharge and charge using operando Li-EELS. (**A**) Bright-field TEM image showing the LTO nanoparticles selected for obtaining Li-EELS spectra (marked by a black dashed circle). Scale bar, 50 nm. (**B** and **C**) Voltage profiles of LTO nanoparticles and the corresponding EELS spectra selected with an interval of 120 s during the first cycle at 2 C rate [marked by color-coded dots in (B) and letters in (C)]. The vertical dashed black lines indicate the energy positions of the main peaks at ~61.5 eV and pre-peaks M (related to metastable configurations of the intermediate compositions) and S (related to stable configurations in $Li_4Ti_5O_{12}$ and $Li_7Ti_5O_{12}$). (**D**) Intensity map of the Li-EELS spectra at a 2 C rate (fig. S8 shows EELS spectra and intensity maps at 1 C, 3 C, and 8 C rates). Two representative spectra (e) and (j) are displayed to show the pre-edge features, M and S. a.u., arbitrary units. (**E**) Intensity ratio of the two pre-peaks, M and S (I_M/I_S) as a function of Li concentration (x) at different rates (1 C, 2 C, 3 C, and 8 C), in which I_M and I_S are the integrated intensities of pre-peaks M and S, respectively.

Upon discharging and subsequent charging, the pre-peak S remained at nearly the same position. However, a new pre-peak emerged at ~58.0 eV (labeled by "M"), which is absent in the spectra of the two end-members (Fig. 2D) as well as the partially lithiated LTO electrodes in ex situ measurements [section 3 in (*22*)] (fig. S7). Pre-peak M was commonly observed across the electrode, and its intensity was found to be strongly rate-dependent (fig. S8). The ratio of integrated intensity of the two pre-peaks M and S (defined as I_M/I_S) as a function of Li concentration (x) was plotted in Fig. 2E. At low rates (1 C and 2 C), I_M/I_S values are small (only ~0.2) but increase abruptly (to ~2.0) at high rates (3 C and 8 C) [quantitative analysis in section 4 of (*22*)]. As shown below, the evolution of features in the Li-EELS spectra (e.g., I_M/I_S ratio) provides key information about Li occupancy and migration in the metastable intermediates ($Li_{4+x}Ti_5O_{12}$) and its rate-dependent behaviors.

To understand the origin of the pre-edge features in the Li-EELS spectra and their evolution during charge and discharge, we studied the local configurations of Li^+ ions in the two end-members ($Li_4Ti_5O_{12}$ and $Li_7Ti_5O_{12}$) and the intermediates ($Li_{4+x}Ti_5O_{12}$, x = 1 and 2) using density functional theory (DFT) (*22*). In the low-energy structures of $Li_{4+x}Ti_5O_{12}$, one of the $Li_{(8a)}$ tetrahedra shares a three-coordinated oxygen face with a $Li_{(16c)}$ octahedron at the domain boundary between $Li_4Ti_5O_{12}$ and $Li_7Ti_5O_{12}$ (fig. S9, C and D). The face-sharing $Li_{(8a)}$ and $Li_{(16c)}$ polyhedra are highly distorted compared to the unperturbed $Li_{(8a)}$ and $Li_{(16c)}$ polyhedra in $Li_4Ti_5O_{12}$ and $Li_7Ti_5O_{12}$ and are stabilized by neighboring $Li_{(16d)}$ octahedra [section 5 in (*22*)], which is consistent with recent DFT calculations (*12*). Within 100 meV/O_4, a large number of relevant atomic configurations exist with different local environments, especially local motifs of $Li_{(8a)}$-$Li_{(16c)}$ face-sharing polyhedra with various levels of distortion (fig. S10). The DFT-based sampling of the $Li_{4+x}Ti_5O_{12}$ potential landscape reveals that a very large configuration space can be accessed, even within the formation energy of 100 meV/O_4 [the nature of these intermediate configurations is described further in sections 5 and 8 in (*22*)].

After establishing the structural models of $Li_{4+x}Ti_5O_{12}$ at different Li concentrations (x = 0, 1, 2, and 3), we calculated the corresponding Li-EELS spectra for Li at the 8a, 16c, and 16d sites for the most stable configurations of $Li_{4+x}Ti_5O_{12}$ using the Z + 1 approach (where Z refers to the atomic number) to model the core-hole effect [computational details in (*22*)]. Because the pre-edge peaks originate predominantly from the intra-atomic Li 1s to 2p transition, this level of description is adequate. The good agreement in the near-edge region of the end-members between the Z + 1 method and the Bethe-Salpeter equation–based method supports our choice of method (fig. S11). The energy positions of pre-peak S in the computed spectra and their relative intensities compared with those of the main peak for both $Li_4Ti_5O_{12}$ and $Li_7Ti_5O_{12}$ are in good agreement with the experimental results (fig. S12A), validating the Z + 1 method used for computing the Li-EELS spectra. The Li-EELS spectra of $Li_{(8a)}$ in the most stable $Li_5Ti_5O_{12}$ and $Li_6Ti_5O_{12}$ configurations are almost the same as those of $Li_{(8a)}$ in $Li_4Ti_5O_{12}$, especially near the pre-peak region (Fig. 3A). However, a new pre-peak appears at ~58 eV in the EELS spectra of face-sharing $Li_{(16c)}$ that is absent in the computed spectra of $Li_{(16c)}$ in $Li_7Ti_5O_{12}$ and $Li_{(16d)}$ at any composition (Fig. 3B; fig. S12, B and C; and fig. S14). Because this new pre-peak coincides in energy with pre-peak M observed in the operando EELS measurements (Fig. 2B), we assign pre-peak M in the Li-EELS spectra at low rates to the inelastic scattering from distorted face-sharing $Li_{(16c)}$ in the metastable configurations. Owing to the distortion of face-sharing $Li_{(16c)}$, several Li–O bonds are elongated, which breaks the degeneracy of Li–O coupling. A representative configuration is shown in Fig. 3C that demonstrates elongated Li–O bond lengths of 2.33 to 2.50 Å in $Li_5Ti_5O_{12}$, considerably longer than those in $Li_7Ti_5O_{12}$ (2.06 to 2.20 Å). The weakened Li–O bonds effectively pull the antibonding Li–O states to lower energy and cause the pre-peak to split, giving rise to pre-peak M in the EELS spectra. The corresponding partial charge densities are shown in the isosurface plot in Fig. 3C. Whereas the partial charge density associated with pre-peak M concentrates more heavily on O atoms with longer Li–O bonds, the partial charge density of pre-peak S is distributed evenly on all O atoms regardless of Li–O bond

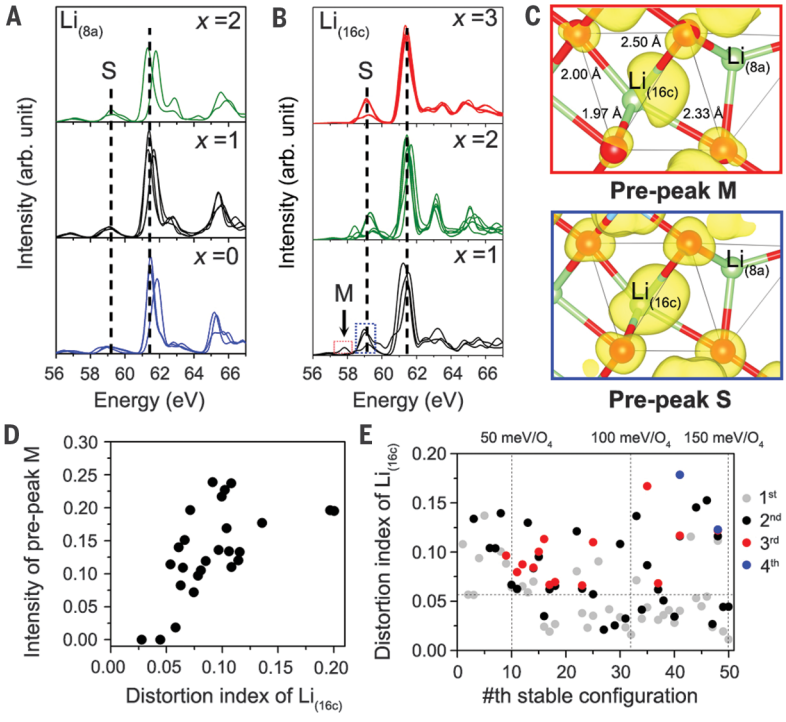


Fig. 3. Identification of Li-EELS fingerprints for Li-polyhedral configurations in $Li_{4+x}Ti_5O_{12}$ ($0 \le x \le 3$) by DFT calculations. (**A** and **B**) Calculated Li-EELS spectra of $Li_{4+x}Ti_5O_{12}$ (x = 0, 1, and 2) for Li at 8a sites and $Li_{4+x}Ti_5O_{12}$ (x = 1, 2 and 3) for Li at 16c sites, respectively. The dashed lines in (A) and (B) mark the energy positions of the main peaks and pre-peak S. The black arrow in (B) indicates the pre-peak M from face-sharing $Li_{(16c)}$ in $Li_5Ti_5O_{12}$ and $Li_6Ti_5O_{12}$, which is not observed for $Li_{(16c)}$ in $Li_7Ti_5O_{12}$. arb. unit, arbitrary unit. (**C**) Isosurface of partial charge density around face-sharing $Li_{(16c)}$ in $Li_5Ti_5O_{12}$ in the energy range of pre-peak M and pre-peak S, as marked with dashed red and blue boxes, respectively, in (B), with 0.015 of isovalue. Bond lengths between $Li_{(16c)}$ and O ions are labeled. (**D**) Intensity of the pre-peak M as a function of distortion index (d) of face-sharing $Li_{(16c)}$ for various Li configurations in $Li_5Ti_5O_{12}$ and $Li_6Ti_5O_{12}$. (**E**) d of face-sharing $Li_{(16c)}$ in the nth stable configuration (ordered by formation energy) of $Li_5Ti_5O_{12}$. The horizontal and vertical dashed lines indicate the approximate value of d above which pre-peak M appears and the formation energy of the nth stable configuration at 50, 100, and 150 meV/O_4. When a single configuration contains multiple face-sharing $Li_{(16c)}$, the different face-sharing $Li_{(16c)}$ are labeled as colored points according to their d values in the ascending order: gray (first lowest distortion index), black (second), red (third), and blue (fourth).

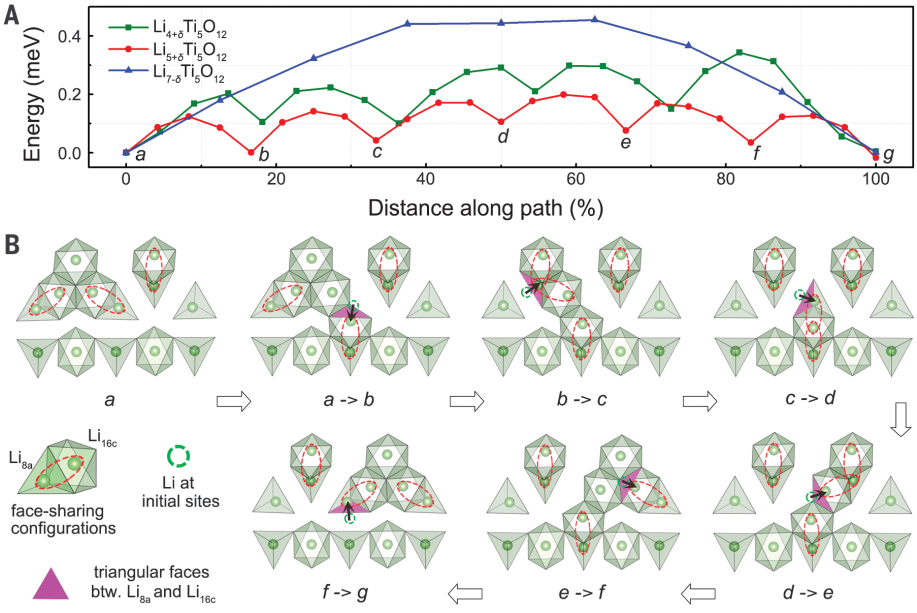


Fig. 4. Li^+-ion migration pathways and the corresponding energy barriers in the intermediates. (**A**) Energy profile of the pathways in $Li_{4+\delta}Ti_5O_{12}$ (green), $Li_{5+\delta}Ti_5O_{12}$ (red), and $Li_{7-\delta}Ti_5O_{12}$ (blue) as a function of distance along the paths. (**B**) Migration pathways involved for each step, from a to g, in one representative intermediate, $Li_{5+\delta}Ti_5O_{12}$ (movie S1 shows the trajectories of Li^+-ion migration in $Li_{5+\delta}Ti_5O_{12}$). The translucent green spots mark the initial Li sites during each substep of migration. The black arrows indicate the migration direction of each substep. The three-coordinated oxygen face through which the Li^+ ions migrate from a Li_{8a} tetrahedron to a Li_{16c} octahedron is colored purple. The Li^+-ion migration pathways in $Li_{4+\delta}Ti_5O_{12}$ and $Li_{7-\delta}Ti_5O_{12}$ are provided in movie S2 and figs. S23 and S24. btw., between.

lengths (see also the projected density of states in fig. S13).

To understand the effect of high (dis)charge rate on pre-peak M in the Li-EELS spectra (Fig. 2E), we also computed the Li-EELS spectra of $Li_5Ti_5O_{12}$ and $Li_6Ti_5O_{12}$ configurations with higher formation energies (which are accessible at higher current rates or large overpotential) . Figure S10 shows that the distortion of face-sharing $Li_{(16c)}$ and $Li_{(8a)}$ polyhedra is affected by the presence of neighboring $Li_{(16d)}$ or $Ti_{(16d)}$ (*22*). The varying local environment results in different intensities and energy levels of pre-peak M [section 7 in (*22*)]. Face-sharing $Li_{(16c)}$ and $Li_{(8a)}$ polyhedra in various $Li_{4+x}Ti_5O_{12}$ with distortion index [*d*; eq. S2 in (*22*)] larger than ~0.06 result in the appearance of pre-peak M, whose intensity tends to increase with *d* [Fig. 3D, fig. S17, and section 7 in (*22*)]. Such high distortion levels reduce the effective coordination number of $Li_{(16c)}$ and $Li_{(8a)}$. On the basis of the coordination number weighting scheme of (*23*), $Li_{(16c)}$ and $Li_{(8a)}$ with d ~ 0.06 are found to be undercoordinated with effective coordination numbers of 4.7 and 3.6 (fig. S18), respectively, rather than the expected 6 and 4. Similar trends showing a more pronounced pre-peak with undercoordinated local environments are well established in transition-metal K-edge x-ray absorption near-edge structures (*24*). As observed in Fig. 3E and fig. S19, highly distorted face-sharing $Li_{(16c)}$ and $Li_{(8a)}$ polyhedra with *d* higher than 0.06 appear more frequently as the formation energy of the intermediate $Li_5Ti_5O_{12}$ and $Li_6Ti_5O_{12}$ increases. This is consistent with the presence of more face-sharing $Li_{(16c)}$ and $Li_{(8a)}$ polyhedra in $Li_5Ti_5O_{12}$ configurations with higher formation energies, as indicated by the points with different colors in Fig. 3E. Highly distorted face-sharing $Li_{(8a)}$ tetrahedra are observed much less often than highly distorted $Li_{(16c)}$ octahedra in configurations with formation energy less than 100 meV/O_4 (Fig. 3E and fig. S19). Therefore, the appearance of pre-peak M is mainly attributed to face-sharing $Li_{(16c)}$ octahedra at low current rates but to both face-sharing $Li_{(16c)}$ and $Li_{(8a)}$ polyhedra at high current rates.

Phase transformation through a solid-solution path is favorable for high-rate performance in many electrodes, such as $LiFePO_4$ (LFP) (*3*, *4*). However, LTO undergoes a first-order phase transition (*6*, *12*) because the macroscopic solid solution is largely inaccessible owing to its high formation energy, in contrast to the low formation energies of the metastable solid solution in LFP [sections 5 and 8 in (*22*)]. Instead, (de)lithiation proceeds by means of a two-phase reaction, involving face-sharing $Li_{(8a)}$–$Li_{(16c)}$ local motifs at the boundaries between the phase domains (*12*, *13*), which are nanosized because of the low formation energy of the interfaces [section 8 in (*22*)]. Hence, whereas the origin of the fast rate in LFP is the low energy of the metastable solid solution, it is the low interfacial energy in LTO that creates nanosized domains of each phase and provides the interfacial pathways for fast Li^+-ion transport. Our results are generally consistent with the two-phase model (*12*, *13*), as we observed mainly interfacial-type configurations at lower energy [sections 5 and 8 in (*22*)]. However, our results further reveal a large variety of face-sharing $Li_{(8a)}$–$Li_{(16c)}$ local motifs in the intermediate $Li_{4+x}Ti_5O_{12}$ configurations. These local motifs, whose number and distortion index display rate dependence, may affect the kinetics in intermediate compositions.

