China's bat coronavirus hunter speaks out p. 487

Avoiding a global food crisis during a global health crisis p. 500

The fetal microglial transcriptome p. 530

Science

$15
31 JULY 2020
sciencemag.org

AAAS

SOLAR FLARES

Predicting the Sun's brightest flashes pp. 504 & 587

CONTENTS

31 JULY 2020 • VOLUME 369 • ISSUE 6503

Rhododendron and other alpine flowers at 4300-meter elevation in the Hengduan Mountains in China, home to the world's oldest alpine plant community.

NEWS

IN BRIEF

484 News at a glance

IN DEPTH

487 Wuhan coronavirus hunter Shi Zhengli speaks out
China's "Bat Woman" denies responsibility for the pandemic, demands apology from Trump *By J. Cohen*
PODCAST

489 Global plan seeks to promote vaccine equity, spread risks
Poor countries fear ending up at the end of the queue *By K. Kupferschmidt*

490 Hidden predictability in winds could improve climate forecasts
Study finds climate models are missing wind patterns, casting doubt on long-term forecasts of rain and storms *By P. Voosen*

491 How an ancient microbial arms race remodeled human cells
Study traces genetic responses to pathogens back to ancestor of Neanderthals and modern humans *By A. Gibbons*

493 Mountains and monsoons created Tibetan biodiversity
The rise of the Hengduan range 30 million years ago kick-started evolution of world's oldest alpine flora *By E. Stokstad*
REPORT p. 578

494 This man can read letters but numbers are a blank
Case exposes complex roots of consciousness *By S. Kean*

495 Altered DNA base could play key role in pregnancy
In mice, methylated adenine enables the placenta to grow *By E. Pennisi*

FEATURES

496 Lizard man
For Jonathan Losos, tiny Caribbean islands and their reptile inhabitants are test tubes of evolution
By E. Pennisi

INSIGHTS

POLICY FORUM

500 COVID-19 risks to global food security
Economic fallout and food supply chain disruptions require attention from policy-makers
By D. Laborde et al.

PERSPECTIVES

503 Mapping the clean air haves and have-nots
The distribution of air pollution has changed little, perpetuating disparities in exposure *By L. Ma*
REPORT p. 575

504 Can we predict solar flares?
A physics-based model may help improve space weather forecasts *By A. M. Veronig*
REPORT p. 587

506 Evolution of the human brain
A human-specific gene is a determinant of the cognitive architecture of the human cerebral cortex *By C. Dehay and H. Kennedy*
REPORT p. 546

507 Fine-tuning receptor–G protein activation and signaling
The activation rate of downstream G proteins imprints receptor signaling *By G. Lebon*
RESEARCH ARTICLE p. 523

509 Ductile van der Waals materials
Normally brittle, single crystals of InSe can be mechanically rolled like a metal into thin films *By X. Han*
REPORT p. 542

510 How does SARS-CoV-2 cause COVID-19?
The viral receptor on human cells plays a critical role in disease progression
By N. J. Matheson and P. J. Lehner

512 William C. Dement (1928–2020)
Father of sleep medicine
By R. Pelayo and P. Mourrain

BOOKS *ET AL.*

513 Confronting illness with empathy
A journalist paints a thoughtful portrait of a disease that defies scientific characterization
By F. R. Davis

PHOTO: JIAN HUANG

Published by AAAS

514 Context is key to olfaction
How we experience smell has more to do with us than with the odor itself *By J. P. McGann*

LETTERS

515 China and India: Toward a sustainable world
By K. S. Bawa et al.

515 Conservation needs a COVID-19 bailout
By R. A. McCleery et al.

516 Extinct-in-the-wild species' last stand
By A. Trask et al.

RESEARCH

IN BRIEF

519 From *Science* and other journals

REVIEW

522 Immunology
The function of the thymus and its impact on modern medicine *J. F. A. P. Miller*
REVIEW SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.ABA2429

RESEARCH ARTICLES

523 Structural biology
Structural insights into differences in G protein activation by family A and family B GPCRs *D. Hilger* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.ABA3373
PERSPECTIVE p. 507

524 Biochemistry
ANGEL2 is a member of the CCR4 family of deadenylases with 2′,3′-cyclic phosphatase activity *P. H. Pinto* et al.

530 Immunogenomics
Human fetal microglia acquire homeostatic immune-sensing properties early in development *L. Kracht* et al.

REPORTS

537 Molecular diffusion
Boosted molecular mobility during common chemical reactions *H. Wang* et al.

542 Materials science
Exceptional plasticity in the bulk single-crystalline van der Waals semiconductor InSe *T.-R. Wei* et al.
PERSPECTIVE p. 509

546 Neurodevelopment
Human-specific *ARHGAP11B* increases size and folding of primate neocortex in the fetal marmoset *M. Heide* et al.
PERSPECTIVE p. 506

550 Ultracold matter
Cooling and entangling ultracold atoms in optical lattices *B. Yang* et al.

554 Structural biology
In-cell architecture of an actively transcribing-translating expressome *F. J. O'Reilly* et al.

557 Organic chemistry
Direct reversible decarboxylation from stable organic acids in dimethylformamide solution *D. Kong* et al.

561 Neurogenomics
Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants *S. Zhang* et al.

566 Biotechnology
DNA capture by a CRISPR-Cas9–guided adenine base editor *A. Lapinaite* et al.

571 Nanomaterials
Rational synthesis of atomically precise graphene nanoribbons directly on metal oxide surfaces *M. Kolmer* et al.

575 Air pollution
Disparities in $PM_{2.5}$ air pollution in the United States *J. Colmer* et al.
PERSPECTIVE p. 503

578 Paleobotany
Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora *W.-N. Ding* et al.
NEWS STORY p. 493

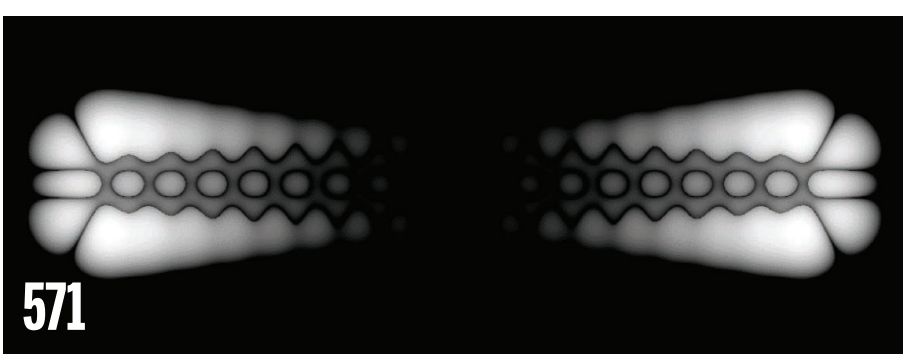
582 Coronavirus
Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California *X. Deng* et al.

587 Solar physics
A physics-based method that can predict imminent large solar flares *K. Kusano* et al.
PERSPECTIVE p. 504

DEPARTMENTS

483 Editorial
Cautious optimism
By H. Holden Thorp
PODCAST

594 Working Life
My womb is my business
By Wen-Jing Lin



CREDIT: M. KOLMER *ET AL.*

ON THE COVER

An extreme ultraviolet image of a large (X-class) solar flare that occurred on 7 March 2012, combining observations taken at 131 Å (blue) and 171 Å (yellow) using the Solar Dynamics Observatory spacecraft. This flare disrupted radio communications on Earth. Predicting when X-class flares will occur will help to forecast Earth's space weather. See pages 504 and 587. *Credit: NASA/GSFC/SDO*

Science Staff 482
AAAS News & Notes 517
New Products 592
Science Careers 593

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, Di Jiang, Marc S. Lavine (Canada), 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, Hilary Stewart (UK), Alana Warnke, Alice Whaley (UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Jeremy Dow, Alexander Kief, Ronmel Navas, 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 **DESIGN EDITOR** 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 **PRODUCTION SPECIALIST** Kristin Wowk **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

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*
Trevor Archer, *NIEHS, NIH*
Paola Arlotta, *Harvard U.*
Johan Auwerx, *EPFL*
David Awschalom, *U. of Chicago*
Clare Baker, *U. of Cambridge*
Nenad Ban, *ETH Zürich*
Franz Bauer, *Pontificia U. 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 Res.*
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*
Ismaila Dabo, *Penn State U.*
Jeff L. Dangl, *U. of North Carolina*
Tom Daniel, *U. of Washington*
Chiara Daraio, *Caltech*
Nicolas Dauphas, *U. of Chicago*
Frans de Waal, *Emory U.*
Claude Desplan, *New York U.*
Sandra Díaz, *U. 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*
Alex L. Kolodkin, *Johns Hopkins U.*
Julija Krupic, *U. of Cambridge*
Thomas Langer, *Max Planck Inst. Cologne*
Mitchell A. Lazar, *U. of Penn.*
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, *QIMR Berghofer*
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, *LSHTM*
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 Sci., Res. & 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)**
Daniel Pauly, *U. of British Columbia*
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*
Shimon Sakaguchi, *Osaka U.*
Miquel Salmeron, *Lawrence Berkeley Nat. Lab*
Nitin Samarth, *Penn. State U.*
Jürgen Sandkühler, *Med. U. of Vienna*
Erica Ollmann Saphire, *La Jolla Inst.*
Alexander Schier, *Harvard U.*
Wolfram Schlenker, *Columbia U.*
Susannah Scott, *U. of California, Santa Barbara*
Rebecca Sear, *LSHTM*
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 Sci.*
V. S. Subrahmanian, *Dartmouth College*
Ira Tabas, *Columbia U.*
Sarah Teichmann, *Wellcome Sanger Inst.*
Rocio Titiunik, *Princeton U.*
Shubha Tole, *Tata Inst. of Fundamental Res.*
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 Sci.*
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 Res.* **(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*

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.

Published by AAAS

EDITORIAL

Cautious optimism

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

The first half of 2020 has seen extraordinary accomplishments in science. The international scientific community has described the genomic sequence of the virus that causes coronavirus disease 2019 (COVID-19) and structures of its important proteins, elucidated principal aspects of the immune response, identified neutralizing antibodies that can serve as therapeutics, and developed promising vaccines. There is much more to learn about COVID-19 and its cause, but the achievements so far are remarkable. So why doesn't this progress feel like the triumph that it is?

Public health guidance is ignored, reopening businesses happens too fast, people fight over wearing masks, and the forces that undermine confidence in vaccines proceed unimpeded. Scientists who burn the midnight oil in academia, government, and industry to decipher COVID-19 are confronted with political leaders who downplay and criticize their tireless efforts. Many are immigrants who hear that they aren't welcome in the United States. President Trump and his allies are sticking their fingers in the eyes of the very people who can lead the world out of this calamity.

> **"With no strategy, a vaccine is the government's best way out of the pandemic crisis."**

There are many reasons to be optimistic about getting a vaccine against the COVID-19 pathogen, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), in record time. Monkey studies of candidate vaccines have shown immune responses that appear linked to protection, and 25 of these have entered human clinical trials. And the U.S. federal government has gambled approximately $6 billion on pharmaceutical companies to produce vaccines in large quantities in the hope that they will perform well in large clinical trials and merit approval by the Food and Drug Administration.

Science also offers reasons to be cautious. We can't know for sure that a vaccine is safe over the long term until we have had more time to observe volunteers in clinical trials. Monkeys don't develop severe disease from SARS-CoV-2, limiting the model's predictive powers. Two doses of the vaccine may be required, which creates a logistical challenge. We won't know whether billions of doses of the vaccine can be produced and distributed until we actually attempt to do so.

But the political and social impediments create even more reason for caution. Despite his recent, hollow backtracking, President Donald Trump has mostly embraced the dogma of the antivaccine movement and cheered on the antimask crowd. Part of the administration's strategy is to undermine Dr. Anthony Fauci, the foremost authority on infectious disease in the United States, which might lead to vulnerable people refusing to get the vaccine that their health and lives may depend on. Even folks who are not against vaccines will have questions about the safety of a vaccine to SARS-CoV-2 given how muddled the messaging has been. It doesn't help that the White House calls the vaccine effort "Operation Warp Speed," which hardly reflects the great care that is being taken to produce a safe and effective vaccine.

The logistical impediments seem most daunting of all. Deciding which vaccines get support from Operation Warp Speed, what signals are required for approval, and who gets the vaccine if it is in limited supply would be challenging under any circumstances. Now a committee assembled by the U.S. National Academies of Sciences, Engineering, and Medicine has been convened to sort through these issues. There is also the Advisory Committee on Immunization Practices that advises the U.S. Centers for Disease Control and Prevention (CDC). Meanwhile, the U.S. National Institutes of Health has organized Accelerating COVID-19 Therapeutic Interventions and Vaccines. How do these committees and agencies interact to make decisions?

Having botched the distribution of diagnostic tests to get ahead of the pandemic, disemboweled the CDC, trampled on its own experts, stoked conspiracy theories about wearing masks and the origins of the virus, pushed an unproven treatment that proved worthless, stepped on the independence of the NIH, and audaciously attacked Fauci, the Trump administration does not inspire confidence in its ability to make sound public health decisions. With no strategy, a vaccine is the government's best way out of the pandemic crisis.

It's not too late to get it right. We need clear decision-making by experts, articulated crisply and without interference. This is not a time for leading with the gut, building up false hope, or making speculative bets. It's time to let the data do the talking.

Science is doing its part. Over to you, Mr. President.

–H. Holden Thorp

10.1126/science.abe0359

PHOTO: CAMERON DAVIDSON

Published by AAAS

90% Accuracy of a new artificial intelligence system trained to identify individual weaver birds, which human birders generally cannot tell apart unless they are tagged (*Methods in Ecology and Evolution*).

IN BRIEF *Edited by* **Jeffrey Brainard**

A drill rig probes the Pebble Mine site, which could become one of the world's largest gold mines.

CONSERVATION

Thumbs-up for Alaskan mine draws fire

A company seeking to build a controversial gold and copper mine in Alaska won a major victory on 24 July when the U.S. Army Corps of Engineers issued an environmental analysis saying the mine wouldn't endanger the world's most productive sockeye salmon fishery. The decision clears the way for the Corps to issue permits needed by promoters of the Pebble Mine, located at the headwaters of two major watersheds that form part of the Bristol Bay salmon runs, just north of the Aleutian Islands. Environmental and Native Alaskan groups and some salmon scientists blasted the new study, saying it understated risks by focusing on the mine's small, initial footprint over 20 years of mining rather than its potential impacts if it expands to become one of the world's largest gold and copper mines, as its promoters hope. Mine backers have said such an expansion would get a closer environmental review later if they pursue it. Scientists have raised concerns that even the smaller mine could have wide impacts, because the resilience of the salmon runs hinges on access to a wide variety of spawning habitats. Environmental groups have vowed to file lawsuits to block the project.

Chinese probe heads for Mars

PLANETARY SCIENCE | China's first independent mission to Mars blasted off from the Xichang Satellite Launch Center on 23 July. To arrive in February 2021, Tianwen-1, a "quest for heavenly truth," comprises an orbiter, lander, and rover. Only the United States and the Soviet Union have successfully landed on Mars. Instruments on the three Tianwen-1 craft will study the planet's magnetic field and atmosphere, map its surface, and characterize its geology. Tianwen-1 is the second in a trio of fresh martian missions: The United Arab Emirates launched its Hope orbiter on 19 July, and NASA planned to launch its Perseverance rover as early as 30 July, after *Science* went to press.

French science bill questioned

FUNDING | A bill in France would increase research spending over the next 10 years and add tenure-track faculty positions, a novelty in France. But critics say the plan's increases would be too small and slow. By 2030, the annual public research budget would rise by about one-third, to €20 billion, toward a goal of lifting overall R&D spending from 2.2% of gross domestic product to 3%. The National Research Agency, which funds researchers through competitive calls, would get €1 billion more over 7 years, reaching about €1.7 billion in 2027, to help raise its grant success rates from 16% to a target of 30%. The new, nonpermanent tenure-track positions would complement the permanent entry-level research positions traditionally offered by the French system, but critics fear the growth may lead to a decline in the permanent ones. Parliament is expected to approve the bill.

Therapy protects preterm babies

DRUG TRIALS | A monoclonal antibody given to babies has strongly protected them from severe disease caused by respiratory syncytial virus (RSV), a leading cause of infant death. As reported this week in *The New England Journal of Medicine*, a placebo-controlled study of nearly 1500 babies born preterm—who are at higher

PHOTO: MICHAEL MELFORD/NATIONAL GEOGRAPHIC IMAGE COLLECTION

Published by AAAS

THREE QS

A magic ride for science

In 1984, artist Bruce Degen met writer Joanna Cole at a publisher's office in New York City to discuss creating a children's book about science. They went on to collaborate and publish 13 colorful, zany books in *The Magic School Bus* series, featuring the ebullient, intrepid teacher Ms. Frizzle (above, right), who takes her students on fantastic adventures into the ocean, across the Solar System, and through the human body, for example. Cole died on 12 July at age 75. But the series continues to teach young readers and their parents about the natural world.

Q: Did you expect to create such a legacy?
A: I was in art school doing very serious art, and I realized that, in my heart of hearts, I wanted to do children's books. In the beginning, it was darn hard work. Some book sketch dummies have five layers of rewrites and reillustrations. The first book was a one-book contract to see if this would work. [The reception] was like the world was waiting for somebody to make this happen. People say, "[As a child,] I [used to] read these books, now I read them to my kids!" I could never have imagined it.

Q: Does scientific accuracy get in the way of storytelling?
A: Frequently. You have to tell kids what is true, but you can't give them all the truth—it's too much. For example, the evolution book goes from now [back] to the beginning of the Earth. I [initially] tried to show every era, year, and life form. It was too complicated. So it ended up as a nice, open spiral with a few representations of each era.

Q: Why use the format of adventures?
A: By following the story, it gave kids a mental filing system—they could retrieve and remember information because it was given to them in a memorable trip.

ILLUSTRATION: BRUCE DEGEN

risk of severe symptoms of RSV—in 23 countries found that a single injection of the antibody before RSV season starts in the fall led to 78.4% fewer hospitalizations for lower respiratory infections associated with the disease. The antibody, being developed by AstraZeneca and Sanofi Pasteur, could replace one now on the market that is rarely used. (It is recommended only for infants at highest risk, requires five shots, and is very expensive.) The companies plan to seek regulatory approval of the new prophylaxis if larger studies now underway in preterm and full-term infants confirm that it is safe and effective.

Astronomy's family tree, mapped

GRADUATE STUDIES | The American Astronomical Society last week launched the Astronomy Genealogy Project, which maps 5000 astronomers to their academic "descendants"—the 28,000 doctorate recipients they supervised. The discipline's family tree, at astrogen.aas.org, stretches back to 1766, but half of the listed doctorates were awarded since 2002. Organizers hope the data will help historians and sociologists of science analyze patterns across countries, universities, and subfields. U.S. universities awarded slightly more than half of the doctorates listed, and about two-thirds of the theses are online.

EPA eyes limits on jet emissions

CLIMATE | Environmental groups last week denounced as weak a plan announced by the U.S. Environmental Protection Agency (EPA) to limit greenhouse gas emissions from aircraft. The new standard would match an existing one adopted in 2016 by a U.N. body, the International Civil Aviation Organization (ICAO), that required emissions cuts by 2028. But recently manufactured planes already meet the standard, and EPA conceded its new rule would not reduce overall airplane emissions. Manufacturers have supported such a U.S. regulation to help them meet ICAO certification requirements. IACO has predicted that even under its standard, airplane emissions will grow by at least 3% a year globally. U.S. aviation accounts for 3% of the country's greenhouse gas emissions.

Oldest microbes found?

EXTREME LIFE | Bacteria from seafloor sediments buried 101 million years ago have been grown in the lab, raising the possibility they are as old as their muddy home. They had somehow survived in an area of the Pacific Ocean almost devoid of organic matter or other

Published by AAAS

nutrients most bacteria need, although the sediments recovered do contain oxygen, the researchers report in *Nature Communications*. The finding pushes back the documented age of bacteria living in marine sediment from 15 million years and provides new insights on the limits of life under extreme conditions. A team led by researchers from the Japan Agency for Marine-Earth Science and Technology harvested the microbes from core samples drilled up to 5700 meters below sea level and took precautions against contaminating them with modern bacteria. The group argues the microbes likely didn't have enough food to keep replicating, and instead may have survived for eons without dividing by repairing age-related cellular damage. The microbes identified are known members of more than eight bacterial groups, many of which are commonly found elsewhere on Earth.

CRISPR beefs up male traits

BIOTECHNOLOGY | Scientists announced last week that they used CRISPR gene editing to modify a cow embryo so that the resulting calf, named Cosmo, should produce more offspring bearing male traits. Bulls are 15% more efficient than cows at converting feed into weight gain, so the new method may allow cattle farmers to raise fewer cattle, benefiting the environment, say the researchers at the University of California, Davis. The researchers inserted a gene called *SRY*, which initiates male development and is normally found on the male sex chromosome, into an embryo's chromosome 17. Next, the researchers plan to determine whether Cosmo's offspring that inherit the *SRY* gene look and grow like males. Fifty percent of the calf's progeny will naturally be male; another 25% will be genetically female but will carry the *SRY* gene.

Florida creates seagrass preserve

CONSERVATION | Florida's governor this month signed a bill to establish a 162,000-hectare marine sanctuary in the Gulf of Mexico and protect one of the state's last remaining stretches of seagrass. Florida's coast boasts the most continuous expanse of seagrass beds in the United States, but these diverse habitats, home to blue crabs and manatees, have been damaged by nutrient-driven algal blooms and boat propellers. Authorities plan to create a management plan for the new Nature Coast Aquatic Preserve to balance protection with ecotourism, boating, and fishing.

COVID-19

Dispatches from the pandemic

Read additional *Science* coverage of the pandemic at sciencemag.org/tags/coronavirus.

U.S. vaccine efficacy trials begin

The first large-scale efficacy trials of COVID-19 vaccines in the United States began last week. On 27 July, the National Institutes of Health, working with Moderna, announced the start of one that aims to recruit 30,000 people. Later that day, a partnership between Pfizer and BioNTech announced separately it was launching a similarly sized study at sites in the United States and elsewhere. Both the Moderna and the Pfizer/BioNTech vaccines contain messenger RNA that prompts cells to make a protein that studs the surface of the COVID-19 virus. If the vaccines work, this viral protein will safely teach the immune system how to battle the virus if a person later is exposed to it. Operation Warp Speed, the Trump administration's push to accelerate development of a COVID-19 vaccine, has committed nearly $3 billion to these two R&D projects, about half its total investment. Other efficacy trials of various COVID-19 vaccines have begun in Brazil, the United Arab Emirates, and the United Kingdom. Results are expected in late fall at the earliest.

CDC slammed over school rules

Guidance issued by the U.S. Centers for Disease Control and Prevention (CDC) last week for safely reopening schools downplays risks that teachers, other staff members, and students will spread or contract COVID-19, many public health specialists say. Provoking claims that CDC's advice had been politicized, the agency revised an earlier draft that President Donald Trump had panned as "very tough and expensive." The nonbinding recommendations, released 23 July, emphasize the social and developmental benefits of in-person schooling and highlight that young children are at low risk for contracting the disease and transmitting the virus that causes it. The document also recommends against screening students for symptoms. Nevertheless, the Trump administration did advise communities with high infection rates to consider not beginning in-person classes. Large school districts, such as those in Atlanta, Houston, Los Angeles, and San Diego, have already announced that they will begin the 2020–21 school year with online instruction only.

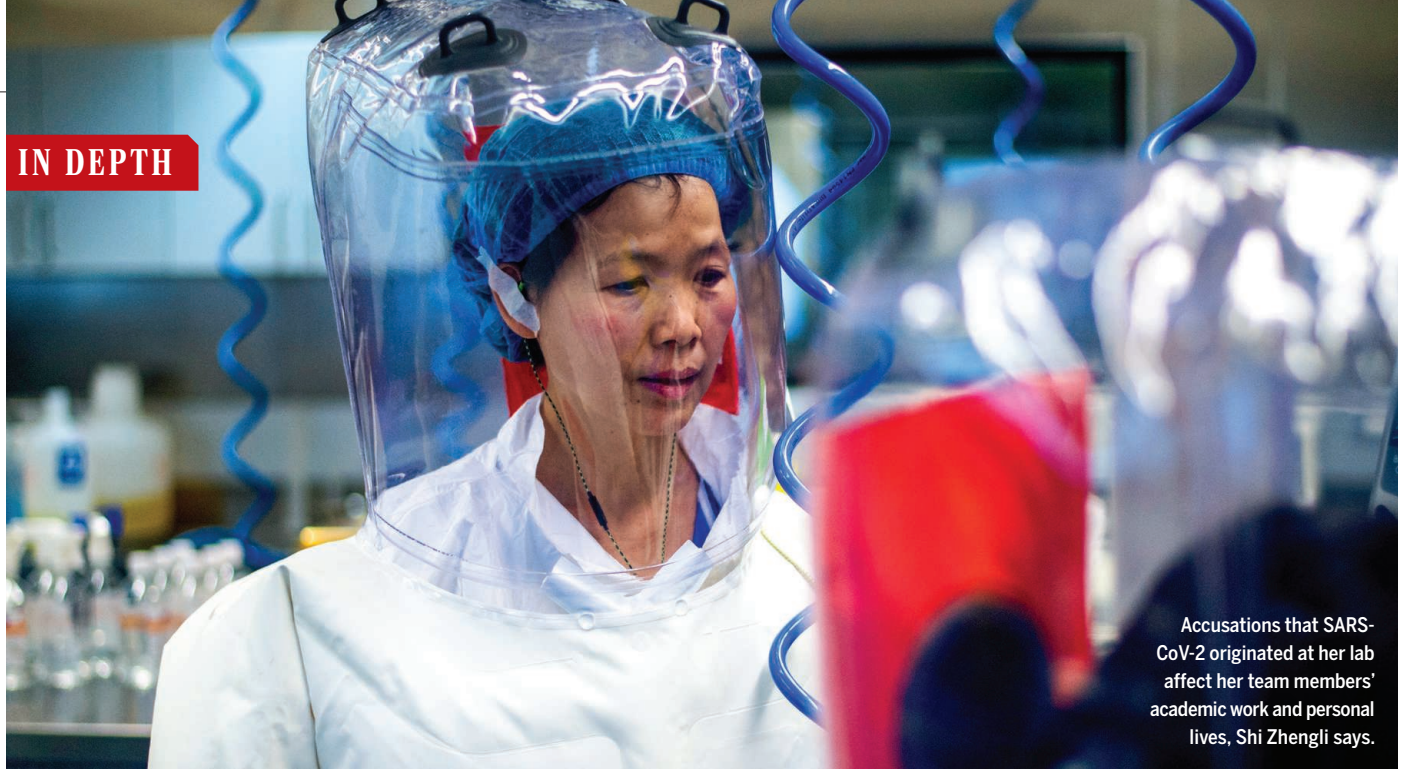
Anti-Fauci TV segment canceled

Following heavy criticism from scientists and others, Sinclair Broadcast Corp. this week canceled plans for its chain of local TV stations to air a segment featuring widely challenged accusations that Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases, intentionally created the virus responsible for COVID-19 and sent it to China. Fauci has helped lead the U.S. effort to control the pandemic despite tangling with President Donald Trump. The allegation came from Judy Mikovits, a virologist and antivaccine activist who appears in a documentary about the coronavirus that was also widely debunked as false and misleading (https://scim.ag/FactCheckMikovits). The Sinclair segment, a new interview with Mikovits, was available online until the company pulled it for review on 25 July, after Media Matters reported its existence. Sinclair announced on 27 July that it would not air the segment on the nearly 200 TV stations it owns or operates in 89 U.S. markets—but not before one in Charleston, West Virginia, had broadcast it.

Students at a summer school session in Monterey Park, California, are offered hand sanitizer.

PHOTO: FREDERIC J. BROWN/AFP VIA GETTY IMAGES

Published by AAAS



Accusations that SARS-CoV-2 originated at her lab affect her team members' academic work and personal lives, Shi Zhengli says.

PHOTO: JOHANNES EISELE/AFP/GETTY IMAGES

COVID-19

Wuhan coronavirus hunter Shi Zhengli speaks out

China's "Bat Woman" denies responsibility for the pandemic, demands apology from Trump

By **Jon Cohen**

The coronavirus pandemic has thrust virologist Shi Zhengli into a fierce spotlight. Shi, nicknamed "Bat Woman," heads a group that studies bat coronaviruses at the Wuhan Institute of Virology (WIV), in the Chinese city where the pandemic began. Many have speculated that SARS-CoV-2, the pathogen that causes COVID-19, accidentally escaped from her lab—a theory promoted by U.S. President Donald Trump. Some have even suggested it could have been engineered there.

China has forcefully rejected such claims, but Shi herself has said very little publicly—until now. On 15 July, Shi emailed *Science* answers to a series of questions about the virus' origin and her research. In them, she hit back at speculation that the virus leaked from WIV. She and her colleagues discovered the virus in late 2019, she says, in samples from patients who had a pneumonia of unknown origin. "Before that, we had never been in contact with or studied this virus, nor did we know of its existence," Shi wrote.

"U.S. President Trump's claim that SARS-CoV-2 was leaked from our institute totally contradicts the facts," she added. "It jeopardizes and affects our academic work and personal life. He owes us an apology."

Shi stressed that over the past 15 years, her lab has isolated and grown in culture only three bat coronaviruses related to one that infected humans: the agent that causes severe acute respiratory syndrome (SARS), which erupted in 2003. The more than 2000 other bat coronaviruses the lab has detected, including one that is 96.2% identical to SARS-CoV-2, are simply genetic sequences that her team has extracted from fecal samples and oral and anal swabs of the animals. She also noted that all staff and students in her lab recently tested negative for SARS-CoV-2, challenging the notion that one of them triggered the pandemic.

Science's COVID-19 reporting is supported by the Pulitzer Center and the Heising-Simons Foundation.

Shi was particularly chagrined about the 24 April decision by the U.S. National Institutes of Health (NIH), made at the White House's behest, to ax a grant to the EcoHealth Alliance in New York City that included bat virus research at WIV. "We don't understand [it] and feel it is absolutely absurd," she said.

Shi's responses—available in full at scim.ag/ShiZhengli—are "a big contribution," says Daniel Lucey of Georgetown University, an outbreak specialist who blogs about SARS-CoV-2 origin issues. "There are a lot of new facts that I wasn't aware of. It's very exciting to hear this directly from her." The answers were coordinated with public information staffers at the Chinese Academy of Sciences, of which WIV is part, and evolutionary biologist Kristian Andersen of Scripps Research suspects they were "carefully vetted" by the Chinese government. "But they're all logical, genuine, and stick to the science," he says.

Richard Ebright, a molecular biologist at Rutgers University, New Brunswick, who has long urged for an investigation into the possibility that a lab accident spawned the pandemic, is unimpressed, however. "Most of these answers are formulaic, almost robotic, reiterations of statements previously made by Chinese authorities and state media," Ebright says.

Shi's responses come at a time when questions about how the pandemic originated are increasingly causing international tensions. Trump frequently calls SARS-CoV-2 "the China virus" and has said China could have stopped the pandemic in its tracks. China, for its part, has added an extra layer of review for researchers who want to publish on the pandemic's origins and has asserted that SARS-CoV-2 may have originated in the United States. Calls for an independent, international probe into the origin are mounting, and two researchers from the World Health Organization are now in China to discuss the scope and scale of a possible mission.

Published by AAAS

Lucey says Shi's answers to *Science*'s questions could help guide the investigation team.

PETER DASZAK of the EcoHealth Alliance, who has long worked with Shi, describes her as social, open, and something of a goodwill ambassador for China at meetings, where she converses in both French and English. (She's also a renowned singer of Mandarin folk songs.) "What I really like about Zhengli is that she is frank and honest, and that just makes it easier to solve problems," he says.

Shi studied at Wuhan University and WIV, then earned a Ph.D. at the University of Montpellier II in France. She returned to WIV in 2000, initially focusing on viruses in shrimp and crabs. A turning point in her career came in 2005, when she published a study in *Science* with Daszak and other researchers from China, Australia, and the United States. The paper reported the first evidence that bats harbored coronaviruses closely related to the lethal virus that jumped from civets to humans and caused the worldwide outbreak of SARS in 2003.

Daszak has continued to work with Shi and her WIV team to sample wild animals and hunt for more coronaviruses. They have published 18 more papers together. Shi "is extremely driven to produce high-quality work," Daszak says. "She will go out in the field, and gets involved in the work, but her real skills are in the lab, and she's one of the best I've worked with in China, probably globally."

Shi told *Science* her lab was thrust into the pandemic on 30 December 2019, the day her team first received patient samples. "Subsequently, we rapidly conducted research in parallel with other domestic institutions, and quickly identified the pathogen," she wrote.

It didn't long take for suspicions and rumors to arise, first on China's social media sites and then in Western media. On 2 February, Shi posted a note on her own social media site saying SARS-CoV-2 was "nature punishing the uncivilized habits and customs of humans," and that she would "bet my life that [the outbreak] has nothing to do with the lab." Partly as a show of support for Shi, Daszak and 26 other scientists from eight countries published a statement of solidarity with Chinese scientists and health professionals in *The Lancet* in February. In a March *Nature* paper that analyzed SARS-CoV-2's genetic makeup, Andersen and other evolutionary biologists argued against it being engineered in a lab.

In her written answers to *Science*, Shi explained in great detail why she thinks her lab is blameless. WIV has identified hundreds of bat viruses over the years, but never anything close to SARS-CoV-2, she says. Although much speculation has centered on RaTG13, the bat virus that most closely resembles SARS-CoV-2, differences in the sequences of the two viruses suggest they diverged from a common ancestor somewhere between 20 and 70 years ago. Shi notes that her lab never cultured the bat virus, making an accident far less likely.

Some suspicions have focused on a naming inconsistency. In 2016, Shi described a partial sequence of a bat coronavirus that she dubbed 4991. That small part of the genome exactly matches RaTG13, leading some to speculate that Shi never revealed the full sequence of 4991 because it actually is SARS-CoV-2. But Shi explained that 4991 and RaTG13 are one and the same. The original name, she says, was for the bat itself, but her team switched to RaTG13 when they sequenced the entire virus. TG stands for Tongguan, the town in Yunnan province where they trapped that bat, she said, and 13 for the year 2013.

Shi Zhengli's team has taken thousands of samples from wild bats but never found the novel coronavirus.

That's "a very logical explanation," says Edward Holmes, an evolutionary biologist at the University of Sydney. Shi's reply also clarified to him why 4991 held such little interest to her team that they didn't even bother to sequence it fully until recently: That short genetic sequence was very different from SARS-CoV, the virus that caused the 2003 outbreak. "In reading this the penny dropped: Of course, they would have been mainly interested in bat viruses closely related to SARS-CoV … not some random bat virus that is more distant," Holmes says.

Shi mentioned other factors that she says exonerate her lab. Their research meets strict biosafety rules, she said, and the lab is subject to periodic inspections "by a third-party institution authorized by the government." Antibody tests have shown there is "zero infection" among institute staff or students with SARS-CoV-2 or SARS-related viruses. Shi said WIV has never been ordered to destroy any samples after the pandemic erupted.

Labs that presumably had strict biosafety rules have had accidents: The SARS virus escaped from several labs after the global outbreak was contained in 2003. And even if everyone in the institute tested negative for the virus today, an infected person could have left WIV months ago. Still, Holmes says, the answers are "a clear, comprehensive, and believable account" of what occurred at WIV.

BUT THEN where did the virus come from? Shi concurs with the scientific consensus that it originated in bats and jumped to humans either directly or, more likely, via an intermediate host. Her lab tested samples from Huanan Seafood Wholesale Market, which Wuhan officials initially fingered as a possible origin because some early patients had links to it, and found RNA fragments from the virus in "door handles, the ground and sewage," she wrote—but not in "frozen animal samples." But the market's role was called into question when two papers revealed that up to 45% of the first confirmed patients—including four of the five earliest cases—did not have any links to it. "The Huanan seafood market may just be a crowded location where a cluster of early novel coronavirus patients were found," Shi says.

Researchers from WIV and Huazhong Agricultural University didn't find the virus in farmed animals and livestock around Wuhan and in other places in Hubei province, she wrote. Years of surveillance in Hubei have never turned up bat coronaviruses close to SARS-CoV-2, she said, leading her to believe the jump to humans happened elsewhere.

Shi provided few details on China's efforts to pin down the origin. "Many groups in China are carrying out such studies," she wrote, using multiple approaches. "We are publishing papers and data, including those about the virus's origins."

Daszak supports the push for an international research effort—which he cautions could take years—and says Shi's group should play a prominent role in it. "I hope and believe that she will be able to help WIV and China show the world that there is nothing to these lab escape theories, and help us all to find the true origins of this viral strain," he says. Shi ended her answers on a similar note. "Here, I would like to make an appeal to the international community to strengthen international cooperation on research into the origins of emerging viruses," she said. "I hope scientists around the world can stand together and work together." ■

PHOTO: ZHANG SHUYI

Published by AAAS

PHOTO: SIMON WOHLFAHRT/AFP/GETTY IMAGES

A lab technician works on COVID-19 tests in Kigali, Rwanda. Health care workers should be first in line for pandemic vaccines, global health experts say.

COVID-19

Global plan seeks to promote vaccine equity, spread risks

Poor countries fear ending up at the end of the queue

By **Kai Kupferschmidt**

As soon as the first COVID-19 vaccines get approved, a staggering global need will confront limited supplies. Many health experts say it's clear who should get the first shots: health care workers around the world, then people at a higher risk of severe disease, then those in areas where the disease is spreading rapidly, and finally, the rest of us. Such a strategy "saves the most lives and slows transmission the fastest," says Christopher Elias, who heads the Bill & Melinda Gates Foundation's Global Development Division. "It would be ludicrous if low-risk people in rich countries get the vaccine when health care workers in South Africa don't," adds Ellen 't Hoen, a Dutch lawyer and public health activist.

Yet money and national interest may win out. The United States and Europe are placing advance orders for hundreds of millions of doses of successful vaccines, potentially leaving little for poorer parts of the world. "I'm very concerned," says John Nkengasong, director of the Africa Centres for Disease Control and Prevention.

To avoid such a scenario, the World Health Organization and other international organizations have set up a system to accelerate and equitably distribute vaccines, the COVID-19 Vaccines Global Access (COVAX) Facility, which seeks to entice rich countries to sign on by reducing their own risk that they're betting on the wrong vaccine candidates. But the idea has been put together on the fly, and it's unclear how many rich countries will join.

Recent history isn't encouraging. A cocktail of powerful antiviral drugs revolutionized HIV treatment in the West in 1996, saving many lives, but it took 7 years for the drugs to become widely available in Africa, the hardest hit continent. "That was catastrophic and that experience is high in my mind," Nkengasong says. During the 2009 H1N1 influenza pandemic, the United States and many European countries donated 10% of their vaccine stocks to poorer countries—but only after it became clear they had enough for their own populations. "Too many had to wait too long for too little," says Richard Hatchett, CEO of the Coalition for Epidemic Preparedness Innovations, a partner in COVAX.

This time, too, rich countries' biggest concern is to protect their own citizens. The U.S. government has signed deals worth more than $6 billion with several vaccine companies as part of Operation Warp Speed, which aims to provide the U.S. population with vaccines by January 2021. Europe's Inclusive Vaccines Alliance, formed by France, Germany, Italy, and the Netherlands, has signed a deal to buy 400 million doses of AstraZeneca's vaccine for use in EU member states. The United Kingdom has inked deals with AstraZeneca and other companies as well. China is developing its own vaccines; it's unclear how early it will be willing to share or whether the shots will come with political strings attached.

The idea behind COVAX is to invest in about 12 different vaccines and ensure early access when they become available. "The goal is to have 2 billion doses by the end of 2021," says Seth Berkley, director of GAVI, the Vaccine Alliance, the third COVAX partner: 950 million for high- and upper middle-income countries, 950 million for low- and lower middle-income countries, and 100 million for "humanitarian situations and outbreaks that are out of control." A first $750 million deal with AstraZeneca for 300 million doses was announced on 4 June.

Berkley accepts that many rich countries will do their own deals with manufacturers. But signing up to COVAX in addition is an insurance policy, he says. If the vaccines they have invested in fail to materialize, they would still have access to others through COVAX, although only enough for 20% of their populations. The money they invest will be used to guarantee lower prices for poorer countries. Separately, the COVAX Advance Market Commitment is collecting donations from the higher income countries. GAVI says COVAX will need $2 billion in donations to pay for the vaccine doses in 90 countries. "We're trying to do everything we can to try to have a global approach because we think it's the right thing to do from a science point of view and from an equity point of view," Berkley says.

COVAX is planning to spread its own risk by investing in a variety of vaccine strategies. Single-dose vaccines could be easier to give in refugee camps, for instance, whereas a double-dose regimen might work fine at a European doctor's office. Some vaccines are based on new technology, creating more uncertainty about regulatory approval and manufacturing capacity. COVAX also hopes to source its vaccines from companies in different locations, so that no single country can stop them from being exported.

COVAX is a clever way to try to hold together the interests of different countries, says David Fidler, an adjunct senior fellow for global health at the Council on Foreign Relations. "Even from the point of view of raw self-interest that governments often have, you can see why this would look attractive," he says. "They're not ignoring the

Published by AAAS

political reality." But he worries countries may hesitate to sign up, and that those that do may renege on the deal once the scramble for doses actually begins.

So far, more than 70 countries that plan to finance their own vaccine have expressed interest in signing up with COVAX. (They have to formally commit by the end of August and provide an advance of 15% of the overall amount.) Whether they will come through is another matter. Some countries in the European Union—which often stresses the importance of global solidarity—have indicated they intend to donate money, but may not order vaccines themselves through COVAX. "Where the procurement of doses for themselves is concerned, we are still actively discussing with them," Hatchett says. "There are models where we could work together."

Nkengasong says Africa needs to explore other avenues as well. "We welcome the COVAX Facility arrangement but we cannot just wait for discussions in Geneva," he says. "We need to take charge of our own destiny." At an African Union meeting in late June, South African President Cyril Ramaphosa called on leaders to secure vaccine supplies for the continent and ensure that vaccines are manufactured there. Nkengasong says African governments are approaching banks to finance deals with pharma companies similar to those signed by the United States. "We need to come together as a continent of 1.3 billion people to not be left behind."

Kate Elder, a vaccines expert at Doctors Without Borders's Access Campaign, sees COVAX as perhaps the best shot at an equitable distribution of vaccines, but says it should be more transparent. "How did they choose AstraZeneca? What conditions are there in the agreement with AstraZeneca if the company doesn't meet volume commitments?" she asks. "None of the conditions around that are known." Both COVAX and governments should also ensure that funded vaccine producers make their data freely available to any company that wants to use them, 't Hoen adds. "I'm very worried that they are doing this without strings attached, that the know-how that is developed with this public funding does not become publicly available for others to use," she says.

What's needed most for COVAX to work is political commitment at the highest levels of government, says Alexandra Phelan, a lawyer at Georgetown University who specializes in global health. Without "a really big international effort," for instance at the United Nations or the G-20, the early doses of vaccines are unlikely to go to those who need them most, Phelan says. "It will be slow, it will be inadequate, and there will be unnecessary deaths." ■

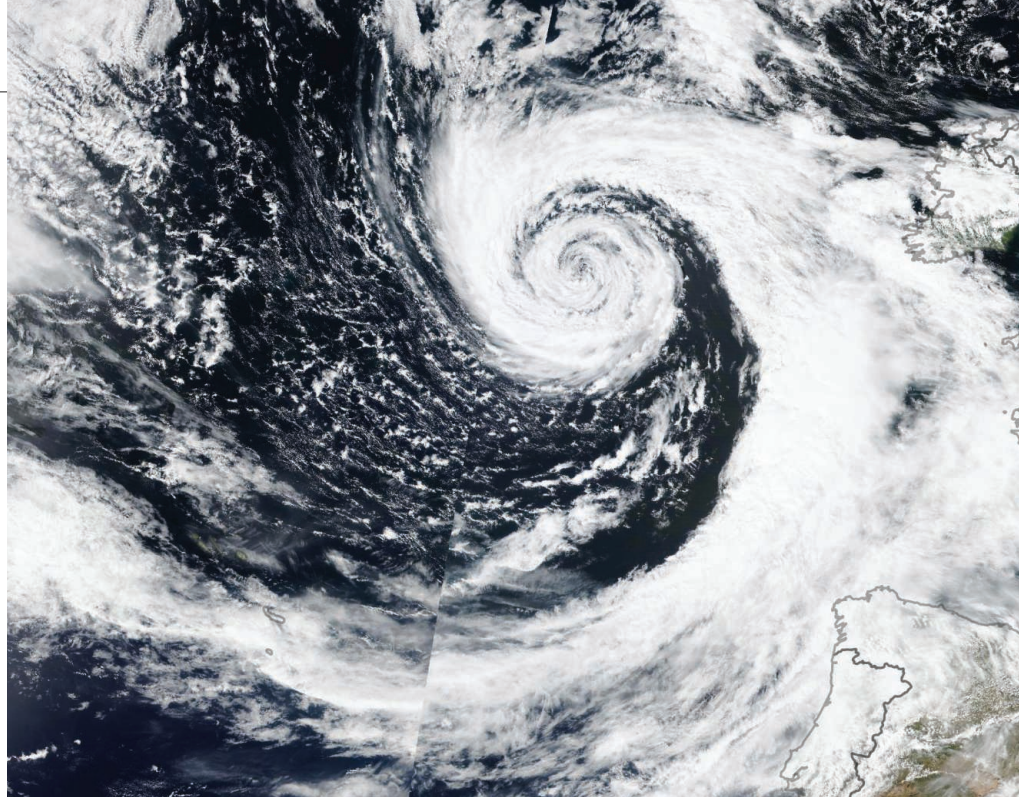


PHOTO: NASA EARTH OBSERVATORY IMAGE BY LAUREN DAUPHIN USING VIIRS DATA FROM THE SUOMI NATIONAL POLAR-ORBITING PARTNERSHIP

ATMOSPHERIC SCIENCE

Hidden predictability in winds could improve climate forecasts

Study finds climate models are missing wind patterns, casting doubt on long-term forecasts of rain and storms

By **Paul Voosen**

Climate scientists can confidently tie global warming to impacts such as sea-level rise and extreme heat. But ask how rising temperatures will affect rainfall and storms, and the answers get a lot shakier. For a long time, researchers chalked the problem up to natural variability in wind patterns—the inherently unpredictable fluctuations of a chaotic atmosphere.

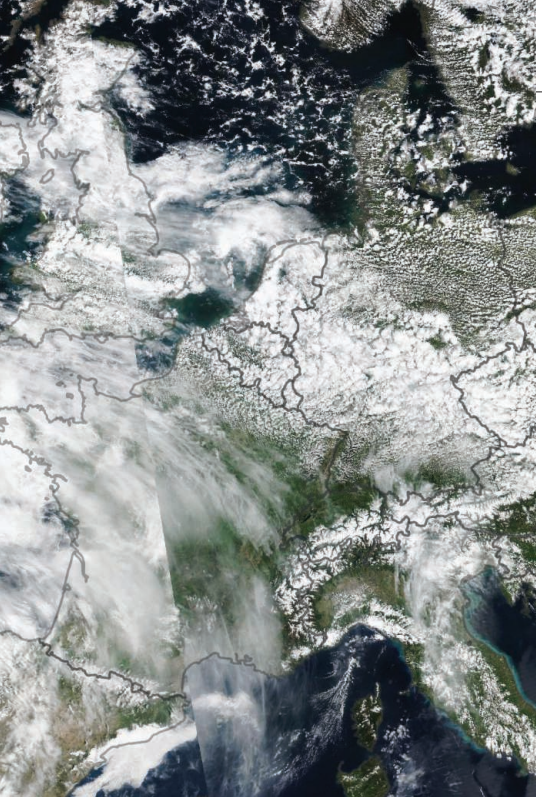
Now, however, a new analysis has found that the problem is not with the climate, it's with the massive computer models designed to forecast its behavior. "The climate is much more predictable than we previously thought," says Doug Smith, a climate scientist at the United Kingdom's Met Office who led the 39-person effort published this week in *Nature*. But models don't capture that predictability, which means they are unlikely to correctly predict the long-term changes that are most influenced by large-scale wind patterns: rainfall, drought, flooding, and extreme storms. "Obviously we need to solve it," Smith says.

The study, which includes authors from several leading modeling centers, casts doubt on many forecasts of regional climate change, which are crucial for policymaking. It also means efforts to attribute specific weather events to global warming, now much in vogue, are rife with errors. "The whole thing is concerning," says Isla Simpson, an atmospheric dynamicist and modeler at the National Center for Atmospheric Research, who was not involved in the study. "It could mean we're not getting future climate projections right."

The study does not cast doubt on forecasts of overall global warming, which is driven by human emissions of greenhouse gases. And it has a hopeful side: If models could be refined to capture the newfound predictability of winds and rains, they could be a boon for farming, flood management, and much else, says Laura Baker, a meteorologist at the University of Reading who was not involved in the study. "If you have reliable seasonal forecasts, that could make a big difference."

The study stems from efforts at the Met Office to predict changes in the North Atlantic

Published by AAAS



Researchers have discovered an untapped predictability in the North Atlantic Oscillation, a wind pattern that drives storms in Europe.

Oscillation (NAO), a large-scale wind pattern driven by the air pressure difference between Iceland and the Azores. The pressure difference reverses every few years, shunting the jet stream north or south; a more northerly jet stream drives warm, wet winters in northern Europe while drying out the continent's south, and vice versa. In previous attempts to project the pattern decades into the future, a single model might yield opposite forecasts in different runs. The uncertainty seemed "huge and irreducible," Smith says.

At first, the Met Office model did no better. But when the team ran the same model multiple times, with slightly different initial conditions, to forecast the NAO a season or a year into the future, a weak signal appeared in the ensemble average. Although it did not match the strength of the real NAO, it did match the overall pattern of its gyrations. But on individual model runs, the signal was drowning in noise.

The new work uses an ensemble of 169 model runs to find the same weak but predictable NAO pattern persisting for up to a decade. For each year since 1960, the team forecasted the NAO pattern 2 to 9 years in the future. When compared with weather records, the ensemble results showed the same pattern, ultimately explaining four-fifths of the NAO's behavior. The massive computational effort suggests changes in the NAO are more predictable than models capture by an order of magnitude, Smith says. It also suggests individual models aren't properly accounting for the ocean or atmospheric forces shaping the NAO.

The missed predictability appears to be universal. "This is being pursued everywhere," says Yochanan Kushnir, a climate scientist at Columbia University, whose team reported last week in *Scientific Reports* that rainfall in the Sahel zone is more predictable than models indicate. In forthcoming work, a group led by Benjamin Kirtman, an atmospheric scientist and model developer at the University of Miami, will flag similar missed predictability in wind patterns above many of the world's oceans.

Kirtman thinks something fundamental is wrong with the models' code. For the time being, he says, "You're probably making pretty profound mistakes in your climate change assessment" by relying on regional forecasts. For example, models predicted that the Horn of Africa, which is heavily influenced by Indian Ocean winds, would get wetter with climate change. But since the early 1990s, rains have plummeted and the region has dried.

The missing predictability also undermines so-called event attribution, which attempts to link extreme weather to climate change by using models to predict how sea surface warming is altering wind patterns. The changes in winds, in turn, affect the odds of extreme weather events, like hurricanes or floods. But the new work suggests "the probabilities they derive will probably not be correct," Smith says.

What's not clear yet is why climate models get circulation changes so wrong. One leading hypothesis is that the models fail to capture feedbacks into overall wind patterns from individual weather systems, called eddies. "Part of that eddy spectrum may simply be missing," Smith says. Models do try to approximate the effects of eddies, but at just kilometers across, they are too small to simulate directly. The problem could also reflect poor rendering of the stratosphere, or of interactions between the ocean and atmosphere. "It's fascinating," says Jennifer Kay, a climate scientist at the University of Colorado, Boulder. "But there's also a lot left unanswered."

While researchers around the globe hunt down the missing predictability, Smith and his colleagues will take advantage of the weak NAO signal they have in hand. The Met Office and its partners announced this month they will produce temperature and precipitation forecasts looking 5 years ahead, and will use the NAO signal to help calibrate regional climate forecasts for Europe and elsewhere.

But until modelers figure out how to confidently forecast changes in the winds, Smith says, "We can't take the models at face value." ■

HUMAN EVOLUTION

How an ancient microbial arms race remodeled human cells

Study traces genetic responses to pathogens back to ancestor of Neanderthals and modern humans

By **Ann Gibbons**

At a recent symposium on the evolution of infectious diseases, University of California, San Diego (UCSD), pathologist Nissi Varki noted that humans suffer from a long list of deadly diseases—including typhoid fever, cholera, mumps, whooping cough, and gonorrhea—that don't afflict apes and most other mammals. All of those pathogens follow the same well-trodden pathway to break into our cells: They manipulate sugar molecules called sialic acids. Hundreds of millions of these sugars stud the outer surface of every cell in the human body—and the sialic acids in humans are different from those in apes.

Varki and an international team of researchers have now traced how evolution may have scrambled to construct new defenses after that molecular vulnerability emerged in our distant ancestors. By analyzing modern human genomes and ancient DNA from our extinct cousins, the Neanderthals and Denisovans, the researchers detected a burst of evolution in our immune cells that occurred in an ancestor of all three types of human by at least 600,000 years ago.

As the researchers report in the current issue of *Genome Biology and Evolution*, these genetic changes may have sharpened the body's defenses against the pathogens that evolved to exploit sialic acids—but created new vulnerabilities. In an added irony, they note, humans' distinctive sialic acids were themselves once a defense against disease. The evolutionary saga is a vivid illustration of the competition between humans and microbes, says microbiologist Christine Szymanski of the University of Georgia, Athens, who is not a co-

Published by AAAS

author. "This gives us a human perspective on how we have to keep changing to keep pace."

The arena for this evolutionary arms race is the glycocalyx, a sugar coating that protects the outer membrane of all cells. It consists of a forest of molecules that sprout from the cell membrane. The sialic acids are at the tip of the tallest branches, sugar chains called glycans, which are rooted to fats and proteins deeper in the membrane.

Given their prominence and sheer number, sialic acids are usually the first molecules that invading pathogens encounter. Human cells are coated with one type of sialic acid, N-acetylneuraminic acid (Neu5Ac). But apes and most other mammals also carry a different one, N-glycolylneuraminic acid (Neu5Gc).

More than 2 million years ago, according to multiple molecular clock methods that estimate when mutations arose, a mutation in a gene on chromosome six made it impossible for human ancestors to make Neu5Gc anymore; instead, they made more of another sialic acid, Neu5Ac (*Science*, 4 September 1998, p. 1432). "We now know we had an ancient complete makeover of the surface of the human cells," says evolutionary biologist Pascal Gagneux of UCSD, a coauthor of the new paper. Birds, some bats, ferrets, and New World monkeys all separately made the same evolutionary change.

The change likely evolved as a defense against malaria, says UCSD physician-scientist Ajit Varki, senior author of the paper and Nissi Varki's spouse. Malarial parasites that infect chimpanzees were no longer able to bind with the altered sialic acids on our red blood cells (*Science*, 24 September 2010, p. 1586).

But in the next million years or so, that mutation became a liability, as Neu5Ac became a portal for a flurry of other pathogens. At the infectious disease symposium organized by UCSD's Center for Academic Research and Training in Anthropogeny, researchers described how multiple diseases evolved to use Neu5Ac to enter cells or to evade immune cells.

Coronaviruses appear to be no exception. "Most coronaviruses infect cells in two steps—first by recognizing abundant sialic acids as binding sites to gain a foothold, and then seeking out the higher affinity protein receptors like ACE2," Ajit Varki says. "Think of it like an initial handshake or introduction that is required before one can ask for a date." Two preprints suggest the novel coronavirus, SARS-CoV-2, also docks with sialic acids before binding with the ACE2 receptor to pierce human cells.

In past studies, Ajit Varki and Gagneux suggested the makeover of the cell and the loss of Neu5Gc may have even contributed to the origin of a new species in our genus *Homo*. If a woman with only Neu5Ac sialic acids mated with a man who still expressed Neu5Gc, her immune system may have rejected that man's sperm or the fetus that developed from it. This fertility barrier might have helped divide *Homo* populations into different species more than 2 million years ago, the researchers speculated.

But the sialic acid change also sparked a new arms race between pathogens and our ancestors. In the new study, the researchers scanned DNA for immune genes in six Neanderthals, two Denisovans, and 1000 humans, and looked at dozens of chimps, bonobos, gorillas, and orangutans as well. They found evolutionary changes that "markedly altered" one class of proteins—sialic acid-binding immunoglobulin-type lectins, or Siglecs—that usually sit on the surface of human immune cells and recognize sialic acids.

Siglecs are molecular sentries: They probe sialic acids to see whether they are familiar parts of our own bodies or foreign invaders. If Siglecs spot sialic acids that are damaged or missing, they signal immune cells to activate, rousing an inflammatory army to attack potential invaders or clean up damaged cells. If sialic acids instead appear to be normal parts of our own cells, other, inhibitory Siglecs throttle back immune defenses so as not to attack our own tissues (see graphic, below).

The researchers identified functional changes in the DNA of eight out of 13 Siglecs encoded by genes on chromosome 19 in humans, Neanderthals, and Denisovans. This hot spot of evolution appears only in Siglec gene variants, not in nearby genes on the chromosome, suggesting natural selection favored these changes, presumably because they helped fight pathogens that target Neu5Ac.

Apes did not show these changes, says first author Naazneen Khan, an evolutionary biologist now at the University of Kentucky. Given the mutations' presence in archaic hominins, this burst of evolution must have happened before our lineages diverged 600,000 years ago, but after the mutation in that altered sialic acid arose more than 2 million years ago, perhaps in *Homo erectus*, thought to be an ancestor of modern humans and Neanderthals.

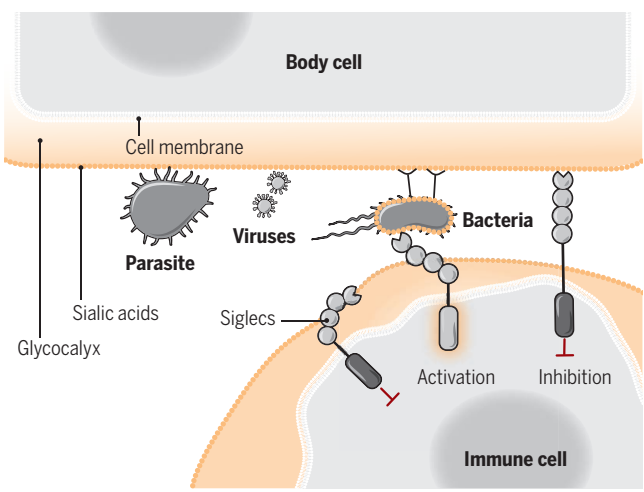
Most Siglecs are found on immune cells, but in the new paper, the team reports that several of the human Siglecs that underwent evolutionary changes are expressed in other types of human cells, including some in the placenta, cervix, pancreas, gut, and brain. Siglec changes may have been a side effect of intense battles with pathogens that infected these tissues, Nissi Varki suggests.

Although the recently mutated Siglecs protect us from pathogens, they may also contribute to other diseases. Some of the genetically changed Siglecs are associated with inflammation and autoimmune disorders such as asthma and with meningitis. The researchers suggest the altered Siglecs are constantly on high alert and do not dampen immune responses against our own tissues; they may even make some individuals more prone to the runaway inflammation seen in severe COVID-19.

Other researchers say the work underscores broad evolutionary principles. "This nicely shows that ... natural selection is not always going for the optimal solution, because the optimal solution is changing all the time," says Rita Gerardy-Schahn, a glycobiologist at Hannover Medical School in Germany, who was not part of the new work. "What is best for natural selection in the short run may be the wrong selection tomorrow." ■

Battle at the cell surface

Some pathogens use sialic acids, which sit on the outer edge of the cell membrane, to invade a cell. Pathogens sometimes coat themselves in humanlike sialic acids to trick signaling molecules called sialic acid-binding immunoglobulin-type lectins (Siglecs) into inhibiting immune responses. But other Siglecs can instead turn on an immune response if they sense sialic acids on pathogens.



GRAPHIC: PASCAL GAGNEUX/UCSD, ADAPTED BY N. DESAI/*SCIENCE*

Published by AAAS

A showy *Rhododendron* is one of some 3000 alpine plant species found in the Hengduan Mountains.

PHOTO: YAOWU XING/XTBG/CAS

EVOLUTION

Mountains and monsoons created Tibetan biodiversity

The rise of the Hengduan range 30 million years ago kick-started evolution of world's oldest alpine flora

By **Erik Stokstad**

The Hengduan Mountains adjoining Asia's massive Tibetan Plateau are a place of superlatives. Four major rivers, including the Mekong and the Yangtze, flow through deep valleys. In winter, snow blankets the peaks, and in summer fierce monsoon rains pelt them. And one of the world's richest alpine plant communities, boasting more than 3000 species, cloaks the slopes.

"It's a fascinating place, especially botanically," says Richard Ree, an evolutionary biologist at the Field Museum. At first glance, he notes, the alpine meadows on the high slopes of the Hengduan peaks look like those found in other ranges, such as North America's Rocky Mountains. "But then you realize that there are 10 times more species." The *Rhododendrons*, *Primulas*, and *Gentians* are particularly diverse.

The plant community is also much, much older, Ree and colleagues reveal on p. 578. It originated some 30 million years ago, tens of millions of years before those blanketing other mountains. That makes it the world's oldest continuous alpine ecosystem "by a long shot," Ree says, which gave it far more time to diversify. Evolution also had a potent driver here, they conclude: intense monsoon rains, which provided essential moisture but also promoted erosion that carved up the landscape, dividing and isolating plant populations that ultimately diverged into new species.

The findings have implications both for understanding how geological and meteorological forces can mold plant communities and for predicting how they might cope with future climate and landscape changes, researchers say. "If we can understand how species and communities responded in the past," says botanist Pam Soltis of the Florida Museum of Natural History at the University of Florida, "we might have a better handle on how fragile they may be over the next century."

Geologists have long puzzled over the complex history of the Hengduan range, which runs north to south mostly through southwestern China. Not long ago, they believed the peaks, which rise to 4500 meters, formed about 5 million years ago. But more recent studies, including new radiometric dating, suggest they arose some 30 million years ago.

To find out whether the region's plant history also stretched back that far, Ree and Wen-Na Ding, a Ph.D. student at the Xishuangbanna Tropical Botanical Garden in China, used DNA sequences to create evolutionary trees for each of the 18 plant groups found on the Tibetan Plateau, Himalayas, and Hengduan Mountains. They also used plant fossils to calibrate a timeline of when new species appeared and link it to key events in the region's geological history. The findings allowed them to estimate speciation rates, and how fast new species spread from one region to others.

The data pointed to "a surprising result," Ree says: Some of the plant groups in the plateau originated in the Hengduan Mountains about 30 million years ago, far earlier than is known for other alpine flora. More new species arose as major geological events, such as the uplift of mountain chains, isolated plant populations some 19 million to 17 million years ago. Plant speciation accelerated again about 15 million years ago, when geologic evidence indicates monsoons intensified, reshaping the landscape. "Definitely, the monsoon has an enormous role to play," says Adrien Favre, an evolutionary biology at the Senckenberg Research Institute and Natural History Museum. But he notes that global cooling at the time also would have spurred the evolution of alpine plants.

Along with time, topography, and weather, the diversity of Hengduan's flora may reflect good fortune. The region avoided the intense glaciation that wiped out ancient plant communities in other mountains. The range's north-south orientation of the mountains also might have helped, by giving plants an escape route to warmer climes, with their seeds carried south by animals, wind, or water.

Now, the question is how the Hengduan's ancient flora will cope with human activities, including a new climate threat. If upheavals long ago, rather than today's conditions, account for the region's superlative diversity, traditional approaches to predicting how plants will react to warming might not work, says Carsten Rahbek, a macroecologist at the University of Copenhagen. "The foundation of how we do biodiversity forecasting of climate change, especially in mountains, is probably on very shaky ground," he says.

There is also a more pressing threat: development, in the form of new roads, hydroelectric dams, and growing settlements and farms. On the bright side, the stunning wildflower displays that appear each summer on the slopes of the Hengduan Mountains have made the region a growing draw for ecotourism, and prompted China's government to protect parts of this ancient cradle of plant diversity. ■

Published by AAAS

A man attempts to draw what he sees when looking at an 8. Numerals fail to register in his brain.

NEUROSCIENCE

This man can read letters but numbers are a blank

Case exposes complex roots of consciousness

By **Sam Kean**

In the video, the man sounds creeped out. "This is too strange for words," he mutters. He's holding a plate-size, green foam 8. When upright, it looks to him like an incoherent jumble. But when he rotates it 90°, the shape snaps into focus; it looks like "a mask." He begins to rotate the numeral back and forth, watching it melt and cohere over and over. He finally hands it to a nearby scientist, saying, "You gotta take that away."

This man, known as RFS, is the subject of a new case study that would have made neurologist Oliver Sacks proud. RFS can read words and letters just fine. But as researchers report this month in the *Proceedings of the National Academy of Sciences*, he cannot see numerals at all—at least not consciously. His amazingly specific deficit could help neuroscientists understand how conscious awareness arises in the brain. "What it tells me," says Christof Koch, a neuroscientist at the Allen Institute who specializes in consciousness, "is that … you can get dissociation between cognition and consciousness."

RFS, an engineering geologist in his late 60s, began to experience headaches, amnesia, tremors, and difficulty walking in October 2010. Doctors couldn't determine the cause—they suspected a stroke, then discovered that he had a disease called corticobasal syndrome, which kills off brain cells.

Then, numbers began to look strange to RFS. The 4 on a clock might flip backward, for instance. Eventually, numerals deteriorated into messy, unrecognizable "spaghetti" blobs—a disaster for someone who did math all the time. And it wasn't just his work that suffered. He couldn't read price tags or speed limit signs. At hotels, he had to mark the doorframe of his room with a magic marker.

Yet he could still do mental arithmetic and perform other mathematical operations. And strangely, although the digits 2 through 9 were scrambled, 0 and 1 looked normal—perhaps because those digits resemble letters, which RFS could read, or perhaps because they have associations with deep concepts such as absence and unity, which might allow his brain to process them. He eventually mastered an entirely new digit system (where ⌊ stood for 2, ⌈ for 8, etc.); determined to keep working, he had his computer rigged to present the new numerals onscreen.

In 2011, RFS was referred to a team of neuroscientists at Johns Hopkins University (JHU) led by Michael McCloskey and his then–graduate students Teresa Schubert and David Rothlein. One test the trio ran with RFS involved the green foam 8. Schubert, now at Harvard University, says the team hoped that touching the numeral would help RFS see it, but no. He could feel the figure's curves, but its image remained stubbornly scrambled—probably, Schubert says, because the brain prioritizes sight over other senses. And feeling one thing but seeing another left RFS disturbed, McCloskey says. "He understood that he could contribute to scientific knowledge, and he was willing to put up with the testing. But it was not pleasant."

The tests also revealed that RFS's deficit is not a simple visual malfunction. After all, he could see the foam 8's shape clearly in certain orientations. Rather, the deficit depended on his interpretation: As soon as his unconscious brain circuits registered a number, everything went haywire. That he could still interpret letters, Schubert says, lends support to the idea that the brain has a specialized module for processing numbers. And his ability to recognize and manipulate new figures as representations of numbers suggests his higher level math abilities remained intact.

Perhaps more important, RFS's deficit could shed light on how conscious awareness arises. In another test, the JHU scientists showed him large numbers and letters with tiny drawings of faces embedded inside them. When viewing the letters with embedded faces, RFS reported seeing both. Moreover, an electroencephalogram (EEG) recorded a characteristic brain wave called N170 that's strongly associated with seeing faces.

In contrast, when shown numbers with embedded faces, the number's effect apparently swamped that of the face: RFS reported seeing neither; everything looked like spaghetti. Yet an EEG still showed the characteristic N170 spike for registering faces. Somehow, his brain was still processing and identifying a face—a fairly high-level skill—even though his conscious mind was oblivious. This deficit shows high-level cognitive processing and consciousness are distinct, Koch says. "You can get one without the other."

Sara Ajina, a neuroscientist at the University of Oxford who studies visual awareness deficits such as blindsight—a residual, unconscious "sight" in people with damage to the brain's visual system—says she's often skeptical about how much scientists can learn from single case studies. But she praised the work with RFS for moving beyond just his basic number deficit to exploring how it affected face recognition and other higher level abilities.

Still, Ajina wishes the JHU team had embedded a wider variety of stimuli in the numbers. For instance, could RFS have noticed movement within the numbers, as opposed to static drawings? Given the human brain's emphasis on sight, perhaps movement could break through the number tangles and penetrate his conscious awareness.

Sadly, such questions will likely be impossible to answer: RFS's health has deteriorated recently, limiting his ability to speak and move. But his unique deficits and willingness to endure unsettling tests have already taught researchers a lot. "It's a very nice demonstration of all that awareness requires," Schubert says—as well as a reminder of just how fragile that gift is. ■

PHOTO: SCHUBERT *ET AL.*, *PNAS*, 117, 16055 (2020)

Published by AAAS

GENOMICS

Altered DNA base could play key role in pregnancy

In mice, methylated adenine enables the placenta to grow

By **Elizabeth Pennisi**

The genetic code in mammals may have gained another letter—or at least a significant footnote. Textbooks speak of four DNA building blocks, whose order specifies genes. But researchers in the field of epigenetics have also identified four chemically modified versions of these nucleotide bases, which affect how genes are expressed. Three of them are altered versions of cytosine, one of the original four bases. The last, methylated adenine, is mired in controversy about whether it exists in mammals.

Now, researchers at Yale University have not only identified plenty of this modified base in mouse embryonic cells, but also found it plays a key role in the development of the placenta. The methylated base—methyladenine for short—gives cells another epigenetic tool for turning genes on or off during normal development or in disease. "This is a very important study for our understanding of the role of [methylated adenine] in gene regulation," says Peng Jin, a geneticist at Emory University who was not part of the work.

Not all of an organism's genes are active at once, and when, where, and how strongly they turn on is what leads to specialized cells. Chemical modification of bases, known as epigenetic modification, helps pattern this gene activity. Animal cells often add methyl or other chemical groups to a cytosine to help switch a gene on or off, whereas bacteria rely heavily on methylation of adenines. But in the past 5 years, researchers have also detected methyladenines in fruit flies, nematode worms, and frogs. Hints of the modified base turned up in mammals, as well. "If [they] exist, it's an exciting new layer of biology," says Eric Greer, an epigeneticist at Harvard University.

But the reported levels were minuscule, and Greer and others found that many "detections" were the result of bacterial DNA contaminating samples, equipment, and even commercially available enzymes used in experiments. "It's difficult to rule out contamination by bacterial DNAs," agrees molecular biologist Ying Liu of Peking University, who has studied methyladenine in nematodes.

In addition, a sequencing technique that uses antibodies to pinpoint methylated adenines sometimes picks them up in RNA as well as in DNA. "The reality is, the techniques we have are not robust enough," says Colm Nestor, a molecular geneticist at Linköping University.

But Yale epigeneticist Andrew Xiao persisted. In 2018, his team found methyladenine in high amounts in human glioblastoma cells, where it seemed to affect the growth of this brain cancer. The researchers showed that when they inhibited a protein that removes adenine's methyl groups, causing methyladenine to accumulate, tumor growth slowed.

Earlier this month in *Nature*, Xiao, epigeneticist Haitao Li at Tsinghua University, and their colleagues reported they have found the modified base in normal mouse cells: the trophoblast stem cells, which eventually give rise to the placenta. The methyladenines are prevalent at places in DNA called M/SAR regions, which help create temporary "compartments" dividing active and inactive regions of the genome. As these compartments form, DNA's double helix unwinds briefly.

Using antibodies but taking precautions to avoid picking up methyladenine in RNA, Xiao and his colleagues discovered methyladenines piled up in unwound DNA. Other experiments revealed that these methyladenines prevented the binding of the SATB1 protein needed to rewind the DNA. The unwinding in turn blocks the expression of genes that would make trophoblast stem cells differentiate and stop growing. Instead, the cells multiply, ensuring there are enough of them to make the full placenta. Later, the methyls are removed—the researchers don't know how—and the cells start to specialize to make the placenta.

The same process may govern placental development in humans, says Indira Mysorekar, a reproductive biologist at Washington University School of Medicine in St. Louis. If so, she adds, flaws in the process could explain problems in embryonic growth and development. "The paper is exciting."

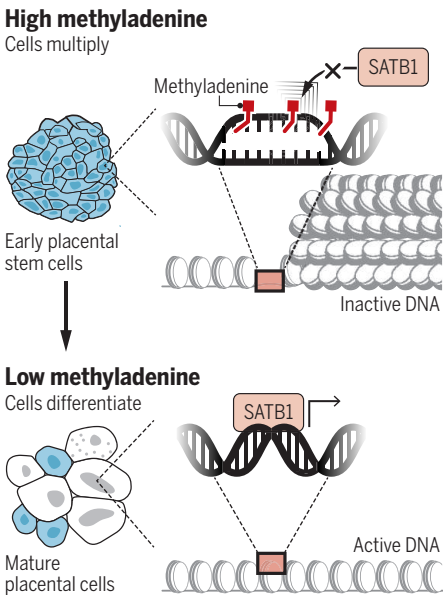
Researchers pursuing methyladenine also see it as vindication. "It provides one of the best supports so far of [the importance of methyladenine] in a mammalian system," says Chuan He, a biochemist at the University of Chicago. In the May issue of *Molecular Cell*, his group reported that methyladenine increases more than 1000-fold in mammalian mitochondrial DNA when cells are starved of oxygen—an earlier hint that it plays a biological role.

Nestor admires Xiao's careful experiments, but stresses that "the devil is in the details." Methyladenine, he says, "could be one of the epigenetic discoveries of the last few decades, or it could be just a cautionary tale," about researchers being too hopeful when contamination and imprecise methods prevail. Xiao is confident others will replicate his work with careful control of growth conditions. "If you follow these procedures, you shouldn't have any problems detecting these modifications," he says.

But even if the discovery is confirmed, Peng says he is leery of calling the modified adenine an addition to the genetic code. "Maybe [part] of the 'epigenetic code' instead?" ■

A switch for placental growth

In mice, a modified DNA base called methyladenine keeps DNA unwound by blocking the SATB1 protein, which spurs placental stem cells to multiply. Removal of the methyladenine allows the double helix to close and causes the cells to differentiate.