To obtain a mechanistic understanding of fast Li diffusion in LTO, we performed nudged–elastic-band (*25*, *26*) calculations that account for distorted face-sharing Li polyhedra (Fig. 4 and figs. S23 and S24). The activation energies of Li^+-ion migration in the lowest-energy $Li_4Ti_5O_{12}$ and $Li_5Ti_5O_{12}$ (with an interstitial Li^+) and $Li_7Ti_5O_{12}$ (with a vacancy) configurations are ~343, ~216, and ~455 meV, respectively (Fig. 4A). The low activation energy of $Li_5Ti_5O_{12}$ is in line with the low migration barriers previously obtained from NMR measurements (*10*) and ab initio molecular dynamics (*12*). Our results indicate that facile Li^+ migration occurs at the two-phase boundaries containing face-sharing Li polyhedra, which is associated with lower activation barriers compared with those in the end-members. In general, the Li^+-ion diffusion pathway involves Li hopping from face-sharing tetrahedral (octahedral) Li sites to octahedral (tetrahedral) Li sites (Fig. 4B and fig. S23, B to H). Along this path, although face-sharing Li^+ ions change position, the number of face-sharing Li^+ ions (three to four) remains nearly constant in the $Li_4Ti_5O_{12}$ (always two) and $Li_5Ti_5O_{12}$ (from three to four to three) pathways. As a result, there is no abrupt increase of the energy of the system. In the higher-energy pathway in $Li_7Ti_5O_{12}$, however, the number of face-sharing Li^+-ions changes substantially from zero to two and back to zero (fig. S24, B to D). The low migration barrier for Li^+ ions in the LTO system can be attributed to two important factors: (i) The number of face-sharing Li polyhedra is smaller in the transition state (when the migrating Li^+ ions are in the triangular face in Fig. 4B) than in the initial and final states within each step [for example, there are three instances of face-sharing in states *a* and *b* (marked in Fig. 4) but only two between *a* and *b* in Fig. 4B]. The reduction in Li^+ – Li^+ repulsion in the transition state is likely to lower the activation barrier. (ii) Because local distortion helps to reduce the effective coordination number of Li (fig. S18), the change in Li coordination is minimized

during Li^+-ion migration through the three-coordinated oxygen face, further lowering the activation barrier (*27*). Both factors minimize changes in the energy, resulting in a relatively flat energy landscape along the migration path, as described in Fig. 4A. Our analysis implies that the improved kinetics at higher rate results from the increased amount of face-sharing Li^+, and thus mobile carriers, and the more highly distorted Li polyhedra seen in the high-energy $Li_{4+x}Ti_5O_{12}$ configurations accessible at high rates.

In this study, we developed an electrochemically functional cell operating inside a TEM that allows operando characterization of electrode materials under real electrochemical conditions. Facile Li^+-ion migration routes were revealed in intermediate configurations involving face-sharing Li polyhedra. The associated low Li^+-ion migration barriers of the face-sharing motifs along with the low formation energy of the interfaces reconcile the apparent contradiction between the high-rate capability of LTO and the poor Li^+-ion conductivity of its end-member phases. Findings from this study, particularly on the improved kinetics originating from the face-sharing Li motifs in intermediate configurations that become accessible at high rates, provide insights into designing electrode materials for fast-charging batteries.

REFERENCES AND NOTES

1. J. Maier, *Nat. Mater.* **4**, 805–815 (2005).
2. C. Zhu, R. E. Usiskin, Y. Yu, J. Maier, *Science* **358**, eaao2808 (2017).
3. H. Liu *et al*., *Science* **344**, 1252817 (2014).
4. R. Malik, F. Zhou, G. Ceder, *Nat. Mater.* **10**, 587–590 (2011).
5. T.-F. Yi, S.-Y. Yang, Y. Xie, *J. Mater. Chem. A Mater. Energy Sustain.* **3**, 5750–5777 (2015).
6. S. Scharner, W. Weppner, P. Schmid-Beurmann, *J. Electrochem. Soc.* **146**, 857–861 (1999).
7. M. Wagemaker *et al*., *Adv. Mater.* **18**, 3169–3173 (2006).
8. B. Ziebarth, M. Klinsmann, T. Eckl, C. Elsässer, *Phys. Rev. B Condens. Matter Mater. Phys.* **89**, 174301 (2014).
9. H. Hain *et al*., *Solid State Nucl. Magn. Reson.* **42**, 9–16 (2012).
10. W. Schmidt *et al*., *Chem. Mater.* **27**, 1740–1750 (2015).
11. M. Wilkening *et al*., *Phys. Chem. Chem. Phys.* **9**, 6199–6202 (2007).
12. S. Ganapathy, A. Vasileiadis, J. R. Heringa, M. Wagemaker, *Adv. Energy Mater.* **7**, 1601781 (2017).
13. W. Zhang *et al*., *J. Am. Chem. Soc.* **139**, 16591–16603 (2017).
14. C. P. Grey, J. M. Tarascon, *Nat. Mater.* **16**, 45–56 (2016).
15. F. Wang *et al*., *Nanotechnology* **24**, 424006 (2013).
16. O. Dolotko *et al*., *Solid State Sci.* **36**, 101–106 (2014).
17. R. F. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope* (Springer Science & Business Media, 2011).
18. F. M. Ross, *Science* **350**, aaa9886 (2015).
19. F. Wang *et al*., *Nat. Commun.* **3**, 1201 (2012).
20. M. Armand, F. Endres, D. R. MacFarlane, H. Ohno, B. Scrosati, *Nat. Mater.* **8**, 621–629 (2009).
21. F. Wang *et al*., *ACS Nano* **5**, 1190–1197 (2011).
22. Supplementary materials.
23. K. Robinson, G. V. Gibbs, P. H. Ribbe, *Science* **172**, 567–570 (1971).
24. T. Yamamoto, *XRay Spectrom.* **37**, 572–584 (2008).
25. G. Henkelman, H. Jónsson, *J. Chem. Phys.* **113**, 9978–9985 (2000).
26. D. Sheppard, R. Terrell, G. Henkelman, *J. Chem. Phys.* **128**, 134106 (2008).
27. Z. Rong *et al*., *Chem. Mater.* **27**, 6016–6021 (2015).

ACKNOWLEDGMENTS

We thank H. Zhao, X. Hu, and Y. Celebi for help with electrode fabrication and discussion on data interpretation. We thank T. Bowman for the graphic design (Fig. 1A). **Funding:** This work was partially supported by the Laboratory Directed Research and Development program at the Brookhaven National Lab (BNL) (contract no. DE-SC0012704). This research used resources of the Center for Functional Nanomaterials, which is a U.S. Department of Energy (DOE) Office of Science facility, and the Scientific Data and Computing Center, a component of the BNL Computational Science Initiative, at BNL (contract no. DE-SC0012704). This work was supported in part by the Assistant Secretary for Energy Efficiency and Renewable Energy, Vehicle Technologies Office, of the U.S. DOE (contract no. DE-AC02-05CH11231), under the Advanced Battery Materials Research (BMR) Program and by the National Science Foundation Graduate Research Fellowship (grant no. DGE 1106400). Efforts by F.W. involving writing and revision of the manuscript were supported by U.S. DOE, Office of Energy Efficiency and Renewable Energy, Vehicle Technologies Office (contract no. DE-SC0012704). Efforts on EELS data analysis by L.W. and Y.Z. were supported by the U.S. DOE, Office of Basic Energy Science, Division of Materials Science and Engineering (contract no. DE-SC0012704). T.C. thanks the National Science Foundation's Extreme Science and Engineering Development Environment (XSEDE) supercomputer Stampede2 (allocation TG-DMR970008S), which is supported by the National Science Foundation (grant no. ACI-1548562), for providing computing resources. **Author contributions:** F.W. initiated the project. F.W. and W.Z. conceived the experiments. W.Z. conducted the experiments and analyzed the data. L.W. and Y.Z. assisted with the EELS data analysis; and D.-H.S., T.C., and G.C. performed the DFT, nudged–elastic-band, and EELS simulations. M.T. and D.L. performed the spectral simulations using the OCEAN package. W.Z., D.-H.S., T.C., D.L., G.C., and F.W. wrote the paper, and all the authors were involved in revising the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and the supplementary materials. Additional data related to the computational work in this paper are available on the Materials Cloud website through the following link: https://archive.materialscloud.org/2020.0006/v1.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/1030/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S24
References (*28*–*40*)
Movies S1 and S2

15 March 2019; accepted 30 January 2020
10.1126/science.aax3520

ECOLOGICAL SPECIATION

Ecologically diverse clades dominate the oceans via extinction resistance

Matthew L. Knope[1]*, Andrew M. Bush[2]†, Luke O. Frishkoff[3]†, Noel A. Heim[4]†, Jonathan L. Payne[5]†

Ecological differentiation is correlated with taxonomic diversity in many clades, and ecological divergence is often assumed to be a cause and/or consequence of high speciation rate. However, an analysis of 30,074 genera of living marine animals and 19,992 genera of fossil marine animals indicates that greater ecological differentiation in the modern oceans is actually associated with lower rates of origination over evolutionary time. Ecologically differentiated clades became taxonomically diverse over time because they were better buffered against extinction, particularly during mass extinctions, which primarily affected genus-rich, ecologically homogeneous clades. The relationship between ecological differentiation and taxonomic richness was weak early in the evolution of animals but has strengthened over geological time as successive extinction events reshaped the marine fauna.

More than half a century ago, Hutchinson (*1*) asked "Why are there so many kinds of animals?" and answered that adaptive ecological divergence permits coexistence of species and leads to taxonomic diversification. Indeed, ecological differentiation and taxonomic diversity are tightly correlated in many modern ecosystems (*2*–*5*). Ecological differentiation can permit the exploration of additional niches, reducing competition and fueling speciation (*4*–*7*). However, a similar correlation between ecological differentiation and taxonomic diversity would also be generated if causation were reversed, and a greater number of speciation events provided additional opportunities for ecological divergence (*6, 7*). Furthermore, taxonomic diversification occurs as a result of the difference between origination and extinction rates (Fig. 1A), but the contribution of extinction to the coupling between ecological differentiation and taxonomic diversification has received relatively little attention (*8*). Thus, the nature of the relationship between ecological differentiation and taxonomic diversification remains a central unanswered question in biology.

Here, we use a comprehensive dataset of both living and fossil marine animal genera to examine how the coupling between ecological and taxonomic diversity has evolved over the past 500 million years, leading to the modern relationship between ecological differentiation and taxonomic diversity on a global scale. The fossil record provides a distinctive perspective on this question; the relationship between the two parameters can be tracked through time, and the influences of origination and extinction can be explicitly separated.

In the modern oceans, taxonomic diversity (referred to as genus richness) is highly correlated with differentiation into varied ecological modes of life (referred to as ecological diversity) (Fig. 2A), which are defined by a genus's tiering position relative to the seafloor, degree of motility, and feeding mode (*9*, *10*) (class level: coefficient of determination $R^2 = 0.64$, $P < 0.0001$, Fig. 2A and fig. S1; phylum level: $R^2 = 0.90$, $P < 0.0001$, fig. S2). The slope of the regression relationship has increased over time (Fig. 2A and tables S1 and S2), and Cenozoic stages (66 million years ago to present) are not significantly different from the modern fauna (Fig. 2B). This steepening is primarily attributable to shifts associated with successive mass extinction events (Fig. 2 and fig. S3) rather than a simple continuous increase over time (multiple linear regression; continuous year-before-present effect: $F = 0.06$, $P = 0.80$; interval effect: $F = 4.66$, $P = 0.002$; full model $R^2 = 0.75$) (tables S1 and S2). These shifts indicate that the cumulative effects of mass extinctions played a key role in generating the currently strong correlation between ecological diversity and genus richness. This change in slope over time is a primary biological signal and not attributable to age-dependent properties of the fossil record or effects of clade age (fig. S4). The quantity of unmetamorphosed marine sedimentary rocks does not increase toward the present (*11*, *12*), and sampling standardization does not remove the drastic Cenozoic biodiversification event from raw diversity estimates (*13*), nor can the increase in slope be attributed to a greater proportion of genera assigned to ecological modes toward recent time intervals (*14*) or to the taxonomic scale of inquiry (compare Fig. 2 with fig. S3). Thus, the relationship that characterizes the living marine fauna and the more recent fossil record is a geologically recent development, shaped over hundreds of millions of years of evolutionary dynamics (Fig. 2).

In the Cenozoic, the most genus-rich classes are associated with lower origination rates (averaged across all prior time intervals) rather than higher rates, and their overall greater taxonomic net diversification rates result from their even lower extinction rates (Fig. 1B). However, before the end-Cretaceous mass extinction (66 million years ago), the most genus-rich classes tended to have much higher origination and extinction rates with greater taxonomic turnover (Fig. 1B). The most important determinant of genus richness in classes has therefore shifted over time from a propensity toward origination to resistance against extinction.

What led to this long-term shift in the relationship between diversity, origination, and extinction? It is often suggested that ecological differentiation either results from or promotes speciation (*6, 7*), but the results from the fossil record of marine animals contradict this hypothesis in several ways. First, ecologically diverse higher taxa tend to be characterized by lower, not higher, instantaneous probabilities (fig. S5 to S7) and rates of genus origination across time (Figs. 1C and 3). Thus, there is no evidence that high ecological diversity drives greater origination on these time scales (Fig. 3). Ecologically diverse clades also display lower instantaneous probabilities (fig. S5 to S7) and rates of genus extinction (Figs. 1C and 3). Across the Phanerozoic, greater ecological diversity is significantly associated with higher net diversification rates, whereas greater genus richness is significantly associated with lower net diversification rates across time (fig. S8), ultimately leading to the lowered taxonomic turnover rates in genus-rich classes in the Cenozoic (Fig. 1B).

The era-bounding mass extinctions resulted in massive loss of taxonomic diversity but little loss of functional groups (fig. S9) (*14*–*16*). The most genus-rich functional groups were hit hardest on a per-genus basis (*14*), whereas the most ecologically diverse groups were relatively spared (Fig. 1 and figs. S5 to S7), resulting in the evening out of taxonomic diversity across ecological modes (fig. S10). Further, although ecological diversity and taxonomic diversity within classes were not significantly correlated during the Paleozoic (541 million to 252 million years ago), standing ecological diversity during the Paleozoic does predict standing genus richness for almost every stage for the past ~150 million years (fig. S11) as well as aggregate genus richness across the Phanerozoic (*14*). In contrast, standing genus richness

[1]Department of Biology, University of Hawaii, Hilo, Hilo, HI 96720, USA. [2]Department of Geosciences and Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA. [3]Department of Biology, University of Texas at Arlington, Arlington, TX 76019, USA. [4]Department of Earth and Ocean Sciences, Tufts University, Medford, MA 02115, USA. [5]Department of Geological Sciences, Stanford University, Stanford, CA 94305, USA.
*Corresponding author. Email: knope@hawaii.edu
†These authors contributed equally to this work.

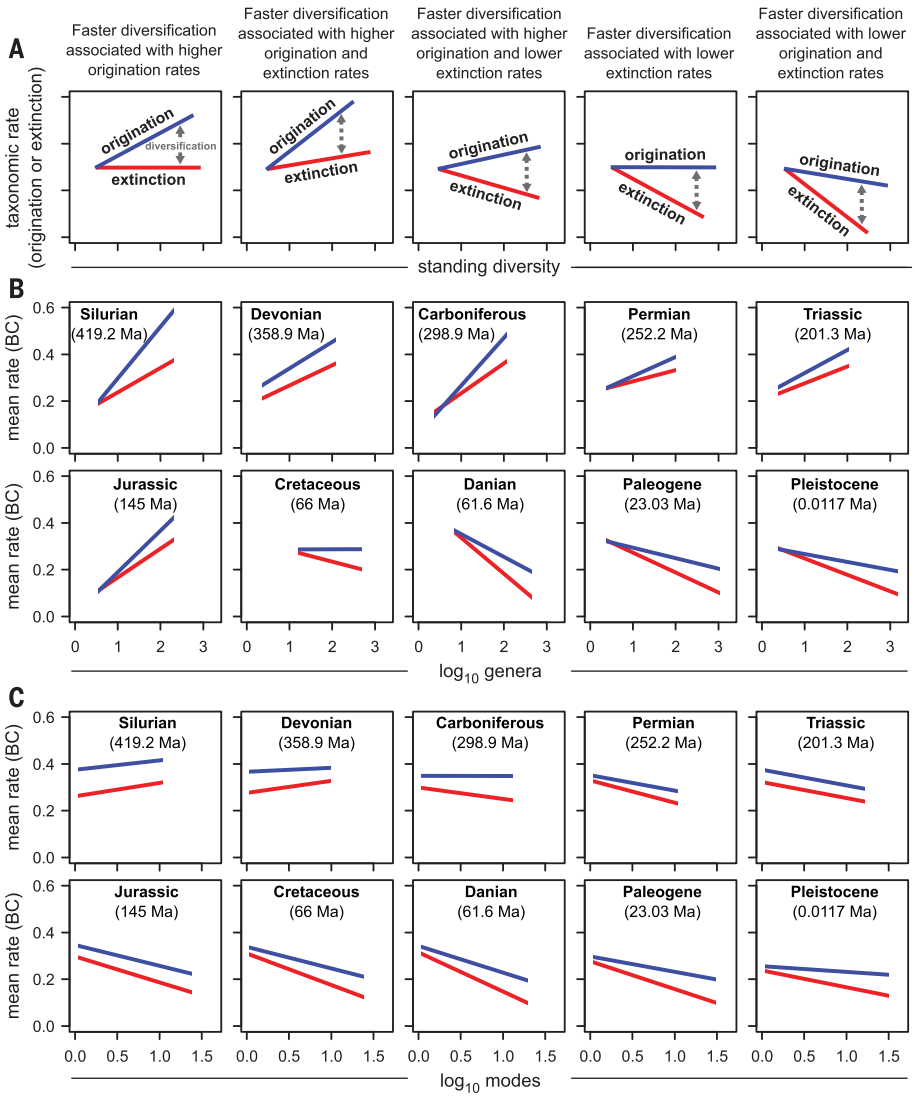


Fig. 1. For clades with similar origination times, greater standing diversity will be associated with faster net diversification rates. The net diversification rate represents the difference between the rates of origination and extinction. Consequently, greater net diversification can be associated with a higher or lower origination rate, depending on the relationship between extinction rate and diversity (**A**). Mean origination and extinction rates across all prior stages versus standing taxonomic richness (**B**) and ecological diversity (**C**) in the final stage of each period. BC, boundary-crosser extinction metric; Ma, million years ago. See figs. S13 and S14 for all underlying data points and net diversification rates.

during the Paleozoic does not predict either coeval or later ecological diversity (fig. S11), indicating that high genus richness does not lead to greater evolution of new modes of life. Rather, clades that had greater ecological diversity early were positioned to achieve greater taxonomic richness across time, because they weathered mass extinctions better than ecologically homogeneous clades did (Fig. 1 and figs. S5 to S7).

The dynamics revealed in the analysis of the fossil record suggest an alternative mechanism by which the modern correlation between ecological diversity and taxonomic richness developed through time. Clades that have high probabilities of origination on geologic time scales also have high probabilities of extinction (*17*, *18*) (figs. S5 to S7), a correlation that has strengthened over time (fig. S12). Taxa that contain many modes of life tend to experience low intensity of both origination and extinction (Figs. 1C and 3) and display low volatility (Fig. 4), meaning that their diversity tends to change more slowly. Volatile taxa (i.e., those with higher origination and extinction rates) are more likely to go completely extinct, either through random chance (*17*) or during mass extinctions (*18*), when extinction intensity increases for all taxa and can be extremely high for volatile clades. Several highly volatile taxa diversified rapidly during the Paleozoic but were hit hard during mass extinctions, leading to a long-term shift in dominance toward low-volatility taxa (*18*), which tend to be ecologically diverse.

Why do ecologically diverse taxa tend to have low volatility? To investigate the possibility that ecologically diverse classes have physiological or ecological traits that are protective against extinction at the genus level, we coded all genera by their physiological buffering capacity (the sophistication of the respiratory and circulatory systems, which are documented to have differential sensitivity to well-known ocean acidification and anoxic events in the geologic record) (*19*, *20*), motility, habitat, and feeding mode (*9*). Extinction probability (in a general linear mixed-effects model) (*10*) was significantly lower for genera that were fully motile ($P = 0.005$), predatory ($P = 0.01$), pelagic ($P = 0.01$), or physiologically buffered and motile (buffered, fully motile: $P < 0.001$; buffered, facultatively motile: $P = 0.02$). Even after controlling for the effects of these traits, genera belonging to ecologically diverse classes still had a lower extinction risk (table S3), indicating the existence of additional, unidentified reasons that ecologically diverse clades have constituent members that are more resistant to extinction.

Motility and physiological buffering are associated with lower extinction risk during a number of major and minor mass extinctions, possibly because they impart a greater ability to cope with physiological stressors, such as ocean acidification and warming (*21*) (motility and buffering are related because unbuffered

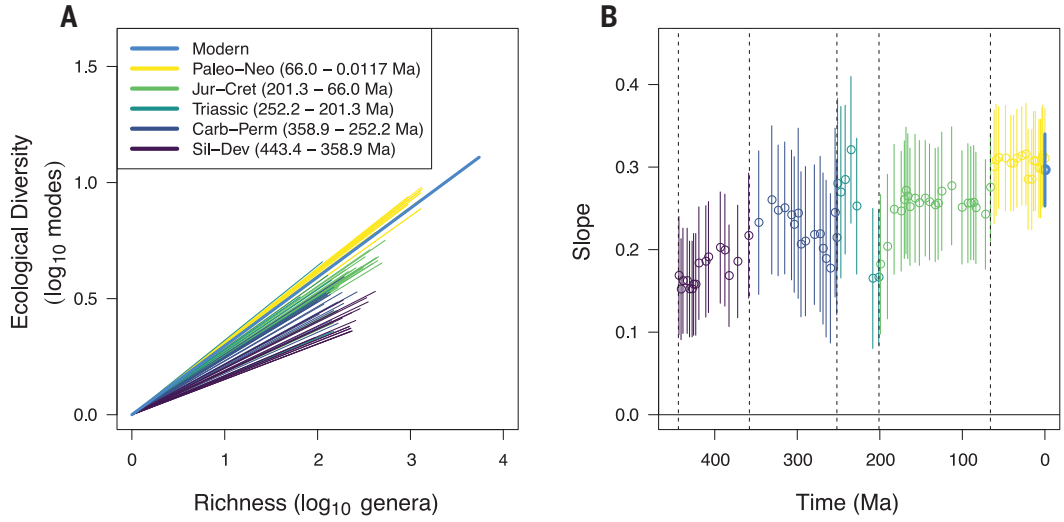


Fig. 2. The relationship between taxonomic richness and ecological diversity in marine animal Linnaean classes in modern oceans and over the past 444 million years. (**A**) The abovementioned relationship for living and fossil faunas from individual geologic stages. Living marine animals demonstrate a strong, positive, and statistically significant power-law relationship [R^2 = 0.64; P < 0.0001; slope 0.31; 95% confidence interval (CI), 0.27 to 0.36]. Paleo, Paleogene; Neo, Neogene; Jur, Jurassic; Cret, Cretaceous; Carb, Carboniferous; Perm, Permian; Sil, Silurian; Dev, Devonian. (**B**) The slopes of the regression lines in (A), with 95% CI plotted across time. Dashed vertical lines indicate mass extinction events.

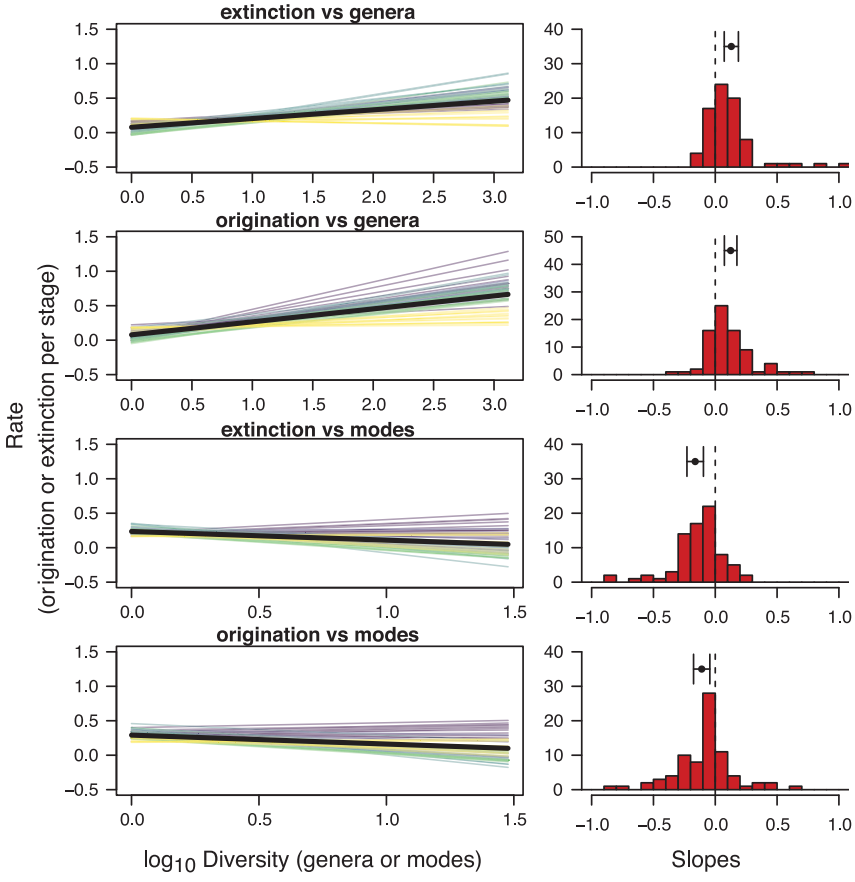


Fig. 3. The slopes of the regression models for all stages since the end of the Ordovician period of the (log_{10}) number of modes or genera versus extinction and origination rates. Histograms on the right display the frequency of stage-specific slopes with mean and 95% CI and examine the origination or extinction rate of single stages (thus maintaining statistical independence for hypothesis testing). Lines on left are cumulative curves of the time-averaged origination or extinction rate for all stages up to the focal stage for illustration of trends if assessed at different points in time, and the slope across all stages is indicated by a thick black line. See Fig. 2A for color legend.

physiology is often associated with a heavy carbonate skeleton, which limits motility). Motility is also necessary for many ecological modes of life, and movement permits access to food and habitats that are unavailable to nonmotile animals (*9*). Additionally, many nonmotile animals may have been negatively affected by increasing disturbance and predation intensity as motile animals diversified (*22*, *23*). Thus, the morphological and physiological traits that permit a clade to explore greater amounts of ecospace also convey to its individual members added resilience against some of the most common and severe environmental stressors in the oceans.

The classes that had high taxonomic richness and low ecological differentiation during the Paleozoic, such as rhychonelliform brachiopods and crinoids, consisted largely or entirely of nonmotile suspension feeders that mostly cannot occupy infaunal or pelagic habitat tiers. In contrast, the classes and phyla that are genus-rich in the Cenozoic (e.g., mollusks, arthropods, and vertebrates) are generally motile, feed in a variety of ways, live across many habitat tiers, have more control over gas exchange with the environment, and have weathered mass extinctions well.

Contrary to the hypothesis that ecological differentiation is required to explain increases in taxonomic diversity via increased origination rates (*6*, *7*), the classes that are the most ecologically diverse and taxonomically rich in the modern oceans have not had significantly higher average origination rates over time (Fig. 1). Instead, ecologically diverse classes have become genus-rich owing

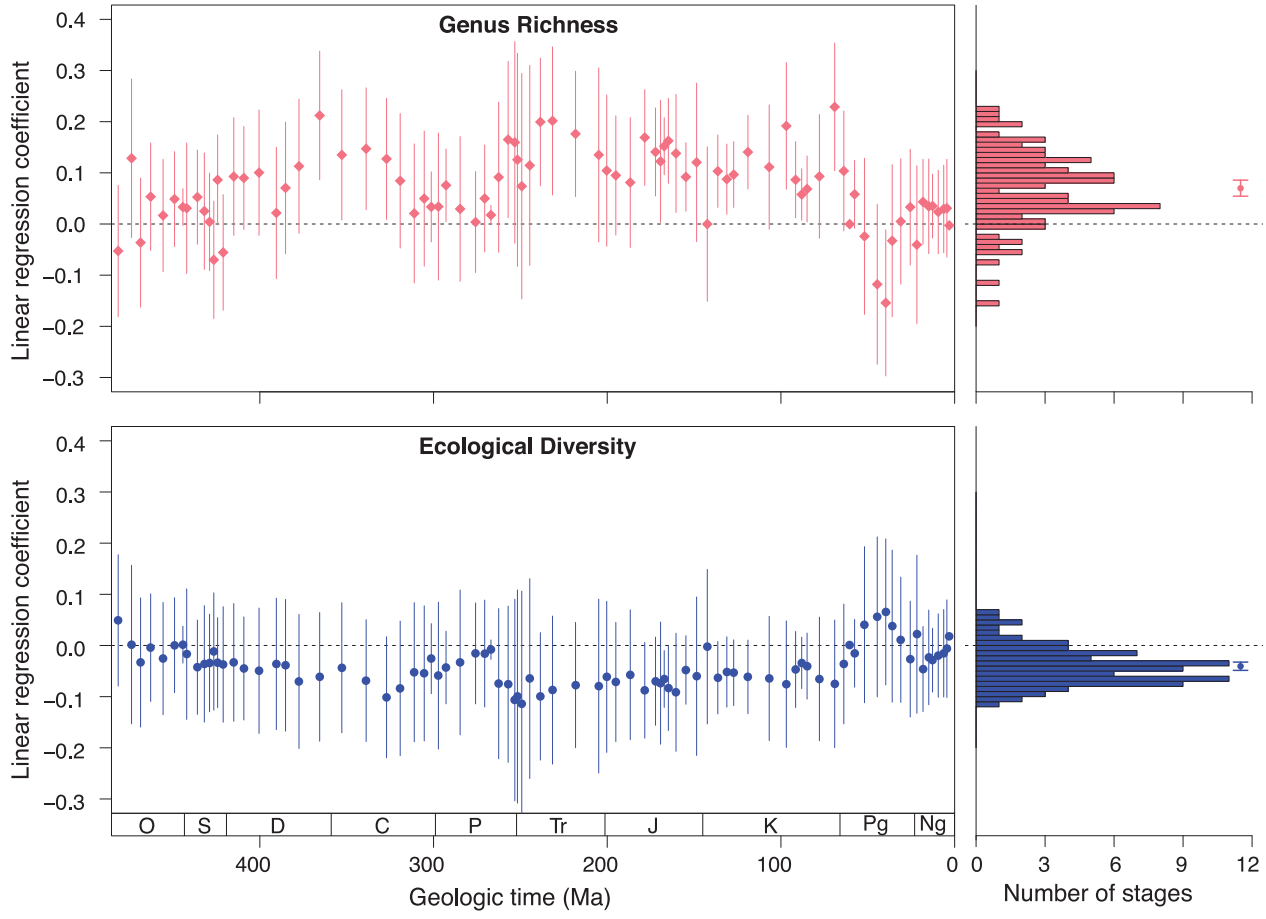


Fig. 4. Volatility across time as a function of the number of genera and the number of ecological modes in Linnaean classes. Histograms (right) display distributions of correlation coefficients for volatility across all geologic stages. Overall Phanerozoic mean plotted with 95% CI to far right for volatility versus genus richness = 0.07 ± 0.02 (top panel) and overall Phanerozoic mean plotted with 95% CI to far right for volatility versus ecological modes = −0.04 ± 0.01 (bottom panel). C, Carboniferous; D, Devonian; J, Jurassic; K, Cretaceous; Ng, Neogene; O, Ordovician; P, Permian; Pg, Paleogene; S, Silurian; Tr, Triassic.

to lower background rates of evolutionary turnover and resistance to mass extinction (Figs. 1 and 3 and figs. S5 to S7). The fable of the tortoise and the hare applies well to marine animal diversification; some clades jumped out to an early diversity lead in the Paleozoic, only to be surpassed by those that combined low volatility, steady diversification, and strong physiological resistance to mass extinctions.