GRAPHIC: ZHENG LI AND ANDREW Z. XIAO, ADAPTED BY V. ALTOUNIAN/*SCIENCE*

Published by AAAS

LIZARD MAN

For Jonathan Losos, tiny Caribbean islands and their reptile inhabitants are test tubes of evolution

By **Elizabeth Pennisi**

The morning of 17 October 1996 started as usual for Jonathan Losos. The evolutionary biologist donned a broad hat and slathered on sunscreen, then headed by boat to several unnamed islets off Great Exuma Island in the Bahamas. Three years earlier, he and ecologist David Spiller had introduced local lizard species there to learn how they would compete in a once lizardless place. The pair spent the day snaring lizards, noting their exact locations, and taking stock of the insects, spiders, and vegetation. They were worried about reports of an impending hurricane, but the locals seemed confident it would veer off and spare the islands, as usual.

Not this time, however. The next day, Losos and Spiller helped their hotel owner board up the windows of their beachfront cottage on Great Exuma as Hurricane Lili bore down on the island. As the wind picked up and the first squalls dumped rain, they scurried to a cinder block building up a hill. That night, the wind blew off parts of the roof and felled palm trees. A 4-meter storm surge flooded the streets, and 2 days later they found their rented motorboat stuck in a tree.

The lizards had it even worse. When Losos and Spiller finally made it back out to their most exposed study sites, the islands were stripped nearly bare of brush and all the lizards were gone. But the setback for Losos's project was the start of a new chapter in his

PHOTO: DAY'S EDGE PRODUCTIONS

Published by AAAS

research on how the animals adapt to the varied, changeable environments on islands in and around the Caribbean.

Since Lili, a half-dozen other hurricanes have inundated islets and swept away animals relocated there by Losos, who is based at Washington University in St. Louis (WashU), and his team. But he and his colleagues have persevered, collecting data on how the animals adapt to predators, storm damage, and other challenges—natural and those contrived by the researchers. A lifelong reptile enthusiast, Losos is driven in part by his passion for a group of lizards called anoles, which thrive in South and Central America and throughout the Caribbean.

He also views them as an opportunity. Almost half of the 400 anole species live on islands, and their diverse lifestyles, habitats, and histories have proved to be a vehicle for exploring some of evolution's biggest questions. "Jonathan's islands are like giant test tubes, and he is the ultimate tinkerer," says Martha Muñoz, an evolutionary biologist at Yale University.

Losos's research on anoles has shown that evolution can happen faster than most scientists had assumed, and that—contrary to what some leading thinkers have proposed—it is often predictable. Faced with similar challenges, separate populations often evolve similar solutions. Along the way, Losos has mentored dozens of young scientists, and some are now carrying his work in new directions. "Beyond his many contributions to the field, Jonathan has also changed the course of science simply by being who he is," Muñoz, a former student, says. "He is proof that success is richer and more rewarding when accompanied by kindness and humility."

Called the Pinocchio anole (*Anolis proboscis*), this South American lizard is one of Jonathan Losos's favorites and was once thought to be extinct.

PHOTO: LUCAS BUSTAMANTE

ODDLY ENOUGH, THE 1950S TV show *Leave it to Beaver* started Losos down this path. When 7-year-old Beaver brought home a pet alligator, young Losos asked his parents whether he, too, could get one. His mom was against it, but his father said he would ask a family friend, the deputy director of the St. Louis Zoo, for advice. A successful businessman, the senior Losos also loved animals, taking his family on nature vacations, joining the zoo's board, and even financing the zoo's acquisition of a baby elephant from Thailand, which he named Carolyn in honor of his wife.

To everyone's surprise, the director heartily approved, saying that having an alligator as a childhood pet was how he got his start in herpetology. So the junior Losos acquired several baby caimans, which lived in a baby pool in the basement in winter and in a horse trough in the yard the rest of the year. Only a few times did the animals escape and terrorize the neighbors. Losos worked summers at the zoo until partway through college, eventually donating his caimans to a zookeeper. "Jonathan started off as a little kid loving nature, endlessly pestering staff at his local zoo, catching lizards on family vacations, and he's never lost that spark," says Harry Greene, herpetologist emeritus at Cornell University and Losos's graduate school adviser.

As an undergraduate at Harvard University, Losos fell under the tutelage of herpetologist Ernest Williams. Sometimes referred to as the father of anole biology, Williams had recognized that anoles on different Caribbean islands evolved independently. Yet on each island he'd found a similar set of body types or "ecomorphs"—one specialized for living in the brush, another for gripping twigs, and still others for life high in the trees. These parallels suggested that where circumstances were similar, evolution would converge on the same set of traits and form communities with similar sets of species.

Williams's lab had already produced several leading evolutionary biologists, and Losos figured the field of anole research was getting too crowded. But no other species both captured his interest and was easy to study. "I went through a dozen failed Ph.D. projects," he recalls. At a low point, he seriously considered law school, but his dad convinced him that the world needed herpetologists more than lawyers.

Losos eventually realized that anoles were perfect for applying new tools in evolutionary biology. Researchers were just beginning to build family trees and trace evolution based on protein variations among species. For his Ph.D., Losos compared proteins in Caribbean anoles and verified that Williams's ecomorphs had indeed evolved independently to form similar communities on different islands (*Science*, 27 March 1998, p. 2115). That insight alone—support for an idea called convergent evolution—"was a really important breakthrough," says Frank Burbrink, a herpetologist at the American Museum of Natural History.

Meanwhile, other researchers were calling for more rigor in evolution studies by requiring evidence that supposedly adaptive traits really give an organism an advantage. So Losos began to study different anole ecomorphs, with legs and toepads of varying sizes (see graphic, p. 499). In the lab, he ran them down miniature racetracks and assessed how well they clung to smooth, vertical surfaces. He found that lizards living near the ground, close to predators, had longer legs that made them fast, whereas those living higher in brush and trees had bigger toepads to stick to leaves and smooth bark. By combining these data with his family tree studies, he got a clearer sense of the lizards' evolutionary history. He "was really one of the first people to move the field into doing evolution by integrating ecology and morphology and getting the bigger picture," Burbrink says.

Inspired by experiments in which researchers monitored evolutionary changes in guppies in Trinidad after relocating them to different streams (*Science*, 24 August 2012, p. 904), Losos began to wonder whether similar studies could be done in Caribbean anoles. And he realized that Thomas Schoener, one of Williams's protégés, had already laid the groundwork. In the 1980s, Thomas and Amy Schoener (they were once married) introduced local lizards to tiny lizardless islands in the Bahamas to investigate how different vegetation affected the reptiles' ability to thrive.

A decade later, Losos teamed up with Thomas Schoener, by then a renowned ecologist at the University of California (UC), Davis, to revisit those sites. Consistent with Williams's and Losos's earlier findings, lizards living in scrubby vegetation had shorter legs and larger toepads than their ancestors, which had lived in tall, broad trees. These adaptations enabled them to cling to tiny twigs as they chased down insects to eat, and the changes had taken just a few generations. "Evolution can happen very quickly when natural selection is very strong," Losos says.

The idea is now well-accepted, but at the time it went against the entrenched belief that evolution was a slow process. "This is

Published by AAAS

one of the few things that [Charles] Darwin got wrong," Losos says. He decided to make anoles his life's work.

HE SOON HAD TO RECKON with hurricanes. Losos and Spiller, now retired from UC Davis, had chosen the islets off Great Exuma to study the effects of competition. On some, they introduced two local species, the green and brown anoles, and on others, just a single species. In the first 3 years, they noticed that on islands with both kinds, the brown lizards were driving the green anoles higher into the bushes, where they were struggling. That's when Lili hit, ruining the experiment before they could see whether the green anoles would go extinct.

"It would have been so easy, I'm sure, to pack it all in and give up," says Luke Harmon, one of Losos's former students and now an evolutionary biologist at the University of Idaho. Instead, Losos and Spiller used the disaster to their advantage. They documented Lili's great, but also patchy, impact. Islands southwest of Great Exuma felt the brunt of the storm surge and were devoid of lizards and vegetation. Life there would have to start over. On islands to the north, the wind and rain snapped twigs and ripped off leaves but a few lizards remained, they reported in the first of several papers about hurricanes (*Science*, 31 July 1998, p. 695).

The work challenged a widespread assumption that extreme events such as hurricanes do not drive evolution because they are rare and have random, unpredictable impacts on plants and animals. The Losos group discovered instead that storms can be agents of natural selection.

For example, in 2017 Losos's postdoc Colin Donihue; functional morphologist Anthony Herrel, now with the French national research agency CNRS at the National Museum of Natural History; and colleagues visited two cays in the Turks and Caicos to measure the body proportions of the anoles living there. Four days after they left, two almost back-to-back hurricanes hit the area with winds of more than 200 kilometers per hour. When the team returned a few weeks later and remeasured the lizards, they found that the survivors tended to have bigger toepads, longer forelimbs, and shorter hindlimbs.

Back in the lab, the researchers tested how these traits affect the lizards' ability to hold onto a perch. In a strong wind, anoles hang on with their forelimbs, but they lose their grip with the hind legs. Cranking up an air-blower, the researchers found that those with longer hind legs (and more surface area for the wind to catch) got blown off their perches onto a padded surface more readily. Conversely, animals with shorter hind limbs and bigger toepads hung on. The hurricanes had apparently selected for those traits, the team reported in 2018. The following year, they found that offspring of the survivors also had big toepads, suggesting the adaptation was genetic and not just a reaction to holding on tight.

The team has since measured toepad size in 188 lizard species across the Caribbean. The more hurricanes an island has experienced, the bigger the toepads of the lizards living there, they reported on 27 April in the *Proceedings of the National Academy of Sciences*. Hurricanes seem to have had a long-term evolutionary effect.

Jonathan Losos (front right), with a Harvard University herpetology class in Costa Rica, has made a name as a mentor as well as a researcher.

LOSOS HAD BEEN A PROFESSOR at WashU for 13 years when Harvard came calling in 2005, seeking to recruit him to its evolutionary biology department. A St. Louis native and a hardcore St. Louis Cardinals baseball fan, he hesitated. He even did a yearlong sabbatical at Harvard before finally accepting, in large part because the position included a curatorship at the Harvard Museum of Comparative Zoology. "That was the one thing St. Louis didn't have," he recalls.

There, he continued to build on a reputation for being a kind, enthusiastic mentor. "I have seen him give high school students the same attention and respect that he gives his closest colleagues," says Melissa Kemp, a former postdoc now at the University of Texas, Austin. "He seems to always be focused on his work, but he also has a whimsical sense of fun at the same time," says Michele Johnson, a former student and an evolutionary biologist at Trinity University. Losos sports a watch with an anole he photographed as its face and is not above lecturing undergraduates while dressed as a platypus—one of his favorite animals since childhood.

Those traits and a firm belief that "there is no one-size-fits-all in terms of how to interact with and mentor students" have helped Losos launch the careers of 59 graduate students and postdocs. They include at least eight Black, Latino, and Native American scholars, in a field that lacks diversity. (Although 3% of U.S. biologists are African American or Black, for example, only 0.3% of evolutionary biologists are.)

Ambika Kamath, now a postdoctoral researcher at UC Berkeley, says Losos backed her completely when her studies challenged the long-held idea that male lizards hold territories to corral their mates. She argued instead that females move around and play a role in mate choice. "It would have been much harder for me to do that work without his excitement," she recalls. Losos worked hard with her to get the paper just right and was eager to be a co-author. "Otherwise it would have just been the work of this young brown woman who could have easily been dismissed as an angry feminist."

Kamath and other students praise Losos for pushing them intellectually without undermining their confidence. Harmon jokes that Losos would never dismiss an idea from his students, no matter how wacky. Instead, he would just pause and say "interesting." "Eventually I figured out that maybe I should think things through a bit more, if Jonathan thought they were 'interesting,'" Harmon says.

LOSOS AND HIS TEAM keep testing their ideas about ecology and evolution on Caribbean islands. In one recent project, Robert Pringle, now at Princeton University, and Losos tested a key principle in ecology—that introducing a top predator tends to increase biodiversity. The researchers added a predatory ground-dwelling lizard to islands with brown and green anoles. To escape this new threat, the brown anoles began to hang out higher in the foliage, displacing the green anoles that normally lived there and driving them toward extinction. Contrary to conventional wisdom, the predator appeared to be pushing the islands toward lower biodiversity, they reported on 5 June in *Nature*.

PHOTO: COURTESY OF JONATHAN LOSOS

Published by AAAS

Another recent study, led by one of Losos's former postdocs, examined the impacts of an invasive anole species on Dominica. Until 20 years ago, the island was home to a single anole species. Then a lumber shipment introduced a second species that is gradually spreading. To study how the native and invader species interact, behavioral ecologist Claire Dufour, now at the University of Montpellier, used robotic lizards as stand-ins for the invader. The robots did pushups and extended a flap of fake skin under the chin, mimicking the aggressive displays of real lizards. In response, native lizards familiar with the invaders postured more aggressively, suggesting the invaders are forcing the natives to expend more energy defending their territory, the group reported on 27 March in the *Journal of Animal Ecology*. "Our biggest conclusion is that the species do compete and have negative consequences on each other," Losos says.

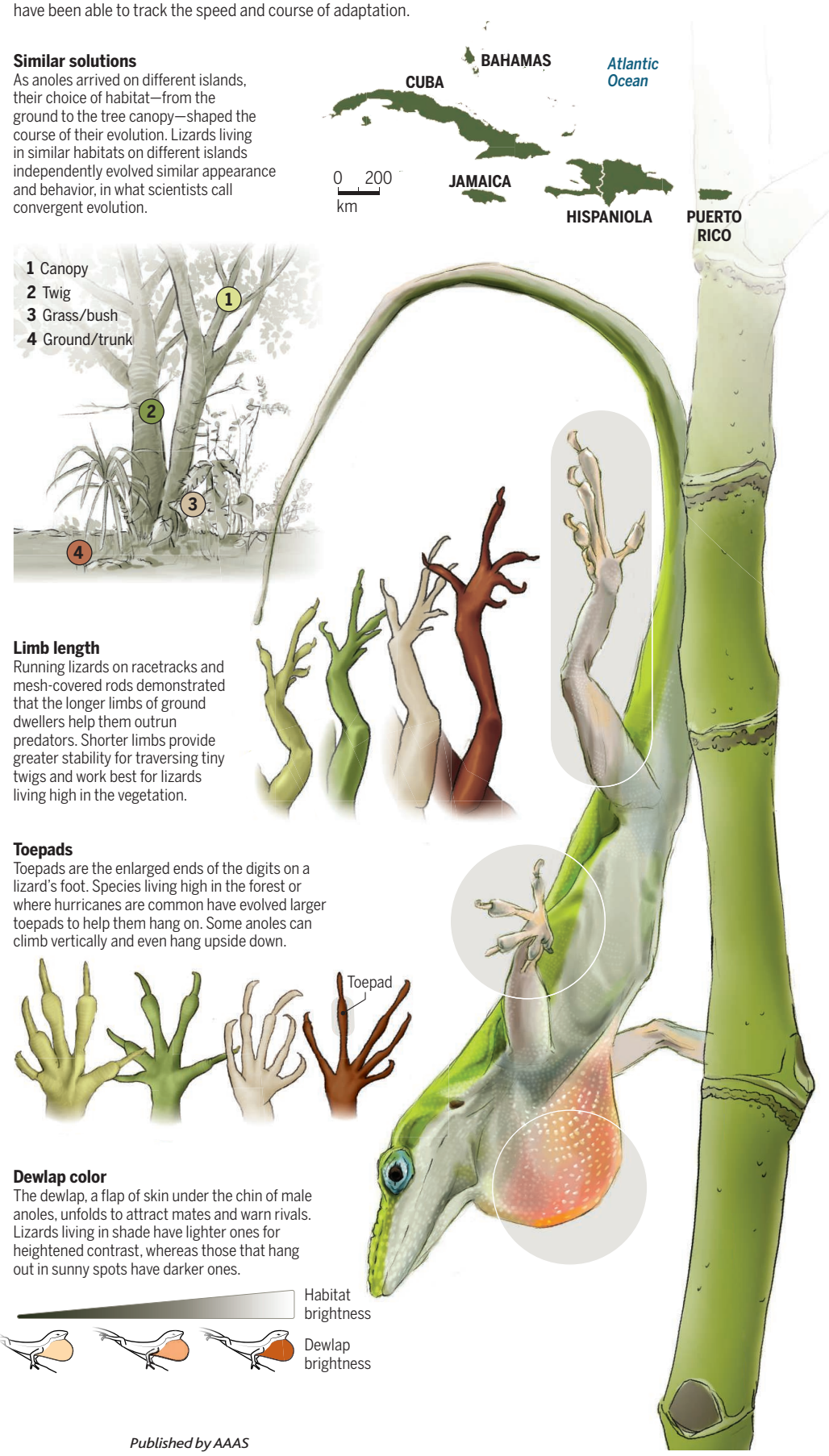
Even as his group continues to churn out papers, Losos is assessing what he has learned so far. In his book *Improbable Destinies: Fate, Chance, and the Future of Evolution*, published in 2017, he challenged a major contention of one of the field's great thinkers, Stephen Jay Gould, the Harvard paleontologist who argued that chance plays such a big role in determining nature's course that evolution would never take the same path twice. Anoles offer evidence to the contrary, Losos wrote: In similar habitats, they have repeatedly evolved similar body shapes, sizes, and behavior. The book was written for the general public, but it made an impression even on his peers. "I've been studying evolution for 30-plus years, and this book made me rethink some things I thought I knew about biology and evolution," says Christopher Austin, an evolutionary biologist and herpetology curator at the Louisiana State University Museum of Natural Science.

Losos left Harvard in 2018, lured by a new job at WashU and the prospect of returning to his hometown, his cats, and his wife, who has a successful real estate career and did not follow him to Massachusetts. He now heads the Living Earth Collaborative, a biodiversity research initiative that unites experts at the Missouri Botanical Garden, the St. Louis Zoo, and WashU. He is working on a book about evolution in the house cat, another of his favorite species.

And he is still dodging hurricanes. Losos and colleagues have been trying to assess the long-term evolutionary impacts of predatory lizards they've introduced to some islands in the Bahamas. "I don't know if we will ever get there," he says. Every few years a hurricane comes through and blows the evolving lizards away. ■

Evolution's stamp on island-dwelling lizards

On islands in and around the Caribbean, 173 species of anole lizards face an array of different environments, predators, and competitors, along with periodic storms. The result is a laboratory of evolution, where scientists have been able to track the speed and course of adaptation.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

Published by AAAS

INSIGHTS

Farm workers using the H-2A temporary agricultural worker visa maintain safe distance on 27 April 2020 in Greenfield, California.

PHOTO: BRENT STIRTON/GETTY IMAGES

POLICY FORUM

FOOD SECURITY: COVID-19

COVID-19 risks to global food security

Economic fallout and food supply chain disruptions require attention from policy-makers

By **David Laborde, Will Martin, Johan Swinnen, Rob Vos**

As the COVID-19 pandemic progresses, trade-offs have emerged between the need to contain the virus and to avoid disastrous economic and food security crises that hurt the world's poor and hungry most. Although no major food shortages have emerged as yet, agricultural and food markets are facing disruptions because of labor shortages created by restrictions on movements of people and shifts in food demand resulting from closures of restaurants and schools as well as from income losses. Export restrictions imposed by some countries have disrupted trade flows for staple foods such as wheat and rice. The pandemic is affecting all four pillars of food security (*1*): availability (is the supply of food adequate?), access (can people obtain the food they need?), utilization (do people have enough intake of nutrients?), and stability (can people access food at all times?). COVID-19 is most directly and severely impacting access to food, even though impacts are also felt through disruptions to availability; shifts in consumer demand toward cheaper, less nutritious foods; and food price instability. We outline the main threats COVID-19 poses to food security and suggest critical responses that policy-makers should consider to prevent this global health crisis from becoming a global food crisis.

FOOD ACCESS

COVID-19 threatens access to food mainly through losses of income and assets that prejudice ability to buy food. The poorest households spend around 70% of their incomes on food and have limited access to financial markets, making their food security particularly vulnerable to income shocks (*2*). As the economic costs of social distancing have become

International Food Policy Research Institute, Washington, DC, USA. Email: r.vos@cgiar.org

Published by AAAS

more evident, global economic forecasts have become increasingly pessimistic. In its most recent forecast, the International Monetary Fund (IMF) projects a 5% decline of the world economy in 2020 (*3*)—a much deeper global recession than during the global financial crisis of 2008–2009. The economic fallout in the initial epicenters of the pandemic (China, Europe, and the United States) is also hurting low- and middle-income countries through declines in trade, oil, and other commodity prices and restrictions on international travel and freight, compounding the economic costs of poorer nations' own COVID-19–related restrictions.

Lacking up-to-date household surveys for most countries, no precise estimates can be made yet regarding the impacts on global poverty and food insecurity. Model-based simulations suggest, however, that between 90 million and 150 million people could fall (or may already have fallen) into extreme poverty (*2*, *4*). Although any such estimate is highly uncertain given the rapid evolution of the pandemic, both of these projections involve substantial increases in global poverty, between 15 and 24% from existing estimated levels. Most of the poverty increases would be in sub-Saharan Africa and South Asia.

Declines in incomes and increases in poverty of this magnitude would have large impacts on food security and nutrition. People in extreme poverty do not have enough resources to buy the food they need to avoid hunger and undernourishment, and both poor and near-poor people will switch to cheaper and less nutritious foods. Even if the recession is short-lived, the impacts through inadequate nutrition could be long lasting, especially for young children, whose growth and cognitive development tend to be affected by undernutrition. Recent telephone survey evidence from Ethiopia (*5*) confirms many of these expectations, particularly that the main challenges to vulnerable households are resulting from income declines rather than food shortages. Although this survey points to rundown of savings as a key coping strategy up to June 2020, only 20% of households were found to have enough savings to meet their food needs for a month or more.

FOOD AVAILABILITY AND STABILITY

This pandemic poses several major threats to food availability and stability.

Agricultural production

Food security crises are often due to sharp declines in food production. Epizootic pandemics, such as avian flu or African swine fever, directly reduced animal-sourced food output. COVID-19 is likely different. It will probably have smaller direct impacts on agricultural production than those shocks and will affect food security mostly in different ways, differing by product and region.

In rich countries, production of staple crops (especially maize, wheat, and soybeans) tends to be highly mechanized, with much inherent social distancing of workers. Most farms deploy large-scale machinery and little labor for land preparation, sowing, and harvesting. Large-scale mechanization is more difficult or too costly for many nonstaple foods, such as fruits and vegetables, requiring human hands for planting, weeding, and/or harvesting. These more labor-intensive parts of agriculture often require changes in practices to reduce the risk of disease transmission, such as by avoiding concentration of workers on the field through staggered shifts. Other labor impacts come from restrictions on movement of seasonal farm workers of the type that have left food unharvested in Europe. These restrictions affect food production and can also have adverse food security effects by constraining the ability of workers from poorer countries to earn income (*6*).

In poor countries, farm production is mostly much more labor intensive, with many processes such as rice planting and harvesting of staple crops bringing workers close together. Although farmers in poorer countries are generally younger than in rich countries, health systems are usually weaker, and preexisting health challenges may increase people's vulnerability to COVID-19.

Supply chain disruptions

The vulnerability of food supply chains differs strongly across food systems, depending on the priority they are afforded and on their structure. We emphasize four features.

(i) Governments worldwide have placed high priority on ensuring that staple foods can be moved to consumers. Global supply chains for staple foods appear to have held up reasonably well so far, with relatively few cases of substantial supply disruptions even in countries with strict social distancing requirements. Evidence from China shows that such disruptions can be reduced by creating "green lanes" that exempt transport, production processes, and distribution of agricultural inputs and food products, as well as movements of food-sector workers, from COVID-19 lockdown measures.

(ii) Labor-intensive "traditional" value chains (mostly in poor countries) are more affected than capital-intensive "modern" food value chains (mostly in high-income countries or in richer parts of low- and middle-income countries). For example, in Ethiopia, which has poorly developed infrastructure and mostly relies on traditional distribution networks, the supply of vegetables has been affected by disruptions in transport and in the supply of key farm inputs (*7*). Also, impacts on food distribution in low-income countries tend to be worse for tens of millions of informal small- and medium-sized food-sector operations. These are typically labor intensive, with many people having to work close together in dense areas and crowded marketplaces, where the risk of COVID-19 transmission is extremely high, while social distancing measures directly affect the operation of food businesses and markets.

(iii) Even modern food supply chains and systems can be seriously affected. In the United States and Europe, more than 30,000 workers in food-processing plants have contracted COVID-19, causing meat processing plants to close or slow production. The near-total closure of international passenger aviation has seriously disrupted supply chains of specialized products that rely on air freight, such as high-value horticultural exports from Africa (*8*). The need to adjust the mix of products and packaging when demand shifts from restaurants to households drives declines in consumption of higher-quality meats, dairy products, and vegetables.

(iv) COVID-19 has affected public food distribution systems. For example, schools closing under India's national lockdown resulted in suspension of school feeding programs—one of the country's largest safety nets. School closures are also depriving many poor U.S. children of publicly provided meals. Farmers and other suppliers have found difficulty finding market outlets to replace institutional outlets such as restaurants and schools, resulting in substantial wastage of milk and other nutrient-rich foods. Other safety nets are also affected, including community nutrition programs for pregnant women and lactating mothers (*9*). Public food relief programs must also manage the risk of exposing more people to the virus by attracting large crowds at distribution points.

Trade restrictions

Although the vast majority of the world's food is produced and consumed in the same country, food trade within and between countries allows diversification of supplies that helps reduce vulnerability to food market shocks. However, policies can make it difficult for trade to play this stabilizing role.

During food crises in 2008 and 2010, many major producer countries imposed export restrictions on staple foods, especially rice and wheat, that caused world market prices to rise (*10*). Policy-makers often respond to a fear of imminent shortage or sharp price increase of a main food product by restricting its export to protect domestic consumers. Although such restrictions may serve a national interest in the short run, they reduce the supply to world markets, putting upward pressure on world prices.

Published by AAAS

This problem reemerged in March 2020, and by 6 July, 21 countries had announced or introduced (temporary) export restrictions covering almost 4% of the caloric value of globally traded food. Fortunately, most restrictions have been lifted, and by early July, only two countries continued to impose these measures, on a very small share of trade. However, the situation could again deteriorate [see (*11*) for up-to-date counts].

A key problem with food export restrictions is that they can create the upward spiral in world prices that they are intended to prevent (*12*). Almost all of the imposed measures limit the amounts of covered products that can be exported. Such quantitative restrictions destabilize prices because they reduce the ability of markets to adjust to production shocks through changes in exports. Expectations about the introduction and subsequent removal of these restrictions add to instability in food availability, with exports increasing before their imposition and stocks accumulating before their removal.

Since the onset of the pandemic, world wheat prices have been quite volatile (*13*), but prices have declined by around 10% between January and early July. By contrast, world market prices of rice rose around 20% between January and April and became highly volatile in May (*14*). The world market price for rice declined after the end of Vietnam's ban on rice exports to roughly their January 2020 levels by early July. Higher food prices benefit farmers but hurt consumers, and vice versa for lower food prices. However, price volatility tends to hurt all because it induces uncertainty in supply, whisking away investments that can improve productivity or food quality.

DIETS AND NUTRITION

Food shortages and lower incomes affect dietary choices. Analysis of 300,000 households in low- and middle-income countries reveals that poor people spend more than a quarter of their total income on staple foods such as wheat, rice, or maize, whereas nonpoor households spend only 14% (*2*). Poor households spend nearly 50% of their incomes on unprocessed nonstaple foods such as fruits, vegetables, and animal-source products. Declines in incomes are likely to force many poor households to cut back on these nonstaples. A recent global economic model scenario assessment suggests that these shifts away from more nutrient-rich nonstaples toward starchy staples are likely to be substantial (*2*). Recent survey evidence from Ethiopia confirms this, finding that reductions in household food consumption were mainly in nutrient-dense foods such as fruit, meat, eggs, and dairy (*7*). The shifts limit declines in calorie intake but increase deficiencies of micronutrient consumption, with lasting adverse consequences for human health and development. The greater disruptions in the supply of fruits, vegetables, milk, and meat products, relative to less disrupted staple foods, are reinforcing the income-related reduction in consumption of these foods, especially by poor households. This reduces dietary diversity, intake of micronutrients, and nutritional status, increasing the risk of adverse health consequences (*15*, *16*).

FOOD SECURITY DURING A HEALTH CRISIS

Because the most important impact of the pandemic on food security is through income declines that put food access at risk, social safety-net policies are particularly suited to the problem. By June, no fewer than 195 countries had planned or introduced additional social protection measures in response to COVID-19. Most are in the form of temporary (typically 3 months) but substantial enhancements of cash transfer programs (*17*). Cash transfers are easy to scale up and allow consumers choices on how to best meet their nutritional needs. Targeting of assistance is particularly important to ensure that the benefits reach those most in need. Targeting benefits to women helps improve nutritional outcomes. Given the fiscal challenges that face low- and middle-income countries and given the strong international spillover effects of the economic consequences of COVID-19, it will be important for high-income countries and international organizations to contribute as much as they can to support the responses of poor countries in financial need. Doing so would aid global economic recovery and avoid enormous humanitarian cost that a global food crisis would imply.

It is critical that agricultural inputs, farms, food processing, and distribution are declared essential and exempted from lockdown measures, so that food can flow in adequate amounts from farm to fork. Health protocols are needed to protect workers in food chains and to help contain COVID-19. Incentives and support for food transport and logistics, including deliveries to needy areas and for the sick, are also important. Likewise, governments should engage with market participants to ensure the smooth functioning of agricultural input markets (seeds, fertilizer, labor, and credits), especially for time-critical inputs for planting and harvesting. Allowing movement of seasonal and migrant labor is also important in many contexts. The European Union, for example, has encouraged its member countries to consider all workers (including seasonal and migrant) in fruits and vegetable production as critical. Governments should avoid further use of disruptive policies, such as export restrictions on food, and keep trade channels open consistent with the multilateral rules and regulations as agreed through the World Trade Organization. In addition, they should ease trade transactions, including through electronic issuance of permits and certificates, and ensure that inspection requirements are compatible with social distancing.

Last, but not least, COVID-19 has highlighted the importance of early detection of new infectious diseases, 70% of which have their source in animals. Improving surveillance systems for zoonotic diseases arising from animals used in the food chain is vitally important for avoiding future catastrophes. ■

REFERENCES AND NOTES

1. Committee on World Food Security (CFS), "Reform of the Committee on World Food Security," CFS:2009/2 Rev.2 (Food and Agriculture Organization of the United Nations, 2009); http://www.fao.org/3/a-k7197e.pdf.
2. D. Laborde, W. Martin, R. Vos, "Estimating the poverty impact of COVID-19: The MIRAGRODEP and POVANA frameworks," IFPRI Technical Note (IFPRI, 2020); https://tinyurl.com/y9fazbzf.
3. IMF, *World Economic Outlook Update: A Crisis Like No Other, An Uncertain Recovery* (IMF, June 2020). https://tinyurl.com/yazvtk7d.
4. D. Gerszon Mahler, C. Lakner, R. Castaneda Aguilar, H. Wu, *World Bank Blog*, 8 June 2020; https://tinyurl.com/yddap4vx.
5. G. Abate, A. de Brauw, K. Hirvonen, Ethiopia Strategy Support Program Paper 145, June (International Food Policy Research Institute, 2020); https://ebrary.ifpri.org/digital/collection/p15738coll2/id/133766/rec/25.
6. International Organization for Migration (IOM), Migrants and global food supply, COVID-19 Analytical Snapshot #18 (IOM, 2020); https://tinyurl.com/y7ocan7d.
7. S. Tamru, K. Hirvonen, B. Minten, *IFPRI COVID-19 blog series*, 13 April 2020; www.ifpri.org/blog/impacts-covid-19-crisis-vegetable-value-chains-ethiopia.
8. N. Bhalla, E. Wuilbercq, "No bed of roses: East Africa's female flower workers lose jobs as coronavirus hits exports" *Reuters*, 11 April 2020; https://tinyurl.com/y7yvm69h.
9. M. Hidrobo, N. Kumar, T. Palermo, A. Peterman, S. Roy, "Gender-sensitive social protection: A critical component of the COVID-19 response in low- and middle-income countries," IFPRI Issue Brief, April 2020; http://ebrary.ifpri.org/utils/getfile/collection/p15738coll2/id/133701/filename/133912.pdf.
10. K. Anderson, M. Ivanic, W. Martin, in *The Economics of Food Price Volatility*, J.-P. Chavas, D. Hummels, B. Wright, Eds. (Univ. Chicago Press, 2014); www.nber.org/chapters/c12818.pdf.
11. IFPRI, COVID-19 food trade policy tracker; https://tinyurl.com/ya5p9lfb.
12. A. Bouët, D. Laborde Debucquet, *Weltwirtsch. Arch.* **148**, 209 (2012).
13. Food Security Portal, Hard Wheat excessive food price variability early warning system; www.foodsecurityportal.org/hard-wheat-price-volatility-alert-mechanism.
14. Food Security Portal, Rice excessive food price variability early warning system; https://tinyurl.com/y83b68by.
15. D. Headey, M. Ruel, *IFPRI COVID-19 blog series*, 23 April 2020; https://tinyurl.com/yctvfcef.
16. S. McAuliffe *et al.*, *BMJ Nutr. Prevent. Health* 10.1136/bmjnph-2020-000100 (2020).
17. U. Gentilini *et al.*, "Social protection and jobs responses to COVID-19: A real-time review of country measures," World Bank living paper version 11 (12 June 2020); https://tinyurl.com/yd4g4z45.

ACKNOWLEDGMENTS

This paper benefitted from comments by five anonymous reviewers. Funding: This work was undertaken as part of the CGIAR Research Program on Policies, Institutions, and Markets (PIM) led by the International Food Policy Research Institute (IFPRI) and the Ceres2030 Project on achieving SDG2, also led by IFPRI in collaboration with Cornell University, the International Institute for Sustainable Development (IISD), and other partners.

10.1126/science.abc4765

Published by AAAS

Though air quality has improved overall, such as from 1988 (left) to 2006 (right) in Los Angeles, many communities remain particularly exposed to pollution.

PHOTOS: (LEFT TO RIGHT) NIK WHEELER/GETTY IMAGES; EKASH/GETTY IMAGES

AIR POLLUTION

Mapping the clean air haves and have-nots

The distribution of air pollution has changed little, perpetuating disparities in exposure

By **Lala Ma**

Environmental quality can affect a host of health and economic outcomes, and the heterogeneous distribution of pollution potentially explains a substantial portion of the gaps in outcomes between groups of different socioeconomic statuses (*1*). Founded on the Civil Rights Movement of the 1950s and 1960s and the antitoxics movement in the late 1970s, the environmental justice movement emerged to address the fact that exposure of minorities to pollution was not necessarily random (*2*). Yet, despite documentation of environmental disparities between different racial and ethnic groups (*3*, *4*), we have little systematic evidence about the evolution of environmental disparities over time and the factors contributing to their change. On page 575 of this issue, Colmer *et al.* (*5*) document trends in the spatial distribution of airborne particulate matter smaller than 2.5 μm in diameter ($PM_{2.5}$). They find that although $PM_{2.5}$ concentration has substantially decreased in North America over the past three decades, the relative ranking of $PM_{2.5}$ across the United States has remained notably stable: The most and least polluted tracts in 1981 remain the most and least polluted in 2016, with disadvantaged communities more likely to have higher pollution ranks at any given time.

Since 1981, U.S. federal policy-making has required the use of benefit-cost analysis (BCA) to evaluate economically important regulations. In 1994, the U.S. government officially recognized environmental justice as a necessary component in federal decision-making. Yet, although regulatory analysis requires investigation of a policy's distributional impact, the degree to which distributional effects are factored into decision-making is unclear when economic efficiency is the primary framework for policy evaluation. BCA tests whether a policy's winners gain enough to compensate the losers (the Kaldor-Hicks criterion), but it does not outline how (or require that) these compensations will be made in practice. Other policy tools (e.g., income transfers) could also arguably be used to achieve distributional goals in a more direct and efficient manner, and pragmatic reasons of political opposition, lack of technical guidance for distributional analysis, and data constraints can prevent BCA from incorporating distributional analysis (*6*). Furthermore, although theories about the causes of environmental injustices exist—from firm siting decisions to individual location choice and the interplay between these forces as mediated through (Coasian) bargaining and regulatory intervention—there is still considerable uncertainty as to the weight that each pathway holds in contributing to inequitable exposure (*7*). This creates difficulty in crafting policies that produce the intended consequences.

Colmer *et al.*'s examination of the persistence in neighborhood pollution rankings is an initial step toward understanding broader trends in the distribution of environmental well-being. They combine newly available data on annual $PM_{2.5}$ estimates at 0.01°-by-0.01° grids from (*8*) with census-tract boundaries. The paper's application of fine-grained pollution data to neighbor-

Department of Economics, University of Kentucky, Lexington, KY, USA. Email: lala.ma@uky.edu

Published by AAAS

hood characteristics improves the measurement of neighborhood exposure because local pollutants can vary greatly over short distances and yield disparate health implications within communities. The focus on ambient air quality, rather than on a particular type of polluting facility, broadens our perspective on air pollution exposure, because it captures emissions from both stationary and mobile sources. When comparing a tract's pollution rank with the characteristics of its residents, the authors find that changes in rank are not strongly correlated with changes in population size or aggregate demographic characteristics.