REFERENCES AND NOTES

1. G. E. Hutchinson, *Am. Nat.* **93**, 145–159 (1959).
2. C. Darwin, *On the Origin of Species* (John Murray, 1859).
3. P. R. Grant, B. R. Grant, *Science* **313**, 224–226 (2006).
4. D. Lack, *Auk* **62**, 644–645 (1945).
5. R. H. MacArthur, *Ecology* **39**, 599–619 (1958).
6. G. G. Simpson, *The Major Features of Evolution* (Columbia Univ. Press, 1953).
7. D. Schluter, *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
8. T. H. G. Ezard, T. Aze, P. N. Pearson, A. Purvis, *Science* **332**, 349–351 (2011).
9. R. K. Bambach, A. M. Bush, D. H. Erwin, *Palaeontology* **50**, 1–22 (2007).
10. Materials and methods are available as supplementary materials.
11. A. Ronov, *Am. J. Sci.* **294**, 777–801 (1994).
12. S. E. Peters, N. A. Heim, in *Comparing the Geological and Fossil Records: Implications for Biodiversity Studies*, A. J. McGowan, A. B. Smith, Eds. (Geological Society of London, 2011), pp. 95–104.
13. A. M. Bush, R. K. Bambach, *Geology* **43**, 979–982 (2015).
14. M. L. Knope, N. A. Heim, L. O. Frishkoff, J. L. Payne, *Nat. Commun.* **6**, 6455 (2015).
15. W. J. Foster, R. J. Twitchett, *Nat. Geosci.* **7**, 233–238 (2014).
16. S. M. Edie, D. Jablonski, J. W. Valentine, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 732–737 (2018).
17. N. L. Gilinsky, *Paleobiology* **20**, 445–458 (1994).
18. S. M. Stanley, *Macroevolution: Pattern and Process* (W. H. Freeman, 1979).
19. J. L. Payne *et al.*, *Biol. Lett.* **12**, 20160202 (2016).
20. A. H. Knoll, R. K. Bambach, J. L. Payne, S. Pruss, W. W. Fischer, *Earth Planet. Sci. Lett.* **256**, 295–313 (2007).
21. M. E. Clapham, *Glob. Change Biol.* **23**, 1477–1485 (2017).
22. C. W. Thayer, *Science* **203**, 458–461 (1979).
23. G. J. Vermeij, *Evolution and Escalation: An Ecological History of Life* (Princeton Univ. Press, 1987).
24. M. L. Knope, Ecologically diverse clades dominate the oceans via extinction resistance, Version 3, Dryad (2019); https://doi.org/10.5061/dryad.280gb5mkw.

ACKNOWLEDGMENTS

We thank E. Datlof for assistance with figure preparation, P. Feinberg and A. Romero for assistance collecting data, and D. Koizumi and T. Chong for administrative support. **Funding:** This Report is based on work supported by the U.S. National Science Foundation under grant 1345247 to M.L.K. Support for this research was also provided by the U.S. National Science Foundation (EAR-1151022 to J.L.P.) and the School of Earth, Energy, and Environmental Sciences at Stanford University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. **Author contributions:** All authors designed the study, collected and analyzed the data, and wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Raw data files and code used for all analyses are permanently archived in the Dryad Digital Repository (*24*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/1035/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S5
References (*25*–*33*)

10 April 2019; accepted 14 January 2020
10.1126/science.aax6398

CELL BIOLOGY

Polymerization in the actin ATPase clan regulates hexokinase activity in yeast

Patrick R. Stoddard[1,2], Eric M. Lynch[3], Daniel P. Farrell[3,4], Annie M. Dosey[3], Frank DiMaio[3,4], Tom A. Williams[5], Justin M. Kollman[3], Andrew W. Murray[1,2]*, Ethan C. Garner[1,2]*

The actin fold is found in cytoskeletal polymers, chaperones, and various metabolic enzymes. Many actin-fold proteins, such as the carbohydrate kinases, do not polymerize. We found that Glk1, a *Saccharomyces cerevisiae* glucokinase, forms two-stranded filaments with ultrastructure that is distinct from that of cytoskeletal polymers. In cells, Glk1 polymerized upon sugar addition and depolymerized upon sugar withdrawal. Polymerization inhibits enzymatic activity; the Glk1 monomer-polymer equilibrium sets a maximum rate of glucose phosphorylation regardless of Glk1 concentration. A mutation that eliminated Glk1 polymerization alleviated concentration-dependent enzyme inhibition. Yeast containing nonpolymerizing Glk1 were less fit when growing on sugars and more likely to die when refed glucose. Glk1 polymerization arose independently from other actin-related filaments and may allow yeast to rapidly modulate glucokinase activity as nutrient availability changes.

The actin adenosine triphosphatase (ATPase) clan (*1*) is a diverse group of structurally similar protein families found in all domains of life (*2*). Several of the actin ATPase families form polymers, but the metabolic enzymes, such as hexose kinases, do not (*3*). It is unclear whether polymerization evolved several times within this clan, or whether these polymerizing families descend from a single ancient polymerizing ancestor, a hypothesis suggested by phylogenetic (*4*) and structural studies (*5*).

Cells use several mechanisms to adjust enzyme activity in response to environmental changes, including allosteric and posttranslational regulation. Enzymes can also change their physical state, assembling into filaments or gels, which serves as a sensitive, tunable way to control activity. Enzyme polymerization can regulate flux through pathways (*6*), store enzymes during starvation (*7*), and measure and signal cellular states (*8*).

Hexokinases and glucokinases of fungi belong to a single family (the hexokinase family) within the actin ATPase clan. Fungal glucokinases phosphorylate glucose, mannose, and glucosamine, whereas the fungal hexokinases also phosphorylate fructose. The glucokinases have a higher substrate affinity and lower maximum velocity (V_{max}) than the hexokinases (*9*). *Saccharomyces cerevisiae* has three hexokinase family proteins: a glucokinase (Glk1) and two hexokinases (Hxk1 and Hxk2). Hxk2 is expressed in glucose-rich environments and regulates the expression of the other two enzymes; Hxk1 and Glk1 repression is eliminated without glucose (*10*) (fig. S1).

To probe the cell biology of these isozymes, we made monomeric superfolder green fluorescent protein (msfGFP) fusions to each at their native loci and examined their behavior in cells. Without glucose, Glk1-msfGFP was diffuse. When glucose-starved cells were refed

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. [2]Center for Systems Biology, Harvard University, Cambridge, MA 02138, USA. [3]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. [4]Institute for Protein Design, University of Washington, Seattle, WA 98195, USA. [5]School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, UK.
*Corresponding author. Email: awm@mcb.harvard.edu (A.W.M.); egarner@g.harvard.edu (E.C.G.)

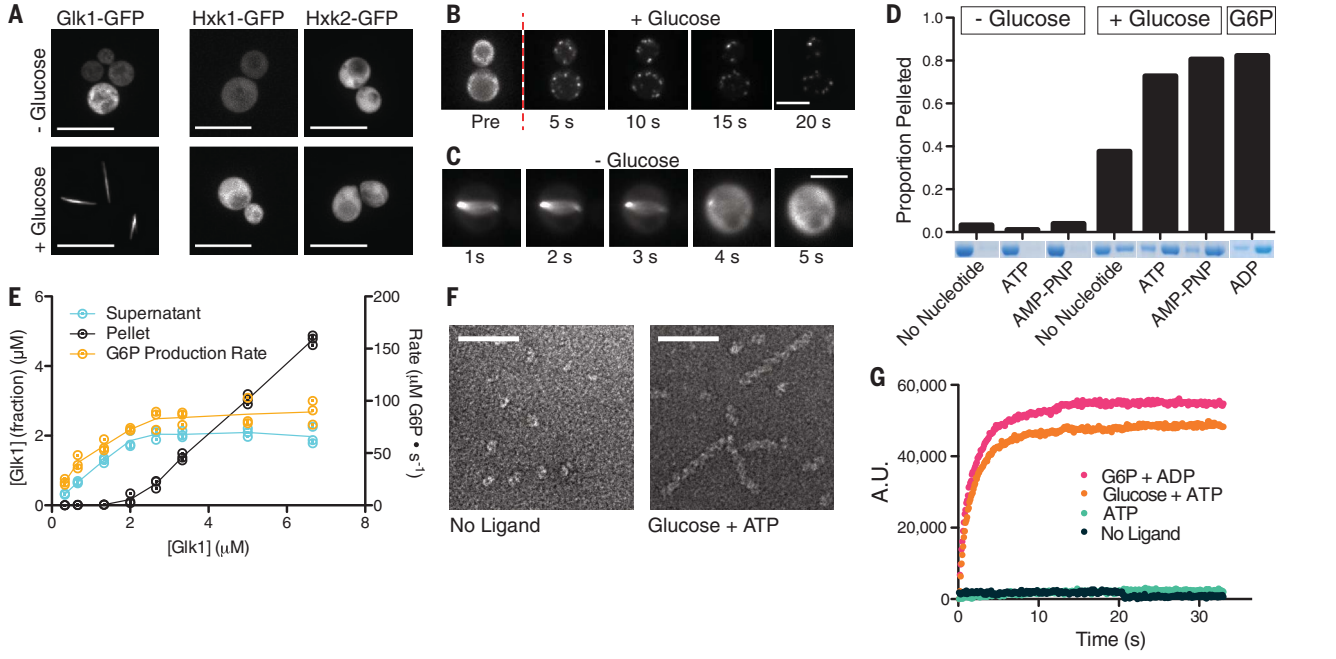


Fig. 1. Glk1 forms filaments in response to its substrates at high enzyme concentration. (**A**) Fluorescence images of stationary-phase Glk1-msfGFP (left), Hxk1-msfGFP (middle), and Hxk2-msfGFP (right) cells before (top) or after (bottom) glucose addition. Scale bars: 10 μm. (**B** and **C**) Fluorescence images of Glk1-msfGFP cells from a time-lapse video, as glucose is added (top) or removed (bottom). Puncta in (B) eventually coalesce into a single filamentous structure, as in (A) and (C). Scale bars: 5 μm. (**D**) Purified Glk1 was ultracentrifuged at 436,000*g* for 30 min with different ligand combinations. The supernatant (left) and pellet (right) of each condition were subjected to SDS–polyacrylamide gel electrophoresis. AMP-PNP: adenylyl-imidodiphosphate, a nonhydrolyzable ATP analog. (**E**) Purified Glk1 concentration was varied in saturating glucose and ATP and assayed for enzyme activity (glucose-6-phosphate production rate) or ultracentrifuged. The concentration of Glk1 in the supernatant and pellet was measured (n = 3 technical replicates). Means are connected by lines. (**F**) Electron micrographs of negatively stained samples: 7.5 μM Glk1 in the absence of ligands (left) or in saturating glucose and ATP (right). Scale bars: 50 nm. (**G**) 90° light scattering of 7.5 μM Glk1 mixed with different ligand combinations. A.U., arbitrary units.

glucose, Glk1-msfGFP formed filamentous structures, which rapidly disassembled upon glucose removal (Fig. 1, A to C, and movies S1 and S2). Glk1-msfGFP also polymerized upon refeeding with other Glk1 substrates: mannose or glucosamine (figs. S2 and S3). Hxk1-msfGFP and Hxk2-msfGFP did not oligomerize (Fig. 1A).

To understand what regulates Glk1 polymerization, we studied it in vitro. Other enzymes form condensates in low pH (*7*), but Glk1 does not (fig. S4). We found that Glk1 polymerized in the presence of its substrates [adenosine triphosphate (ATP) and glucose, mannose, or glucosamine] or its products [adenosine diphosphate (ADP) and sugar-6-phosphate]. Modest polymerization also occurred with Glk1 inhibitors (Fig. 1D and figs. S5 and S6) (*9*). Although fructose and galactose induced Glk1 polymerization in vivo, they did not do so in vitro, suggesting that, in vivo, polymers assemble when cells convert these sugars into glucose-6-phosphate (G6P) (figs. S2, S3, and S7) (*11*, *12*).

Like actin, Glk1 exhibits a critical concentration (CC) for polymerization. Below 2 μM Glk1, there was no polymerization. Above 2 μM, the concentration of Glk1 polymer increased, while unpolymerized Glk1 remained constant (Fig. 1E). This is consistent with the lack of polymers in fermenting cells where Glk1 expression is suppressed by glucose. Indeed, Glk1-msfGFP polymers were observed in glucose when Glk1-msfGFP was expressed from a strong promoter (fig. S9). In vitro Glk1 polymerization reached steady state in a matter of seconds (Fig. 1G), which is consistent with the rapid polymerization observed in vivo.

Polymerization can either activate or inhibit enzyme activity (*6*, *13*), so we measured Glk1 activity as we varied the concentration of Glk1. Below Glk1's CC, the G6P production rate increased with increasing concentration of Glk1. Above the CC, the rate of product formation was constant (Fig. 1E). Thus, polymerization inhibits Glk1 activity, and the monomer-polymer equilibrium caps monomer concentration, thereby keeping net enzymatic activity constant.

We used electron microscopy of negatively stained samples to examine Glk1 oligomers. Glk1 formed helical filaments in the presence of substrates (Fig. 1F). The micrometer-scale structures seen in vivo are likely driven by crowding (*14*), filament binding proteins (*15*, *16*), or dimerization of the fluorescent tag (*17*). Similar polymers were observed when Glk1 was fused to other monomeric fluorescent proteins (fig. S8).

To investigate why Glk1 polymerizes but Hxk1 and Hxk2 do not, we solved the crystal structure of Glk1 (table S1) (*18*). Comparing this structure to that of *S. cerevisiae* Hxk2 (*19*) revealed differences in key regions: the N and C termini and two loops (Fig. 2A and figs. S10 and S11). We then used cryo–electron microscopy (cryo-EM) to determine the structure of Glk1 filaments (table S2) (*20*). This 3.8-Å-resolution structure (Fig. 2B) revealed that Glk1 formed two-stranded, antiparallel filaments. Similar to other actin-like filaments (*4*), subunits were in the closed state and ligand-bound (Fig. 2C and fig. S13) (*21*) but, unlike actin, did not flatten. Glk1 homologs alternate between open and closed states during their catalytic cycle (*21*), thus Glk1 inhibition likely arises from the inability to transition between states within the polymer.

The interactions between Glk1 subunits along a strand differ from the conserved interactions in other actin ATPase clan polymers (*5*) (fig. S14). Along the strand, the N-terminal, solvent-exposed Phe^3 of one subunit inserts into the hydrophobic pocket at the C terminus

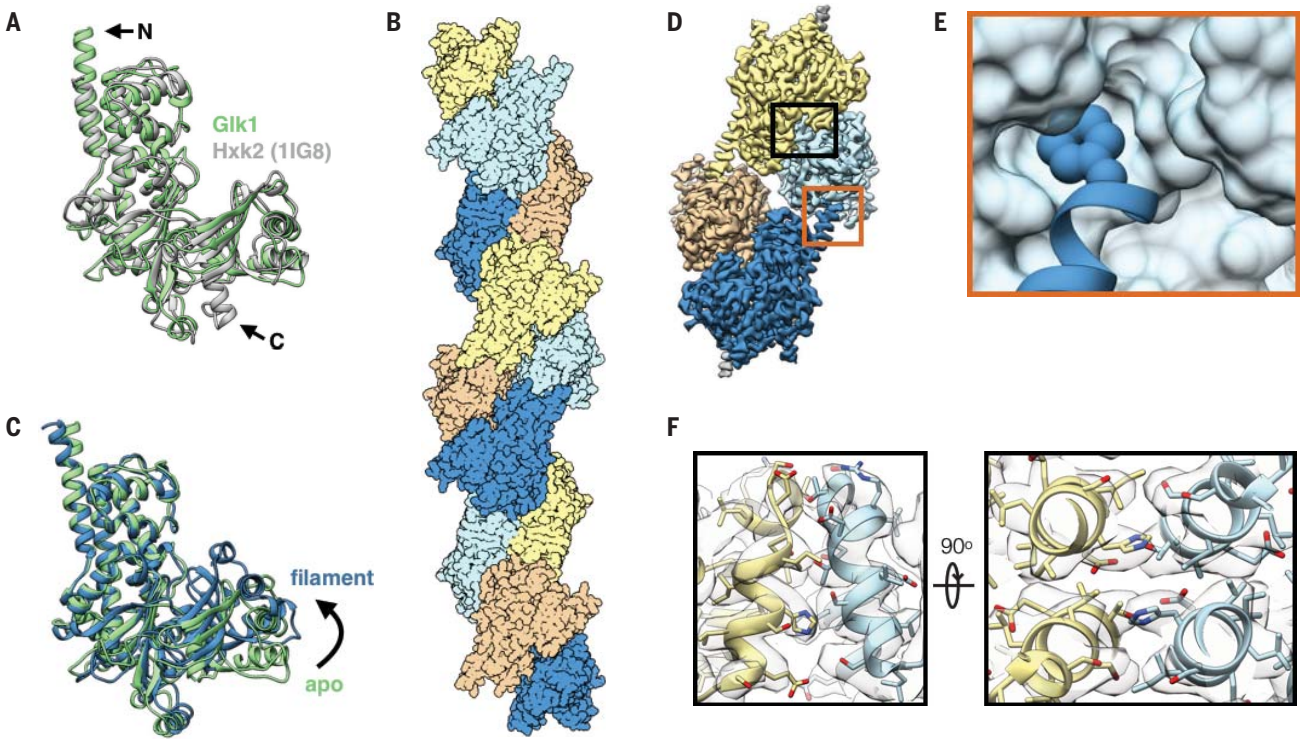


Fig. 2. Glk1 forms antiparallel, two-stranded filaments in its closed state. (**A**) Superimposition of Glk1 crystal structure (green; residues 1 to 500) with Hxk2 (PDB ID 1IG8) (*19*) (white; residues 18 to 486). The N-terminal helix of Glk1 extends farther than that of Hxk2 (arrow labeled N), whereas the C-terminal helix of Hxk2 extends farther than that of Glk1 (arrow labeled C). (**B**) Surface representation of a model of Glk1 filaments reconstructed from cryo-EM (3.8-Å resolution). Glk1 filaments are two-stranded, antiparallel helices. Subunits along each strand are either orange and yellow or blue and cyan. (**C**) Superimposition of the Glk1 crystal structure (green) with the Glk1 filament conformation from the cryo-EM reconstruction (blue). The crystal structure is not ligand-bound and is in the open state, while the filament form is ligand-bound and is in the closed state. (**D**) Cryo-EM map of four subunits in the Glk1 filament colored by subunit. Longitudinal contact is indicated with an orange box, and lateral contact with a black box. (**E**) Close-up of longitudinal contact with Phe^3 represented as van der Waals spheres, and the next subunit represented as a surface model. Phe^3 of one subunit inserts into the hydrophobic pocket near the next subunit's C terminus. (**F**) Two orthogonal close-ups of lateral filament contact. The helix-loop-helix from residue 371 to 393 of one subunit (yellow) binds antiparallel to the same region of the adjacent subunit (blue). Cryo-EM map is transparent gray.