Colmer *et al.*'s findings highlight outstanding issues relevant to implementing environmental justice in practice. It is, for instance, unclear what rank persistence for locations implies about the trend in distributional equity. Rank persistence may be of second-order importance for environmental equity if considered alongside the magnitude of air quality improvements across the entire pollution distribution. Regulation, with its focus on efficiency, is likely to target the dirtiest places first, which would drive pollution reductions in the most disadvantaged communities. Thus, air quality could be improved for all people and disproportionately for people of lower socioeconomic status even if the relative positions of groups within the pollution distribution remained the same. Conversely, the persistence of pollution rankings and fall in pollution across locations may not translate to reduced exposure for populations. Census tracts, with populations of between 1200 and 8000, vary greatly in their pollution sources, and populations are unevenly distributed within tracts. Aggregating exposure to broader communities can obscure the relationship between pollution and socioeconomic status (*9*). In the face of air quality changes, people may respond (by migrating) in ways that mitigate or amplify disparities in exposure without being reflected in aggregate data (*10*).

The implications of behavioral responses become even more relevant after factoring in interactions with markets. People are heterogeneous in their willingness and ability to pay for clean air. If air quality improvements drive up demand for particular areas and are reflected in higher housing prices, then residents may find themselves paying a higher price for cleaner air, even if they would rather spend their resources on other things. Relocation is potentially a costly option; however, with the persistence in rankings, those living in the highest pollution-ranked communities may not have this option. Forces such as environmental gentrification can alter welfare (and gaps in welfare) as a result of the equilibrium market price effects that accompany improvements in environmental quality. Concern that low-income and other vulnerable populations are being "priced out" are reflected in calls to consider the effects of social displacement in the evaluation of climate resilience programs (*11*). Other forces such as stigma (*12*) or endogenous neighborhood change resulting from the location choices of initial residents (*13*) may prevent prices from responding to environmental improvements and reduce the risks of gentrification and displacement.

Air pollution, and $PM_{2.5}$ in particular, is only one facet of environmental health. The story becomes more complex when we consider cumulative exposure that combines environmental assaults from alternative pollution pathways. More broadly, increasing global climate risks have implications for inequitable exposure to a host of other threats, such as hurricanes, floods, and heat waves. These risks will also have implications for distributional equity through differential access to insurance and social safety nets that mitigate losses and facilitate adaptation (*14*). Further investigation to determine what caused the downward shift in the pollution distribution and the preservation of rankings over time could inform policies that avoid undesirable distributional consequences. Implementing environmental justice is therefore ultimately inseparable from assessing the mechanisms that created the inequities in the first place. ■

REFERENCES AND NOTES

1. J. Currie, *Am. Econ. Rev.* **101**, 1 (2011).
2. L. W. Cole, S. R. Foster, *From the Ground Up: Environmental Racism and the Rise of the Environmental Justice Movement* (NYU Press, 2001).
3. U.S. General Accounting Office (GAO), "Siting of hazardous waste landfills and their correlation with racial and economic status of surrounding communities" (Technical Report RCED-83-168, GAO, 1983).
4. Commission for Racial Justice, United Church of Christ, "Toxic wastes and race in the United States: A national report on the racial and socio-economic characteristics of communities with hazardous waste sites" (Technical Report, United Church of Christ, 1987).
5. J. Colmer *et al.*, *Science* **369**, 575 (2020).
6. L. A. Robinson, J. K. Hammitt, R. J. Zeckhauser, *Rev. Environ. Econ. Policy* **10**, 308 (2016).
7. S. Banzhaf, L. Ma, C. Timmins, *J. Econ. Perspect.* **33**, 185 (2019).
8. J. Meng *et al.*, *Environ. Sci. Technol.* **53**, 5071 (2019).
9. P. Mohai, R. Saha, *Demography* **43**, 383 (2006).
10. B. Depro, C. Timmins, M. O'Neil, *J. Assoc. Environ. Resour. Econ.* **2**, 439 (2015).
11. I. Anguelovski *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 26139 (2019).
12. K. Messer, W. Schulze, K. Hackett, T. Cameron, G. McClelland, *Environ. Resour. Econ.* **33**, 299 (2006).
13. H. S. Banzhaf, R. P. Walsh, *J. Urban Econ.* **74**, 83 (2013).
14. S. L. Cutter, *Hazards Vulnerability and Environmental Justice* (Routledge, 2012).

ACKNOWLEDGMENTS

I thank G. Blomquist and C. Timmins for helpful comments and suggestions.

10.1126/science.abb0943

SOLAR PHYSICS

Can we predict solar flares?

A physics-based model may help improve space weather forecasts

By **Astrid M. Veronig**

Flares from the Sun are the strongest explosions in our Solar System. They can cause severe space weather disturbances, posing a hazard to astronauts and technological systems in space and on the ground. Solar flares have an immediate impact in the form of enhanced radiation and energetic particles in as little as 8 min after the start of the event. Reliable prediction methods for flares are needed to provide longer warning times. However, pinning down the flare onset conditions is necessary for reliable predictions and is still a struggle (*1*). On page 587 of this issue, Kusano *et al.* (*2*) introduce a method to predict and successfully test for large imminent flares.

Since their discovery more than 160 years ago by Carrington and Hodgson (*3*, *4*), flares have been associated with sunspots on parts of the Sun with strong magnetic fields called active regions. The vast amount of flare energy is stored in complex (nonpotential) active region magnetic fields. The energy is impulsively released by magnetic reconnection, a fundamental plasma physics process that changes the topology of the magnetic field and converts magnetic energy into kinetic energy, thermal energy, and the acceleration of high-energy particles (*5*).

Solar flares have been extensively studied for many decades. Now, regular observations are made at various wavelengths from a large number of ground- and space-based observatories. Regular measurements are made of the magnetic field and its vector components in the photosphere, which is considered the "surface" of the Sun. Despite extensive observations, the specific onset conditions and what triggers a flare are not understood. The lack of magnetic field measurements in the corona, where the field reconfiguration causing the sudden energy release in flares takes place, is a major limitation. Analytical and numerical methods are used instead to reconstruct the coronal magnetic field (*6*), model the instability and evolution of the field using magnetohydrodynamics (*5*, *7*, *8*), and indirectly infer magnetic reconnection

Published by AAAS

signatures from the spatial and thermal distribution of the flaring plasma (*9*).

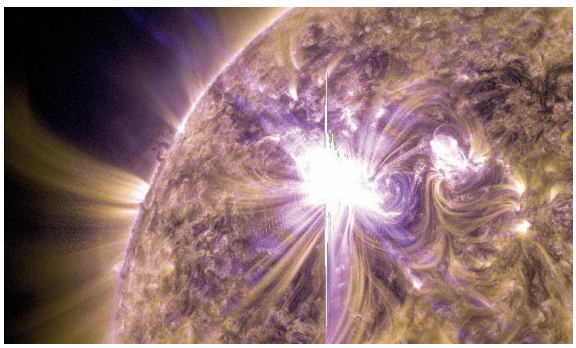
Magnetohydrodynamic models are an important means for understanding the physics of solar eruptions and their onset conditions, but they cannot be used to predict the time of a flare. Current flare prediction schemes are mostly empirical and based on parameterizations of the surface magnetic field, such as the total magnetic flux of active regions, strong field gradients, and so on (*10*, *11*). However, the forecast accuracy of empirical flare prediction schemes is still rather low, as measured by metrics like the skill score. Kusano *et al*. chose a different, physics-based approach for predicting imminent large flares through a critical condition of a magnetohydrodynamic instability that is triggered by magnetic reconnection.

The method of Kusano *et al*. is based on "double-arc instability" (*12*) that allows a stability assessment of the sigmoidal field that is formed by reconnection between two sheared magnetic flux systems. The instability starts with magnetic reconnection on small spatial scales, called "trigger-reconnection," between the sheared magnetic loops to create a double-arc loop, which contains the magnetic free energy that can be released during a flare. When the instability grows, the double-arc loops move upward, causing further magnetic reconnection that provides a positive feedback to the instability (see the figure).

The double-arc instability is initiated when the ratio of the poloidal flux in the double arc to the flux of the stabilizing overlying magnetic field exceeds a limit. Based on this criterion, Kusano *et al*. determine the critical length-scale for the trigger-reconnection to destabilize the double arc and estimate the energy that can be released. The predictive model was tested on 200 active regions during solar cycle 24 that contained the largest sunspots but which did not produce large flares, in comparison with all seven active regions that produced large flares above class X2. X-class flares are powerful enough to trigger long-lasting radiation storms. Magnetic field vector measurements of the solar photosphere by the Helioseismic and Magnetic Imager (*13*) onboard NASA's Solar Dynamics Observatory were used for data. The authors found that the location and the time evolution of the identified critical regions provide a precursor to large flares, with lead times of 1 to 24 hours.

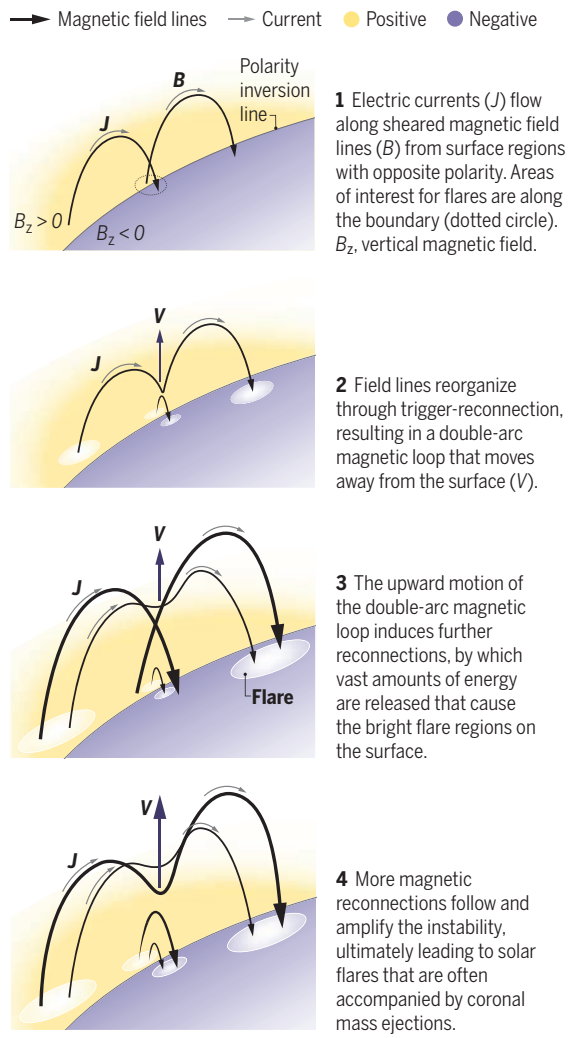
The two large flares for which this scheme did not work were from a very specific large active region (AR12192) present on the Sun in October 2014. AR12192 was the source of numerous large flares, including six X-class flares, but none of them associated with a large cloud of plasma and embedded magnetic field ejected from the Sun known as a coronal mass ejection (*14*). This is a strong exception to the general statistics, because >90% of all large flares are accompanied by coronal mass ejections (*15*). However, the failure of the Kusano *et al*. prediction scheme for these events is also relevant because it shows that flare prediction methods have specific difficulties for large flares that are not accompanied by coronal mass ejections. This may be related to flare reconnection occurring relatively high in the corona or to strong overlying fields preventing an eruption (*2*, *14*).

Predicting solar flares is a very challenging task. The physics is complex and covers a large range of spatial scales, and key observables like the coronal magnetic field are lacking. Finally, the potential that flares are inherently stochastic processes cannot be ruled out. Nonetheless, tackling this issue has occurred by working along different paths. The diverse set of approaches will gain enormously from the upcoming observations of the 4-m Daniel K. Inouye Solar Telescope (DKIST), which had first light in December 2019. DKIST will provide improved resolution of the solar magnetic field fine structure and its dynamics and provide measurements of the coronal magnetic field. These key data are vital for a better understanding and probing of the onset and physics of solar flares. ■



Forecasting solar flares

A physics-based model helps better estimate where and when large solar flares (shown above) will erupt.



1 Electric currents (*J*) flow along sheared magnetic field lines (*B*) from surface regions with opposite polarity. Areas of interest for flares are along the boundary (dotted circle). B_z, vertical magnetic field.

2 Field lines reorganize through trigger-reconnection, resulting in a double-arc magnetic loop that moves away from the surface (*V*).

3 The upward motion of the double-arc magnetic loop induces further reconnections, by which vast amounts of energy are released that cause the bright flare regions on the surface.

4 More magnetic reconnections follow and amplify the instability, ultimately leading to solar flares that are often accompanied by coronal mass ejections.

Institute of Physics and Kanzelhöhe Observatory for Solar and Environmental Research, University of Graz, Graz, Austria. Email: astrid.veronig@uni-graz.at

CREDITS: (PHOTO) NASA/GSFC/SDO; (GRAPHIC) N. CARY/*SCIENCE*

REFERENCES AND NOTES

1. L. Green, T. Török, B. Vršnak, W. Manchester IV, A. Veronig, *Space Sci. Rev.* **214**, 46 (2018).
2. K. Kusano *et al.*, *Science* **369**, 587 (2020).
3. R. A. Carrington, *Mon. Not. R. Astron. Soc.* **20**, 13 (1859).
4. R. Hodgson, *Mon. Not. R. Astron. Soc.* **20**, 15 (1859).
5. E. R. Priest, T. E. Forbes, *Astron. Astrophys. Rev.* **10**, 313 (2002).
6. C. J. Schrijver *et al.*, *Astrophys. J.* **675**, 1637 (2008).
7. T. Török, B. Kliem, *Astrophys. J.* **630**, L97 (2005).
8. T. Amari, A. Canou, J.-J. Aly, *Nature* **514**, 465 (2014).
9. Y. Su *et al.*, *Nat. Phys.* **9**, 489 (2013).
10. C. J. Schrijver, *Astrophys. J.* **655**, L117 (2007).
11. K. D. Leka *et al.*, *Astrophys. J. Suppl. Ser.* **243**, 36 (2019).
12. N. Ishiguro, K. Kusano, *Astrophys. J.* **843**, 101 (2017).
13. P. H. Scherrer *et al.*, *Sol. Phys.* **275**, 207 (2012).
14. J. K. Thalmann, Y. Su, M. Temmer, A. M. Veronig, *Astrophys. J.* **801**, L23 (2015).
15. S. Yashiro, S. Akiyama, N. Gopalswamy, R. A. Howard, *Astrophys. J.* **650**, L143 (2005).

ACKNOWLEDGMENTS

A.M.V. acknowledges the Austrian Science Fund (FWF): P27292-N20 and I4555-N.

10.1126/science.abb6150

Published by AAAS

NEUROSCIENCE

Evolution of the human brain

A human-specific gene is a determinant of the cognitive architecture of the human cerebral cortex

By **Colette Dehay**[1] *and* **Henry Kennedy**[1,2]

Since early hominids emerged 5 million years ago, humans have evolved sizable brains to support higher cognitive functions. In particular, the human cerebral cortex is greatly expanded, allowing accommodation of the evolutionary increases in the number of cortical areas, the functional modules that subserve perception, attention, motor control, cognition, memory, and learning. Duplicated genes specific to the *Homo* lineage have played key roles in human speciation, particularly in the development of the highly complex human brain (*1*) and the circuits of the cerebral cortex (*2*). On page 546 of this issue, Heide *et al.* (*3*) identify *ARHGAP11B* [Rho guanosine triphosphatase (GTPase) activating protein 11B], a human-specific duplicated gene, as a regulator of human cerebral cortex development. By expressing *ARHGAP11B* in marmosets, a smooth-brained primate, this study explores the influence of the gene on expansion of the primate cortex.

The human neocortex is marked by an important increase in surface area and its radial dimension, the latter due to the selective enlargement of the supragranular layers (*4*). Supragranular neurons have an important role in the integration of ascending and descending cortico-cortical pathways that underlie information transfer and processing between the numerous hierarchically organized cortical areas in primates. Therefore, the specific expansion of supragranular neurons contributes to the cognitive functions of primates, culminating in humans (*4*). Much of the origin of this expansion can be attributed to primate-specific features of corticogenesis, including an expanded progenitor pool in the developing primate cerebral cortex: the outer subventricular zone (OSVZ) (*5*), which includes specialized progenitors called basal radial glial cells (bRGs) (*6*). bRGs are endowed with extensive proliferative capacities and generate mostly supragranular neurons (*7*).

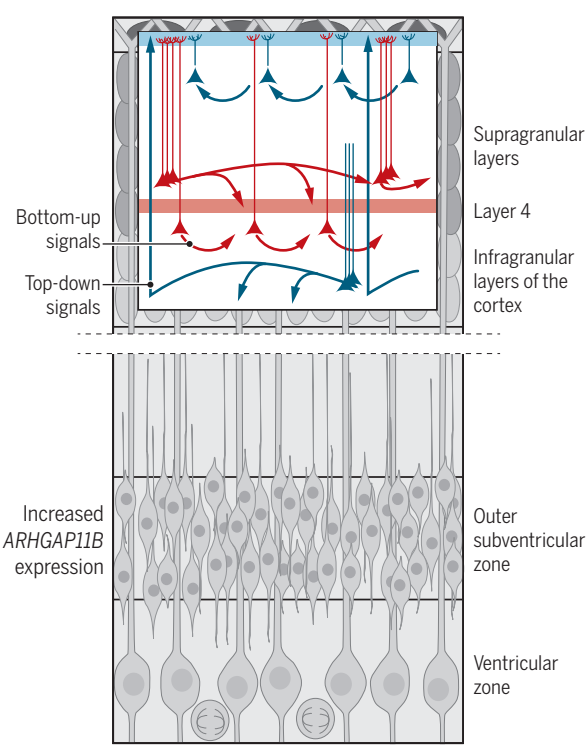
ARHGAP11B has received much attention because it is specifically enriched in cortical bRGs (*8*). When locally overexpressed in mouse or ferret cortex, *ARHGAP11B* boosts bRG proliferation and increases the numbers of cortical neurons (*8*, *9*). These observations suggest that this gene could link specific aspects of primate corticogenesis and characteristic features of the adult primate cortex architecture. To test this, Heide *et al.* expressed *ARHGAP11B* in the developing cortex of the embryonic marmoset. When *ARHGAP11B* is expressed under the control of the human promoter and upstream regulatory sequences, the transgenic mid-gestation marmoset exhibits an enlarged developing cortex with signs of folding. The crucial observation is that there is a selective increase in the numbers of neurons in the supragranular layers. This "humanization" of the marmoset fetal cortex demonstrates that expression of *ARHGAP11B* in bRGs in a primate substrate has the capacity to contribute to neocortical expansion and supragranular complexification during human evolution. *ARHGAP11B*-induced expansion of the cortical progenitor pool is mediated by metabolic changes in mitochondria, particularly increased glutaminolysis, a characteristic of highly mitotically active cells (*10*). This illustrates how cell metabolism—one of the most ancient of biological networks—participates in shaping the human lineage.

How does the increased rate of supragranular neuron production, resulting from *ARHGAP11B* expression in OSVZ progenitors, affect the functional architecture of the cortex? And do these effects provide evolutionary insights? In the cortical hierarchy, areas are linked by a dense network of ascending (or bottom-up) and descending (or top-down) pathways forming a highly distributed hierarchy (*11*). Current theories of hierarchical processing of information in the cortex, including predictive coding theory, postulate that top-down messages signaling expectations interact in the supragranular layers with bottom-up activity from the sensory periphery, thereby enabling the brain to actively infer the causes of sensory stimulus (*12*). Recent structural analysis reveals that in the supragranular layers, top-down and bottom-up connections form two opposing streams, thereby constituting a counterstream architecture (*11*) (see the figure).

During evolution, there is a marked increase in the numbers of cortical areas, so that larger numbers of human supragranular

Shaping the human cortex

Heide *et al.* show that *ARHGAP11B* [Rho guanosine triphosphatase (GTPase) activating protein 11B] boosts proliferation in the outer subventricular zone, leading to increased production of cells destined for the supragranular layers. The counterstream architecture of the supragranular layers comprises feedback projections carrying top-down signals (blue arrows) that interact with feedforward projections (red arrows) carrying bottom-up sensory signals. The integration of these two pathways into the local microcircuit is a key feature of hierarchical processing in the primate cortex and will be favored by increased numbers of supragranular neurons.



GRAPHIC: MELISSA THOMAS BAUM/*SCIENCE*

[1]University of Lyon, Université Claude Bernard Lyon 1, Inserm, Stem Cell and Brain Research Institute U1208, 69500 Bron, France. [2]Institute of Neuroscience, State Key Laboratory of Neuroscience, Chinese Academy of Sciences Key Laboratory of Primate Neurobiology, Shanghai 200031, China. Email: colette.dehay@inserm.fr; henry.kennedy@inserm.fr

Published by AAAS

neurons are required to integrate cortico-cortical circuits compared to non-human primates. In addition, because of the specific coding properties of the supragranular layers, increases in the number of supragranular neurons are expected to increase the circuit efficiency of these layers (*13*), which, along with their complexification (*14*), could drive gains in computational power and the capacity to integrate top-down and bottom-up signals. The architecture of the primate brain has therefore evolved for the computational mechanisms that affect human perception and sense of self; this also has implications for the evolution of memory and learning.

The findings of Heide *et al*. illuminate how a molecular mechanism driving cortical development can scale up phylogenetically ancestral primate brains to the complexity of the human brain. ARHGAP11B-mediated humanization of the marmoset fetal cortex demonstrates the involvement of a human-specific duplicated gene in the expansion of the supragranular layers. The effect of *ARHGAP11B* expression on OSVZ progenitors and their cortical progeny reinforces the importance of recent findings showing that human-specific regulatory elements are enriched in the OSVZ and the adult supragranular layers (*15*). Together with the role of *SRGAP2* (Slit-Robo-GTPase activating protein 2), a human-specific duplicated gene that acts on cortical neuron complexity and synaptic circuitry (*2*), these findings point to crucial evolutionary adaptations converging on the cardinal structural features of the human cortex that underlie its unrivaled computational and cognitive performance. Future studies will need to address the effect of *ARHGAP11B* expression at different time points in corticogenesis, its potential role in determining human specific cell types in the brain, and its intersection with the etiology of neurological disorders (*4*, *14*, *15*). ■

REFERENCES AND NOTES

1. M. Y. Dennis, E. E. Eichler, *Curr. Opin. Genet. Dev.* **41**, 44 (2016).
2. C. Charrier *et al.*, *Cell* **149**, 923 (2012).
3. M. Heide *et al.*, *Science* **369**, 546 (2020).
4. A. M. M. Sousa, K. A. Meyer, G. Santpere, F. O. Gulden, N. Sestan, *Cell* **170**, 226 (2017).
5. I. H. Smart, C. Dehay, P. Giroud, M. Berland, H. Kennedy, *Cereb. Cortex* **12**, 37 (2002).
6. D. V. Hansen, J. H. Lui, P. R. Parker, A. R. Kriegstein, *Nature* **464**, 554 (2010).
7. M. Betizeau *et al.*, *Neuron* **80**, 442 (2013).
8. M. Florio *et al.*, *Science* **347**, 1465 (2015).
9. N. Kalebic *et al.*, *eLife* **7**, e41241 (2018).
10. T. Namba *et al.*, *Neuron* **105**, 867 (2020).
11. J. Vezoli *et al.*, bioRxiv 2020.04.08.032706 (2020).
12. R. P. Rao, D. H. Ballard, *Nat. Neurosci.* **2**, 79 (1999).
13. K. D. Harris, T. D. Mrsic-Flogel, *Nature* **503**, 51 (2013).
14. J. Berg *et al.*, bioRxiv 2020.03.31.018820 (2020).
15. H. Won, J. Huang, C. K. Opland, C. L. Hartl, D. H. Geschwind, *Nat. Commun.* **10**, 2396 (2019).

10.1126/science.abd1840

STRUCTURAL BIOLOGY

Fine-tuning receptor–G protein activation and signaling

The activation rate of downstream G proteins imprints receptor signaling

By **Guillaume Lebon**

G protein–coupled receptors (GPCRs) are eukaryotic plasma membrane receptors that are organized into four classes in humans: A, B, C, and Frizzled. They internalize extracellular stimuli by activating a common pool of intracellular signaling partners such as the heterotrimeric G proteins (composed of Gα, β, and γ subunits) that subsequently induce an appropriate cellular response. Recent advances in cryo–electron microscopy (cryo-EM) enables challenging structures of GPCR signaling complexes to be solved, providing unprecedented insights about the molecular basis of their signal transduction (*1*). Qiao *et al*. (*2*) reported two cryo-EM structures of the class B human glucagon receptor (GCGR) G_s and G_i complexes, which helped clarify GCGR G protein selectivity. On page 523 of this issue, Hilger *et al*. (*3*) report a cryo-EM structure of a GCGR-G_s complex and reveal the effect of conformational changes on GCGR signaling properties. These studies support a common mechanism for class B receptor activation.

Activation of G proteins by GPCRs trigger nucleotide exchange and hydrolysis occurring in sequential transition of conformational states from inactive guanosine disphosphate (GDP)-bound forms, to the intermediate nucleotide-free state, then to the active guanosine triphosphate (GTP)–bound G proteins that activate downstream signaling. Accordingly, GPCRs function as guanine nucleotide exchange factors (GEFs). GCGR physiology and signaling are challenging our knowledge of GPCR activation mechanisms. For example, circulating glucagon generated by pancreatic α-cells activates the GCGR and controls glucose homeostasis in the liver (*4*). Once glucagon activates the GCGR-G_s signaling complex in hepatocytes, the G_s protein interacts with adenylate cyclase, which in turn induces production of the second messenger cyclic adenosine monophosphate (cAMP). GCGR has a specific functional signature when considering the amplitude and time scale of glucagon-induced cAMP production compared with other hormones such as epinephrine, which activates the class A GPCR β2-adrenergic receptor (β2AR). It is also common that a GPCR interacts with several G proteins—such as G_s, G_i, and G_q—and GCGR is no exception: It can also activate G_q and G_i proteins (*5*).

But how do GPCRs achieve the functional selectivity to activate specific G proteins? The two cryo-EM structures of GCGR-G_s and GCGR-G_i complexes solved by Qiao *et al*. reveal the large movement of the intracellular tip of transmembrane helix 6 (TM6), which is tilted outward, to accommodate the α5 helix from both G_s and G_i. TM6 movement is a molecular signature of class A and B GPCR active conformations. GCGR TM5 and TM7 undergo a similar motion, but to a lesser extent. This creates a large cavity for the G protein to bind. Conformational changes leading to GCGR activation occur in a different manner than for class A receptors. The conserved residues of the PXXG motif in TM6 are repositioned, and TM6 locally unwinds, which results in a sharp kink that tilts the straight intracellular tip of TM6 away from the receptor core (*2*, *3*). Such a kink was reported for several class B GPCRs (*6*), and the superposition of the corresponding receptor signaling complexes clearly illustrates this conserved feature compared with class A receptors, for which there is no kink; instead, TM6 simply bends over, with some degree of variability, depending on the receptor (*1*).

The G protein–binding site of type B GPCRs compares well with class A GPCRs coupled to G_s, although it is larger and accommodates a carboxyl-terminal (C-terminal) extremity of the α5 helix from both G_s and G_i (see the figure). There are substantial differences in the sequences of G_s and G_i α5 helices, which is indicative of the requirement of a larger binding site for G_s. The overall contact surface resulting from G_s and G_i subunit engagement with GCGR also differs. This molecular interface is mainly mediated by the α5 helix C-terminal extremity and is larger for G_s, 802 Å^2, compared with 551 Å^2 for G_i.

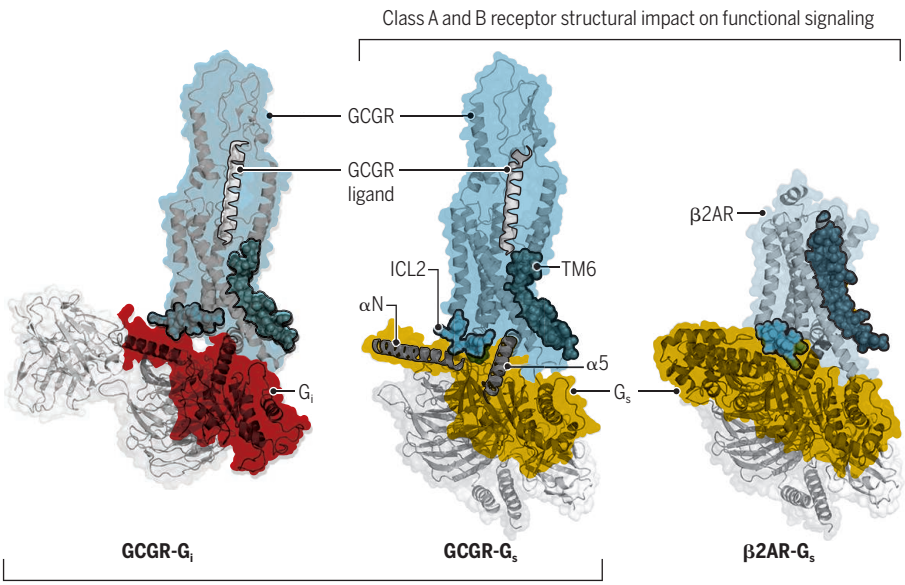
Another difference is the contribution of the receptor intracellular loop 2 (ICL2) in

Institut de Génomique Fonctionnelle (IGF), University of Montpellier, CNRS, INSERM, Montpellier, France. Email: guillaume.lebon@igf.cnrs.fr

Published by AAAS

Structural features of G protein activation and selectivity

Structural analysis of the class B glucagon receptor (GCGR) reveals the contribution of α5, αN, and transmembrane helix 6 (TM6) helices and intracellular loop 2 (ICL2) conformation to the structural mechanism of G protein selectivity. The activated conformation of GCGR-G_s leads to sustained G protein activation compared with type A β2-adrenergic receptor (β2AR)–G_s signaling. The sharp kink in TM6 of active GCGR illustrates a common mechanism of G protein activation for class B receptors.



the GCGR–G protein complexes. The positioning of the G_i αN helix relative to the receptor stresses the ICL2 conformation, making its contribution to the interaction less important for the GCGR-G_i complex, whereas ICL2 makes extended molecular interactions with the αN helix, the β1 strand, and the α5 helix of G_s. The resulting difference in total contact surfaces established between GCGR with G_s and G_i provides an explanation for the lower efficiency of G_i coupling but also for the G_s selectivity over G_i. Disruption of these interfaces by means of site-directed mutagenesis further highlights the importance of the shape and the size of the G protein binding cavity, also stressing the important contribution of ICL2 for G protein selectivity.

Once stimulated by glucagon, the GCGR-G_s complex induces a sustained production of cAMP over time compared with β2AR-G_s, suggesting a difference in the kinetics of cAMP accumulation and G protein activation (*7*). Although the structure of the receptor and receptor signaling complexes provide a snapshot of the receptor in action, they do not indicate how the receptor conformation influences G protein activation kinetics. Hilger *et al*. compare cryo-EM structures of GCGR-G_s to previous β2AR-G_s structures and present an elegant analysis that connects structural differences with differential rates of G protein activation. The cryo-EM structure of GCGR bound to the synthetic glucagon derivative ZP3780 reveals the molecular intricacies involved in the agonist-induced structural changes of TM6, which lead to its intracellular outward shift and the opening of the cavity that allows the G protein to bind. However, ZP3780 is not sufficient to induce the same outward shift of TM6 as observed for β2AR and class A GPCR agonists. Instead, ZP3780 triggers discrete conformational changes in the environment of TM6, stabilizing a GCGR-GDP–bound G_s. This observation correlates with the weak interaction of G_s with GCGR ICL2 compared with the β2AR-G_s complex, which likely slows down GDP release and prevents the formation of the fully activated state and nucleotide-free G protein conformation.

In addition, in vitro biophysical characterization of purified receptors shows a large difference in GCGR GEF activity, which catalyzes nucleotide exchange 70 times slower than β2AR. Hilger *et al*. also report slower recruitment of G protein and that GTP binding is three times slower by GCGR, which confirms the reduced rate of G protein activation and nucleotide hydrolysis.

It is surprising that the conformation of the TM6 tilt, which is characteristic of the GCGR active conformation, maintains the G protein pocket in an open conformation, even though the G protein has dissociated. This differs from the fast relaxation of the bent TM6 associated with class A GPCRs, which allows them to return to a more energetically stable conformation after G protein activation. Hilger *et al*. propose that a higher energy barrier separates the GCGR inactive conformation from the intermediate and fully activated states, which potentially accounts for the slower kinetic rate of activation and also for the sustained production of cAMP. In addition to conserved structural features of the activated state, three other class B GPCRs display a slower G protein dissociation rate as compared with that of class A GPCRs, which is indicative of a conserved activation mechanism for class B GPCRs.

Our understanding of the complexity and diversity of GPCR activation mechanisms, their coupling selectivity, and all the parameters that may govern their signaling in time and space is only just beginning. GPCR conformational diversity is likely to translate into subtle tuning of intracellular signaling partners such as G proteins, reinforcing the multiplicity of possible biological outputs despite the limited numbers of intracellular partners. In this respect, the composition of the heterotrimeric G protein, namely the β and γ subunits, and other additional regulatory proteins or signaling partners may also affect the G protein activation rate and will require further investigation.

More structural information is also needed from different receptor classes to achieve a deeper understanding of their signaling activation mechanism. GPCRs are challenging allosteric machines and are incredibly important drug targets; more than 30% of approved drugs target GPCRs. Class B GPCRs are validated drug targets, and GCGR as well as glucagon–like peptide 1 receptor are important drug targets for type 2 diabetes (*8*). In the future, it will be necessary to combine and integrate structural information with the dynamic properties of receptor conformation and kinetic parameters of GPCR signal transduction when engaging in drug design and the development of new pharmaceuticals. ■

REFERENCES AND NOTES

1. J. García-Nafría, C. G. Tate, *Mol. Cell. Endocrinol.* **488**, 1 (2019).
2. A. Qiao *et al.*, *Science* **367**, 1346 (2020).
3. D. Hilger *et al.*, *Science* **369**, eaba3373 (2020).
4. K. E. Mayo *et al.*, *Pharmacol. Rev.* **55**, 167 (2003).
5. D. Wootten, L. J. Miller, C. Koole, A. Christopoulos, P. M. Sexton, *Chem. Rev.* **117**, 111 (2017).
6. Y.-L. Liang *et al.*, *Mol. Cell* **77**, 656 (2020).
7. J. H. Exton, G. A. Robison, E. W. Sutherland, C. R. Park, *J. Biol. Chem.* **246**, 6166 (1971).
8. Y. M. Cho, C. E. Merchant, T. J. Kieffer, *Pharmacol. Ther.* **135**, 247 (2012).

ACKNOWLEDGMENTS

This work was supported by the University of Montpellier, Centre National de la Recherche Scientifique, and Institut National de la Santé et de la Recherche Médicale.

10.1126/science.abc9291

GRAPHIC: ADAPTED BY A. KITTERMAN/*SCIENCE* FROM G. LEBON

Published by AAAS

MATERIALS SCIENCE

Ductile van der Waals materials

Normally brittle, single crystals of InSe can be mechanically rolled like a metal into thin films

By **Xiaodong Han**

Van der Waals (vdW) materials are made up of strongly bonded two-dimensional (2D) layers that are bound in the third dimension through weaker dispersion forces. Graphite is a vdW material broadly used in industry in electrodes, lubricants, fibers, heat exchangers, and batteries. Its single-layer form, graphene (*1*), ushered in the new era of 2D materials that have tunable thickness and bandgap, as well as electronic confinement effects (*2*). Although these thin materials are inherently flexible (*3*), most vdW materials other than silver sulfide (Ag_2S) (*4*) are intrinsically brittle in their bulk form, which can limit how they are processed, as well as their applications. On page 542 of this issue, Wei *et al.* (*5*) report that indium selenide (InSe), in its single-crystalline form, exhibits exceptional plasticity under compressive stress.

The strong bonding within the 2D plane, and the weak vdW force between the layers, make plastic deformation of a vdW material intrinsically difficult because there is no favorable route to accommodate the atomic motion within and across interatomic layers under tensile or compressive stress. Traditional mechanical processing such as forging of these materials is normally impossible, and indeed, 2D thin flakes can be exfoliated from 3D single crystals (*1, 2, 6*). The exfoliate routes are efficient and quick, but the resulting flakes are difficult to precisely reproduce, easily broken, and small in size.

A thermal-mechanical processing route, similar to rolling thin metal sheets, has been very appealing for large-scale production of 2D materials but requires more ductile bulk materials that undergo plastic deformation. Thus, the discovery by Wei *et al.* of exceptional plasticity in InSe alloys is noteworthy. The room-temperature compressive strain can exceed ~80%. A flake containing ~10^5 layers of InSe was still highly bendable and exhibited flexibility previously achieved only in single-layer graphene (*7*). These properties of InSe single crystals present a distinctive malleability among vdW materials even when compared to Ag_2S (*4*).

Bulk InSe has a direct bandgap of about 1.26 eV at room temperature (*8*) with a hexagonal honeycomb crystalline structure. Each layer forms through Se-In-In-Se intraplane and interplane covalent bonds. These layers are held together through Se-Se vdW interactions with a spacing of 0.8 nm (*8*). Unlike the sp^2 character of graphite, which has a weak vdW interlayer link that makes graphite brittle at room temperature, the traversing-layer dislocation gliding seen in InSe is indicative of stronger vdW interlayer bonding than in regular vdW materials. This stronger interaction could be the result of long-range Coulomb interaction that makes the vdW cleavage plane of InSe much stronger and enables cross-layer sliding. The cross-layer dislocation gliding adds a plastic deformation mode absent in most vdW materials.

Large-scale fabrication of 2D materials through thermal-mechanical rolling is attractive, as it can greatly reduce costs. Large-area or long belts of 2D materials made through this rolling process would have orders-of-magnitude greater area than the original plate (see the figure). The vdW flakes formed through thermal-mechanical processing with controlled thickness could be used for a variety of applications, including flexible electronics, data storage, energy storage, and light sensing.

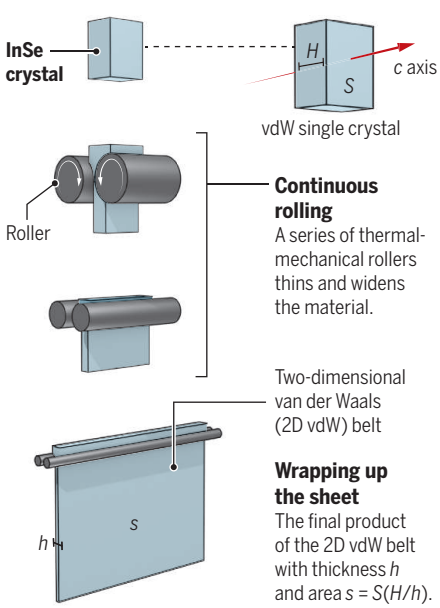
Single-crystalline InSe rolled into nanometer-thick flakes could be used to fabricate free-standing InSe-based 2D heterostructures with other low-dimensional materials (M). Mechanical combination with other 2D crystals could create vertical stacks, but InSe nanosheets could also be integrated with nanowires, quantum dots, or even single metallic atoms (*9*) or metal oxides (*10*). Such free-standing InSe-M heterostructures would allow a greater number of combinations than would traditional exfoliation and evaporation techniques. The bandgap of thin-layer InSe can be tuned to be direct or indirect by changing the material thickness or applying strain. The 2D InSe-M heterostructures could be combined with 2D materials that generate polaritons (*11*), have magneto-optical properties (*12*), or exhibit thermoelectric properties (*13*).

Many bulk single crystals of vdW materials have been successfully grown and fabricated in recent years with outstanding physical properties, such as polarized thermoelectric properties (*13*, *14*). Expanding research in this field and formulating general rules for designing ductile vdW materials are highly desired. By bridging the traditional wisdom of metallurgical technologies and vdW semiconductor materials, numerous new opportunities may emerge for 2D field and flexible electronics. ■

Van der Waals materials on a roll

Wei *et al.* show that indium selenide (InSe) single crystals are ductile and can be rolled into continuous thin sheets.

Orientation
An InSe single crystal, oriented with its layers parallel to the rolling direction (perpendicular to crystal *c* axis), has a thickness *H* and surface area *S*.



GRAPHIC: C. BICKEL/*SCIENCE*

Beijing University of Technology, Institute of Microstructure and Property of Advanced Materials, Beijing 100124, China. Email: xdhan@bjut.edu.cn

REFERENCES AND NOTES

1. K. S. Novoselov *et al.*, *Science* **306**, 666 (2004).
2. K. S. Novoselov *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10451 (2005).
3. J. Peng, G. J. Snyder, *Science* **366**, 690 (2019).
4. X. Shi *et al.*, *Nat. Mater.* **17**, 421 (2018).
5. T.-R. Wei *et al.*, *Science* **369**, 542 (2020).
6. S. Z. Butler *et al.*, *ACS Nano* **7**, 2898 (2013).
7. H. Chen *et al.*, *Science* **365**, 1036 (2019).
8. J. Camassel, P. Merle, H. Mathieu, A. Chevy, *Phys. Rev. B Condens. Matter* **17**, 4718 (1978).
9. B. Qiao *et al.*, *Nat. Chem.* **3**, 634 (2011).
10. C. Wang *et al.*, *Adv. Mater.* **31**, 1903491 (2019).
11. D. N. Basov, M. M. Fogler, F. J. García de Abajo, *Science* **354**, aag1992 (2016).
12. D. A. Bandurin *et al.*, *Nat. Nanotechnol.* **12**, 223 (2017).
13. K. L. Peng *et al.*, *Energy Environ. Sci.* **9**, 454 (2016).
14. W. He *et al.*, *Science* **365**, 1418 (2019).

ACKNOWLEDGMENTS

X.H. thanks the National Science Foundation of China, program nos. 51988101 and 91860202, for support.

10.1126/science.abd4527

Published by AAAS

VIEWPOINT: COVID-19

How does SARS-CoV-2 cause COVID-19?

The viral receptor on human cells plays a critical role in disease progression

By **Nicholas J. Matheson**[1,2] *and* **Paul J. Lehner**[1]

Viruses enter cells and initiate infection by binding to their cognate cell surface receptors. The expression and distribution of viral entry receptors therefore regulates their tropism, determining the tissues that are infected and thus disease pathogenesis. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the third human coronavirus known to co-opt the peptidase angiotensin-converting enzyme 2 (ACE2) for cell entry (*1*). The interaction between SARS-CoV-2 and ACE2 is critical to determining both tissue tropism and progression from early SARS-CoV-2 infection to severe coronavirus disease 2019 (COVID-19). Understanding the cellular basis of SARS-CoV-2 infection could reveal treatments that prevent the development of severe disease, and thus reduce mortality.

As with all coronaviruses, SARS-CoV-2 cell entry is dependent on its 180-kDa spike (S) protein, which mediates two essential events: binding to ACE2 by the amino-terminal region, and fusion of viral and cellular membranes through the carboxyl-terminal region (*2*). Infection of lung cells requires host proteolytic activation of spike at a polybasic furin cleavage site (*3*). To date, this cleavage site is found in all spike proteins from clinical SARS-CoV-2 isolates, as well as some other highly pathogenic viruses (e.g., avian influenza A), but it is absent from SARS-CoV and is likely to have been acquired by recombination between coronaviruses in bats. Cleavage by the furin protease therefore expands SARS-CoV-2 cell tropism and may have facilitated transmission from bats to humans (*1*). Membrane fusion also requires cleavage by additional proteases, particularly transmembrane protease serine 2 (TMPRSS2), a host cell surface protease that cleaves spike shortly after binding ACE2 (*3*). SARS-CoV-2 tropism is therefore dependent on expression of cellular proteases, as well as ACE2.

Other proteins that enable SARS-CoV-2 cell entry are also emerging, including neuropilin 1 (NRP1), a receptor that binds the carboxyl-terminal RXXR motif in spike that is exposed after furin cleavage. How NRP1 promotes cell entry is unclear, but it may further increase the cell types infected (*4*).

Of the seven known human coronaviruses, three are highly pathogenic [SARS-CoV, SARS-CoV-2, and Middle East respiratory syndrome (MERS)-CoV], and the remaining four (HCoV-NL63, HCoV-229E. HCoV-OC43, and HCOV-HKU1) are less virulent, causing "common colds." SARS-CoV, SARS-CoV-2, and HCoV-NL63 use ACE2 as their cell entry receptor. MERS-CoV binds DPP4 (dipeptidyl peptidase 4) and HCoV-229E uses CD13 (aminopeptidase N) (*2*). No host protein receptors have been identified for the other two viruses. It seems a notable coincidence that all known human coronavirus receptors are cell surface peptidases, particularly because the interactions do not involve the endopeptidase active site (*2*). The presence of a specific region within ACE2, targeted by three coronaviruses, is particularly noteworthy (*1*). Conversely, the receptor-binding domain of spike is encoded by the most variable part of the coronavirus genome (*1*). This implies that diversification of these viruses generated different sequences that converged on the same region of the same protein using alternative structural solutions. What, then, is so special about ACE2?

ACE2 is a transmembrane protein best characterized for its homeostatic role in counterbalancing the effects of ACE on the cardiovascular system (*5*). ACE converts angiotensin I to angiotensin II, a highly active octapeptide that exhibits both vasopressor (vascular contraction to increase blood pressure) and proinflammatory activities. The carboxypeptidase activity of ACE2 converts angiotensin II to the heptapeptide angiotensin-(1-7), a functional antagonist of angiotensin II. Because ACE is highly expressed in vascular endothelial cells of the lungs, angiotensin II is also likely to be high in the pulmonary vasculature. Indeed, *Ace2* deletion in mouse models of acute lung injury results in more severe disease, suggesting a protective role for ACE2 in lung tissue (*5*). In many host-virus interactions, expression of the viral receptor is down-regulated in infected cells, and expression of ACE2 in the lungs of mice was reduced by SARS-CoV infection. Depletion of ACE2 may thus play a causative role in the lung injury caused by SARS-CoV and SARS-CoV-2, and high plasma angiotensin II is reported in patients with COVID-19. However, MERS-CoV causes a similar lung disease without targeting ACE2, so other factors must also be important.

As a respiratory virus, SARS-CoV-2 must initially enter cells lining the respiratory tract. Single-cell sequencing and RNA in situ mapping of the human respiratory tract show ACE2 and TMPRSS2 expression to be highest in ciliated nasal epithelial cells, with lesser amounts in ciliated bronchial epithelial cells and type II alveolar epithelial cells (*6*). This translates to greater permissivity of upper versus lower respiratory tract epithelial cells for SARS-CoV-2 infection in vitro and fits disease pathology: Upper respiratory tract symptoms are common early in disease, with nasopharyngeal and throat swabs positive for SARS-CoV-2 at clinical presentation (*7*). In contrast to SARS-CoV, infectivity of patients with SARS-CoV-2 peaks before symptom onset (*8*). Indeed, presymptomatic transmission makes SARS-CoV-2 impossible to contain through case isolation alone and is a key driver of the pandemic (*8*). This alteration in the pattern of disease may relate to the acquisition of the furin cleavage site in spike or increased binding affinity for ACE2 in SARS-CoV-2, compared with SARS-CoV (*9*).

If the main role of ACE2 is to cleave angiotensin II, it is unclear why expression in lung tissue is more prominent in epithelial than in endothelial cells. Furthermore, the Human Cell Atlas highlights ACE2 expression in intestinal enterocytes, rather than in the lungs. This distribution may reflect non-enzymatic roles of ACE2, such as chaperoning amino acid transporters. Indeed, SARS-CoV-2 infection of the gastrointestinal (GI) tract is common, with viral RNA detectable in stool in up to 30% of COVID-19 patients. This likely contributes to the frequency of GI symptoms. Conversely, whereas fecal-oral transmission of coronaviruses is thought to be prominent among bats, it appears to be a minor transmission route for SARS-CoV-2 in humans, perhaps because colonic fluid inactivates the virus. Whether extrapulmonary ACE2 expression and concomitant viral infection account for other clinical manifestations of SARS-CoV-2 is unclear. The association between SARS-CoV-2 infection and anosmia (loss of smell) may reflect ACE2 and TMPRSS2 expression in sustentacular cells, which maintain the integrity of olfactory sensory neurons. Olfactory epithelial cells also express NRP1 and could provide a direct route to the brain (*4*).

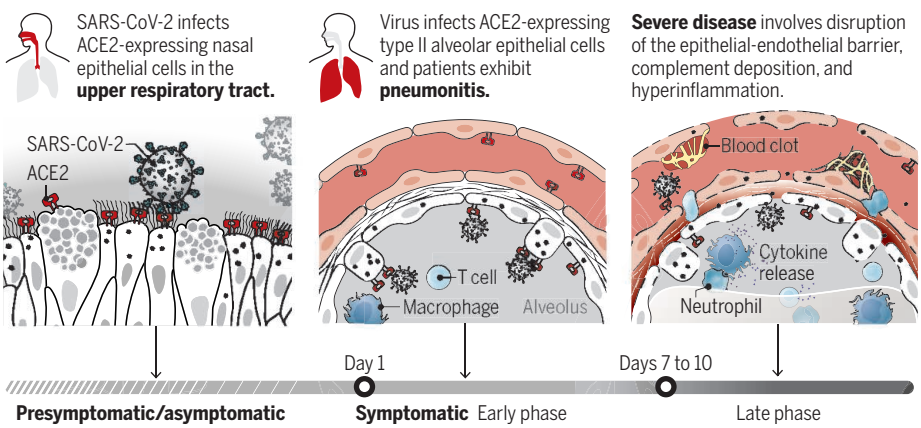
[1]Department of Medicine, Cambridge Institute for Therapeutic Immunology and Infectious Disease (CITIID), University of Cambridge, Cambridge, UK. [2]NHS Blood and Transplant, Cambridge, UK. Email: pjl30@cam.ac.uk; njm25@cam.ac.uk

Published by AAAS

Although a substantial proportion of SARS-CoV-2–infected individuals report few, if any, symptoms and recover completely, chest computed tomography (CT) evidence of viral pneumonitis is present in >90% of symptomatic cases within 3 to 5 days of onset (*10*). This presumably reflects viral replication in the lower respiratory tract, with infection of type II pneumocytes and accompanying inflammation (see the figure). Lung pathology in this early phase is poorly reported because most patients recover. Histopathology from cynomolgus monkeys, 4 days after inoculation with SARS-CoV-2, shows a viral pneumonitis with alveolar edema, capillary leakage, inflammatory cell infiltration, interstitial thickening, and cell fusion (a feature of coronavirus infection), with viral spike expression on alveolar epithelial cells (*11*).

Key phases of disease progression

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) binds to angiotensin-converting enzyme 2 (ACE2). Initial infection of cells in the upper respiratory tract may be asymptomatic, but these patients can still transmit the virus. For those who develop symptoms, up to 90% will have pneumonitis, caused by infection of cells in the lower respiratory tract. Some of these patients will progress to severe disease, with disruption of the epithelial-endothelial barrier, and multi-organ involvement.



Around 80% of patients with COVID-19 pneumonitis recover without specific treatment (*12*). However, ~20% of patients deteriorate, often rapidly, ~7 to 10 days after symptom onset. This is when patients are most frequently admitted to hospital, with worsening fever, hypoxia, lymphopenia, rising inflammatory markers [C reactive protein (CRP), interleukin-1 (IL-1), and IL-6], coagulopathy, and cardiovascular involvement. About 25% of these patients will require mechanical ventilation, which is associated with high mortality (50 to 80%). The demographic of this "at risk" group is reproduced across many countries: older men with hypertension, diabetes, and obesity, as well as a less well defined contribution of ethnicity (*12*). Similar factors regulate ACE2 expression, which may therefore contribute to disease severity. Nonetheless, the amount and distribution of ACE2 expression cannot be the only factor affecting disease progression, because the three human coronaviruses that use ACE2 for cell entry exhibit markedly different pathogenicity.

What causes the sharp deterioration that leads to severe systemic COVID-19? The lung pathology in severe disease is different from the earlier pneumonitis, with progressive loss of epithelial-endothelial integrity, septal capillary injury, and a marked neutrophil infiltration, with complement deposition, intravascular viral antigen deposition, and localized intravascular coagulation (*13*). If the earlier viral pneumonitis reflects direct ACE2-mediated infection of type II pneumocytes, what drives this next, potentially deadly phase of acute lung injury, with the concomitant breakdown of the respiratory epithelial barrier, endothelial damage, and patient deterioration (*14*)? Most importantly—how can it be stopped?

The timing of deterioration suggests a role for the adaptive immune system, either antibodies or T cells, and has many hallmarks of immune-driven inflammation. Endothelial injury may result from immune-mediated damage, through complement activation, antibody-dependent enhancement, and/or cytokine release. Attention has therefore focused on the use of immunomodulatory therapies in patients with severe disease. Nonetheless, the most obvious culprit for severe disease is the virus itself, either alone or with immune pathology. Breakdown of the lung epithelial-endothelial barrier might trigger endothelial damage and viral dissemination, with more widespread infection (*14*). Studies documenting the time course of SARS-CoV-2 RNA shedding are mainly limited to mild disease (*7*) and typically show a progressive decline after a peak around symptom onset. However, viral load from lung swabs may correlate with disease severity (*15*), and patients with severe lung disease remain RNA-positive for longer. It is critical to determine how long active viral replication really persists in the lungs of patients with severe disease, and how frequently viral replication occurs at extrapulmonary sites where ACE2 (or other receptors) is expressed, such as vascular endothelium.

Although there are huge efforts to understand and treat severe COVID-19, it would be preferable to prevent the development and progression of clinical disease. How might this be achieved? Vaccine candidates are mainly aimed at eliciting neutralizing antibodies, to prevent the binding of spike to ACE2. The same rationale underpins the use of passive immunization, with convalescent plasma or monoclonal antibodies, or the administration of recombinant, soluble ACE2. Alternatively, antiviral drugs may be used to target essential viral enzymes such as the RNA-dependent RNA polymerase. Experience from other infections, such as influenza, emphasizes that treatment with antiviral agents is most effective when administered as early as possible in infection. Therefore, it is essential to identify individuals with early SARS-CoV-2 infection who are at high risk of progression to severe disease, and test antiviral therapies to prevent viral entry and replication. It should not be too difficult to identify these "at risk" patients who are in danger of progressing to severe disease through contact tracing and testing, even prior to symptom onset. Conversely, delaying candidate antiviral treatment until patients are hospitalized with severe lung injury may be too late, and combination with immune modulation is likely to be required. ■

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES AND NOTES

1. K. G. Andersen *et al.*, *Nat. Med.* **26**, 450 (2020).
2. F. Li, *J. Virol.* **89**, 1954 (2015).
3. M. Hoffmann *et al.*, *Mol. Cell* **78**, 779 (2020).
4. L. Cantuti-Castelvetri *et al.*, *bioRxiv* 2020.06.07.137802 (2020).
5. K. Kuba *et al.*, *J. Mol. Med.* **84**, 814 (2006).
6. Y. J. Hou *et al.*, *Cell* 10.1016/j.cell.2020.05.042 (2020).
7. R. Wölfel *et al.*, *Nature* **581**, 465 (2020).
8. X. He *et al.*, *Nat. Med.* **26**, 672 (2020).
9. D. Wrapp *et al.*, *Science* **367**, 1260 (2020).
10. A. Bernheim *et al.*, *Radiology* **295**, 200463 (2020).
11. B. Rockx *et al.*, *Science* **368**, 1012 (2020).
12. Z. Wu, J. M. McGoogan, *JAMA* **323**, 1239 (2020).
13. C. Magro *et al.*, *Transl. Res.* **220**, 1 (2020).
14. L. A. Teuwen *et al.*, *Nat. Rev. Immunol.* **20**, 389 (2020).
15. J. Chen *et al.*, *J. Infect.* 10.1016/j.jinf.2020.03.004 (2020).

ACKNOWLEDGMENTS

The authors are supported by the Medical Research Council (CSF MR/P008801/1 to N.J.M.), NHS Blood and Transplant (WPA15-02 to N.J.M.), and the Wellcome Trust (PRF 210688/Z/18/Z to P.J.L.).

10.1126/science.abc6156

Published by AAAS

RETROSPECTIVE

William C. Dement (1928–2020)

Father of sleep medicine

By **Rafael Pelayo** and **Philippe Mourrain**

William C. Dement, one of the founders of sleep science and medicine, died on 17 June. He was 91. Bill devoted his life to helping the public appreciate the importance of sleep health. His pioneering work, which included the discovery of rapid eye movement (REM) sleep, touched countless lives.

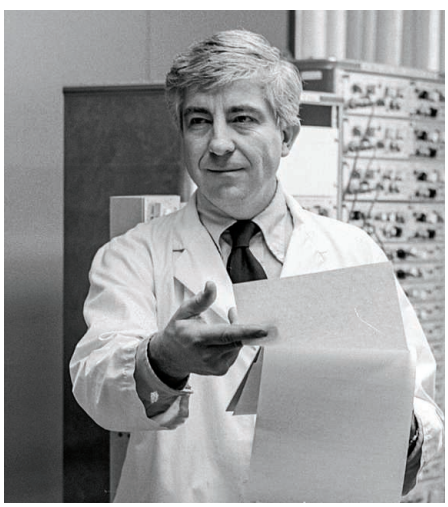
Born in Wenatchee, Washington, in 1928, Bill earned a bachelor's degree from the University of Washington in Seattle in basic medical science in 1951. He then started medical school at the University of Chicago, where he conducted groundbreaking work even before receiving his M.D. in 1955 and his Ph.D. in neurophysiology in 1957. He earned his medical license at Mount Sinai Hospital in New York City and then stayed on to continue his sleep research. In 1963, Bill joined the psychiatry department at Stanford University in Stanford, California, where he remained for the rest of his career.

Bill first encountered sleep research at the University of Chicago, where he joined the lab of physiologist and sleep researcher Nathaniel Kleitman. A fellow graduate student, Eugene Aserinsky, had begun to study eye movements in sleep in 1953. Bill helped discover and describe REM sleep, a term he later coined. From 1954 through 1957, he described the relationship between REM sleep and dreaming, established the all-night sleep patterns in human beings, discovered REM sleep in animals and newborn babies, and demonstrated that the patterns of specific rapid eye movements are related to the visual dream experience.

Upon completion of his internship at Mount Sinai Hospital, Bill obtained a research grant to establish a sleep laboratory, which he housed in his apartment. His ingenuity and persuasive powers were evident in the grant stipulations: Because the sleep research took place at night and he needed to be near his young family, the funding covered half of his apartment rent. Among the research subjects coming to his sleep research lab were members of the Radio City Rockettes. In 1960, he published a paper in *Science* based on that work, in which he described the effect of dream deprivation.

Sleep Medicine Division, Stanford University School of Medicine, Stanford, CA, USA. Email: pelayo@stanford.edu

In 1963, Bill developed a sleep research program at Stanford that drew international attention. He collaborated with many of the scientists who would move the new field forward, including neurophysiologists Michel Jouvet and Christian Guilleminault. In 1970, he opened the Stanford Sleep Disorders Clinic. Bill invented the term polysomnography to convince insurance companies to reimburse patients for the costs of clinical sleep studies. This coding change, from experimental treatment to sanctioned medical procedure, marked the beginning of the modern field of clinical sleep medicine.



Bill helped create the multiple sleep latency test, now the international standard for evaluating all forms of hypersomnia, and he established the use of clinical sleep studies in pharmacology trials. Sleep-onset REM periods, another of Bill's discoveries, became a pathophysiological marker of narcolepsy. His narcoleptic dog colony at Stanford eventually helped establish the cause of narcolepsy in humans. REM sleep phenomena have now been uncovered in dragon lizards, fishes, and even cephalopods, broadening the influence of Bill's work to more than half a billion years of animal evolution.

As the field's founder, Bill worked to provide sleep research with the necessary resources. He helped found the Sleep Research Society in 1961 and the Society for Neuroscience in 1969. He published *Sleep*, the first journal devoted to the field. The founding president of the American Sleep Disorders Association (now the American Academy of Sleep Medicine), he served in that role for 12 years. He created Stanford's incredibly popular university course Sleep and Dreams. Out of necessity, he wrote the first college textbook on the subject. Bill was such a pioneer that he had to sign his own certificate when he became board certified in sleep medicine.

An intellectually curious and bold man who took advantage of every opportunity, Bill anticipated the great potential for the sleep field to improve public health and worked tirelessly to promote it. He chaired the National Commission on Sleep Disorders Research, advocating awareness of sleep conditions, and spurred Congress to establish the National Center on Sleep Disorders Research at the National Institutes of Health. His work raising awareness about the dangers of drowsy driving and drowsiness in the workplace has undoubtedly saved innumerable lives.

I (R.P.) met Bill in 1993 when I applied for a fellowship at Stanford. When the famous Dr. Dement burst into the room and extended his hand, he immediately put me at ease. From that day on, Bill always made me smile. Later, we co-taught his Sleep and Dreams course, traveling together to class in his golf cart. Bill was just as welcoming to me (P.M.). When I met him in 2005, I was shocked that he asked me, a junior faculty member, to share my opinions about how to improve the definition of sleep. Kind and generous to colleagues and students alike, Bill had a positive energy and good humor that motivated everyone who worked with him.

While stationed in Japan as a journalist for the army after World War II, Bill performed professionally as a jazz musician. Surrounded by talented young musicians such as Ray Charles and Quincy Jones, Bill later would say that he realized he could "make a better living as a mediocre physician than as a mediocre musician." However, he never lost his love of performing. When asked, he would sing a cappella to the students in class. His musical performances were often the highlight of our professional conferences.

Bill served as a role model by always keeping an open mind and embracing scientific inquiry with the goal of helping others. A living legend, beloved by his friends and family, he enjoyed his celebrity status. He died secure in the knowledge that his colleagues and students would honor and carry on his work. Sleep researchers have lost a giant, but Bill's legacy will guide the field as it continues to grow. ■

10.1126/science.abd6855

PHOTO: CHUCK PAINTER/STANFORD NEWS SERVICE

Published by AAAS

Photographer Thilde Jensen, shown here in a self-portrait in a makeshift tent, documented the daily struggle faced by sensitives in *The Canaries* (2013).

PHOTO: THILDE JENSEN

ENVIRONMENTAL TOXICOLOGY

Confronting illness with empathy

A journalist paints a thoughtful portrait of a disease that defies scientific characterization

By **Frederick Rowe Davis**

What is it about a road trip that so effectively locates a good story in time and place? In his new book, *The Sensitives*, Oliver Broudy recounts the story of an unusual road trip on which he recently embarked that provides a frame for a discussion of a mysterious ailment that has thus far defied scientific characterization.

By way of introduction, Broudy notes that there are, at present, 85,000 synthetic chemicals circulating in our daily environments, including 9000 food additives, the vast majority of which have never been tested for potential effects on humans. In 2017 alone, 3.9 billion pounds of chemicals were dumped into the environment, he notes. Meanwhile, billions of metric tons of plastics (8.3 billion in 2015) accumulate around us. Cancer rates have also spiked, notes Broudy, and he cites drastic increases in other conditions, from autism to allergies.

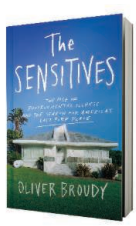
Enter the "sensitives," as they call themselves, who report an array of chronic symptoms, most commonly fatigue, memory impairment, and muscle aches, which they attribute to an increased vulnerability to chemicals and other environmental triggers. But despite the high incidence of environmental illness (EI)—Broudy indicates that as much as 30% of the population may exhibit some environmental hypersensitivity—many nonsufferers, particularly in the medical and scientific communities, wonder if it is even real. To date, no one has identified a definitive biomarker for the disease. The symptoms that sufferers report vary widely, and uncertainty prevails, especially in the published literature. This, in Broudy's view, makes EI the defining disease of our time.

The EI experience can come across as both specific to the individual and highly abstract. Yet Broudy refuses to dismiss the sensitives' accounts of their symptoms. Throughout his exploration, he remains skeptical yet empathetic.

To guide his journey toward understanding EI and those it afflicts, Broudy embarks on a road trip with a sensitive named James, who knows both the disease and the places where sensitives seek relief from their symptoms. Woven throughout the story of their travels, extended reflections reveal the state of past and present research on EI. We learn, for example, about Claudia Miller, a professor of environmental and occupational medicine at the University of Texas Health Science Center in San Antonio, who has developed a model to understand EI, which she calls toxicant-induced loss of tolerance (TILT). "Miller defined TILT as something like the inverse of addiction," Broudy writes. "[I]t commenced with an exposure, and the exposure led to sensitivity—the opposite of tolerance. Thereafter, even minimal exposures could provoke outsized effects." He notes parallels with the "rain barrel" theory, advanced by many sensitives, which holds that each individual has a personal toxicity threshold. Exceed the threshold, and illness follows.

How did we come to live in a world so saturated in synthetic chemicals? To answer this question, Broudy recounts the story of British chemist William Henry Perkin, who, in 1856, discovered how to synthesize a deep-purple dye. Other chemists followed suit, producing a colorful palette of synthetic dyes. "From this rainbow our modern world of industrial chemicals emerged," writes Broudy. Pharmaceuticals followed from the labs of Bayer and Monsanto, and in 1910 Fritz Haber successfully synthesized ammonia, a feat that won him the Nobel Prize in Chemistry and ushered in the era of agricultural chemistry.

As the road trip proceeds, Broudy encounters additional environmental risks: mining operations and power plants fired by coal, for example. He uses this to transition the narrative toward a history of risk assessment, which breezes from the birth of modern insurance to the dawn of nuclear armaments. Here, he draws striking comparisons between EI and other diseases with diffuse etiologies, such as Gulf War illness, posttraumatic stress disorder, and neurasthenia.



The Sensitives
Oliver Broudy
Simon and Schuster, 2020. 352 pp.

In Snowflake, Arizona, Broudy meets a community of sensitives, most of whom—like consumptives in the time of tuberculosis—were drawn to the warm, arid desert air. In representing the diversity of experience and symptomology, Broudy reveals EI to be a highly diffuse disease. His comparison of EI to tuberculosis before Robert Koch's discovery of the bacterium is apt.

Traveling with Broudy, we learn much about an inscrutable disease, the people who live with it, the places where they find relief, and the discursive path through history and environmental health that has brought us to new levels of risk and uncertainty. With a tone that is both thoughtful and humorous, *The Sensitives* provides a model for how history, philosophy, research, clinical practice, and—above all—patient experience inform meaningful discourse about a disease that resists characterization at all stages. ■

10.1126/science.abd0342

The reviewer is at the Department of History, Purdue University, West Lafayette, IN 47907, USA. Email: frdavis@purdue.edu

Published by AAAS

NEUROSCIENCE

Context is key to olfaction

How we experience smell has more to do with us than with the odor itself

By **John P. McGann**

Smell is an underappreciated part of life. It affects how our food tastes, how we are feeling, and to whom we are attracted. But smell is different from other senses because the connection between a smell and its physical source is fluid. The smell of your dinner persists even after it has been consumed, for example, but your eyes do not continue to see the food after the meal is finished. Similarly, although we have little trouble perceiving individual voices in a group conversation, a spice rack does not smell like a group of distinct herbs.

In her debut book, *Smellosophy: What the Nose Tells the Mind*, A. S. Barwich melds a philosophical perspective with a rich history of olfactory science, tackling big questions with layers of perceptual, psychological, and neurobiological explanations. She argues that olfaction does not reflect objective reality but rather a mental construct with modest relation to the chemical objects in the world. She synthesizes disparate lines of evidence in consideration of this argument, and along the way she embeds her philosophy in a larger narrative of how our current understanding of smell evolved from earlier beliefs and expectations.

Odors are experienced within a contextual tapestry of situation and experience. Lemon scent smells "fresh" to many of us and is often added to cleaning products, for example, but would be an unwelcome note in a perfume. Perfume, meanwhile, often contains fecal tones that combine with a wearer's natural scent to evoke an earthy intimacy. If such a scent were to be smelled in a bathroom, however, it might make one wish for a lemon-scented cleaning product.

Barwich uses such observations to argue that the affective quality of odors is thus found not in their chemical structures but inside the head and history of the perceiver. She offers rich discussions of olfactory perception, the conscious and subconscious impacts of smell on behavior and emotion, and the physical and behavioral details that determine what odors we inhale, furnishing broad insights into the psychology of olfaction.

Neuroscientists studying the olfactory brain have typically operated on the assumption that chemistry is destiny for an odor, looking for spatial and temporal patterns of neural activity that "represent" the molecular structure and concentration of odorous chemicals in the nose. Barwich argues that this is a naïve approach. The olfactory circuit, she writes, begins with a mapping of odor receptors in the nose onto the surface of the brain's olfactory bulb. Spatial patterns of neural activity across this structure capture the chemical properties of an odor. The cerebral cortex, however, seems to abandon this information, favoring instead a complex distribution of neural connections that process smells in a manner that has little relationship to odor chemistry.

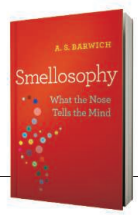
Smell is more mental construct than objective reality, argues Barwich.

Barwich suggests that contextual clues, including semantic content, cultural associations, and affective value, are more essential to capturing the true experience of smell. However, the challenges of testing this hypothesis are manifold. Modern research methods permit uniquely powerful experimentation in the brains of flies and mice, but if smells mean different things to lab animals than they do to human subjects—as Barwich suggests—it may be difficult to extrapolate research findings across species. By identifying the need for a more cognitive approach to olfactory neuroscience, the book sets up the practical challenge of linking human and animal research.

Barwich's educated prose balances visceral stimulation (in one passage, she contrasts the smell of an armpit on the subway with that of one's lover in bed) with academic language ("smells *qua* smells"). It is cheeky at times, alleging, for example, that perfumery is "one of the two oldest professions," and occasionally slips in snarky asides about Kant and his ilk. The author clearly delights in putting the reader in her shoes as she recounts stories of scientists arguing, joking, and teasing each other in conversations over a beer.

Smellosophy generally resists, and at times even disassembles, clichéd smell narratives, relying instead on close consideration of the scientific literature. In chapter 4, for example, Barwich undercuts the oft-cited "madeleine incident," in which a childhood memory is evoked by a tea-soaked cake in Marcel Proust's *In Search of Lost Time*, or *Remembrance of Things Past*. This example is frequently used to illustrate smell's special ability to evoke rich, emotionally laden memories. However, there is little scientific support for this claim, she reveals, noting that Proust never actually mentions the smell of the madeleine anyway. The book also sets a high bar when it comes to biological detail but deftly guides the lay reader through the minefield of terminology and competing scientific frameworks. ■

Smellosophy: What the Nose Tells the Mind
A. S. Barwich
Harvard University Press, 2020. 384 pp.



The reviewer is at the Department of Psychology and the Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ 08854, USA.
Email: john.mcgann@rutgers.edu

10.1126/science.abc6831

PHOTO: ALVAREZ/ISTOCK.COM

Published by AAAS

An Indian army convoy drives near the disputed Himalaya border between India and China.

PHOTO: MUKHTAR KHAN/ASSOCIATED PRESS

Edited by **Jennifer Sills**

China and India: Toward a sustainable world

In light of current political tensions at the China-India border (*1*), the two countries should consider their shared purpose and the mutual benefits that would result from cooperation. Both countries face environmental challenges at global, regional, and local levels, and both have the capabilities to be global leaders in providing environmental solutions. We urge the governments of China and India to facilitate science diplomacy, starting with the Himalaya region.

Both countries have made environmental progress a priority in recent years. China and India rank first and second in greening the Earth as measured by increases in leaf area index since 2000 (*2*), and both are highly ranked in solar and wind-electricity generation (*3*). China has implemented a series of programs to assess and conserve biodiversity and ecosystems (*4*), and India has launched a national mission on biodiversity and human well-being (*5*).

Despite these potential opportunities, new restrictions have been placed on scientific collaboration between China and India, and the exchange programs between their national academies have ground to a halt (*6*). Cultural and linguistic differences, as well as political tensions such as the unresolved border disputes, create barriers to collaboration even at international meetings where both countries are represented (*7*).

Now is the time to boost bilateral collaborations. China and India should put policies in place to facilitate collaborative environmental research that moves them closer to meeting their United Nations Sustainable Development Goals, as they did when they agreed to promote a circular economy at the Sixth China-India Strategic Economic Dialogue (*8*). Organizations from both sides of the Himalaya should come together and establish joint research centers on the environment of the region and climate change adaptation and mitigation, similar to the coordination between institutes in Kunming and Guwahati (*9*). Such initiatives need to be scaled up through renewed partnerships between the national academies for joint long-term research programs as well as through coordinated dialogue in international forums, where other parties that have good relations with both countries can help foster their cooperation.

The shared Himalaya, a region with extraordinarily rich biodiversity, ice fields, and water, is threatened by some of the highest hydroelectric-dam densities and climate change rates in the world (*10*, *11*). Long-term ecological security is more important than ongoing border disputes over desolate, high-altitude lands that may be best suited as peace parks (*7*) or nature reserves (*12*). Science diplomacy on environmental issues in the Himalaya could increase the possibility of sustained peace along the international border and allow the two superpowers to lead the world toward a sustainable future.

Kamaljit S. Bawa[1,2], Eben Goodale[3]*, Wambura Mtemi[3], You-Fang Chen[3], Ranjit Barthakur[4], Uromi Manage Goodale[3], Jianguo Liu[5], Aiwu Jiang[3], Christos Mammides[3], Madhava Meegaskumbura[3], Maharaj K. Pandit[6], Kun-Fang Cao[3]

[1]Department of Biology, University of Massachusetts, Boston, Boston, MA 02125, USA. [2]Ashoka Trust for Research in Ecology and the Environment, Bengaluru, Karnataka, India. [3]Guangxi Key Laboratory of Forest Ecology and Conservation, College of Forestry, Guangxi University, Nanning Guangxi, China. [4]Balipara Foundation, Guwahati, Assam, India. [5]Center for Systems Integration and Sustainability, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, USA. [6]Department of Environmental Studies, University of Delhi, Delhi, India.
*Corresponding author.
Email: ebengoodale@gxu.edu.cn

REFERENCES AND NOTES

1. R. Goldman, "Indian-China border dispute: A conflict explained," *The New York Times* (2020).
2. C. Chen *et al.*, *Nat. Sustain.* **2**, 122 (2019).
3. International Energy Agency, *Clean Energy Transitions Programme: Annual Report 2019* (2020).
4. J. Liu *et al.*, *Annu. Rev. Environ. Resour.* **43**, 1 (2018).
5. Office of the Principal Scientific Adviser to the Government of India, "National biodiversity mission" (2019); http://psa.gov.in/pmstiac-mssions/national-biodiversity-mission.
6. S. Kumar, *Science* 10.1126/science.aaz8953 (2019).
7. K. S. Bawa *et al.*, *Science* **327**, 1458 (2010).
8. L. Xia, "China, India vow to deepen cooperation at 6th Strategic Economic Dialogue held in India," *Xinhua* (2019).
9. "India and China join hands for community conservation," *NorthEastNow News* (2018).
10. K. S. Bawa, S. Kadur, *Himalaya: Mountains of Life* (Ashoka Trust for Research in Ecology and the Environment, 2013).
11. M. K. Pandit, *Life in the Himalaya: An Ecosystem at Risk* (Harvard University Press, 2017).
12. M. K. Pandit, *Nature* **583**, 9 (2020).