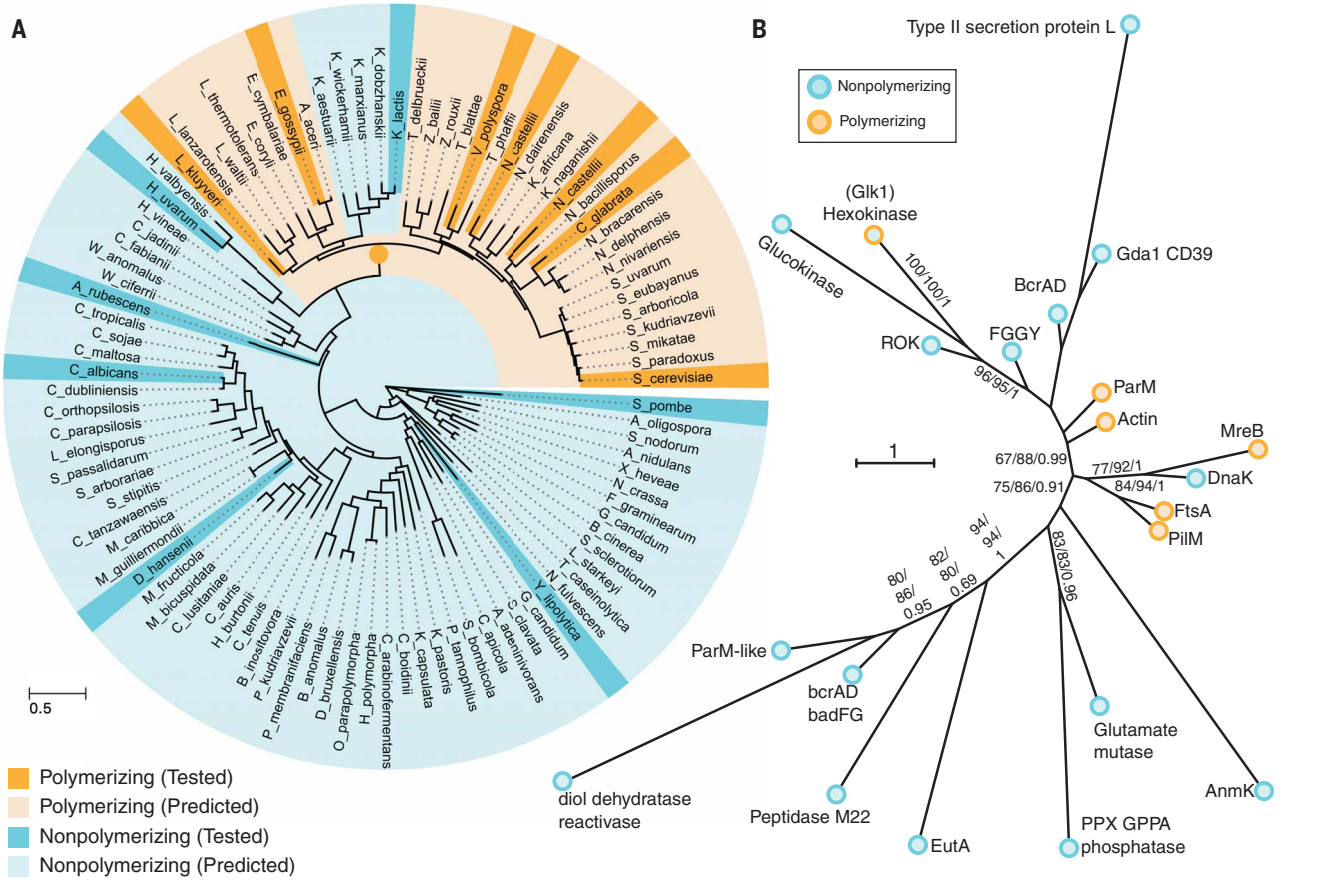


Fig. 3. Glucokinase polymerization evolved independently of other actin-related polymers. (**A**) Tree of ascomycetes, calculated by (*23*), indicating which species Glk1 homologs do (dark orange) and do not (dark cyan) polymerize. Species whose Glk1 homologs are predicted to polymerize on the basis of conserved motifs (pale orange) and those predicted to not polymerize (pale cyan) are also indicated. Dark-orange circle marks the likely origin of Glk1 polymerization. Scale bar: 0.5 expected changes per site. (**B**) Phylogeny of actin ATPase families, summarizing phylogenetic analysis of 802 sequences from actin ATPase protein families. A maximum likelihood (ML) tree was inferred under the LG+C20 substitution model in IQ-Tree (*28*). This displays the backbone structure of that ML tree with each family collapsed. Support values indicated are ultrafast bootstrap/SH-aLRT/aBayes. Much of the backbone is uncertain; bootstrap supports shown when SH-aLRT (middle value) > 70. This tree suggests that the hexokinase family, which contains Glk1, forms a clade with ROKs and glucokinases and is only distantly related to other polymer-forming actin families. Families that do not polymerize are cyan, while families that do polymerize are orange. Scale bar: 1 expected change per site. The full phylogeny is available online (see supplementary materials).

of the next (Fig. 2, D and E, and fig. S15). Lateral associations between strands are mediated by the helix-loop-helix between residues 371 and 393, which binds antiparallel to the same region on the adjacent subunit (Fig. 2F).

Phylogenetically, the fungal glucokinases and hexokinases form separate clades. The group of yeast containing *S. cerevisiae* (the Saccharomycetaceae) arose ~200 million years ago. The Glk1 homologs in most Saccharomycetaceae contain four conserved motifs that are missing in both the other Glk1 homologs and all Hxk1 and Hxk2 homologs. These motifs are at, or near, filament contacts: the N and C termini, loop 230 to 243, and loop 438 to 444. To test whether these motifs predict polymerization, we purified different Glk1 and Hxk1 or Hxk2 homologs and tested their ability to polymerize. Only Glk1 homologs that contained all four motifs polymerized (figs. S10 and S16). Phylogenetic logistic regression showed these motifs correlate significantly with polymerization (P = 0.018). These results suggest that Glk1 polymerization arose ~200 million years ago. One ascomycete lineage, the Kluveromycetes, lost this ability (Fig. 3A) (*22*, *23*).

The hexokinase family, which contains Glk1, segregates from the known polymer-forming actin ATPase families (Fig. 3B). This pattern is consistent with pairwise similarity between the hidden Markov models of each family (fig. S17 and table S3), and the following lines of evidence suggest that Glk1 polymerization evolved independently of other actin-fold polymers: (i) polymerizing Glk1 sequences form a subclade within the hexokinase family (Fig. 3A); (ii) the broader hexokinase family is monophyletic in the global actin-fold tree [100% bootstrap, 100% Shimodaira-Hasegawa approximate likelihood-ratio test (SH-aLRT), 1 aBayes] (Fig. 3B); (iii) hexokinases robustly group with the nonpolymerizing glucokinases and the repressor, ORF, kinase family (ROKs) (96% bootstrap, 95% SH-aLRT, 1 aBayes); and (iv) the monophyly of Glk1 with other polymerizing actins was rejected by an approximately unbiased test (P = 0.00106).

Next, we examined how disrupting Glk1's polymerization affected enzymatic activity and cell physiology. To create nonpolymerizing Glk1 (NonPol-Glk1), we mutated the N-terminal phenylalanine involved in intersubunit contacts to serine (Glk1-F3S; F3S, Phe^{3}→Ser). This mutation eliminated polymerization both in vitro and in vivo (fig. S18, A and B). NonPol-Glk1 was enzymatically active but lacked the concentration-dependent inhibition of wild-type Glk1 (fig. S18C).

To distinguish between the cellular effects of a lack of inhibition and the absence of polymers, we mutated the catalytic lysine (K182A) (fig. S13B) (*24*) to create catalytically dead Glk1 (CatDead-Glk1). We combined these mutations (F3S and K182A) to create nonpolymerizing, catalytically dead Glk1 (NonPolCatDead-Glk1). CatDead-Glk1 formed polymers in vivo and in

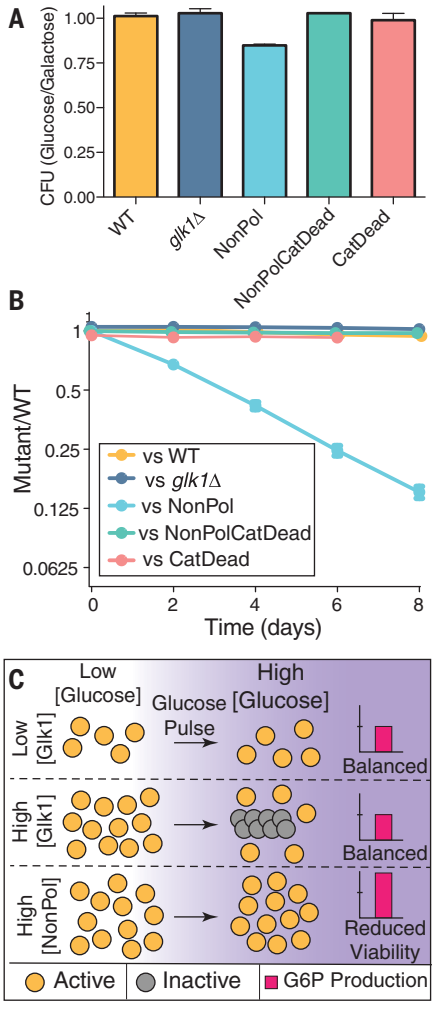


Fig. 4. Elimination of Glk1 polymerization reduces fitness. (**A**) Cells were preconditioned in citrate buffered synthetic (CBS) medium with galactose and refed either glucose or galactose. The ratios of the resulting colonies (CFU, colony-forming units) are reported here. Mean ± SD (n = 4 biological replicates). (**B**) Wild-type cells expressing mCherry were competed against cells expressing GFP with different Glk1 genotypes through growth and dilution cycles in synthetic medium with glucose. The proportion of strains was measured after dilution by flow cytometry. Mean ± SD (n = 5 biological replicates). (**C**) Schematic of how Glk1 polymerization affects glucokinase activity. When Glk1 concentration is high and glucose increases, Glk1 polymerizes until the monomer concentration equals the CC. Glk1 polymers lack enzyme activity; regardless of Glk1 concentration, the concentration of active enzyme is the same after glucose addition. When Glk1's ability to polymerize is disrupted, its glucokinase activity is unconstrained, leading to fitness and viability defects.

vitro but did not generate G6P. NonPolCatDead-Glk1 neither polymerized nor produced G6P (fig. S18, A to C).

When starved yeast are refed glucose, excess sugar kinase activity is toxic owing to an imbalance between the early steps of glycolysis, which consume ATP, and the late steps, which generate ATP (*25*). Hxk1 and Hxk2 activity is inhibited by trehalose-6-phosphate, a transiently accumulating metabolite (*26*). Because Glk1 is not inhibited by trehalose-6-phosphate, we hypothesized that Glk1 polymerization limits activity when starved cells are refed glucose. When cells grown in galactose were refed glucose, 15% of NonPol-Glk1 cells died (Fig. 4A), which is consistent with this model. This death was caused by unregulated Glk1 activity and not by lack of Glk1 polymers; cells with GLK1 deleted (*glk1*Δ), CatDead-Glk1, and NonPolCatDead-Glk1 behaved indistinguishably from wild-type cells. Thus, Glk1 polymerization limits the rate of glucose phosphorylation during glucose refeeding.

Unregulated Glk1 activity is detrimental to fitness during the entire growth cycle. We used differential fluorescent labeling to compare the fitness of wild-type cells with that of each mutant. We grew mixed cultures to saturation, diluted them into fresh medium every 48 hours, and measured the proportion of strains by flow cytometry. NonPol-Glk1 cells had a substantial fitness defect in these conditions, averaging to a fitness cost of 6% (Fig. 4B) (*27*). In contrast, no growth defect was observed when mixed cultures were maintained in a glucose-rich environment (fig. S19), suggesting that environmental changes are required to expose the growth defects of NonPol-Glk1 cells. Similar effects were observed when competing these strains on other sugars (fig. S20). Cells lacking Glk1 activity (*glk1*Δ, CatDead-Glk1, NonPolCatDead-Glk1) showed minor growth defects in acetate. Thus, Glk1 activity is important for growth on non-sugar sources, and Glk1 polymerization prevents toxic overactivity during refeeding.

Glk1 polymerization governs its bulk rate of catalysis, with the Glk1 CC setting the upper limit of flux throughout the entire Glk1 pool (Fig. 4C). This mode of self-regulation is robust to growth state and cell-to-cell variation in protein concentration, and it allows rapid adaptation to transient perturbations. In this sense, Glk1 polymerization behaves as a molecular surge protector, defending the cell against nutrient spikes.