10.1126/science.abd4723

Conservation needs a COVID-19 bailout

Our ongoing extinction crisis (*1*) poses an existential threat to civilization (*2*), yet almost all biodiversity conservation strategies are underfunded (*3*). As the coronavirus disease 2019 (COVID-19) pandemic exposes the fragile foundation of global conservation (*4*), the ability of current conservation strategies to prevent the accelerated loss of biodiversity has become even more precarious.

COVID-19–induced reductions in trade and travel may temporarily reduce threats

Published by AAAS

to biodiversity (*4*). However, the nearly complete cessation of ecotourism and other income sources to many conservation areas and agencies is likely to drastically reduce biodiversity management and anti-poaching activities (*5*) and escalate unsanctioned resource extraction as economic hardships threaten livelihoods (*5*, *6*). Globally, governments have responded to pandemic impacts in other sectors of the economy by allocating almost US$11 trillion (*7*) of economic stimulus. However, conservation has yet to receive such stimulus packages, even though conservation is at the core of the United Nations' Sustainable Development Goals (*8*).

One of the most effective mechanisms for successful conservation of biodiversity is the establishment, expansion, and effective management of protected areas (*3*) and associated conservation lands (*9*). Ideal for economic stimulus funding, protected areas provide both direct short- and long-term economic benefits (*10*), assist vulnerable communities, and address policy needs (*11*). With clear multiplier effects, stimulus investments in protected areas will immediately promote job creation and economic activity, while subsequently bolstering national economies, maintaining vital ecosystem services, and mitigating climate change (*3*, *6*). Unlike traditional economic sectors, protected areas also have the potential to advance social development agendas, including just employment, sustainable food production, social inclusion, and access to education, safe drinking water, and dignified sanitation (*11*, *12*).

As pandemic-induced losses in biodiversity are unlikely to be recovered (*3*), the time to act is now. We urgently advocate for future stimulus packages to include funds for conservation and protected areas. These investments should be contingent on creating more resilient and sustainable models of conservation funding, increasing ownership rights, and promoting equitable development that strengthens livelihoods and benefits society as a whole.

Robert A. McCleery[1]*, Robert J. Fletcher Jr.[1], Laurence M. Kruger[2,3], Danny Govender[4], Sam M. Ferreira[4]
[1]Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL 32603, USA. [2]Organization for Tropical Studies, Skukuza 1350, South Africa. [3]Department of Biological Sciences, University of Cape Town, Cape Town, 7701, South Africa. [4]Scientific Services, SANParks, Skukuza, South Africa.
*Corresponding author. Email: ramccleery@ufl.edu

REFERENCES AND NOTES

1. E. S. Brondizio, J. Settele, S. Díaz, H. T. Ngo, "Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services" (IPBES Secretariat, 2019).
2. G. Ceballos, P. R. Ehrlich, P. H. Raven G., *Proc. Natl. Acad. Sci. U.S.A.* 10.1073/pnas.1922686117 (2020).
3. J. E. Watson, N. Dudley, D. B. Segan, M. Hockings, *Nature* **515**, 67 (2014).
4. R. T. Corlett, *Biol. Conserv.* **246**, 10.1016/j.biocon.2020.108571 (2020).
5. M. Hockings *et al.*, *Parks* **26**, 7 (2020).
6. G. Wittemyer, *Conserv. Biol.* **25**, 1002 (2011).
7. V. Gaspar, G. Gopinath, "Fiscal policies for a transformed world," *IMF Blog* (2020); https://blogs.imf.org/2020/07/10/fiscal-policies-for-a-transformed-world/.
8. United Nations, Department of Economic and Social Affairs, Sustainable Development (https://sdgs.un.org/#goal_section).
9. S. H. M. Butchart *et al. Conserv. Lett.* **8**, 329 (2015).
10. S. D. Gaines, C. White, M. H. Carr, S. R. Palumbi, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18286 (2010).
11. A. S. Pullin *et al.*, *Environ. Evidence* **2**, 19 (2013).
12. O. Venter *et al.*, *PLOS Biol.* **12**, 10.1371/journal.pbio.1001891 (2014).

COMPETING INTERESTS

R.A.M. receives funding from the U.S. National Institute of Food and Agriculture, Hatch project FLA-WEC-005125. R.A.M. and R.J.F. receive funding from a National Science Foundation International Research Experiences for Students grant (No. 1459882).

10.1126/science.abd2854

Extinct-in-the-wild species' last stand

The lockdowns and closures enacted in response to the coronavirus disease 2019 (COVID-19) pandemic have led to massive income losses for zoos, aquariums, and botanical gardens worldwide (*1*–*4*). Insufficient funding could affect the ability of these institutions to support wildlife conservation, which could lead to extinctions. There are currently at least 77 species of plants and animals that are extinct in the wild and exist only in zoological and botanical collections, where they rely on human care for survival (*5*).

Most extinct-in-the-wild species exist in small, closed populations, vulnerable to stochastic demographic processes and genetic threats associated with inbreeding (*6*, *7*). For example, the sihek (*Todiramphus cinnamominus*), a kingfisher endemic to Guam, was extirpated by 1988 as a result of predation by introduced brown tree snakes (*Boiga irregularis*) (*8*). Only 29 individuals were rescued, and they have subsequently been managed in Association of Zoos and Aquariums institutions across the United States, in addition to a facility on Guam (*9*). However, because not all of the captured birds bred successfully, the current sihek population of fewer than 140 individuals descends from only 16 genetic founders. It also suffers sex-ratio imbalances. The population therefore remains at risk, and further declines through loss of zoological institution support would hamper recovery efforts (*10*).

Ideally, extinct-in-the-wild populations can increase in captivity to the point that it is safe to release them back into their natural habitat. The successful reintroduction of the ko'ko' (Guam rail, *Hypotaenidia owstoni*) onto a small island near Guam has been heralded as a major conservation success. In 2019, the ko'ko' was reclassified from extinct in the wild to critically endangered (*11*). Such successes depend on full support for the zoos, aquariums, and gardens that struggle to maintain these collections. Further population declines will jeopardize recovery and increase extinction risks. We call for urgent funding to ensure that breeding, propagation, and holding facilities have the resources to care for extinct-in-the-wild species during the COVID-19 pandemic and beyond.

Amanda Trask[1]*, Stefano Canessa[2,3], Axel Moehrenschlager[4,3], Scott Newland[5], Suzanne Medina[6], John Ewen[1,3]
[1]Institute of Zoology, Zoological Society of London, Regents Park, London, NW1 4RY, UK. [2]Wildlife Health Ghent, Department of Pathology, Bacteriology, and Avian Diseases, Faculty of Veterinary Medicine, Ghent University, Merelbeke, Belgium. [3]IUCN Species Survival Commission, Conservation Translocation Specialist Group, Calgary, AB T2E 7V6, Canada. [4]Centre for Conservation Research, Calgary Zoological Society, Calgary, AB T2E 7V6, Canada. [5]Sedgwick County Zoo, Wichita, KS 67212, USA. [6]Guam Department of Agriculture, Division of Aquatic and Wildlife Resources, Mangilao, Guam.
*Corresponding author.
Email: amanda.trask@ioz.ac.uk

REFERENCES AND NOTES

1. Association of Zoos and Aquariums (AZA), "COVID-19 update: Taking action in unprecedented times" (2020).
2. British and Irish Association of Zoos and Aquariums, "Permanent closure during coronavirus" (2020).
3. World Association of Zoos and Aquariums, "Zoo and aquarium emergency operating funds" (2020).
4. French Association of Zoological Parks, "The Species Survival Commission of the International Union for Conservation of Nature (IUCN) calls for support for zoos weakened by health crisis" (2020); www.zoonaute.net/2020/06/17/commissioniucnzoo/ [in French].
5. The IUCN Red List of Threatened Species, Search results for Global, Extinct in the Wild, Species (2020); www.iucnredlist.org/search?redListCategory=ew.
6. B. A. Melbourne, A. Hastings, *Nature* **454**, 100 (2008).
7. P. Hedrick, A. Garcia-Dorado, *Trends Ecol. Evol.* **31**, 940 (2016).
8. J. A. Savidge, *Ecology* **68**, 660 (1987).
9. S. M. Haig, J. D. Ballou, N. J. Casna, *J. Hered.* **86**, 423 (1995).
10. S. Newland, K. Hundgen, G. Ferrie, "Guam kingfisher (*Todiramphus cinnamominus*) AZA Species Survival Plan® Yellow Program" (2020).
11. BirdLife International, *Hypotaenidia owstoni* (IUCN Red List of Threatened Species, 2019).

COMPETING INTERESTS

A.T. is funded by the Guam Department of Agriculture and a U.S. Fish and Wildlife Service Endangered Species Section 6 grant (award numbers F16AF01007 and F18AF01285) to facilitate a conservation translocation plan for the Guam kingfisher. A.M. is a research associate for the Wildlife Conservation Research Unit at Oxford University.

10.1126/science.abd4560

Published by AAAS

The "Cosmic Reef" image released by NASA on Hubble's 30th anniversary shows a star-forming region 163,000 light-years away.

PHOTO: NASA, ESA, AND STSCI

Bonds between AAAS and Hubble span three decades

As NASA's Hubble Space Telescope celebrates 30 years, a look at its scientists' AAAS connections

By **Adam D. Cohen**

On New Year's Day, 1925, Henry Russell, director of the Princeton University Observatory, presented to the joint meeting of the American Association for the Advancement of Science and the American Astronomical Society a research paper that would change humanity's understanding of the Universe and our place in it. Busy with ongoing observations at the Mount Wilson Observatory near Los Angeles, the paper's author, Edwin Hubble, mailed his work to the meeting rather than traveling to Washington, D.C., to present it himself.

Using Mount Wilson's 100-inch telescope—the world's largest from 1917 to 1949—Hubble had taken images of the Andromeda Nebula, determining that it is a galaxy of its own, composed of individual stars. The observation proved that our Milky Way does not take up nearly the entire Universe, as many astronomers of the time believed, but instead is one galaxy of many.

"This paper is the product of a young man of conspicuous and recognized ability in a field which he has made peculiarly his own," Russell and Joel Stebbins, secretary of the American Astronomical Society, wrote in a letter to the AAAS Committee on Awards. "It has already expanded one hundred-fold the known volume of the material universe."

Of approximately 1,700 scientists who shared research at the meeting, the committee chose two—Hubble and a zoologist named L. R. Cleveland—to share the 1924 Thousand Dollar Prize, an award honoring the most noteworthy contributions to science presented at the AAAS Annual Meeting each year. Now called the Newcomb Cleveland Prize, the yearly award goes to the author or authors of a particularly impactful paper published in *Science*.

"The Association Prize will never be awarded to a more appreciative student than your present choice," Hubble wrote to AAAS secretary Burton Livingston in February 1925. "The occasion of the reward, however, must be regarded as a triumph of modern instruments rather than a personal achievement.... This was accomplished by the use of the largest telescopes in existence."

Decades later, the National Aeronautics and Space Administration launched into Earth orbit an instrument exponentially larger and more precise than anything available to Hubble, and it bore his name. In April 1990, the Hubble Space Telescope became the first major optical telescope to be placed in space. Free from the atmospheric distortion and light pollution faced by land-based telescopes, Hubble has provided data leading to more than 17,000 peer-reviewed publications on topics including star formation, galaxy mergers, and dark matter, the largely unseen mass that occupies most of the Universe.

NASA is celebrating a series of the Hubble mission's 30th-anniversary milestones this year, from the launch in April to its "first light" image in May and its observations of Supernova 1987A in August. Throughout the mission's first three decades, AAAS has honored its scientists with awards and fellowships and used their expertise to

Published by AAAS

guide AAAS programs. AAAS and Hubble also overlap in their vision for the science, technology, engineering, and mathematics (STEM) enterprise, as both work to enhance public engagement, educate and diversify the STEM workforce, and foster international collaboration.

"Hubble, slam dunk, aligns with almost all of them," astronomer Kathryn Flanagan said of AAAS's goals.

Prior to retiring in March, Flanagan oversaw the Hubble mission as deputy director and interim director of the Baltimore-based Space Telescope Science Institute (STScI), a nonprofit science center operated for NASA and responsible for Hubble's science operations and public outreach. She also chaired AAAS's astronomy section in 2016 and was inducted as an elected AAAS fellow this year, the latest of dozens of Hubble-affiliated scientists to receive the honor.

"Hubble was, when I assessed it a few years back, contributing as a recommended or required component in half the state departments of education in the U.S.," Flanagan said, referencing the prevalence of the telescope's findings in public school curricula. "It reached half the public middle school students in the country."

Flanagan also highlighted the Hubble mission's dedication to international collaboration through its partnership with the European Space Agency in development and operations, publicly available archives, research time allotted to scientists from around the world, and a double-anonymous peer-review process, which ensures that unconscious bias does not play a role in determining which research proposals are selected. Many AAAS programs, including Science in the Classroom, the Center for Science Diplomacy, and SEA Change, have a similar focus on science education, international collaboration, and achieving equity in STEM disciplines.

Carol Christian, Hubble Space Telescope outreach project scientist at STScI, works to share the mission's discoveries with the world. Christian is part of the Hubble news team, highlighting science results from hundreds of refereed papers each year. Additionally, she leads a program that brings Hubble data to individuals with blindness and visual impairments using 3D printing.

Hubble was deployed from the space shuttle *Discovery* on 25 April 1990.

PHOTO: NASA

Christian also has served as a screener for the AAAS Kavli Science Journalism Awards since 2010. Each year, approximately 70 scientists volunteer to assess the scientific accuracy of the award program's print, digital, and video reporting candidates. Christian said that the review panels consist of researchers from diverse disciplines, including geology and the life sciences.

"You learn a lot from the people that are your fellow reviewers," she said.

Prior to working with the Kavli Awards, Christian brought her expertise in satellite imagery to the U.S. Department of State, where she worked from 2003 to 2007 through the AAAS Science and Technology Policy Fellowships program, which places more than 250 scientists each year in policy roles across all three branches of the federal government. She enjoyed putting her scientific knowledge to practical use during her time as a fellow.

"Suddenly, to be doing something that actually has import—policy may depend on it, economics may depend on it—that was pretty interesting," Christian said.

Margaret Burbidge, an influential astrophysicist who died in April at the age of 100, was perhaps the first scientist to contribute to both the Hubble mission and AAAS. During her term as president of AAAS in 1983, she was on the team that developed the Faint Object Spectrograph, one of the five science instruments on Hubble when it launched. Burbidge and her colleagues designed the instrument to detect the physical and chemical properties of faint objects in galaxies beyond our own, and it provided the first strong observational evidence for the existence of a supermassive black hole in the core of another galaxy.

"She was one of the names that we all knew of as pioneers in the field of modern astronomy," astrophysicist Jennifer Wiseman said following Burbidge's death. "An inspirational person as a brilliant scientist and in particular as a courageous woman scientist who opened up the field for many of us coming later on."

Like Burbidge did for a time, Wiseman works with the Hubble Space Telescope and is also active in AAAS activities. As Hubble's senior project scientist, Wiseman works with a team of scientists and engineers at NASA's Goddard Space Flight Center to keep the telescope as scientifically productive as possible.

Prior to her NASA career, Wiseman served as an American Physical Society congressional science fellow, a role under the umbrella of the AAAS Science and Technology Policy Fellowships. She brought her subsequent connection with Hubble to a broad AAAS audience by co-organizing a symposium on the Hubble Space Telescope mission—its scientific discoveries, image processing feats, and future promise—at the AAAS 2015 Annual Meeting, and she has given Hubble overview talks at subsequent AAAS gatherings.

Since 2010, Wiseman has also directed the AAAS Dialogue on Science, Ethics, and Religion (DoSER) program, which works to foster communication between scientific and religious communities.

"Being an active scientist while attentive to public science engagement helps me to recognize how scientists and the other communities that DoSER works with—including ethicists and religious communities—can better understand each other and our common interests in science, technology, and their positive contributions to society," Wiseman said.

"Astronomy in general gives everyone a sense of humility, awe, wonder, and curiosity," she added. "I enjoy sharing with many types of public audiences the amazing things we're discovering with telescopes and other scientific facilities. I'm very humbled and encouraged by how such discoveries can inspire deeper discussions of meaning and purpose and of using science to help one another, as we all ask questions together of who we are as human beings in an awesome Universe and how we can use our knowledge to uplift the human spirit."

Published by AAAS

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

ARCHAEOLOGY

Source of Stonehenge's sarsens

Despite centuries of debate, the geological origin of the massive standing stones of Stonehenge—the sarsens—has never been conclusively identified. Previous inference suggested a site 30 kilometers to the north as the source. Nash *et al*., using a combination of analytical techniques, tested this assertion and discovered that 50 of the 52 sarsens had a consistent geochemistry, thus signaling a common source. Additional testing revealed that a location only 25 kilometers north of Stonehenge is the most probable source of the sarsens. These findings help to clarify competing reconstructions about how the stones were transported to the site and lend support to hypotheses that the sarsens were erected around the same time, roughly 2500 BCE. —MA

Sci. Adv. 10.1126/sciadv.abc0133 (2020).

Geochemical evidence suggests a precise origin for some of the larger stones used in Stonehenge.

PHOTO: © ENGLISH HERITAGE, ANDRE PATTENDEN

HOST AND MICROBE

This is the brain on probiotics

The gut microbiome plays an important role in the gut-brain axis, a system of bidirectional communication between the central and enteric nervous systems. Zhang *et al*. analyzed the presence and abundance of brain peptides—the peptidome—along with the gut microbiomes of mice fed two different species of probiotics. Probiotic treatment altered the abundances of many neuropeptides in a manner that correlated with changes in the composition of the gut microbiome, varied by brain region, and depended on the probiotic species, whether the probiotic was live or heat-killed, and the duration of treatment. —AMV

Sci. Signal. **13**, eabb0443 (2020).

ORGANIC CHEMISTRY

Simple swaps of CO_2

The loss of carbon dioxide (CO_2) from carboxylic acids is a common reaction in both biochemical and synthetic contexts, but it has generally involved catalysis or prolonged heating. Kong *et al*. now report that certain polar solvents, such as dimethylformamide, promote reversible CO_2 loss all by themselves from carboxylates bridged by one carbon to aromatic rings. With electron-withdrawing substituents on the ring, isotopically labeled CO_2 can be efficiently swapped in even at room temperature. Alternatively, reaction with aldehydes leads to alcohol formation. —JSY

Science, this issue p. 557

PALEOBOTANY

Origins of an alpine flora

The evolution of high mountain floras is strongly influenced by tectonic and climatic history. Ding *et al*. document the timing, tempo, and mode by which the world's most species-rich alpine flora, that of the Tibet-Himalaya-Hengduan region, was assembled. Alpine assemblages in the region are older than previously thought, with lineages tracing their alpine ancestry to the early Oligocene—older than any other modern alpine system. Alpine species diversified faster during periods of orogeny and intensification of the Asian monsoon, and the Hengduan Mountains—the most species-rich area in this region—played a key biogeographic role as the location of the earliest pulse of alpine diversification in the Oligocene. —AMS

Science, this issue p. 578

IMMUNOGENOMICS

The development of microglia

Microglia are the brain's immune cells, and they play important roles in health and neurodegenerative disease. Kracht *et al*. performed single-cell analysis of human microglial gene expression and chromatin

Published by AAAS

accessibility and compared the results with those of other studies of human and mice microglial development. By using in situ validation, these data identify fetal microglial subsets that appear to be distinct from adult human microglia, suggesting functional differences between the developing and mature brain. —LMZ

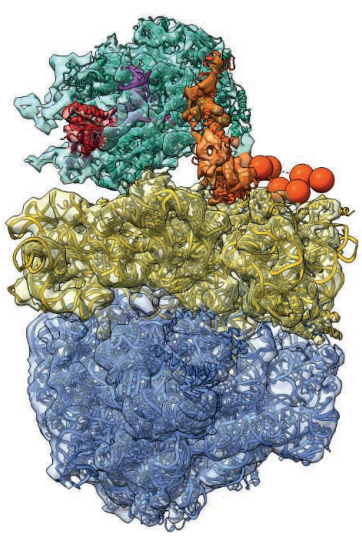
Science, this issue p. 530

STRUCTURAL BIOLOGY

Integrative in-cell structural biology

In bacteria, RNA polymerases can associate with ribosomes to form transcription-translation units called expressomes. Multiple models based on structural data of in vitro reconstitution assays have been proposed for how the two machineries interface with one another. Understanding this bacteria-specific coupling mechanism offers insight regarding the central dogma of molecular biology and might be leveraged for antibiotic development. O'Reilly *et al*. found that the NusA protein interfaces between the two complexes. The authors combined cryo–electron tomography and cross-linking mass spectrometry to produce an integrative model of the transcribing, translating expressome of *Mycoplasma pneumoniae* obtained entirely from in-cell data. This approach contributes to the development of in-cell structural biology. —SMH

Science, this issue p. 554



A complex of RNA polymerase and the ribosome termed the expressome, as visualized in bacterial cells by cryo–electron tomography

BIOTECHNOLOGY

Secrets of a fast base editor

CRISPR-Cas9 base editors comprise RNA-guided Cas proteins fused to an enzyme that can deaminate a DNA nucleoside. No natural enzyme deaminates adenine in DNA, and so a breakthrough came when a natural transfer RNA deaminase was fused to Cas9 and evolved to give an adenine base editor (ABE) that works on DNA. Further evolution provided the enzyme ABE8e, which catalyzes deamination more than 1000 times faster than early ABEs. Lapinaite *et al*. now present a 3.2-angstrom resolution structure of ABE8e bound to DNA in which the target adenine is replaced with an analog designed to trap the catalytic conformation. The structure, together with kinetic data comparing ABE8e to earlier ABEs, explains how ABE8e edits DNA bases and could inform future base-editor design. —VV

Science, this issue p. 566

ULTRACOLD MATTER

Lowering the entropy

Atoms loaded in an optical lattice can convincingly mimic the behavior of electrons in solids. However, reaching very low temperatures, where the most interesting quantum phases are expected to occur, is tricky. Yang *et al*. introduced a clever experimental technique to reduce the entropy of their sample. They created an array consisting of alternating rows of sample atoms and reservoirs. The entropy was transferred from the sample atoms to the adjacent reservoirs, which were then removed. The resulting low-entropy system can be used as a basis for quantum simulation and information. —JS

Science, this issue p. 550

IN OTHER JOURNALS

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

***Sphagnum* moss, which forms peat and is a major carbon store, is at risk of drying out under climate change.**

ORGANIC CHEMISTRY

A switch in time saves ligand

Isolating just one of two mirror-image products, or enantiomers, is essential in pharmaceutical synthesis. Often distinct enantiomeric ligands must be appended to a catalyst to steer the reaction one way or the other. Tu *et al*. report an intriguing case in which a single catalyst produces either product enantiomer, depending on reaction time. An iridium-catalyzed allylation of hydroxyisoquinolines yields one product within 10 minutes via kinetic resolution of the allyl precursors. Over 10 more hours, this initial product comes apart via a catalyzed reaction with the solvent, whereas its enantiomer steadily builds up and stays intact. —JSY

Nat. Chem. 10.1038/s41557-020-0489-1 (2020).

CELL BIOLOGY

Too much of a good thing

The membrane remodeling machinery known as ESCRT-III plays a role in maintaining nuclear integrity. Chromosomal disruption and the formation of micronuclei in a process known as chromothripsis can lead to pathologies, including cancer. Using a combination of live-cell imaging and electron tomography in tissue culture cells, Vietri *et al*. found that the nuclear integrity surveillance system paradoxically drives micronuclear catastrophe. The authors characterized the role of the cytosolic ESCRT-III binding protein CHMP7 and its intranuclear partner LEMD2, which together detect ruptures in the nuclear envelope and then trigger envelope repair. Ruptured micronuclei, however, were unable to control this system, which led to the accumulation of ESCRT-III subunits. This accumulation drove membrane aberrations and thereby torsional stress–induced DNA damage. These findings elucidate the chromosome-shattering process associated with chromothripsis that potentially drives cancer development. —SMH

Nat. Cell Biol. **22**, 856 (2020).

CREDITS (FROM LEFT): O'REILLY *ET AL.*; BLICKWINKEL/ALAMY STOCK PHOTO

Published by AAAS

BIOGEOGRAPHY

Shrubification of peatlands

Soil is a key player in the global carbon cycle. Peatlands are important because they contain more than half of global soil carbon, much of which is sequestered in slow-decaying *Sphagnum* moss. Hence, there is a need to understand how peatlands will respond to the rapidly changing climate. Malhotra *et al*. assessed the response of plant roots to experimental warming in a peatbog ecosystem in the state of Minnesota, United States. The growth of fine roots of shrubs and trees over two growing seasons rapidly increased by 130% for each 1°C temperature increase, mainly because of the drying of the soil and the increase in the duration of the growing season. The authors suggest that over the longer term, drying will allow replacement of *Sphagnum* moss by other plants and hasten the loss of soil carbon. —AMS

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.2003361117 (2020).

ELECTRONICS

From my hand to your ears

Sign language—in the form of hand, face, and body gestures—is a broadly used technique for communicating with people who are partially or fully deaf. However, most nondeaf people do not know sign language, making communication between the two groups challenging. Zhou *et al*. developed a wearable device that can convert hand motions or facial expressions into sound. At its core are two dissimilar soft materials with opposite triboelectric polarizations woven into a yarn that can effectively convert a small tensile force or pressure into electricity. Thus, finger or facial movements generate specific voltage pulses that can be recognized as individual letters or words or phrases. —MSL

Nat. Electron. 10.1038/s41928-020-0428-6 (2020).

ANTHROPOLOGY

Inside the Paleolithic mind

It is not often that we gain insight into the cognitive functions of our forebears. A very rare type of tool made from a hippopotamus bone has been discovered at the Konso Formation in southern Ethiopia. This artifact represents a technological breakthrough achieved 1.75 million years ago. Bone is tricky to shape and requires considerable judgment and skill to form successfully. Sano *et al*. show how this large (>10 centimeter) fragment was intentionally shaped by a controlled knapping technique and turned into a handaxe. This breakthrough represents a step-change from the conservative thinking of previous tool makers. By analyzing the scarring around the edges of the superbly preserved tool, the authors inferred that the maker, probably *Homo erectus*, was able to adjust the thickness on both sides of the material by a distinctive flaking technique. This allowed it to be used for precise purposes, such as butchering animal carcasses. —CA

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.2006370117 (2020).

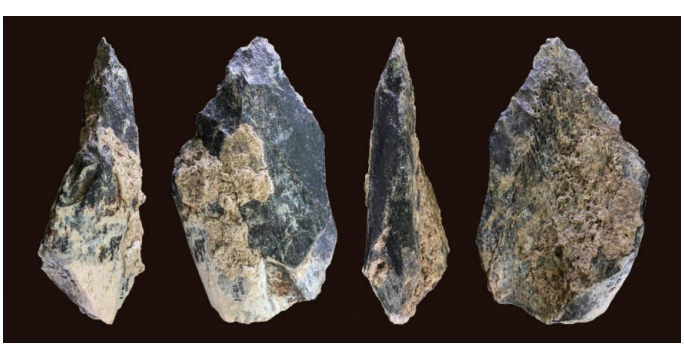


PHOTO: SANO *ET AL*., *PROC. NATL. ACAD. SCI. U.S.A.* 10.1073/PNAS.2006370117 (2020)

A rare handaxe tool made from hippopotamus bone indicates highly skilled workmanship by *Homo erectus* 1.75 million years ago.

NEUROSCIENCE

Autobiographical subnetworks

The brain is organized into large networks that can be identified during brain scanning. Not every region of the human brain is functionally connected with all the other regions. One of the best-studied networks is the so-called default-mode network, which is commonly associated with autobiographical memory and other types of internally oriented cognition. Gordon *et al*. identified nine subnetworks within the default-mode network that were found consistently in human participants. These results create an organizational framework for the investigation of the default-mode network in individual subjects. The authors hypothesize that such a subnetwork structure is likely a general feature of all human brain networks. —PRS

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.2005238117 (2020).

LIGHT POLLUTION

Dimming Tucson's street lighting

Poorly designed artificial lighting wastes energy and produces light pollution, with damaging effects on the natural environment, human health, and astronomy. Dimming or extinguishing street lights can reduce light pollution, but its efficacy has been difficult to measure. Barentine *et al*. monitored the sky brightness over the city of Tucson, United States, during a test of a new municipal street lighting system. The city authorities dimmed their streetlights at midnight, allowing the authors to determine that street lighting now contributes only 14% of Tucson's total light pollution. Further reductions must focus on other sources, such as private businesses and illuminated advertising. —KTS

J. Quant. Spectrosc. Radiat. Transf. **253**, 107120 (2020).

Published by AAAS

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

IMMUNOLOGY

How the thymus shaped immunology

The function of the thymus was discovered by Jacques Miller in 1961 and laid a foundation for immunology and modern medicine. Until that time, researchers mistakenly believed the thymus merely represented a remnant of defunct lymph tissue, something of an immune cell graveyard. In a Review, Miller recounts the seminal experiments and conceptual thinking that led to uncovering the critical function of the thymus and provides insights that can be learned from those early days. How knowledge of thymus function subsequently spawned the field of T cell biology and the impact it has had on immune cell interactions, vaccination, cancer immunotherapy, and the microbiome are also discussed. —PNK

Science, this issue p. 522

CORONAVIRUS

How does COVID-19 develop?

Coronavirus disease 2019 (COVID-19) is a heterogeneous disease, with patients experiencing diverse symptoms ranging from none to fatal respiratory distress. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes COVID-19, infects numerous cell types, which is determined by the cellular receptor, or receptors, it binds. In a Perspective, Matheson and Lehner discuss how receptor binding determines the pattern of COVID-19 disease symptoms, the unanswered questions about the biology of infection, and whether infection can be prevented to attenuate severe disease. —GKA

Science, this issue p. 510

STRUCTURAL BIOLOGY

Revealing family differences

In response to low blood glucose concentrations, both the glucagon receptor (GCGR)—a family B G protein–coupled receptor (GPCR)—and the β_2 adrenergic receptor (β_2AR)—a family A GPCR—are activated and act through the cyclic adenosine monophosphate signaling pathway to increase glucose production. The kinetics of the response is different for the two receptors. Based on structural and spectroscopic data, Hilger *et al*. show that the conformation of transmembrane helix 6 in the activated state is a key differentiator (see the Perspective by Lebon). In β_2AR, the helix moves toward its active conformation when an agonist binds, but in GCGR, both agonist and G protein binding are required. This likely explains why activation of its partner G protein is slower for GCGR than for β_2AR. —VV

Science, this issue p. 523; see also p. 507

BIOCHEMISTRY

Deadenylating RNA molecules

Transfer RNA (tRNA) and messenger RNA molecules often acquire a terminal 2′,3′-cyclic phosphate group when processed in the cell. These cyclic phosphates provide attachment points for tRNA ligases and must be removed to recycle tRNAs from stalled ribosomes. Pinto *et al*. identified a deadenylase from human tissue culture cells that can do the job. Biochemical characterization and analysis of a crystal structure reveal ANGEL2 as a 2′,3′-cyclic phosphatase with functions for RNA processing and modification. —BAP

Science, this issue p. 524

NEUROGENOMICS

Effects of allele-specific open chromatin

Genetic variants in noncoding regions of the genome may underlie the development of disease. However, we are just beginning to tease apart the function of such variants associated with neuropsychiatric disease. Using five types of neural progenitor cells derived from 20 human induced pluripotent stem cell lines, Zhang *et al*. looked at allele-specific open chromatin (ASoC) variants. Many ASoC variants overlapped with genomic elements, such as transcription factor binding sites, and loci identified in genome-wide association studies for neurological traits. From the experimental and computational analyses, they identified single-nucleotide polymorphisms and illuminate how one schizophrenia-associated variant affects neurodevelopment. —LMZ

Science, this issue p. 561

NEURODEVELOPMENT

Neocortex in the fetal brain

Along the path of human evolution, gene duplication and divergence produced a protein, ARHGAP11B, that is found in humans but not nonhuman primates or other mammals. Heide *et al*. analyzed the effects of *ARHGAP11B* gene expression, under control of its own human-specific promoter, in the fetal marmoset (see the Perspective by Dehay and Kennedy). In the early weeks of fetal growth, the gene drove greater elaboration of neural progenitors and neocortex than is evident in the normal fetal marmoset. ARHGAP11B expression may be one cause of the more robust neocortex that characterizes the human brain. —PJH

Science, this issue p. 546; see also p. 506

MOLECULAR DIFFUSION

Reactions give solvents a kick

During a chemical reaction, the reorganization of solvent molecules not directly in contact with reactants and products is normally viewed as a simple diffusion response. Wang *et al*. studied molecular diffusion in six common reactions—including the copper-catalyzed click reaction and the Diels-Alder reaction—with pulsed-field gradient nuclear magnetic resonance. They observed a boost in mobility relative to Brownian diffusion that was stronger for the catalyzed reactions that were studied. The mobilities for the click reaction were verified with a microfluidic gradient method. They argue that energy release produces transient translational motion of reacting centers that mechanically perturbs solvent molecules. —PDS

Science, this issue p. 537

NANOMATERIALS

Graphene nanoribbons made on oxides

Atomically precise nanographenes and nanoribbons have been synthesized on metal surfaces that catalyze cyclodehydrogenation of precursors. However, for use in devices, these structures usually must be transferred to insulating or semiconducting surfaces. Kolmer *et al*. synthesized precise graphene nanoribbons on the surface of rutile titanium dioxide (TiO_2) that assisted the cyclodehydrofluorination of specifically designed precursor molecules through a series of thermally triggered transformations. Scanning tunneling microscopy and spectroscopy confirmed the formation of well-defined zigzag ends of the nanoribbons as well as their weak interaction with the substrate. —PDS

Science, this issue p. 571

Published by AAAS

MATERIALS SCIENCE

Deformable semiconductors

Semiconductors are usually brittle and do not deform easily. Wei *et al*. found that bulk single crystals of indium selenide instead have excellent flexibility (see the Perspective by Han). The deformability comes from the compliant intralayer bonding between indium and selenium. The authors used these observations along with a previously discovered silver sulfide to determine a deformability factor for materials that may help find other deformable semiconductors. —BG

Science, this issue p. 542; see also p. 509

AIR POLLUTION

Cleaner skies

Particulate air pollution in the contiguous United States has decreased considerably over recent decades, but where exactly has that progress been made? Colmer *et al*. analyzed 36 years of data and found that the spatial distribution of fine particulate matter concentrations has remained largely unchanged over that interval (see the Perspective by Ma). Although, fine particulate pollution levels have dropped overall, those areas that were most and least polluted in 1981 remain so today. We may have made important strides in pollution control, but we have been less successful in addressing disparities of exposure between communities. —HJS

Science, this issue p. 575; see also p. 503

CORONAVIRUS

Epidemic in Northern California

Genome sequencing of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) outbreaks is valuable for tracing the sources and perhaps for drawing lessons about preventing future outbreaks. Genomic analysis by Deng *et al*. revealed that Northern California experienced a complex series of introductions of the virus, deriving not only from state-to-state transmission but also from international travel by air and ship. The study highlights the importance of being able to rapidly test and trace contacts of positive cases to enable swift control. —CA

Science, this issue p. 582

SOLAR PHYSICS

Predicting large solar flares

The sudden release of magnetic energy on the Sun drives powerful solar flares, which are difficult to predict. Kusano *et al*. derived physics-based thresholds for the onset of large solar flares and show how they can be predicted from routine solar observations (see the Perspective by Veronig). They tested their method using observations of the Sun from 2008 to 2019. In most cases, the method correctly identifies which regions will produce large flares within the next 20 hours, although there are some false positives and false negatives. The method also provides the exact location where each flare will begin and limits on how powerful it will be. Accurate predictions of solar flares could improve forecasts of space weather conditions around Earth. —KTS

Science, this issue p. 587; see also p. 504

CARDIOMYOPATHY

Lovastatin for cardiolaminopathy

Mutations in the gene *LMNA*, which encodes nuclear envelope proteins, can cause dilated cardiomyopathy associated with arrhythmia and sudden cardiac death. To understand the mechanisms contributing to this disease, Sayed *et al*. studied induced pluripotent stem cell–derived endothelial cells (iPSC-ECs) from a family harboring an *LMNA* mutation. They found down-regulation of a protein involved in mechanotransduction, which caused endothelial dysfunction. Lovastatin could induce this protein in iPSC-ECs, improving cardiomyocyte function in coculture and clinical endothelial cell function in two patients treated with the drug. This study demonstrates a workflow for identifying and validating potential drug treatments for patients with cardiolaminopathy. —CC

Sci. Transl. Med. **12**, eaax9276 (2020).

IMMUNODEFICIENCY

A human immune system sans MAIT cells

Mucosal-associated invariant T (MAIT) cells represent ~5% of circulating human T cells, reside in high frequency in mucosal tissues, and provide enhanced immunity against bacterial and viral infections. Howson *et al*. set out to determine the genetic basis of immunodeficiency in a patient with persistent tattoo-associated human papilloma virus–positive warts. The gene *MR1* encodes a major histocompatibility complex–like molecule that presents bacteria-derived riboflavin metabolites to MAIT cells. Whole-exome sequencing led the researchers to identify a rare point mutation in both copies of the *MR1* gene. By solving the structure of the mutant MR1 protein, they demonstrated that the mutation disrupted the ligand-binding domain of MR1. Consistent with this finding, the patient lacked circulating MAIT cells. —AB

Sci. Immunol. **5**, eabc9492 (2020).