REFERENCES AND NOTES

1. S. El-Gebali *et al.*, *Nucleic Acids Res.* **47**, D427–D432 (2019).
2. B. Wickstead, K. Gull, *J. Cell Biol.* **194**, 513–525 (2011).
3. P. Bork, C. Sander, A. Valencia, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7290–7294 (1992).
4. P. R. Stoddard, T. A. Williams, E. Garner, B. Baum, *Mol. Biol. Cell* **28**, 2461–2469 (2017).
5. F. van den Ent, T. Izoré, T. A. Bharat, C. M. Johnson, J. Löwe, *eLife* **3**, e02634 (2014).
6. R. M. Barry *et al.*, *eLife* **3**, e03638 (2014).
7. I. Petrovska *et al.*, *eLife* **3**, e02409 (2014).
8. T. M. Franzmann *et al.*, *Science* **359**, eaao5654 (2018).
9. P. K. Maitra, *J. Biol. Chem.* **245**, 2423–2431 (1970).
10. A. Rodríguez, T. De La Cera, P. Herrero, F. Moreno, *Biochem. J.* **355**, 625–631 (2001).
11. P. A. Frey, *FASEB J.* **10**, 461–470 (1996).
12. J. B. Green, A. P. Wright, W. Y. Cheung, W. E. Lancashire, B. S. Hartley, *Mol. Gen. Genet.* **215**, 100–106 (1988).
13. E. M. Lynch *et al.*, *Nat. Struct. Mol. Biol.* **24**, 507–514 (2017).
14. F. Hilitski *et al.*, *Phys. Rev. Lett.* **114**, 138102 (2015).
15. L. M. Griffith, T. D. Pollard, *J. Biol. Chem.* **257**, 9135–9142 (1982).
16. H. Takatsuki, E. Bengtsson, A. Månsson, *Biochim. Biophys. Acta* **1840**, 1933–1942 (2014).
17. D. Landgraf, B. Okumus, P. Chien, T. A. Baker, J. Paulsson, *Nat. Methods* **9**, 480–482 (2012).
18. P. R. Stoddard, E. C. Garner, A. W. Murray, PDB ID 6P4X (2020); https://doi.org/10.2210/pdb6P4X/pdb.
19. P. R. Kuser, S. Krauchenco, O. A. Antunes, I. Polikarpov, *J. Biol. Chem.* **275**, 20814–20821 (2000).
20. E. M. Lynch, A. M. Dosey, D. P. Farrell, P. R. Stoddard, J. M. Kollman, PDB ID 6PDT (2020); https://doi.org/10.2210/pdb6PDT/pdb.
21. E. B. Kuettner *et al.*, *J. Biol. Chem.* **285**, 41019–41033 (2010).
22. A. Hagman, T. Säll, J. Piškur, *FEBS J.* **281**, 4805–4814 (2014).
23. X.-X. Shen *et al.*, *G3* **6**, 3927–3939 (2016).
24. J. Zhang *et al.*, *PLOS ONE* **4**, e6304 (2009).
25. J. H. van Heerden *et al.*, *Science* **343**, 1245114 (2014).
26. M. A. Blázquez, R. Lagunas, C. Gancedo, J. M. Gancedo, *FEBS Lett.* **329**, 51–54 (1993).
27. F. Duveau *et al.*, *eLife* **7**, e37272 (2018).
28. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, *Mol. Biol. Evol.* **32**, 268–274 (2015).

ACKNOWLEDGMENTS

Special thanks to Q. Justman for gifting strains and for extended discussions of this work. We thank R. Gaudet for sharing beamtime and resources and for crystallographic advice, along with C. Zimanyi, L. Bane, and E. May. We thank J. Wilhelm for helpful discussions. We thank the Arnold and Mabel Beckman Cryo-EM Center at the University of Washington for access to electron microscopes. This work used NE-CAT beamlines (GM124165) at the APS (DE-AC02-06CH11357). **Funding:** This work was supported by NIH grants DP2AI117923-01 to E.C.G., R01GM043987 to A.W.M., R01GM118396 to J.M.K., R01GM123089 to F.D., F31GM116441 to P.R.S., and by the NSF-Simons Center for Mathematical and Statistical Analysis of Biology at Harvard (1764269) and the Harvard Quantitative Biology Initiative. E.C.G. and P.R.S. were also supported by Wellcome Grant 203276/Z/16/Z and by the Volkswagen Foundation. T.A.W. is supported by a Royal Society University Research Fellowship. **Author contributions:** P.R.S. cloned *S. cerevisiae* strains; performed fluorescence microscopy; measured the relative expression of Glk1 by flow cytometry; carried out plasmid construction and protein purification, in vitro measurements of polymerization and enzymatic activity, x-ray crystallography, Glk1 crystal structure refinement, and negative-stain EM; and measured the viability of cells during glucose refeeding and the fitness of competed strains using flow cytometry. Cryo-EM sample preparation and data collection were performed by E.M.L. and A.M.D. Cryo-EM data were analyzed by E.M.L. and J.M.K., and the atomic model of Glk1 filaments was built by E.M.L., D.P.F., and F.D. Phylogenetic analysis of actin ATPase families and regression analysis of polymerization-associated motifs were performed by T.A.W. This project was conceived of by P.R.S., E.C.G., and A.W.M. P.R.S., E.M.L., and T.A.W. generated figures for this work. The paper was written by P.R.S., E.C.G., and A.W.M. and edited by P.R.S., E.C.G., A.W.M., E.M.L., J.M.K., and T.A.W. **Competing interests:** The authors have no competing interests. **Data and materials availability:** Atomic coordinate files of the Glk1 crystal structure (PDB ID 6P4X) and the cryo-EM filament structure (PBD ID 6PDT) are hosted online at the RCSB PDB (www.rcsb.org). The EM map of the Glk1 cryo-EM filament structure is hosted online at the EM Data Resource (www.emdataresource.org/) under accession number EMD-20309. All other data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/1039/suppl/DC1
Materials and Methods
Figs. S1 to S20
Tables S1 to S4
References (*29*–*60*)
Movies S1 and S2
Tree Description File for Fig. 3B

View/request a protocol for this paper from *Bio-protocol*.

26 June 2019; accepted 27 January 2020
10.1126/science.aay5359

NEUROSCIENCE

Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody

Philippe Albouy[1,2,3]*, Lucas Benjamin[1], Benjamin Morillon[4]†, Robert J. Zatorre[1,2]†

Does brain asymmetry for speech and music emerge from acoustical cues or from domain-specific neural networks? We selectively filtered temporal or spectral modulations in sung speech stimuli for which verbal and melodic content was crossed and balanced. Perception of speech decreased only with degradation of temporal information, whereas perception of melodies decreased only with spectral degradation. Functional magnetic resonance imaging data showed that the neural decoding of speech and melodies depends on activity patterns in left and right auditory regions, respectively. This asymmetry is supported by specific sensitivity to spectrotemporal modulation rates within each region. Finally, the effects of degradation on perception were paralleled by their effects on neural classification. Our results suggest a match between acoustical properties of communicative signals and neural specializations adapted to that purpose.

Speech and music represent the most cognitively complex, and arguably uniquely human, use of sound. To what extent do these two domains depend on separable neural mechanisms? What is the basis for such specialization? Several studies have proposed that left hemisphere neural specialization of speech (*1*) and right hemisphere specialization of pitch-based aspects of music (*2*) emerge from differential analysis of acoustical cues in the left and right auditory cortices (ACs). However, domain-specific accounts suggest that speech and music are processed by dedicated neural networks, the lateralization of which cannot be explained by low-level acoustical cues (*3*–*6*).

Despite consistent empirical evidence in its favor, the acoustical cue account has been computationally underspecified: Concepts such as spectrotemporal resolution (*7*–*9*), time integration windows (*10*), and oscillations (*11*) have all been proposed to explain hemispheric specializations. However, it is difficult to test these concepts directly within a neurally viable framework, especially using naturalistic speech or musical stimuli. The concept of spectrotemporal receptive fields (*12*) provides a computationally rigorous and neurophysiologically plausible approach to the neural decomposition of acoustical cues. This model proposes that auditory neurons act as spectrotemporal modulation (STM) rate filters, based on both single-cell recordings in animals (*13*, *14*) and neuroimaging in humans (*15*, *16*). STM may provide a mechanistic basis to account for lateralization in AC (*17*), but a direct relationship among acoustical STM features, hemispheric asymmetry, and behavioral performance during processing of complex signals such as speech and music has not been investigated.

We created a stimulus set in which 10 original sentences were crossed with 10 original melodies, resulting in 100 naturalistic a cappella songs (Fig. 1) (stimuli are available at www.zlab.mcgill.ca/downloads/albouy_20190815/). This orthogonalization of speech and melodic domains across stimuli allows the dissociation of speech-specific (or melody-specific) from nonspecific acoustic features, thereby controlling for any potential acoustic bias (*3*). We created two separate stimulus sets, one with French and one with English sentences, to allow for reproducibility and to test generality across languages. We then parametrically degraded each stimulus selectively in either the temporal or spectral dimension using a manipulation that decomposes the acoustical signal using the STM framework (*18*).

We first investigated the importance of STM rates on sentence or melody recognition scores in a behavioral experiment (Fig. 2A). Native French (n = 27) and English (n = 22) speakers were presented with pairs of stimuli and asked to discriminate either the speech or the melodic content. Thus, the stimulus set across the two tasks was identical; only the instructions differed. The degradation of information in the temporal dimension impaired sentence recognition ($t_{(48)}$ = 13.61 < 0.001, one-sample t test against zero of the slope of the linear fit relating behavior to the degree of acoustic degradation) but not melody recognition ($t_{(48)}$ =

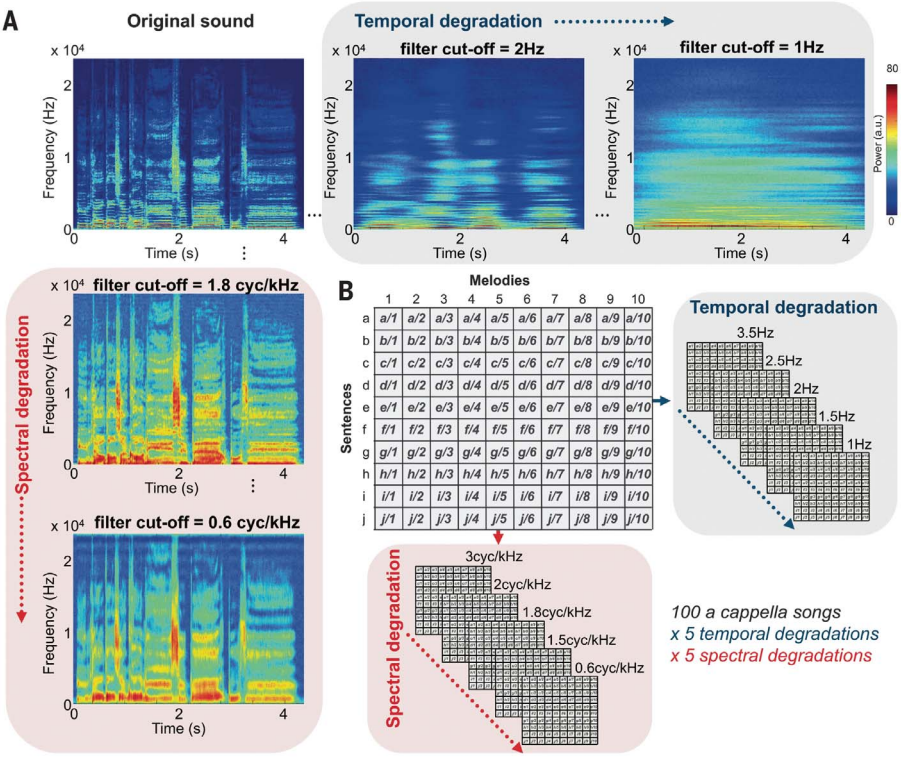


Fig. 1. Spectrotemporal filtering and stimulus set. (**A**) Spectral and temporal degradations applied on an original a cappella song. (**B**) One hundred a cappella songs in each language were recorded following a 10 × 10 matrix with 10 melodies (number code) and 10 sentences (letter code). Stimuli were then filtered either in the spectral or in the temporal dimension with five filter cutoffs, resulting in 1000 degraded stimuli for each language.

[1]Cognitive Neuroscience Unit, Montreal Neurological Institute, McGill University, Montreal, QC, Canada. [2]International Laboratory for Brain, Music and Sound Research (BRAMS); Centre for Research in Brain, Language and Music; Centre for Interdisciplinary Research in Music, Media, and Technology, Montreal, QC, Canada. [3]CERVO Brain Research Centre, School of Psychology, Laval University, Quebec, QC, Canada. [4]Aix Marseille University, Inserm, INS, Institut de Neurosciences des Systèmes, Marseille, France.
*Corresponding author. Email: philippe.albouy@psy.ulaval.ca
†These authors contributed equally to this work.

0.62, $p = 0.53$), whereas degradation of information in the spectral dimension impaired melody recognition ($t_{(48)} = 8.24 < 0.001$) but not sentence recognition ($t_{(48)} = -1.28$, $p = 0.20$; Fig. 2, B and C). This double dissociation was confirmed by a domain-by-degradation interaction (2 × 2 repeated-measures ANOVA: $F_{(1,47)} = 207.04$, $p < 0.001$). Identical results were observed for the two language groups (see fig. S2 and the supplementary results for complementary analyses).

We then investigated the impact of STM rates on the neural responses to speech and melodies using functional magnetic resonance imaging (fMRI). Blood oxygenation level–dependent (BOLD) activity was recorded while 15 French speakers who had participated in the behavioral experiment listened to blocks of five songs degraded either in the temporal or spectral dimension. Participants attended to either the speech or the melodic content (Fig. 3A). BOLD signal in bilateral ACs scaled with both temporal and spectral degradation cutoffs [i.e., parametric modulation with quantity of temporal or spectral information; $p < 0.05$ familywise error (FWE) corrected; Fig. 3B and table S1]. These regions were located lateral to primary ACs and correspond to the ventral auditory stream of information processing, covering both parabelt areas and the lateral anterior superior temporal gyrus [parabelt and auditory area 4 (A4) regions; see (*19*)], but there was no significant difference in the hemispheric response to either dimension (whole-brain two-sample t tests; all $p > 0.05$).

To investigate more fine-grained encoding of speech and melodic contents, we performed a multivariate pattern analysis on the fMRI data. Ten-category classifications (separately for melodies and sentences) using whole-brain searchlight analyses (support vector machine, leave-one-out cross-validation procedure, cluster corrected) revealed that the neural encoding of sentences significantly depends on neural activity patterns in left A4 [TE.3; subregion of AC; see (*19*)], whereas the neural decoding of melodies significantly depends on neural activity patterns in right A4 ($p < 0.05$ cluster corrected; Fig. 3, C and D, and table S1; other, subthreshold clusters are reported in fig. S3). To ensure that this effect was generalizable to the population, we performed a complementary information prevalence analysis within temporal lobe masks (see the materials and methods). For the decoding of sentences, a prevalence value of up to 70% was observed in left A4 ($p = 0.02$, corrected), whereas a prevalence value of up to 69% was observed for the decoding of melodies in right A4 ($p = 0.03$, corrected; see table S1). Finally, we tested whether the classification accuracy was better for sentence or melody in the right or the left hemisphere. We computed a lateralization index on accuracy scores [(R – L)/(R + L)] and observed a significant asymmetry in opposite directions for the two domains in region A4 (Fig. 3F, table S1, and fig. S4; $p < 0.05$, cluster corrected at the whole-brain level).

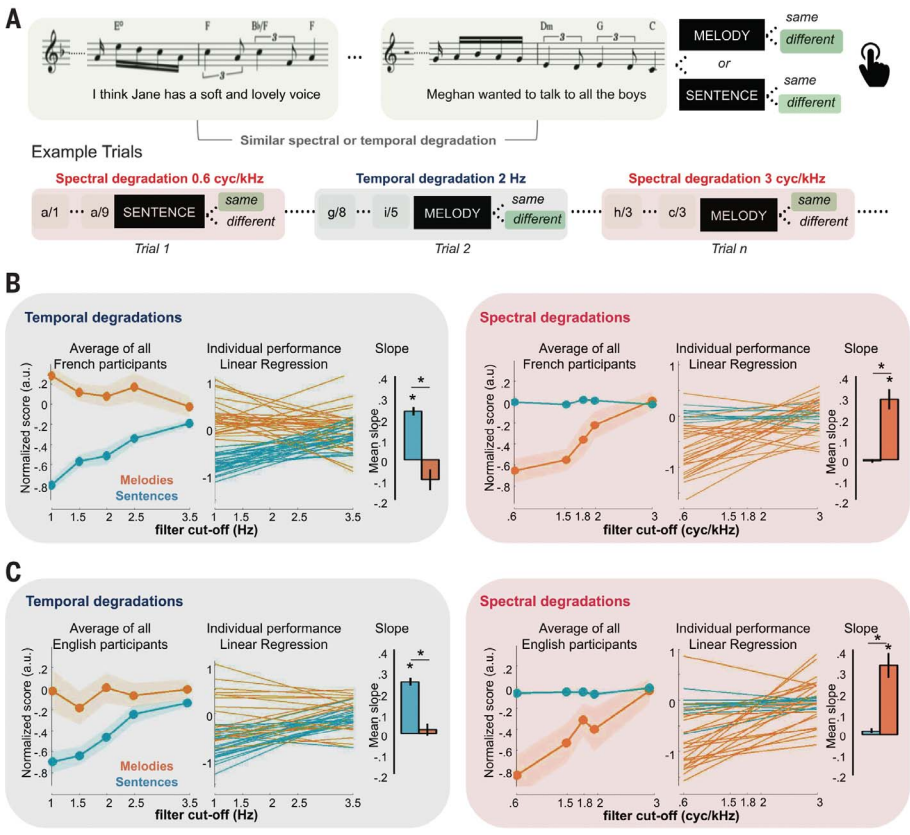


Fig. 2. Behavioral experiment. (**A**) Participants listened to degraded (either in the spectral or temporal dimension) a cappella songs presented in pairs. After the second song, a visual instruction indicated the domain of interest (sentences or melodies). Lower panel shows example trials. (**B**) Behavioral performance of French-speaking listeners. Aqua shading indicates temporal degradations and orange shading indicates spectral degradations. Average performance across participants (95% confidence interval) and individual performance modeled with linear regression are shown for both types of degradations. (**C**) Same as (B) but for English-speaking listeners. Error bars indicate SEM. Asterisks indicate significant differences.

We then tested the relationship between neural specialization of left and right hemispheres for speech and melodic contents and behavioral processing of these two domains. We estimated linear and nonlinear statistical dependencies by computing the normalized mutual information [NMI (*20*)] between the confusion matrices extracted from classification of neural data (whole brain, for each searchlight) and those from behavioral data recorded offline (for each participant and each domain). To investigate the correspondence between neural and behavioral patterns (pattern of errors) instead of mere accuracy (diagonal), these analyses were done after removing the diagonal information (Fig. 4A). NMI was significantly higher in left than right A4 for sentences, whereas the reverse pattern was observed for melodies, as measured by the lateralization index ($p < 0.05$, cluster corrected; see the materials and methods, table S1, and fig. S5).

We next tested whether the origin of the observed lateralization was related to attentional processes by investigating the decoding accuracy and NMI lateralization index as a function of attention to sentences or melodies. Whole-brain analyses did not reveal any significant cluster, suggesting that the previously observed hemispheric specialization is robust to attention and thus is more likely to be linked to automatic than to top-down processes (see fig. S6 and the supplementary results for details).

Finally, we investigated whether the hemispheric specialization for speech and melodic contents was directly related to a differential acoustic sensitivity of left and right ACs to STMs, as initially hypothesized. We estimated the impact of temporal or spectral degradations on decoding accuracy by computing the accuracy change (with negative indicating accuracy loss and positive indicating accuracy gain) between decoding accuracy computed on all trials (all degradation types) and on a

specific degradation type (temporal or spectral). We observed a domain-by-degradation interaction in bilateral ACs (left and right area A4; $p < 0.05$, cluster corrected; Fig. 4C and fig. S7). For sentences, accuracy loss was observed only in the left A4 for temporal as compared with spectral degradations ($p < 0.001$, Tukey corrected; all others, $p > 0.16$), whereas the reverse pattern was observed for melodies only in right A4 ($p = 0.003$, Tukey corrected; all others, $p > 0.29$).

This differential sensitivity to acoustical cues in left and right ACs was also observed in the brain–behavior relationship. We investigated the effect of degradations on the NMI lateralization index. We first show a significant domain-by-degradation interaction observed in area A4 ($p < 0.05$, cluster corrected; Fig. 4D, left; table S1; and fig. S8). The main effect of degradation (temporal > spectral) was then analyzed with two-sample t tests for each domain to reveal that the NMI lateralization index was affected in opposite directions by temporal and spectral degradations (A4 and

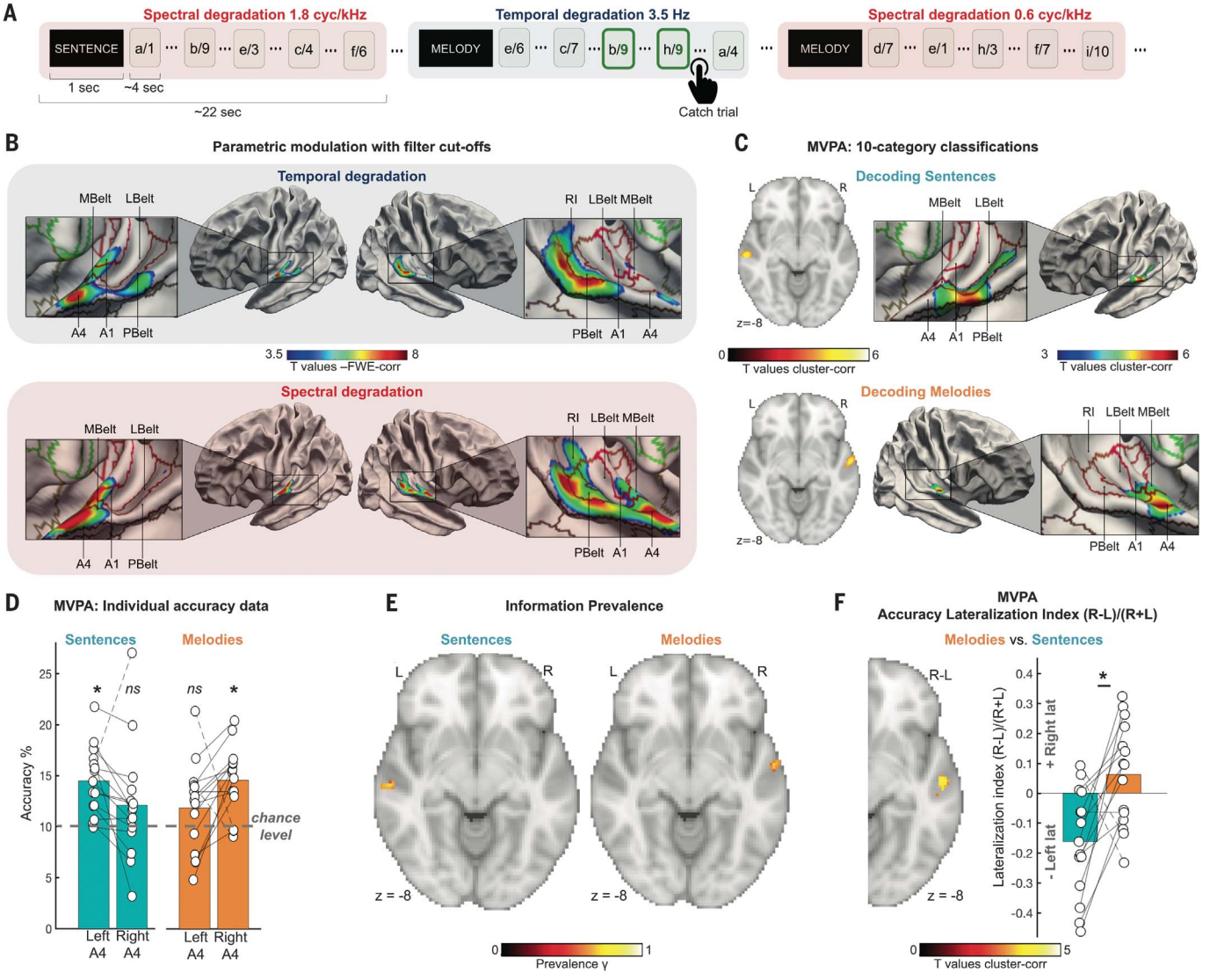


Fig. 3. fMRI experiment. (**A**) fMRI design. BOLD activity was collected while participants listened to blocks of five songs (degraded either in the temporal or spectral dimension). To control for attention, participants were asked to detect two catch trials (with high filter cutoffs containing melody (or sentence) repetition (1-back task). (**B**) Univariate fMRI results ($p < 0.05$, voxelwise FWE corrected). Shown is parametric modulation of BOLD signal with temporal (top) or spectral (bottom) filter cutoffs. (**C**) Multivariate analysis: accuracy (minus chance) maps for 10-category classifications of sentences and melodies ($p < .05$, cluster corrected). t-values are plotted on FsAverage surface (Freesurfer https://surfer.nmr.mgh.harvard.edu/). Regions of interest are extracted from the atlas by Glasser *et al*. (*19*): A1, primary auditory cortex; A4, auditory 4 complex (TE.3); RI, retroinsular cortex; PBelt, parabelt complex; LBelt, lateral belt complex; MBelt, medial belt complex. (**D**) Decoding accuracy in significant clusters of (C) presented as a function of domain and regions (chance level at 10%). Solid/dashed lines: participants showing the expected/reversed effect (sentences: $n = 14/1$; melodies: $n = 13/2$). (**E**) Prevalence analysis for 10-category classifications of sentences and melodies computed in an anatomically defined mask covering the entire temporal lobe (see the materials and methods). Highlighted areas are those where the majority null hypothesis (prevalence $\gamma < 50\%$ of the population) can be rejected at a level of $\alpha = 0.05$ ($p < 0.05$, corrected). (**F**) Left panel: Multivariate analysis for 10-category classification of melodies versus sentences for accuracy values analyzed in terms of lateralization index [$(R - L)/(R + L)$; $p < 0.05$, cluster corrected]. Right panel: Accuracy lateralization values in the significant cluster. Negative values indicate left-lateralized accuracy, whereas positive values indicate right-lateralized accuracy. Solid/dashed lines: participants showing the expected/reversed effect ($n = 13/2$). Bar plots show mean accuracy. White circles indicate individual data. Asterisks indicate significant effects; ns, nonsignificant effects.

superior temporal sulcus dorsal anterior regions, see table S1; $p < 0.05$, cluster corrected; Fig. 4D, right, and fig. S9). Post hoc tests (one-sample *t* tests) revealed that for sentences, NMI was left lateralized for spectral degradations ($t_{(14)} = -2.32$, $p = 0.03$), but the lateralization vanished for temporal degradations ($t_{(14)} = 0.44$, $p = 0.66$). By contrast, for melodies, NMI was right lateralized for temporal degradations ($t_{(14)} = 3.46$, $p = 0.004$) and the lateralization vanished for spectral degradations ($t_{(14)} = -0.24$, $p = 0.80$).

Years of debate have centered on the theoretically important question of the representation of speech and music in the brain (*2*, *6*, *21*). Here, we take advantage of the STM framework to establish a rigorous demonstration that: (i) perception of speech content is most affected by degradation of information in the temporal dimension, whereas perception of melodic content is most affected by degradation in the spectral dimension (Fig. 2, B and C); (ii) neural decoding of speech and melodic contents primarily depends on neural activity patterns in the left and right AC regions, respectively (Fig. 3, C to F, and fig. S4); (iii) in turn, this neural specialization for each stimulus domain is dependent on the specific sensitivity to STM rates of each auditory region (Fig. 4C and fig S7); and (iv) the perceptual effect of temporal or spectral degradation on speech or melodic content is mirrored specifically within each hemispheric auditory region (as revealed by mutual information), thereby demonstrating the brain–behavior relationship necessary to conclude that STM features are processed differentially for each stimulus domain within each hemisphere (Fig. 4D and figs. S8 and S9).

These results extend seminal studies on the robustness of speech comprehension to spectral degradation (*17*, *22*) and are also consistent with observations that the temporal modulation rate of speech samples from many languages is substantially higher than that of music samples across genres (*23*). It remains to be seen whether such a result also applies to other languages, such as tone languages, for which spectral information is arguably more important, and to musical pieces with complex rhythmic and harmonic variations or belonging to musical systems different from the Western tonal melodies used here.

The idea that auditory cognition depends on processing of spectrotemporal energy patterns

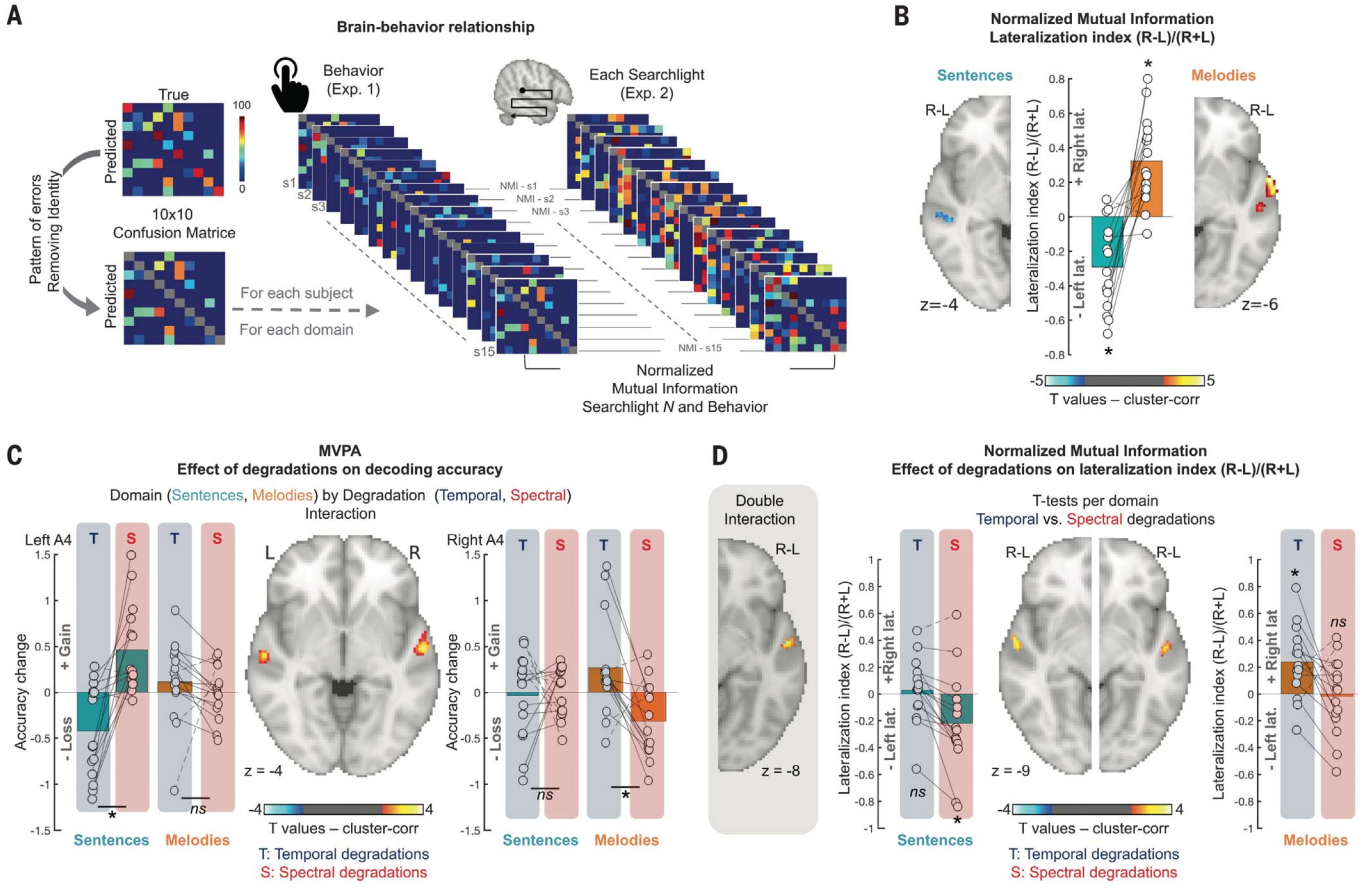


Fig. 4. Effect of degradations on decoding accuracy and NMI. (**A**) NMI computed between confusion matrices extracted from behavioral and fMRI data at the whole-brain level (searchlight procedure) for each participant and for each domain (sentences or melodies). (**B**) NMI results presented in terms of lateralization index [(R − L)/(R + L); $p < 0.05$, cluster corrected]. Light blue clusters indicate a left-lateralized NMI for sentences; red and yellow clusters indicate a right-lateralized NMI for melodies. White circles indicate individual data. Solid/dashed lines: participants showing the expected/reversed effect ($n = 14/1$). (**C**) Accuracy change for 10-category classification of sentences (aqua) and melodies (orange) presented as a function of degradation type (blue-shaded bars: temporal degradation; red-shaded bars: spectral degradation) revealing a domain-by-degradation interaction ($p < 0.05$, cluster corrected). Same conventions as in (B) (solid/dashed lines: $n = 13/2$ for sentences, $n = 12/3$ for melodies). (**D**) Effect of acoustic degradations on NMI lateralization index [(R − L)/(R + L); $p < 0.05$, cluster corrected]. Left panel: Domain-by-degradation interaction. Right panel: Effect of degradation (temporal versus spectral) performed per domain (sentences and melodies). Bar plots indicate the mean lateralization index per domain and degradation type (temporal and spectral). Same conventions as in (B) and (C) (solid/dashed lines: $n = 14/1$ for sentences, $n = 12/3$ for melodies). Asterisks indicate significant lateralization index; ns, nonsignificant lateralization index.

and that these features often trade off against one another is supported by human psychophysics (*17*, *18*), recordings from cat inferior colliculus (*13*), and human neuroimaging (*6*, *7*, *15*–*17*). During passive listening of short, isolated stimuli lacking semantic content, preferences for high spectral versus temporal modulation are distributed in an anterior–posterior dimension of the AC, with relatively weaker hemispheric differences (*6*, *7*, *15*, *16*). Our results suggest that this purely acoustic lateralization may be enhanced during the iterative analysis of temporally structured natural stimuli (*24*) in the most anterior and inferior auditory (A4) patches, which are known to analyze complex acoustic features and their relationships, or sound categories, thus fitting well with their encoding of relevant speech or musical features (*6*, *25*, *26*). We hypothesize that hemispheric lateralization of STM cues scales with the strength of the dynamical interactions between acoustic and higher-level (motor, syntactic, working memory, etc.) processes, which are typically maximized with complex, cognitively engaging stimuli that require decoding of feature relationships to extract meaning (speech or melodic content), as used here.