Published by AAAS

REVIEW SUMMARY

IMMUNOLOGY

The function of the thymus and its impact on modern medicine

Jacques F. A. P. Miller[1,2]*

BACKGROUND: Until the 1960s, the thymus had long remained an enigmatic organ with no known function or was considered to be just a "graveyard" for dying lymphocytes. By contrast, the circulating small lymphocytes found in blood and lymph had been shown in the late 1950s to be long-lived cells able to initiate an immune response when appropriately stimulated by antigen. However, although the thymus was known to be populated with lymphocytes, immunologists had not attributed any immune function to the organ. There were many reasons for this. For example, the cytological hallmarks of an immune response, such as the presence of plasma cells and germinal centers, were not seen in the thymus of healthy animals even after extensive immunization. Thymus lymphocytes, unlike lymphocytes from blood and lymphoid tissues, were unable to mount an immune response on transfer to immunocompromised recipients. Furthermore, thymectomy, which had always been performed in adult animals, was not associated with any immune defects.

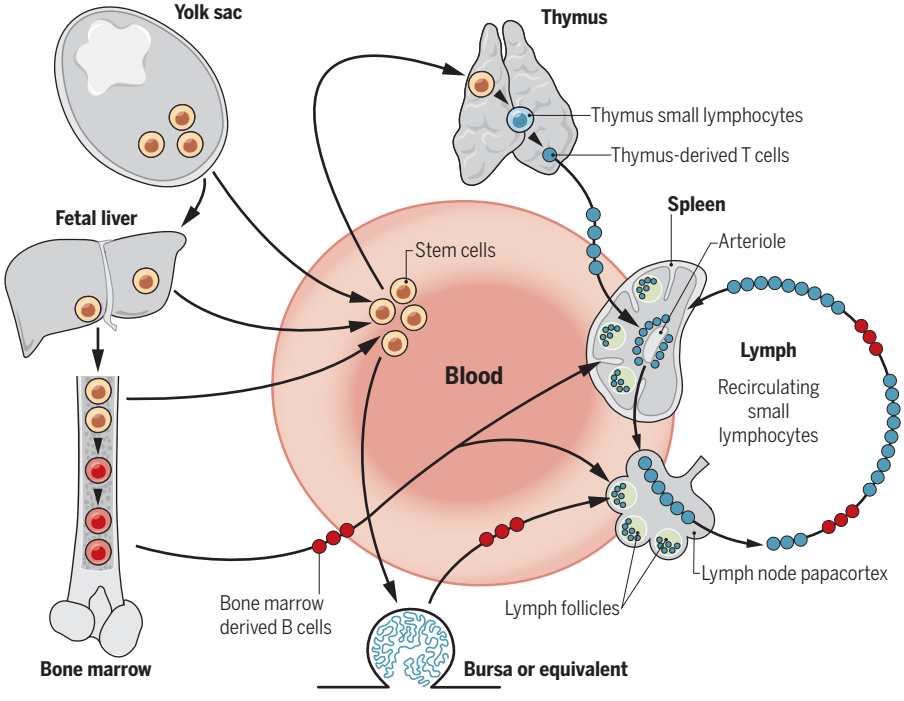
ADVANCES: In 1961, thymectomy was performed in mice during the immediate neonatal period and revealed the critically important function of the thymus in enabling the development of the immune system. Neonatally thymectomized (NTx) mice were highly susceptible to intercurrent infections, deficient in lymphocytes, unable to reject foreign skin grafts or produce antibody to some (though not all) antigens, and prone to developing certain tumors. There was no major immune defect after adult thymectomy unless the lymphoid tissues had been damaged by total body irradiation. Implanting thymus tissue into NTx mice or irradiated adult thymectomized mice restored immune function. When the thymus graft was taken from a foreign strain, the thymectomized recipients were found to be specifically tolerant of the histocompatibility antigens of the donor. This implied that the thymus was responsible not only for the normal development of immune functions but also for imposing tolerance to the body's own tissues.

The thymus was shown to seed the lymphoid system with T lymphocytes that migrated to certain discrete areas of the lymphoid tissues and recirculated from the blood through lymphoid tissues, lymph, and back to the bloodstream. T cells exiting the thymus were found to be responsible for killing infected cells and for rejecting foreign tissue grafts. Therefore, T cells could soon be subdivided into subsets based on function, cell surface markers, and secreted products or interleukins. Thymus lymphoid stem cells were subsequently identified, and the epithelial and dendritic cells in the thymus were shown to greatly influence T cell development. They were able to educate T cells to recognize a great diversity of peptide antigens bound to the body's own markers, major histocompatibility complex molecules, but purged any T cells that strongly reacted against the body's own self-components. The use of irradiated adult thymectomized mice showed that T cells were not the precursors of antibody-forming cells but were essential to help, through some type of collaboration, other lymphocytes originating in bone marrow (B cells) to respond to antigen by producing antibody.

OUTLOOK: The discovery of thymus function and of T and B cell collaboration was a major immunological milestone because it not only opened up the field of immune cell interactions but also changed the course of immunology and medicine. It promoted the need for all immune phenomena, for example, memory, tolerance, autoimmunity, and immunodeficiency, as well as inflammatory and immunopathological disease conditions, to be reassessed in terms of the role played by the two distinct sets of lymphocytes and their subsets. We now know that T cells are involved in the entire spectrum of tissue physiology and pathology and even in situations not considered to be bona fide immunological conditions, such as tissue repair, dysbiosis, eclampsia, senescence, and cancer. ■

ON OUR WEBSITE

Read the full article at https://dx.doi.org/10.1126/science.aba2429



Origin and migration of T and B cells. Hemopoietic stem cells originating in the yolk sac, fetal liver, or bone marrow migrate through the blood to the thymus, where they differentiate into T lymphocytes. Others differentiate within the bone marrow in mammals, or the bursa of Fabricius in birds, to produce B cells. Mature T and B cells circulate in the blood and lymph and colonize the T and B cell compartments of the lymphoid tissues.

[1]The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia. [2]Department of Medical Biology, The University of Melbourne, Parkville, Victoria 3010, Australia.
*Corresponding author. Email: miller@wehi.edu.au
Cite this article as J. F. A. P. Miller, *Science* **369**, eaba2429 (2020). DOI: 10.1126/science.aba2429

REPRINTED WITH PERMISSION FROM THE ROYAL SOCIETY, LONDON

REVIEW

IMMUNOLOGY

The function of the thymus and its impact on modern medicine

Jacques F. A. P. Miller[1,2]*

The lymphoid system is intimately involved in immunological processes. The small lymphocyte that circulates through blood into lymphoid tissues, then through the lymph and back to the blood through the thoracic duct, is able to initiate immune responses after appropriate stimulation by antigen. However, the lymphocytes found in the thymus are deficient in this ability despite the fact that the thymus plays a central role in lymphocyte production and in ensuring the normal development of immunological faculty. During embryogenesis, lymphocytes are present in the thymus before they can be identified in the circulation and in other lymphoid tissues. They become "educated" in the thymus to recognize a great diversity of peptide antigens bound to the body's own marker antigen, the major histocompatibility complex, but they are purged if they strongly react against their own self-components. Lymphocytes differentiate to become various T cell subsets and then exit through the bloodstream to populate certain areas of the lymphoid system as peripheral T lymphocytes with distinct markers and immune functions.

Thymectomy in the immediate neonatal period in mice is associated with a decrease in the population of T cells throughout the body. As a result, neonatal thymectomy impairs various immunological reactivities, but thymectomy performed later in adult life, after the lymphoid system has been constructed, has no major effect unless the lymphocyte population has been depleted by agents such as irradiation.

Peripheral T cells are intimately involved in responding to infections and in reactions such as delayed-type hypersensitivity (e.g., the tuberculin reaction) and foreign tissue graft rejection, but they are not able to produce antibodies. Nevertheless, in many cases, T cells are essential to help, through some type of collaboration, other lymphocytes originating in bone marrow (B cells) to respond to antigen by producing antibody.

The discovery of thymus immune functions and of the interaction between T and B cells has had wide repercussions in many areas of medicine and even in the management of some cancers.

To survive and resist invasion by pathogenic organisms, multicellular species have had to evolve defense systems. In a very general way, one can distinguish between constitutive mechanisms that are innate and nonspecific and adaptive mechanisms that are specific and exemplified by acquired immunity. The latter are a function of the circulating small lymphocyte, as first demonstrated by Jim Gowans (*1*). The adaptive defense system encompasses both "humoral immunity," in which antibodies are secreted as immunoglobulin molecules, and "cellular immunity" (*2*), in which the immune response is carried out by lymphocytes that do not produce antibodies.

The thymus occupies a specialized position in the lymphoid system (*3*) and differs from other lymphoid structures both structurally and functionally. It is situated in the chest behind the upper part of the sternum or breastbone and above the heart and extends upward for a short distance into the neck. It is relatively large in infancy and in humans reaches a size of 35 to 40 g at puberty. Thereafter, it regresses and eventually becomes reduced to little more than a vestigial structure in old age (*4*). It is divided into lobules composed of a central part or medulla and a peripheral part or cortex. Its major cell types are the lymphocytes, epithelial cells, and dendritic cells. The former are densely packed in the cortex, where they outnumber the epithelial cells. The latter are most prominent in the medulla (*5*–*7*) The various steps in thymus lymphocyte differentiation were discovered in detail after the 1960s, as described in this review.

During development in the mouse, lymphocytes appear first in the thymus and only after birth in the circulation, spleen, lymph nodes, gastrointestinal lymphoid tissues, and other tissues (*8*). They show practically no mitotic activity outside the thymus before birth. Lymphocyte proliferation in the thymus cortex exceeds that in any other lymphoid tissues throughout life. Even with advancing age and when the thymus involutes, its lymphocyte mitotic activity is still intense and higher than in lymph nodes (*7*). However, most thymus lymphocytes, unlike lymphocytes elsewhere, are not able to induce immune responses when transferred to immunoincompetent hosts and appropriately stimulated by antigen (*9*, *10*).

It seems reasonable to postulate from these insights that at least some lymphocytes undergo a process of maturation in the thymus and subsequently migrate at various times, before or after birth depending on the species, to populate the rest of the lymphoid system with immunocompetent cells. However, this was not established until after the results of neonatal thymectomy in the mouse were published, as detailed below. We now know how and where lymphocytes in the thymus acquire their ability to distinguish self from nonself and how they differentiate into various T cell subsets with distinct immune functions that migrate out to the rest of the lymphoid system.

A long-neglected organ

For centuries, the thymus has remained an enigmatic organ, and claims and counterclaims mainly involved questions regarding its function. The fact that it is a large mass of tissue in infancy was not appreciated by clinicians in the early part of the 20th century. Autopsies performed after severe illnesses such as diphtheria revealed a shrunken thymus, which was the result of stress during the infection. Whenever death occurred in surgery for stress-unrelated conditions such as congenital heart defects, it was believed that the large thymus had interfered with breathing, rather than that the anesthesia had contributed to death. Some physicians even prescribed irradiation to reduce the size of the thymus (*11*), not realizing that these patients might later develop thyroid tumors or thymomas.

The lymphopoietic function of the thymus was firmly established in the 1950s, yet immunologists were reluctant to accept its role in immunity. Thus, unlike circulating or splenic lymphocytes, thymic lymphocytes failed to adoptively transfer immunological capacity to immunodeficient animals. Antibody-forming plasma cells and germinal centers, so prominent in spleen and lymph nodes, never appeared in the thymus of normal immunized animals. Furthermore, thymectomy, which had always been performed in the adult, had never been associated with immune defects [for review, see (*3*)]. All of these findings were cited as arguments against an immune function for the thymus, which was considered by many to be a vestigial structure, perhaps a "graveyard" for dying lymphocytes. As late as 1963, Sir Peter Medawar stated: "We shall come to regard the presence of lymphocytes in the thymus as an evolutionary accident of no very great significance. Lymphocytes are found in other evolutionary relics, such as the palatine and

[1]The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia. [2]Department of Medical Biology, The University of Melbourne, Parkville, Victoria 3010, Australia.
*Corresponding author. Email: miller@wehi.edu.au

pharyngeal tonsils, which arise, like the thymus, from what used to be the pharyngeal epithelium, and it is the common lot of most of these relic organs to be infiltrated with lymphocytes" (*12*).

The thymus: vestigial no more

In the late 1950s and early 1960s, I was working on a lymphocytic leukemia induced in a low-leukemic strain of mice (e.g., C3H) by inoculation of the Gross virus (given as a filtered extract of neoplastic tissues from high-leukemic AKR strain mice). Leukemia developed but only if the virus was inoculated in newborn C3H mice, not in adults. Thymectomy of neonatally virus-inoculated mice at 8 to 10 weeks of age prevented leukemogenesis (*13*) and thymus grafting and, as late as 6 months after adult thymectomy, was able to restore leukemogenic potential (*14*) (Fig. 1). It was thought that perhaps the virus might have persisted somewhere, and it could indeed be recovered from the nonleukemic brain and spleen tissues of adult thymectomized mice that had been inoculated with virus at birth (*15*). Could it multiply only in neonatal thymus tissue and not elsewhere? One way to answer this question was to give the virus to mice immediately after neonatal thymectomy (NTx) and then check for neoplastic development after the grafting of neonatal thymus tissue at various ages.

The NTx mice fared well until some weeks after weaning, when many lost weight and died. This happened regardless of whether they had been inoculated with virus. Histological examination revealed liver lesions, which were probably due to infection by a mouse hepatitis virus, and marked deficiency of lymphocytes in spleen, lymph nodes, and blood, but only in the NTx mice, not in the healthy, sham-operated mice that were kept in the same cages (*16*, *17*). Such findings were totally unexpected because they had never been recorded after adult thymectomy, and this suggested that "the thymus at birth may be essential to life" (*18*). Because circulating lymphocytes, but not thymic lymphocytes, had been shown by Gowans in the late 1950s and early 1960s (*1*) to be immunologically competent cells, it was evident that NTx mice should be checked for immune deficiencies.

Discovering the role of the thymus in immune function

NTx mice were next challenged with *Salmonella typhi* H agglutinins or sheep erythrocytes before they showed signs of illness. They failed to respond like normal mice (*17*, *19*). They were grafted with skin from different mouse strains and from rats, yet the NTx mice failed to reject skin grafts from H-2–disparate mice and even from rats, and the controls responded like normal mice (*16*, *17*, *20*) (Fig. 2). The neonatal thymus restored the ability of the NTx mice to reject foreign skin. When a thymus from a foreign mouse strain was grafted to NTx mice, they rejected the allogeneic skin grafts but not those derived from the donor of the thymus graft (Fig. 3). This observation prompted my suggestion that "when one is inducing a state of immunological tolerance in a newly-born animal, one is performing a thymectomy, not a complete thymectomy, but a partial, selective or immunological thymectomy… antigenic material might make contact with

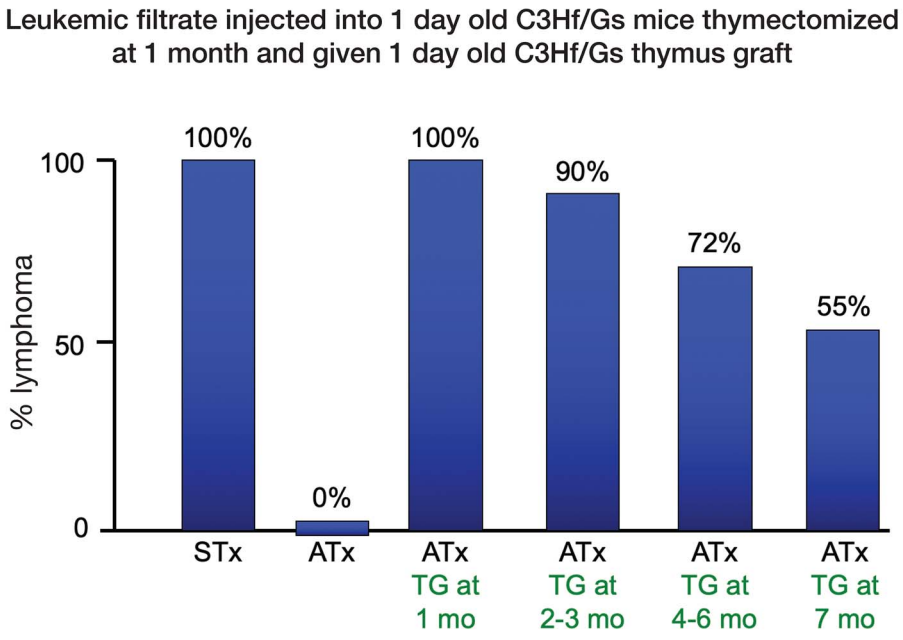


Fig. 1. Lymphoma incidence in leukemic virus–injected mice. Shown is the incidence of lymphomas in C3H mice injected intraperitoneally at birth with leukemic filtrate, thymectomized at 1 month of age, and grafted thereafter with neonatal C3H thymus. STx, sham thymectomized; ATx, adult thymectomized; TG, thymus graft. *n* = 10 to 20 mice per group. [Data are from reference (*14*)]

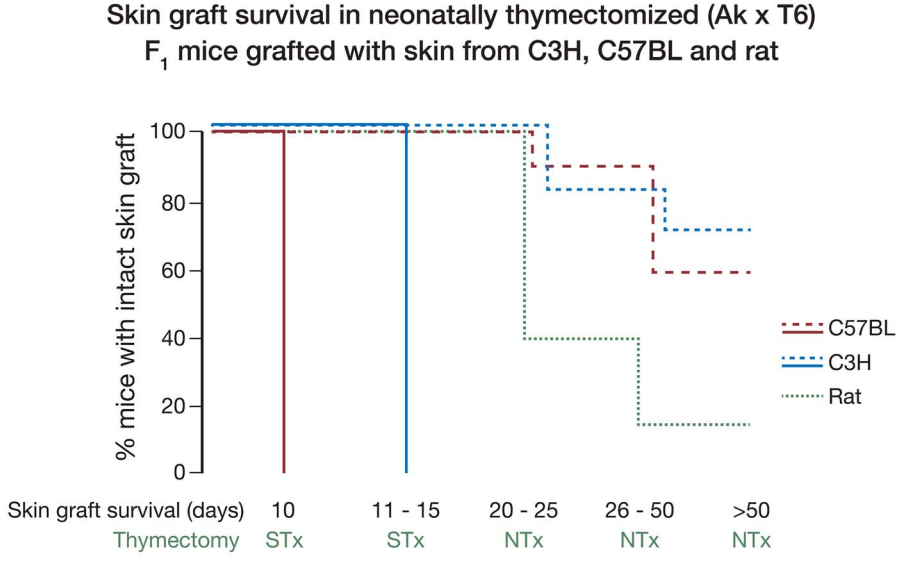


Fig. 2. Allogeneic skin graft survival in neonatally thymectomized mice. Shown is skin graft survival in neonatally thymectomized (AkXT6)F1 mice grafted at 4 to 5 weeks of age with skin from C3H mice, C57BL/6 mice, and rats. [Data are from references (*16*) and (*17*)]

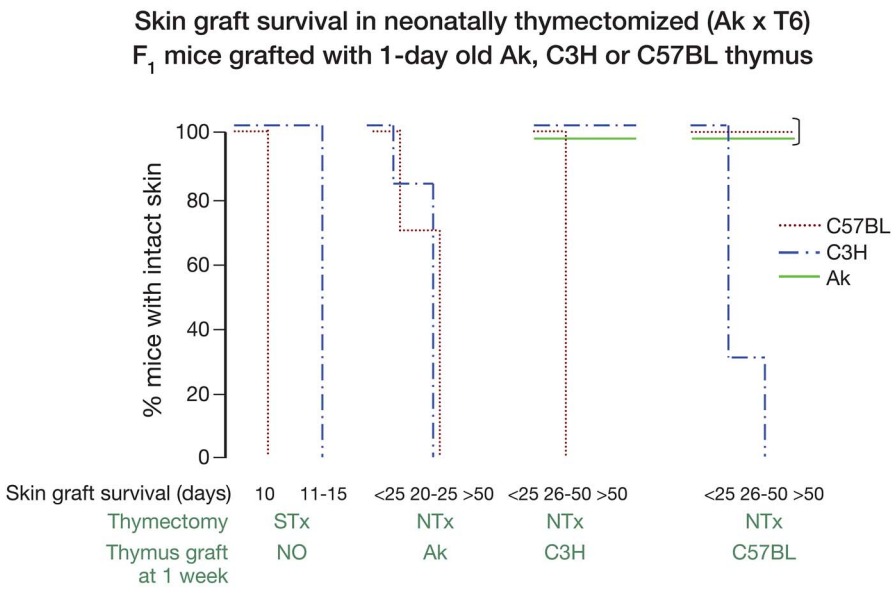


Fig. 3. Allogeneic skin graft survival in neonatally thymectomized mice grafted with thymus tissue. Shown is skin graft survival in 4- to 5-week-old neonatally thymectomized (AkXT6)F1 mice grafted at 1 week of age with thymus from 1-day-old Ak, C3H, or C57BL mice. *n* = 3 to 7 mice per group. [Data are from references (*16*) and (*17*)]

certain cell types differentiating in the thymus and in some way preventing these cells from maturing to a stage where they would be capable of reacting immunologically" (*17*). Thus, tolerance might be learned in the thymus and, by implication, self-tolerance must result from negative selection of self-reacting lymphocytes during thymus lymphopoiesis.

How could these results be reconciled with the well-documented fact that adult thymectomy had no discernible effects on the health of animals or humans [for review, see (*3*)]? Because the neonatal mouse has only few lymphocytes outside the thymus, lymphocytes might have gradually matured intrathymically and eventually left to join and build up the pool of recirculating immunocompetent lymphocytes outside the thymus. If that were the case, then two experiments needed to be done, one to show that thymus lymphocytes do leave the thymus and the other to test the immune capacity of adult thymectomized mice after eradication of the lymphocyte population by irradiation.

In the early 1960s, no fluorescence cell sorter was available and there were no markers that distinguished thymus lymphocytes from other cells. Histocompatibility differences between different mouse strains (e.g., CBA and C57BL/6) could be used, as well as a strain that had a chromosome marker (T6 mice). To determine whether lymphocytes could leave the thymus, NTx (AkXT6)F1 mice were grafted at 7 days of age with Ak or C3H neonatal thymus. After immunization with foreign skin 2 to 4 months later, cytological analysis of spleen cells showed that up to 15% of spleen cells in metaphase lacked the T6 marker and were thus derived from the thymus graft (*17*). Later, when the Thy-1/CD90 cell surface marker became available, the presence of thymus-derived cells in the thoracic duct lymph was demonstrated, proving that the thymus does indeed export cells, now known to be T lymphocytes, to the rest of the lymphoid system (*21*).

To test whether the adult thymus might still have some function, mice were thymectomized at 8 weeks of age and subjected to total body irradiation 2 weeks later. If given high doses of irradiation, they were injected with syngeneic bone marrow as a source of stem cells, which have been shown to repopulate both the myeloid and lymphoid systems (*22*). Euthymic (nonthymectomized) control animals were treated in the same way. The restoration of thymus function in irradiated euthymic mice after bone marrow stem cell influx into the thymus epithelial framework [known to be radioresistant (*23*)] was found to be completed within 4 to 6 weeks after irradiation. These mice rejected allogeneic skin, but 70 to 77% of the thymectomized irradiated mice failed to do so (*24*–*26*). The adult thymus can therefore still function to replace peripheral lymphocytes after they have been depleted. Furthermore, when unirradiated adult mice were thymectomized or sham operated at various ages and challenged with sheep erythrocytes, differences in responsiveness were not found between thymectomized and control mice at 4 months but were significant at 9, 18, and 24 months (*27*).

There is unquestionable evidence that neoantigens may appear during the development of some tumors. This was shown in both virus- and chemical carcinogen–induced tumors in mice (*28*). Experiments were therefore performed to test whether NTx mice might be more susceptible than controls to the carcinogenic activity of 3,4-benzopyrene or 20-methylcholantrene and to polyoma virus. The chemical 3,4-benzopyrene was painted on the shaved backs of young adult mice that had been thymectomized or sham operated. Papillomas occurred in both sets of mice but reached a larger area in the NTx mice (Fig. 4). Most importantly, by 180 days, ~12% of skin tumors in the NTx mice became malignant, compared with only 4% in the STx mice. This led to the conclusion that "Interference with the cellular immune mechanism may be necessary, in some case[*sic*], to allow the full expression of a carcinogenic process" (*29*). In other experiments, 20-methyl-cholantrene was injected intramuscularly in NTx C57BL mice and sham-operated controls at 5 weeks of age. Ten to 15 weeks later, all mice except one sham-operated control developed sarcomata. The latent period of tumor development was shorter in NTx mice and at 14 weeks after injection, tumor size was significantly greater in the thymectomized group (*30*).

The polyoma virus induces multiple neoplasms in most strains of mice, provided they are inoculated at or soon after birth. C57BL/6 mice are not susceptible to the oncogenic activity of this virus. Neonatal thymectomy abolished strain resistance and extended the period of susceptibility, and these mice remained susceptible even when injected at 30 days of age (*31*).

At about the same time, many researchers were pondering the role of the thymus in immunity, and some produced data showing varying degrees of immune defects in animals thymectomized at or soon after birth. These studies have been reviewed in great detail elsewhere [please see (*3*) for details]. For example, Fichtelius *et al.* (*32*) thymectomized and sham-thymectomized young adult guinea pigs after an intravenous injection of *S. typhi* H antigen and reported somewhat lower antibody titers in the thymectomized group. There was no difference between the two groups after a secondary challenge with the same antigen.

Other investigators used neonatally thymectomized mice and rats. One group showed that NTx mice did reject foreign skin grafts from H-2–incompatible strains of mice, though not from donors differing at other weaker histocompatibility gene loci (*33*). Such a discrepancy between their results and mine was later explained as follows: "Careful autopsies performed in the thymectomized animal often revealed minute amounts of residual thymic tissue in these animals. With perfection of our

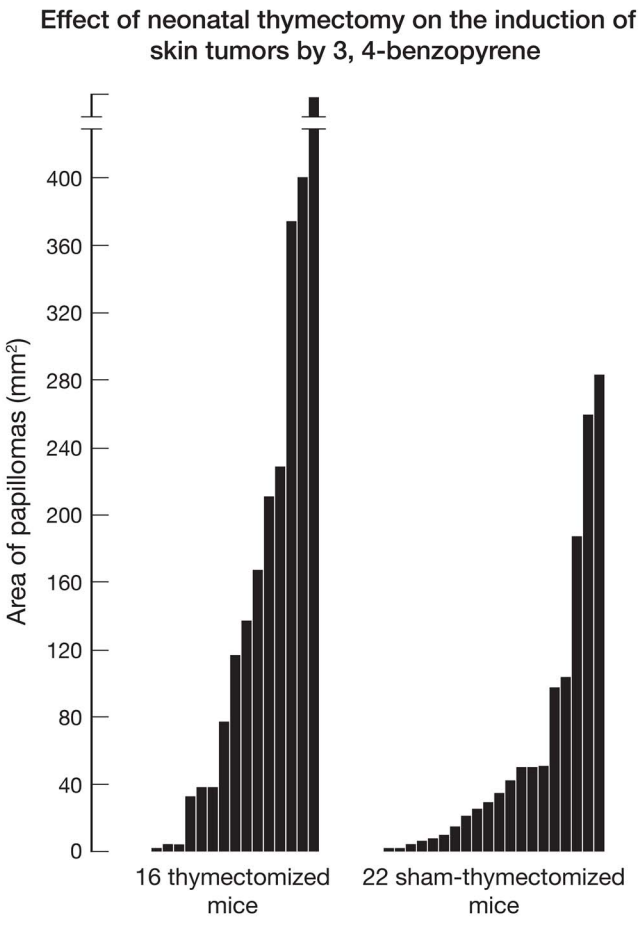


Fig. 4. Induction of tumors in carcinogen-treated neonatally thymectomized mice. Shown is the effect of neonatal thymectomy on the induction of skin tumors 160 days after first application of the carcinogen 3-4-benzopyrene on the skin of 5-week-old mice. [Data are from reference (*29*)]

technique a large proportion of neonatally thymectomized mice accepted H-2 incompatible grafts in contrast to partially thymectomized mice" (*34*). In their follow-up studies, this group provided data showing that their NTx mice failed to reject H-2–disparate skin grafts (*35*). Other researchers using thymectomized Sprague-Dawley rats showed a reduction in the delayed hypersensitivity response to bovine serum albumin and noted that "thymectomy appeared to delay the onset and slow the tempo of rejection of skin homografts from Sherman strain rats" when checked 10 days after grafting (*36*).

How did the immunologic community react to all these findings when most believed that the thymus had become redundant during the course of evolution? They could not fault the data, but they could question the interpretation. The most reasonable criticism was that mice bred in converted horse stables (not specific-pathogen free) must have been exposed to so many intercurrent infections that the additional trauma of neonatal thymectomy or adult thymectomy followed by irradiation precipitated immune deficiency. This prompted a repeat of certain experiments in germ-free tanks (available at that time only at the National Institutes of Health in Bethesda, Maryland). Germ-free C57BL/6 mice were thymectomized or sham operated soon after birth and grafted with H-2–disparate BALB/c skin. None of the mice became sick and none of the thymectomized mice rejected the skin (*37*).

After the publication of this germ-free work, the work from Good's (*35*) and Waksman's (*36*) laboratories, and work on the athymic nude mouse strain in 1970 (*38*), most in the immunological community accepted the notion of thymus immune function.

Identification of T and B cells

Clues about the existence of two separate subsets of lymphocytes became evident in experiments performed in the 1960s in Australia and in the United States. In 1962, Noel Warner and Alex Szenberg in Frank MacFarlane Burnet's laboratory in Melbourne inoculated chickens in ovo with testosterone to impair bursa development. In most such chickens, antibody production and delayed hypersensitivity were reduced but foreign skin was rejected. A few sick birds, in which lymphoid atrophy had extended to the thymus, failed to reject skin allografts (*39*). These results suggested the existence of two separate lymphocyte groups, one coming from the bursa and responsible for antibody production, as had been shown in 1956 (*40*), and the other from the thymus and involved in cell-mediated immunity. Because mammals do not have a bursa, Burnet surmised that "in mammals it is highly probable that the thymus also carries out the function performed by the bursa of Fabricius in the chicken" (*41*).

Max Cooper, then working in Robert Good's laboratory (*42*), surgically thymectomized or bursectomized chickens and used total body irradiation to remove any immune cells that may have arisen before hatching. The results were clear: The bursectomized chickens failed to produce antibodies, whereas the thymectomized birds could do so but were deficient in allograft rejection. These results implied but did not conclusively identify two distinct lymphocyte populations. Claman and colleagues in Denver showed that irradiated mice receiving either bone marrow cells or thymus cells alone did not produce a substantial antisheep erythrocyte antibody response, but mice receiving both sets of syngeneic cells together did produce some antibody. Because no markers were available, the origin of the antibody-forming cells (AFCs), whether from the thymus or the bone marrow, could not be identified (*43*).

Quantitative studies on the recirculating lymphocyte pool of NTx and sham-operated mice were performed by our group at the Walter and Eliza Hall Institute of Medical Research in Melbourne in 1966–1967. This was done by cannulating the mouse thoracic duct for a period of 48 hours, after which no further lymphocytes could be obtained. The cumulative total number of such cells in 6-week-old NTx and sham-operated control mice differed markedly: close to 10^8 cells in controls but slightly more than 10^6 cells in NTx mice (*44*). Injection of thymus or thoracic duct lymphocytes from normal (CBAXC57BL)F1 mice into NTx CBA mice challenged with sheep erythrocytes (sheep red blood cells, or SRBCs) enabled these to produce a normal anti-SRBC AFC response. Anti-CBA serum, which is directed against both the donor and recipient cell types, reduced the number of AFCs by 89 to 97%, as expected, but unexpectedly, anti-C57BL/6 serum (directed against the donor cell type) caused an insignificant reduction (amounting to 0 to 17% in experiments repeated 11 times) (6% in Fig. 5A). These findings showed that the AFC precursors arose from the immunoincompetent NTx host, not from the injected thoracic duct cells (*45*).

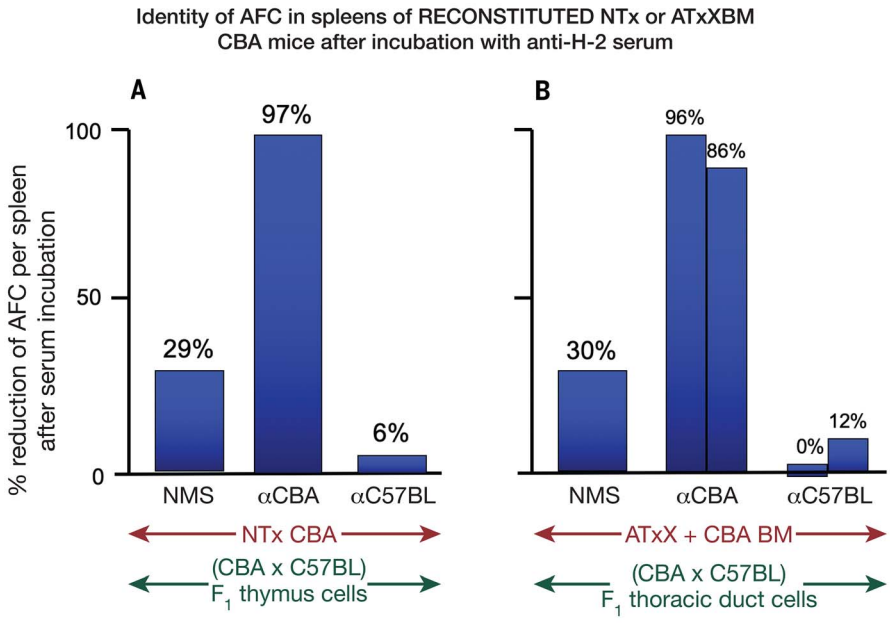


Fig. 5. Identification of T and B cells. Shown is anti–H-2 serum identification of AFCs in NTx CBA mice (**A**) or adult thymectomized, irradiated, and bone marrow–protected CBA mice (**B**) reconstituted with thymus cells or thoracic duct lymphocytes. NMS, normal mouse serum; α, anti. Each group consists of spleens from three to six mice. [Data are from references (*45*) and (*46*)]

To determine the identity of the AFC, adult CBA mice were thymectomized, subjected to a heavy dose of total body irradiation, and protected with CBA bone marrow. After recovery from irradiation, their immune function was restored by an intravenous injection of normal (CBAXC57BL)F1 thoracic duct lymphocytes (of which 80 to 90% were found to be T cells). They were then challenged with SRBCs and their spleen assayed for the number of AFCs against SRBCs. As expected, a normal AFC response was found. Now the time was ripe to determine whether the AFCs were derived from T cells in thoracic duct lymph. Anti-CBA serum reduced the number of AFCs by 86 to 96%, as expected, but anti-C57BL/6 serum reduced it by only 0 to 12% (*46*) (Fig. 5B).

This work showed beyond a doubt and for the first time that AFCs in mice, and presumably other mammals, are derived not from T cells but from the bone marrow, and they were now called B cells. It further showed that B cells in many cases need help from T cells to produce a normal AFC response. This provided a cellular basis for the two arms of the adaptive immune defense system, and it suggested that the bone marrow served as a bursa equivalent in mammalian species, as was eventually found to be the case (*47*). It also explained the existence of separate thymus-dependent and thymus-independent areas in the lymphoid tissues (*48*).

The existence of two distinct lymphocyte subsets was initially regarded with skepticism by many. Burnet questioned "the significance of results obtained in such biological monstrosities as pure line mice thymectomized, lethally-irradiated, and salvaged by injection of bone marrow from another mouse" (*49*). Good claimed to "have evidence that in the rabbit, it [the bursa equivalent] resides in the ilial lymphoid tissue and in the lymphoid tissue of the appendix." He was also "concerned at separating thymus-derived from marrow-derived cells" because the former "are in fact marrow-derived cells" (*50*) despite the fact that, as stated above, bone marrow was known to be a source of stem cells for both the myeloid and lymphoid systems (*22*).

Gowans, who had proven that recirculating lymphocytes in the rat were capable of both humoral and cellular immune responses (*1*) and believed that the same cell could produce both, stated: "Had it not been for Dr Miller's experiments, I would have assumed that a single variety of small lymphocyte was involved in each of our experiments…If we have two cell types that are collaborating, then we have specificity residing in two cell lines, one thymus derived, the other marrow derived. The problem is to bring these two cell lines together. Does this necessity for the two cells to find each other raise problems? It seems an inefficient mechanism if it rests only on chance contacts" (*51*).

The culmination of these immunological findings made during the 1960s is summarized in Fig. 6. Hemopoietic stem cells, first arising in the yolk sac and then in the fetal liver and adult bone marrow, migrate through the bloodstream to various myeloid and lymphoid organs, where the differentiation to mature cells is dictated by the microenvironment of the tissue in which they lodge. Lymphoid stem cells in the thymus undergo differentiation to eventually become immunocompetent T lymphocytes that leave the thymus, recirculate through the blood into the "thymus-dependent" areas of the lymphoid tissues, and exit into the lymph to return to the blood through the thoracic duct. They are responsible for cellular immunity. B cells are generated in the bone marrow in mammals, or in the bursa of Fabricius in birds, and eventually migrate through the bloodstream to "thymus-independent" areas of the lymphoid tissues. Some also recirculate like T cells. These cells take part in humoral immunity (*52*).