More generally, studies across numerous species have indicated a match between ethologically relevant stimulus features and the spectrotemporal response functions of their auditory nervous systems, suggesting efficient adaptation to the statistical properties of relevant sounds, especially communicative ones (*27*). This is consistent with the theory of efficient neural coding (*28*). Our study shows that in addition to speech, this theory can be applied to melodic information, a form-bearing dimension of music. Humans have developed two means of auditory communication: speech and music. Our study suggests that these two domains exploit opposite extremes of the spectrotemporal continuum, with a complementary specialization of two parallel neural systems, one in each hemisphere, that maximizes the efficiency of encoding of their respective acoustical features.

REFERENCES AND NOTES

1. D. Poeppel, *Speech Commun.* **41**, 245–255 (2003).
2. R. J. Zatorre, P. Belin, V. B. Penhune, *Trends Cogn. Sci.* **6**, 37–46 (2002).
3. C. McGettigan, S. K. Scott, *Trends Cogn. Sci.* **16**, 269–276 (2012).
4. I. Peretz, M. Coltheart, *Nat. Neurosci.* **6**, 688–691 (2003).
5. A. D. Friederici, *Philos. Trans. R. Soc. London B Biol. Sci.* **375**, 20180391 (2020).
6. S. Norman-Haignere, N. G. Kanwisher, J. H. McDermott, *Neuron* **88**, 1281–1296 (2015).
7. R. J. Zatorre, P. Belin, *Cereb. Cortex* **11**, 946–953 (2001).
8. J. Obleser, F. Eisner, S. A. Kotz, *J. Neurosci.* **28**, 8116–8123 (2008).
9. M. Schönwiesner, R. Rübsamen, D. Y. von Cramon, *Eur. J. Neurosci.* **22**, 1521–1528 (2005).
10. A. Boemio, S. Fromm, A. Braun, D. Poeppel, *Nat. Neurosci.* **8**, 389–395 (2005).
11. B. Morillon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18688–18693 (2010).
12. T. Chi, P. Ru, S. A. Shamma, *J. Acoust. Soc. Am.* **118**, 887–906 (2005).
13. F. A. Rodríguez, H. L. Read, M. A. Escabí, *J. Neurophysiol.* **103**, 887–903 (2010).
14. J. Fritz, S. Shamma, M. Elhilali, D. Klein, *Nat. Neurosci.* **6**, 1216–1223 (2003).
15. M. Schönwiesner, R. J. Zatorre, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14611–14616 (2009).
16. R. Santoro *et al.*, *PLOS Comput. Biol.* **10**, e1003412 (2014).
17. A. Flinker, W. K. Doyle, A. D. Mehta, O. Devinsky, D. Poeppel, *Nat. Hum. Behav.* **3**, 393–405 (2019).
18. T. M. Elliott, F. E. Theunissen, *PLOS Comput. Biol.* **5**, e1000302 (2009).
19. M. F. Glasser *et al.*, *Nature* **536**, 171–178 (2016).
20. R. A. Ince *et al.*, *Hum. Brain Mapp.* **38**, 1541–1573 (2017).
21. D. Schön *et al.*, *Neuroimage* **51**, 450–461 (2010).
22. R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, M. Ekelid, *Science* **270**, 303–304 (1995).
23. N. Ding *et al.*, *Neurosci. Biobehav. Rev.* **81**, (pt. B), 181–187 (2017).
24. A. M. Leaver, J. P. Rauschecker, *J. Neurosci.* **30**, 7604–7612 (2010).
25. T. Overath, J. H. McDermott, J. M. Zarate, D. Poeppel, *Nat. Neurosci.* **18**, 903–911 (2015).
26. M. Chevillet, M. Riesenhuber, J. P. Rauschecker, *J. Neurosci.* **31**, 9345–9352 (2011).
27. L. H. Arnal, A. Flinker, A. Kleinschmidt, A. L. Giraud, D. Poeppel, *Curr. Biol.* **25**, 2051–2056 (2015).
28. J. Gervain, M. N. Geffen, *Trends Neurosci.* **42**, 56–65 (2019).
29. P. Albouy, L. Benjamin, B. Morillon, R. J. Zatorre, Data for: Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody, Open Science Framework (2020); https://doi.org/10.17605/OSF.IO/9UB78.

ACKNOWLEDGMENTS

We thank S. Norman-Haignere, A.-L. Giraud, and E. Coffey for comments on a previous version of the manuscript; C. Soden for creating the melodies; A.-K. Barbeau for singing the stimuli; and M. Generale and M. de Francisco for expertise with recording. **Funding:** This work was supported by a foundation grant from the Canadian Institute for Health Research to R.J.Z. P.A. is funded by a Banting Fellowship. R.J.Z. is a senior fellow of the Canadian Institute for Advanced Research. B.M.'s research is supported by grants ANR-16-CONV-0002 (ILCB) and ANR-11-LABX-0036 (BLRI) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). **Author contributions:** Conceptualization: B.M., P.A., R.J.Z.; Methodology: P.A., L.B., B.M., R.J.Z.; Analysis: P.A., L.B.; Investigation: L.B., P.A.; Resources: R.J.Z.; Writing original draft: P.A., B.M., R.J.Z.; Writing – review & editing: P.A., L.B., B.M., R.J.Z.; Visualization: P.A.; Supervision: B.M., R.J.Z. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Sound files can be found at www.zlab.mcgill.ca/downloads/albouy_20190815/. A demo of the behavioral task can be found at: https://www.zlab.mcgill.ca/spectro_temporal_modulations/. Data and code used to generate the findings of this study are accessible online (*29*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/1043/suppl/DC1
Materials and Methods
Figs. S1 to S9
Table S1
Supplementary Results
References (*30*–*32*)

View/request a protocol for this paper from *Bio-protocol*.

2 September 2019; accepted 2 January 2020
10.1126/science.aaz3468

UT Health
San Antonio

Postdoctoral Fellowship in Cardiovascular Research

A postdoctoral research position is immediately available in the laboratory of James Stockand, PhD at the University of Texas Health Science Center at San Antonio. The position is funded at a competitive salary by the National Institutes of Health Grant **T32 HL007446**. Highly motivated early-career scientists interested in cardiovascular, cardiorenal and metabolism are encouraged to apply. Priority will be given to those with experience with electrophysiology. At the time of appointment, candidates must hold an advanced professional degree (e.g. MD or PhD), be a citizen or permanent resident of the USA, and no more than 4 years from graduation with their terminal degree. To apply, please send a curriculum vitae and cover letter, which includes a brief statement of interest and qualifications and the names of three professional referees, to Dr. Stockand at **Stockand@uthscsa.edu**.

All postdoctoral appointments are designated as security sensitive positions. The University of Texas Health Science Center at San Antonio is an Equal Employment Opportunity/Affirmative Action Employer, including protected veterans and persons with disabilities.

Where Science Gets Social.

AAAS.ORG/COMMUNITY

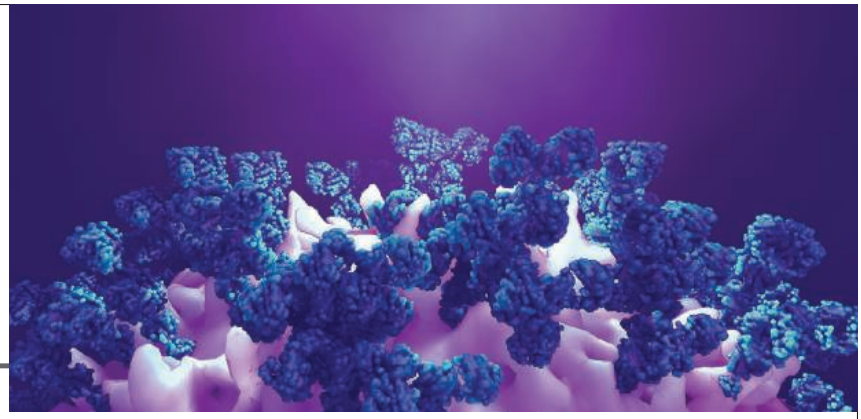
Member COMMUNITY
AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

SPECIAL JOB FOCUS:

Cancer Research

Issue date: April 10
Book ad by March 26
Ads accepted until April 3 if space allows



To book your ad, contact:
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.

Why choose this job focus for your advertisement?

- Relevant ads lead off the career section with a special cancer research banner
- Bonus distributions:
  American Association for Cancer Research, April 24–29, San Diego, CA
  AACR Career Fair, April 25, San Diego, CA.

Science Careers
AAAS
SCIENCECAREERS.ORG

FOR RECRUITMENT IN SCIENCE, THERE'S ONLY ONE *SCIENCE*.

Programme to promote the return of highly qualified young researchers from abroad

Ministerium für
Kultur und Wissenschaft
des Landes Nordrhein-Westfalen

You are currently conducting research abroad and have thought about returning to Germany? You have already achieved results and are now keen to take your career to the next level? Why not come to North Rhine-Westphalia under the state government's Returning Scientists Scheme?

This year's call is all about battery research. North Rhine-Westphalia, already home to a number of leading institutions in this field, aims to expand this position and become an international centre of excellence.

The Returning Scientists Programme will open the door for you to set up and lead an independent junior research team at a university in NRW. You will receive a personal financial commitment of up to 1.25 million euros for a period of five years and enter into negotiations with universities of your choice. You will be able to establish your team of junior researchers at the university that offers you the best outlook, perhaps even beyond the duration of the funding period. A total of six talented young people can be funded this year.

We await your application if you meet the following conditions: You are conducting research outside Germany and have a doctorate from between two and six years ago. You were a German resident before you left the country and you can today show at least twelve months of successful research outside Germany.

For more details and the full invitation to apply, click on

www.rueckkehrprogramm.nrw.de

Please use the online tool to submit your application documents **by April 26, 2020 (time-critical)**.

www.mkw.nrw.de

THE INTERNATIONAL HUMAN FRONTIER SCIENCE PROGRAM ORGANIZATION

CALL FOR NOMINATIONS FOR THE
2021 HFSP NAKASONE AWARD

The HFSP Nakasone Award is awarded to scientists in recognition of pioneering work that has moved the frontier of the life sciences. This may encompass conceptual, experimental or technological breakthroughs. The award recognizes the vision of former Prime Minister Nakasone of Japan in the creation of the Organization.

The competition is open; it is not limited to HFSP awardees and there is no age limit for candidates. In selecting the awardee, the Council of Scientists will pay particular attention to recent breakthroughs by younger scientists. The awardee will receive an unrestricted research grant of 10,000 USD, a commemorative medal and an invitation to deliver the HFSP Nakasone Lecture at the 2021 HFSP Awardees Meeting.

Previous winners of the HFSP Nakasone Award are listed on the HFSP website (see the link below).

Nominations must be received before **24 April 2020** and include the HFSPO nomination form and the nominee's CV (see **http://www.hfsp.org/awardees/hfsp-nakasone-award** for more information).

McMaster University | David Braley Centre for Antibiotic Discovery

Faculty Positions – Antimicrobial Discovery and Resistance

The *David Braley Centre for Antibiotic Discovery* at McMaster University invites applications for up to three faculty in the areas of Antimicrobial Discovery and Resistance. Applications are sought from candidates at all career stages. Research areas of interest include, but are not limited to, chemical biology, computational biology, medicinal chemistry, microbiology, and systems biology.

The *David Braley Centre for Antibiotic Discovery* is embarking on an ambitious program to develop the areas of preclinical antibiotic drug discovery and has outstanding resources in high-throughput biology and chemical screening, natural products, microbiology, and models of infectious disease. The Centre is home to a strong faculty contingent that is working at the interface of chemistry, microbiology, and medicine. We are seeking highly collaborative candidates to expand this initiative with research programs complementary to the research strengths of the group.

Applicants must possess a PhD or MD/PhD with relevant experience and training, and a strong desire to work collaboratively with other faculty members. The successful applicants will develop an active research program and contribute to educational programs at McMaster University through an appointment with a suitable academic Department.

Applicants should submit a curriculum vitae, a statement of research experience, a summary of future plans, teaching dossier and the names of three references by **July 1, 2020** to the search committee Chair Professor Eric Brown, David Braley Centre for Antibiotic Discovery, McMaster University, MDCL 2235, 1280 Main St W, Hamilton, ON L8S 4L8, c/o manneng@mcmaster.ca.

All qualified candidates are encouraged to apply; however Canadian and permanent residents will be given priority. McMaster University is strongly committed to employment equity within its community and to recruiting a diverse faculty and staff. The University encourages applications from all qualified candidates, including women, members of visual minorities, aboriginal persons, members of sexual minorities and persons with disabilities

BRIGHTER WORLD

Science Careers online @sciencecareers.org

By **Luis Batista**

Adapting to my brain

Six months after starting my first faculty position, I lost consciousness while speaking with my students and postdocs. I hit my head on a lab bench as I fell to the floor. A few minutes later, utterly bewildered, I woke up to see the unfamiliar faces of the emergency medical team, as well as the scared faces of my mentees. At the hospital, doctors performed a battery of tests, collecting my spinal fluid and scanning my brain. Later, they told me that I'd suffered a grand mal seizure; it was the first seizure I'd ever had. I was diagnosed with epilepsy at 34 years of age—and my career hasn't been the same ever since.

When I had my first seizure, I'd toiled in academia for 14 years, working my way up the academic ladder. I'd always been laser focused on the next step in what I imagined would be a very linear career trajectory—Ph.D., postdoc, faculty job, tenure. Then, less than 1 year after starting my dream job as a tenure-track professor, health complications disrupted my path.

I decided not to read much about epilepsy. Learning about my condition would be a distraction from my own research. I also wanted to show my trainees and colleagues that all was well. I was on medication, after all, so there was no reason to worry. I thought I could get away with making minimal lifestyle changes—more sleep, no driving—and leave it at that.

"Every seizure took my confidence in my ability to run a lab down a notch."

For 1 year it worked, and life went on as normal. But then, out of nowhere, I started to have recurrent seizures. I fell in my department's bathroom, where a graduate student found me. I had a seizure while traveling home from a lecture in Europe. I had several seizures in front of my 3-year-old daughter at home, where I awoke to see her staring at me and crying.

It's hard to describe the feeling of having a seizure, of coming back to reality and finding yourself confused, lost, and completely vulnerable. I had to take higher doses of my medication, which caused side effects such as fatigue, anxiety, and hostility. The fatigue reduced my focus, at times preventing me from writing grants or even reading papers. Leading my research team became a challenge. Often I was physically unable to stay in the lab for long hours, and mentally unable to focus during meetings with my trainees. Every seizure took my confidence in my ability to run a lab down a notch.

Yet, thanks to an amazing group of trainees and collaborators, my lab remained productive and well-funded. My epilepsy became a challenge we all face together. My trainees now understand that sometimes I won't be in the lab, possibly for a prolonged period. But we stay in constant touch through email and phone calls. I know I can count on them to keep the lab running in my absence. They know I have faith in them and that I am available when they need me.

I am still afraid. Afraid of what's next. Afraid of having a seizure during a conference talk. Afraid of traveling. Afraid of forgetting people's names. Afraid of forgetting deadlines. Afraid of not looking forward to the next step in my career path.

It's difficult. But I've also started to realize that most people outside the orderly world of academia deal with similar fears about the future every day. I've learned that it's OK to give myself some slack—to deviate from the linear career path I once imagined I was on. I work fewer hours. I have to take more notes to avoid forgetting things. I am less efficient. I am still afraid. But that doesn't stop me from traveling, giving talks, and writing grants and papers. After every seizure, I still think of quitting. But I don't. I adapt.

So, if something unexpected derails your ideal career path, find an alternate route. Accept that you have to do things differently. Don't be afraid to ask for help, and don't compare yourself with others. Surround yourself with positive people and learn how to carry on. Linear or not, a career in the sciences is still an amazing privilege that no seizure will erase. ■

Luis Batista is an assistant professor at Washington University in St. Louis, Missouri. Do you have an interesting career story to share? Send it to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

Published by AAAS