Impetus for later research

The discovery of thymus function has had a tremendous impact on further immunological research. Distinct thymus epithelial cells were described in detail (*53*), thymic lymphoid stem cells were characterized (*54*), and major events in thymus T lymphocyte differentiation were mapped (*55*). How the thymus induces self-tolerance was shown to occur within the thymus through negative selection (*56*) and through the activity of the transcription factor AIRE (*57*). Both T and B cells are subjected to several checkpoints during their differentiation from precursor stem cells to ensure that useless cells or cells with self-reacting receptors are deleted. Different apoptotic mechanisms purge these cells (*58*), and failure of apoptosis leads to autoimmunity (*59*). T cells escaping thymus censorship could be made tolerant in the peripheral lymphoid tissues (*60*, *61*) and deleted by a Bcl-2–inhibitable pathway (*62*). Many T cell subsets in the thymus and in the periphery were identified and their activities determined, notably CD4 helper cells, CD8 cytotoxic cells (*63*), and CD4 regulatory T cells inducing suppression of inflammation and of other immune responses (*64*, *65*). Cytokines or lymphokines generated within the thymus and produced by various T cell subsets were shown to exert substantial biological effects in immune responses (*66*). Cells essential for antigen presentation to T cells were discovered (*67*), and it was shown that T cells required two signals for activation (*68*). T cells were found to be essential to allow B cells to switch from immunoglobulin M (IgM) to IgG antibody (*69*, *70*). It became evident that T cells, unlike B cells, do not perceive unprocessed antigenic determinants but rather short peptide fragments (*71*) in association with major histocompatibility complex (MHC) molecules (*72*). This MHC restriction was shown to reflect an intrathymic positive selection (*73*)

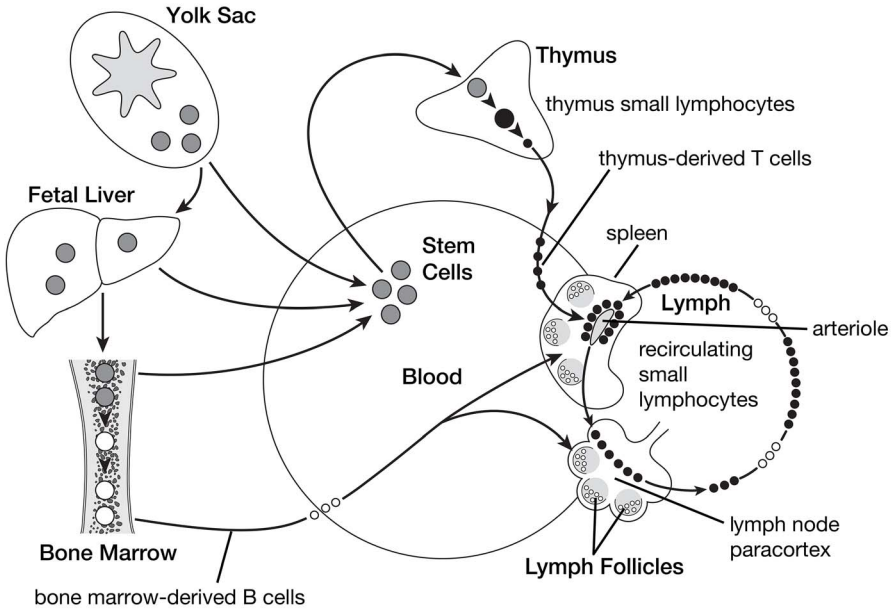


Fig. 6. Origin and migration of T and B cells. Hemopoietic stem cells originating in the yolk sac, fetal liver, or bone marrow migrate through blood to the thymus, where they differentiate into T lymphocytes. Others differentiate within the bone marrow in mammals, or the bursa of Fabricius in birds, to produce B cells. Mature T and B cells circulate in the blood and lymph and colonize the T and B cell compartments of the lymphoid tissues. [Figure reprinted with permission from the Royal Society, London, from reference (*52*)]

of lymphocytes capable of interacting with self-peptide–MHC complexes (*74*). The T cell receptor for antigen was not immunoglobulin, as existed on the B cell membrane, but a distinct molecule able to bind both MHC and processed antigen (*75*, *76*). The genes responsible for generating diversity in B and T cell antigen-specific receptors were mapped and recombination events determined (*77*, *78*).

We now have considerable knowledge about the development of lymphocytes in the thymus, but there is still a lot to be learned. For example, how does the medullary transcription factor AIRE select which set of self-antigens are to be expressed in the thymus? In addition, little is known about the intricate structure of the thymic epithelium in both the cortex and medulla. We now have new technologies to explore this, including two-photon laser-scanning microscopy, which can provide large depth penetration, up to hundreds of micrometers, in a mouse thymus transplanted under the kidney capsule. Two-photon excitation microscopy will also give us a deeper knowledge of thymocyte selection and interplay with both epithelial and dendritic cells and thymocyte population trafficking intrathymically and extrathymically.

Another question relating to the thymus that has yet to be answered satisfactorily is why does it involute at a relatively early age? In humans, the thymus atrophies from infancy, resulting in an exponential decline in T cell production with a half-life of ~16 years. Does the age-related decline in T cell output account for the rising incidence of many infectious diseases and cancer with age, as has been postulated (*4*)?

Relevance to clinical medicine

The identification of two distinct major lymphocyte subsets, T and B cells, necessitated a reevaluation of numerous immunological phenomena and diseases in terms of the roles played by each subset. These roles include the carrier effect (priming to one part of the antigen molecule, the carrier, enhances the antibody response to another, smaller part, the hapten, which on its own cannot elicit a response), immunological memory, immunological tolerance, original antigenic sin (B cells respond faster to antigens from a previous encounter than to a second encounter with antigens of a slightly different version), allergies, inflammatory conditions, dysbiosis (microbial imbalance inside the body, such as impaired microbiota), tissue and organ transplantation, tissue repair, preeclampsia, graft-versus-host reactions, vaccination procedures, infectious diseases, genetically determined unresponsive states, immunodeficiency, autoimmunity, and cancer.

Major parallels between immunodeficiencies in human patients and the results of bursectomy or thymectomy in chickens have been highlighted. Patients with X-linked agammaglobulinemia controlled virus infections but could not make antibodies, whereas those with the X-linked Wiskott-Aldrich syndrome developed herpes infections but could produce antibodies (*79*).

Antigen-specific treatment of autoimmune diseases and allergies to activate regulatory T cells and induce tolerance, or to cause apoptosis of effector T cells, is likely to be accomplished once the specific antigen that triggers these conditions has been isolated.

Influenza vaccination generally relies on products that use inactivated virus, isolated viral envelope hemagglutinin, or neuraminidase proteins. These activate B cells to elicit strain-specific antibodies but are only seasonally effective. If CD8 T cells could be stimulated by relatively conserved peptides from the virus internal elements using a live attenuated virus that can infect dendritic cells and cover major human leukocyte antigen types, then a long-term protection not needing annual vaccination may be achieved (*80*).

There is no doubt that researching the microbiome will have a far-reaching impact on our understanding of immune-mediated conditions such as inflammatory bowel dysfunction and allergies. Dysbiosis has in fact been linked to chronic inflammation and cancer development. Intraepithelial γδ T cells lie in close proximity to the microbiota and comprise up to 50% of the $CD3^+$ T cells in the intestine of mice. These and other "unconventional" T lymphocytes deserve further intensive studies (*81*).

Cancer immunotherapy has made good use of both T and B cells. Monoclonal antibodies (*82*) such as herceptin (*83*) have been used successfully in HER2-positive breast cancers, and chimeric antigen receptor (CAR) T cells (*84*) have shown great promise, particularly for B cell lymphomas. However, loss of the target molecule may occur when tumor cells mutate, rendering these measures ineffective. A broader approach, such as one using checkpoint inhibitors (e.g., CTLA-4, PD-1, or LAG-3) to rev up the immune system (*85*–*87*), perhaps together with deletion of CARMAI1 (*88*), may achieve better results. The future is indeed bright for immunology and immunotherapy.

Conclusions

The thymus and T cells have featured prominently in a great number of immunology studies in the past 50 years. T cells are involved essentially across the entire spectrum of tissue physiology and pathology. Manipulation to enhance, decrease, or suppress the activity of different T cell subsets has already had, and will undoubtedly continue to have, beneficial effects in inflammatory diseases, diseases of immunological aberrations, vaccine production, and malignancies. Perhaps a quantitative framework using mathematics modeling to better understand T cell behavior (*89*) will help to determine or improve therapeutic strategies.

REFERENCES AND NOTES

1. J. L. Gowans, D. D. McGregor, D. M. Cowen, Initiation of immune responses by small lymphocytes. *Nature* **196**, 651–655 (1962). doi: 10.1038/196651a0; pmid: 13949634
2. P. B. Medawar, The Croonian Lecture 1958. The homograft reaction. *Proc. R. Soc. London Ser. B* **149**, 145–166 (1958). doi: 10.1098/rspb.1958.0058
3. J. F. Miller, D. Osoba, Current concepts of the immunological function of the thymus. *Physiol. Rev.* **47**, 437–520 (1967). doi: 10.1152/physrev.1967.47.3.437; pmid: 4864541
4. S. Palmer, L. Albergante, C. C. Blackburn, T. J. Newman, Thymic involution and rising disease incidence with age. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1883–1888 (2018). doi: 10.1073/pnas.1714478115; pmid: 29432166
5. E. Andreasen, S. Christensen, The rate of mitotic activity in the lymphoid organs of the rat. *Anat. Rec.* **103**, 401–412 (1949). doi: 10.1002/ar.1091030307; pmid: 18144438
6. K. Nakamura, D. Metcalf, Quantitative cytological studies on thymic lymphoid cells in normal, preleukaemic and leukaemic mice. *Br. J. Cancer* **15**, 306–315 (1961). doi: 10.1038/bjc.1961.38; pmid: 13727612
7. D. Metcalf, "The thymus and lymphopoiesis," in *The Thymus in Immunobiology*, R. A. Good, A. E. Gabrielsen, Eds. (Harper & Row, 1964); pp. 150–179.
8. J. F. Miller, Immunity in the foetus and the new-born. *Br. Med. Bull.* **22**, 21–26 (1966). doi: 10.1093/oxfordjournals.bmb.a070431; pmid: 5321810
9. J. C. Kennedy, L. Siminovitch, J. E. Till, E. A. McCulloch, A transplantation assay for mouse cells responsive to antigenic stimulation by sheep erythrocytes. *Proc. Soc. Exp. Biol. Med.* **120**, 868–873 (1965). doi: 10.3181/00379727-120-30678; pmid: 5858721
10. O. Vos, M. J. De Vries, J. C. Collenteur, D. W. Van Bekkum, Transplantation of homologous and heterologous lymphoid cells in x-irradiated and non-irradiated mice. *J. Natl. Cancer Inst.* **23**, 53–73 (1959). pmid: 13673270
11. H. W. Brayton, A. C. Heublein, Enlarged thymus gland in infancy and its treatment by radium. *Boston Med. Surg. J.* **181**, 740–743 (1919). doi: 10.1056/NEJM191912251812604
12. P. B. Medawar, in *The Immunologically Competent Cell: Its Nature and Origin*, G. E. W. Wolstenholme, J. Knight, Eds. (Churchill, 1963); pp. 6–16.
13. J. F. Miller, Role of the thymus in murine leukaemia. *Nature* **183**, 1069 (1959). doi: 10.1038/1831069a0; pmid: 13644292
14. J. F. Miller, Fate of subcutaneous thymus grafts in thymectomized mice inoculated with leukaemic filtrate. *Nature* **184** (Suppl 23), 1809–1810 (1959). doi: 10.1038/1841809a0; pmid: 14422530
15. J. F. Miller, Recovery of leukaemogenic agent from nonleukaemic tissues of thymectomized mice. *Nature* **187**, 703 (1960). doi: 10.1038/187703a0; pmid: 14422531
16. J. F. Miller, Immunological function of the thymus. *Lancet* **2**, 748–749 (1961). doi: 10.1016/S0140-6736(61)90693-6; pmid: 14474038
17. J. F. Miller, Effect of neonatal thymectomy on the immunological responsiveness of the mouse. *Proc. R. Soc. London Ser. B* **156**, 415–428 (1962).
18. J. F. Miller, Analysis of the thymus influence in leukaemogenesis. *Nature* **191**, 248–249 (1961). doi: 10.1038/191248a0; pmid: 13770811
19. J. F. Miller, P. M. De Burgh, G. A. Grant, Thymus and the production of antibody-plaque-forming cells. *Nature* **208**, 1332–1334 (1965). doi: 10.1038/2081332a0; pmid: 5870196
20. J. F. Miller, Role of the thymus in transplantation immunity. *Ann. N. Y. Acad. Sci.* **99**, 340–354 (1962). doi: 10.1111/j.1749-6632.1962.tb45319.x
21. J. F. Miller, J. Sprent, Thymus-derived cells in mouse thoracic duct lymph. *Nat. New Biol.* **230**, 267–270 (1971). doi: 10.1038/newbio230267a0; pmid: 5280665
22. A. S. Moore, J. J. T. Owen, Stem cell migration in developing myeloid and lymphoid system. *Lancet* **290**, 658–659 (1967). doi: 10.1016/S0140-6736(67)90693-9
23. L. Hofman, V. Stankovic, N. Allegretti, The effect of total-body x-irradiation on the thymus and the number of its cells. *Radiat. Res.* **15**, 30–38 (1961). doi: 10.2307/3571064; pmid: 13715086
24. J. F. Miller, Immunological significance of the thymus of the adult mouse. *Nature* **195**, 1318–1319 (1962). doi: 10.1038/1951318a0
25. J. F. Miller, S. M. Doak, A. M. Cross, Role of the thymus in the recovery of the immune mechanism in the irradiated adult mouse. *Proc. Soc. Exp. Biol. Med.* **112**, 785–792 (1963). doi: 10.3181/00379727-112-28170
26. A. M. Cross, E. Leuchars, J. F. Miller, Studies on the recovery of the immune response in irradiated mice thymectomized in adult life. *J. Exp. Med.* **119**, 837–850 (1964). doi: 10.1084/jem.119.5.837; pmid: 14157033
27. J. F. Miller, Effect of thymectomy in adult mice on immunological responsiveness. *Nature* **208**, 1337–1338 (1965). doi: 10.1038/2081337a0; pmid: 5870199
28. J. F. Miller, "The thymus in relation to neoplasia," in *Modern Trends in Pathology*, T. Crawford, Ed. (Butterworths, ed. 2, (1967), pp. 140–175.
29. J. F. Miller, G. A. Grant, F. J. Roe, Effect of thymectomy on the induction of skin tumours by 3,4-benzopyrene. *Nature* **199**, 920–922 (1963). doi: 10.1038/199920a0; pmid: 14079914
30. G. A. Grant, J. F. Miller, Effect of neonatal thymectomy on the induction of sarcomata in C57 BL mice. *Nature* **205**, 1124–1125 (1965). doi: 10.1038/2051124a0; pmid: 5833220
31. J. F. A. P. Miller, R. C. Ting, L. W. Law, Influence of thymectomy on tumor induction by polyoma virus in C57BL mice. *Proc. Soc. Exp. Biol. Med.* **116**, 323–327 (1964). doi: 10.3181/00379727-116-29237; pmid: 14189129
32. K. E. Fichtelius, G. Laurell, L. Philipsson, The influence of thymectomy on antibody formation. *Acta Pathol. Microbiol. Scand.* **51**, 81–86 (1961). doi: 10.1111/j.1699-0463.1961.tb00346.x; pmid: 13699308
33. C. Martinez, J. Kersey, B. W. Papermaster, R. A. Good, Skin homograft survival in thymectomized mice. *Proc. Soc. Exp. Biol. Med.* **109**, 193–196 (1962). doi: 10.3181/00379727-109-27149; pmid: 14470489
34. C. Martinez, A. P. Dalmasso, R. A. Good, "Homotransplantation of normal and neoplastic tissue in thymectomized mice," in *The Thymus in Immunobiology*, R. A. Good, A. E. Gabrielsen, Eds. (Harper & Row, 1964), pp. 465–477.
35. A. P. Dalmasso, C. Martinez, R. A. Good, Further studies of suppression of the homograft reaction by thymectomy in the mouse. *Proc. Soc. Exp. Biol. Med.* **111**, 143–146 (1962). doi: 10.3181/00379727-111-27729; pmid: 14024761
36. B. G. Arnason, B. D. Jankovic, B. H. Waksman, Effect of thymectomy on 'delayed' hypersensitive reactions. *Nature* **194**, 99–100 (1962). doi: 10.1038/194099a0; pmid: 13862387
37. K. R. McIntire, S. Sell, J. F. Miller, Pathogenesis of the post-neonatal thymectomy wasting syndrome. *Nature* **204**, 151–155 (1964). doi: 10.1038/204151a0; pmid: 14222260
38. E. M. Pantelouris, Observations on the immunobiology of 'nude' mice. *Immunology* **20**, 247–252 (1971). pmid: 4100989
39. N. L. Warner, A. Szenberg, F. M. Burnet, The immunological role of different lymphoid organs in the chicken. I. Dissociation of immunological responsiveness. *Aust. J. Exp. Biol. Med. Sci.* **40**, 373–387 (1962). doi: 10.1038/icb.1962.42; pmid: 13998953
40. B. Glick, T. S. Chang, R. G. Jaap, The bursa of Fabricius and antibody production. *Poult. Sci.* **35**, 224–225 (1956). doi: 10.3382/ps.0350224
41. F. M. Burnet, The thymus gland. *Sci. Am.* **207**, 50–57 (1962). doi: 10.1038/scientificamerican1162-50
42. M. D. Cooper, D. A. Raymond, R. D. A. Peterson, M. A. South, R. A. Good, The functions of the thymus system and the bursa system in the chicken. *J. Exp. Med.* **123**, 75–102 (1966). doi: 10.1084/jem.123.1.75; pmid: 5323079
43. H. N. Claman, E. A. Chaperon, R. F. Triplett, Thymus-marrow cell combinations. Synergism in antibody production. *Proc. Soc. Exp. Biol. Med.* **122**, 1167–1171 (1966). doi: 10.3181/00379727-122-31353; pmid: 5918139
44. J. F. Miller, G. F. Mitchell, N. S. Weiss, Cellular basis of the immunological defects in thymectomized mice. *Nature* **214**, 992–997 (1967). doi: 10.1038/214992a0; pmid: 6055415
45. J. F. Miller, G. F. Mitchell, Cell to cell interaction in the immune response. I. Hemolysin-forming cells in neonatally thymectomized mice reconstituted with thymus or thoracic duct lymphocytes. *J. Exp. Med.* **128**, 801–820 (1968). doi: 10.1084/jem.128.4.801; pmid: 5691985
46. G. F. Mitchell, J. F. Miller, Cell to cell interaction in the immune response. II. The source of hemolysin-forming cells in irradiated mice given bone marrow and thymus or thoracic duct lymphocytes. *J. Exp. Med.* **128**, 821–837 (1968). doi: 10.1084/jem.128.4.821; pmid: 5691986
47. D. G. Osmond, Population dynamics of bone marrow B lymphocytes. *Immunol. Rev.* **93**, 103–124 (1986). doi: 10.1111/j.1600-065X.1986.tb01504.x; pmid: 3491034
48. D. V. Parrott, M. A. De Sousa, J. East, Thymus-dependent areas in the lymphoid organs of neonatally thymectomized mice. *J. Exp. Med.* **123**, 191–204 (1966). doi: 10.1084/jem.123.1.191; pmid: 5902569
49. F. M. Burnet, *Auto-Immunity and Auto-Immune Disease* (Springer, 1972).
50. R. A. Good, "Immunocompetent cell populations," in *Immunological Tolerance: A Reassessment of Mechanisms of the Immune Response: Proceedings of an International Conference Held at Brook Lodge, Augusta, Michigan September 18–20, 1968*, M. Landy, W. Braun, Eds. (Academic, 1969), p. 136.
51. J. L. Gowans, "Immunocompetent cell populations," in *Immunological Tolerance: A Reassessment of Mechanisms of the Immune Response: Proceedings of an International Conference Held at Brook Lodge, Augusta, Michigan September 18–20, 1968*, M. Landy, W. Braun, Eds. (Academic, 1969), pp. 164–169.
52. J. F. A. P. Miller, The Croonian Lecture, 1992. The key role of the thymus in the body's defence strategies. *Philos. Trans. R. Soc. London B Biol. Sci.* **337**, 105–124 (1992). doi: 10.1098/rstb.1992.0087; pmid: 1355916
53. W. van Ewijk, G. Holländer, C. Terhorst, B. Wang, Stepwise development of thymic microenvironments in vivo is regulated by thymocyte subsets. *Development* **127**, 1583–1591 (2000). pmid: 10725235
54. L. Wu, M. Antica, G. R. Johnson, R. Scollay, K. Shortman, Developmental potential of the earliest precursor cells from the adult mouse thymus. *J. Exp. Med.* **174**, 1617–1627 (1991). doi: 10.1084/jem.174.6.1617; pmid: 1683894
55. A. C. Carpenter, R. Bosselut, Decision checkpoints in the thymus. *Nat. Immunol.* **11**, 666–673 (2010). doi: 10.1038/ni.1887; pmid: 20644572
56. J. W. Kappler, N. Roehm, P. Marrack, T cell tolerance by clonal elimination in the thymus. *Cell* **49**, 273–280 (1987). doi: 10.1016/0092-8674(87)90568-X; pmid: 3494522
57. M. S. Anderson *et al.*, Projection of an immunological self shadow within the thymus by the aire protein. *Science* **298**, 1395–1401 (2002). doi: 10.1126/science.1075958; pmid: 12376594
58. S. R. Daley, C. Teh, D. Y. Hu, A. Strasser, D. H. D. Gray, Cell death and thymic tolerance. *Immunol. Rev.* **277**, 9–20 (2017). doi: 10.1111/imr.12532; pmid: 28462532
59. A. Strasser *et al.*, Enforced BCL2 expression in B-lymphoid cells prolongs antibody responses and elicits autoimmune disease. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8661–8665 (1991). doi: 10.1073/pnas.88.19.8661; pmid: 1924327
60. J. F. Miller *et al.*, Induction of peripheral $CD8^+$ T-cell tolerance by cross-presentation of self antigens. *Immunol. Rev.* **165**, 267–277 (1998). doi: 10.1111/j.1600-065X.1998.tb01244.x; pmid: 9850866
61. L. Zheng *et al.*, Induction of apoptosis in mature T cells by tumour necrosis factor. *Nature* **377**, 348–351 (1995). doi: 10.1038/377348a0; pmid: 7566090
62. G. M. Davey *et al.*, Peripheral deletion of autoreactive CD8 T cells by cross presentation of self-antigen occurs by a Bcl-2-inhibitable pathway mediated by Bim. *J. Exp. Med.* **196**, 947–955 (2002). doi: 10.1084/jem.20020827; pmid: 12370256
63. H. Cantor, E. A. Boyse, Functional subclasses of T-lymphocytes bearing different Ly antigens. I. The generation of functionally distinct T-cell subclasses is a differentiative process independent of antigen. *J. Exp. Med.* **141**, 1376–1389 (1975). doi: 10.1084/jem.141.6.1376; pmid: 1092798
64. S. Hori, T. Nomura, S. Sakaguchi, Control of regulatory T cell development by the transcription factor Foxp3. *Science* **299**, 1057–1061 (2003). doi: 10.1126/science.1079490; pmid: 12522256
65. J. D. Fontenot, M. A. Gavin, A. Y. Rudensky, Foxp3 programs the development and function of $CD4^+CD25^+$ regulatory T cells. *Nat. Immunol.* **4**, 330–336 (2003). doi: 10.1038/ni904; pmid: 12612578
66. V. Paetkau, Lymphokines on the move. *Nature* **294**, 689–690 (1981). doi: 10.1038/294689a0; pmid: 7322199
67. R. M. Steinman, Z. A. Cohn, Identification of a novel cell type in peripheral lymphoid organs of mice. I. Morphology, quantitation, tissue distribution. *J. Exp. Med.* **137**, 1142–1162 (1973). doi: 10.1084/jem.137.5.1142; pmid: 4573839
68. K. J. Lafferty, A. J. Cunningham, A new analysis of allogeneic interactions. *Aust. J. Exp. Biol. Med. Sci.* **53**, 27–42 (1975). doi: 10.1038/icb.1975.3; pmid: 238498
69. P. W. Kincade, A. R. Lawton, D. E. Bockman, M. D. Cooper, Suppression of immunoglobulin G synthesis as a result of antibody-mediated suppression of immunoglobulin M synthesis in chickens. *Proc. Natl. Acad. Sci. U.S.A.* **67**, 1918–1925 (1970). doi: 10.1073/pnas.67.4.1918; pmid: 5275387
70. C. Cheers, J. F. A. P. Miller, Cell-to-cell interaction in the immune response. IX. Regulation of hepten-specific antibody class by carrier priming. *J. Exp. Med.* **136**, 1661–1665 (1972). doi: 10.1084/jem.136.6.1661; pmid: 4118416

71. A. R. M. Townsend *et al*., The epitopes of influenza nucleoprotein recognized by cytotoxic T lymphocytes can be defined with short synthetic peptides. *Cell* **44**, 959–968 (1986). doi: 10.1016/0092-8674(86)90019-X; pmid: 2420472
72. R. M. Zinkernagel, P. C. Doherty, Immunological surveillance against altered self components by sensitised T lymphocytes in lymphocytes choriomeningitis. *Nature* **251**, 547–548 (1974). doi: 10.1038/251547a0; pmid: 4547543
73. P. Kisielow, H. S. Teh, H. Blüthmann, H. von Boehmer, Positive selection of antigen-specific T cells in thymus by restricting MHC molecules. *Nature* **335**, 730–733 (1988). doi: 10.1038/335730a0; pmid: 3262831
74. R. M. Zinkernagel *et al*., On the thymus in the differentiation of "H-2 self-recognition" by T cells: Evidence for dual recognition? *J. Exp. Med.* **147**, 882–896 (1978). doi: 10.1084/jem.147.3.882; pmid: 305459
75. S. M. Hedrick, D. I. Cohen, E. A. Nielsen, M. M. Davis, Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* **308**, 149–153 (1984). doi: 10.1038/308149a0; pmid: 6199676
76. Y. Yoshikai *et al*., Sequence and expression of transcripts of the human T-cell receptor β-chain genes. *Nature* **312**, 521–524 (1984). doi: 10.1038/312521a0; pmid: 6334237
77. N. Hozumi, S. Tonegawa, Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3628–3632 (1976). doi: 10.1073/pnas.73.10.3628; pmid: 824647
78. G. Siu *et al*., The human T cell antigen receptor is encoded by variable, diversity, and joining gene segments that rearrange to generate a complete V gene. *Cell* **37**, 393–401 (1984). doi: 10.1016/0092-8674(84)90369-6; pmid: 6202421
79. M. D. Cooper, H. P. Chae, J. T. Lowman, W. Krivit, R. A. Good, Wiskott-Aldrich syndrome: An immunologic deficiency disease involving the afferent limb of immunity. *Am. J. Med.* **44**, 499–513 (1968). doi: 10.1016/0002-9343(68)90051-X; pmid: 4171085
80. S. Gras *et al*., Cross-reactive $CD8^+$ T-cell immunity between the pandemic H1N1-2009 and H1N1-1918 influenza A viruses. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12599–12604 (2010). doi: 10.1073/pnas.1007270107; pmid: 20616031
81. D. I. Godfrey, A. P. Uldrich, J. McCluskey, J. Rossjohn, D. B. Moody, The burgeoning family of unconventional T cells. *Nat. Immunol.* **16**, 1114–1123 (2015). doi: 10.1038/ni.3298; pmid: 26482978
82. G. Köhler, C. Milstein, Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495–497 (1975). doi: 10.1038/256495a0; pmid: 1172191
83. D. J. Slamon *et al*., Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* **344**, 783–792 (2001). doi: 10.1056/NEJM200103153441101; pmid: 11248153
84. J. F. Miller, M. Sadelain, The journey from discoveries in fundamental immunology to cancer immunotherapy. *Cancer Cell* **27**, 439–449 (2015). doi: 10.1016/j.ccell.2015.03.007; pmid: 25858803
85. S. C. Wei *et al*., Distinct cellular mechanisms underlie anti-CTLA-4 and anti-PD-1 checkpoint blockade. *Cell* **170**, 1120–1133.e17 (2017). doi: 10.1016/j.cell.2017.07.024; pmid: 28803728
86. Y. Ishida, Y. Agata, K. Shibahara, T. Honjo, Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. *EMBO J.* **11**, 3887–3895 (1992). doi: 10.1002/j.1460-2075.1992.tb05481.x; pmid: 1396582
87. F. Triebel *et al*., LAG-3, a novel lymphocyte activation gene closely related to CD4. *J. Exp. Med.* **171**, 1393–1405 (1990). doi: 10.1084/jem.171.5.1393; pmid: 1692078
88. M. Di Pilato *et al*., Targeting the CBM complex causes T_{reg} cells to prime tumours for immune checkpoint therapy. *Nature* **570**, 112–116 (2019). doi: 10.1038/s41586-019-1215-2; pmid: 31092922
89. J. M. Marchingo *et al*., Antigen affinity, costimulation, and cytokine inputs sum linearly to amplify T cell expansion. *Science* **346**, 1123–1127 (2014). doi: 10.1126/science.1260044; pmid: 25430770

10.1126/science.aba2429

RESEARCH ARTICLE SUMMARY

STRUCTURAL BIOLOGY

Structural insights into differences in G protein activation by family A and family B GPCRs

Daniel Hilger*, Kaavya Krishna Kumar*, Hongli Hu*, Mie Fabricius Pedersen, Evan S. O'Brien, Lise Giehm, Christine Jennings, Gözde Eskici, Asuka Inoue, Michael Lerch, Jesper Mosolff Mathiesen†, Georgios Skiniotis†, Brian K. Kobilka†

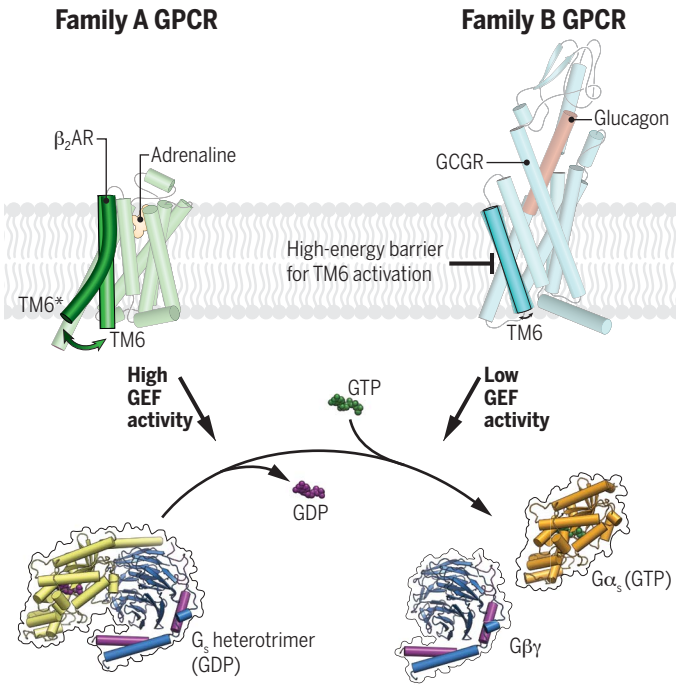
INTRODUCTION: Transmembrane signaling of heterotrimeric guanine nucleotide–binding protein (G protein)–coupled receptors (GPCRs) is mediated by ligand-dependent conformational changes that are transmitted from the extracellular ligand binding site to the intracellular side of the receptor to allow coupling with transducers. One hallmark of GPCR activation is the outward movement of the cytoplasmic end of transmembrane domain 6 (TM6) that opens up an intracellular cavity to accommodate the Gα subunit, leading to nucleotide exchange and activation of the G protein. Comparison of family A and family B receptor-G_s protein complex structures has revealed substantial differences in the conformational changes of TM6 upon activation. In family B GPCRs, TM6 shows a disruption of the helical fold and the formation of a sharp kink. This differs from the gradual bending in TM6 observed in family A GPCRs.

RATIONALE: Despite the recent surge of determined GPCR–G protein complex structures, the activation mechanism of family B receptors remains poorly understood. The missing conserved structural motifs found in family A GPCRs together with the marked differences in the conformation of TM6 in the active state suggest distinct activation mechanisms between family B and family A GPCRs. In particular, the disruption of the helical fold and the unraveling of the extracellular end of TM6 suggest that the energy required to produce the fully active-state of family B GPCRs is higher than for family A GPCRs. We investigated the functional impact of these structural differences by comparing the structure and function of a prototypical family B receptor, the glucagon receptor (GCGR), with the β_2 adrenergic receptor (β_2AR), a family A GPCR.

RESULTS: We present the cryo–electron microscopy structure of the GCGR-G_s complex bound to an engineered soluble glucagon derivative. The structure shows that full activation of GCGR leads to a disruption in the α-helix of TM6 typical for family B GPCRs. Analysis of the functional consequence of this helix break on receptor-mediated G protein dissociation and guanosine triphosphate (GTP) turnover reveals that GCGR exhibits a substantially lower guanine nucleotide exchange factor (GEF) activity in comparison with the family A receptor β_2AR. Characterization of G protein association, guanosine diphosphate (GDP) release, and GTP binding kinetics shows that the receptor-mediated GDP dissociation and GTP binding of G_s are slower for GCGR than for β_2AR. Measurements of ligand-dependent conformational alterations of GCGR by means of fluorescence and double electron-electron resonance spectroscopy show that agonist binding alone is insufficient to promote TM6 opening, in contrast to previously studied family A GPCRs, including β_2AR. The outward movement of TM6 of GCGR is only observed upon interaction with G_s, suggesting that TM6 activation is only triggered by the engagement of the α5 helix of $G\alpha_s$. Furthermore, TM6 of GCGR remains in the active state for a prolonged time after disengagement of G_s, which might contribute to the persistent and sustained cyclic adenosine monophosphate (cAMP) signaling previously observed for this receptor. A comprehensive comparison of the G protein activation kinetics for a number of other family A and family B GPCRs shows that family B receptors are in general less efficient than family A GPCRs in triggering G protein signaling.

ON OUR WEBSITE

Read the full article at https://dx.doi.org/10.1126/science.aba3373



Difference in the activation mechanism of family A and family B GPCRs affects G protein signaling. In contrast to family A GPCRs such as β_2AR, agonist binding alone is not sufficient to induce TM6 outward movement of GCGR, most likely because of the high energy barrier for the formation of the kinked and partially unwound TM6 found in the G protein–coupled state of family B GPCRs. In comparison with family A GPCRs, this leads to a reduced G protein activation rate.

CONCLUSION: Our findings provide evidence for distinct activation mechanisms between family A and family B GPCRs. We propose that formation of the helix break and the sharp kink in TM6 of GCGR requires overcoming a higher energy barrier than the bending and outward movement of TM6 in family A receptors. Because of this kinetic barrier, ligand binding alone is not sufficient to stabilize the outward movement of TM6 but promotes the initial formation of the receptor–G protein complex and subsequent full engagement of the G protein at a slower time scale. Once activated by the insertion of the α5 helix of $G\alpha_s$ into the receptor core, as seen in the nucleotide-free complex structure, TM6 stays in the active conformation long after full disengagement of the G protein. This may be responsible for the previously described sustained and prolonged signaling of GCGR. Our comprehensive comparison of the G protein activation kinetics of family A and family B receptors suggests that the activation mechanism described for GCGR is generalizable to other family B GPCRs. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: jmm@zealandpharma.com (J.M.M.); yiorgo@stanford.edu (G.S.); kobilka@stanford.edu (B.K.K.)
Cite this article as D. Hilger *et al.*, *Science* **369**, eaba3373 (2020). DOI: 10.1126/science.aba3373

RESEARCH ARTICLE

STRUCTURAL BIOLOGY

Structural insights into differences in G protein activation by family A and family B GPCRs

Daniel Hilger[1]*†, Kaavya Krishna Kumar[1]*, Hongli Hu[1,2]*‡, Mie Fabricius Pedersen[3], Evan S. O'Brien[1], Lise Giehm[3], Christine Jennings[4], Gözde Eskici[1,2], Asuka Inoue[5], Michael Lerch[4], Jesper Mosolff Mathiesen[3]§, Georgios Skiniotis[1,2,6]§, Brian K. Kobilka[1]§

Family B heterotrimeric guanine nucleotide–binding protein (G protein)–coupled receptors (GPCRs) play important roles in carbohydrate metabolism. Recent structures of family B GPCR-G_s protein complexes reveal a disruption in the α-helix of transmembrane segment 6 (TM6) not observed in family A GPCRs. To investigate the functional impact of this structural difference, we compared the structure and function of the glucagon receptor (GCGR; family B) with the β_2 adrenergic receptor (β_2AR; family A). We determined the structure of the GCGR-G_s complex by means of cryo–electron microscopy at 3.1-angstrom resolution. This structure shows the distinct break in TM6. Guanosine triphosphate (GTP) turnover, guanosine diphosphate release, GTP binding, and G protein dissociation studies revealed much slower rates for G protein activation by the GCGR compared with the β_2AR. Fluorescence and double electron-electron resonance studies suggest that this difference is due to the inability of agonist alone to induce a detectable outward movement of the cytoplasmic end of TM6.

The cytoplasmic end of transmembrane helix 6 (TM6) of heterotrimeric guanine nucleotide–binding protein (G protein)–coupled receptors (GPCRs) opens to allow the activated receptor to accommodate the α5 helix of Gα, leading to G protein coupling and activation. Recent structures of family B receptor–G_s protein complexes have revealed a disruption of the helical fold at the conserved PxxG motif in the middle of TM6 and an unraveling of the extracellular end of the helix (*1*–*5*). This contrasts the changes observed in family A GPCR–G_s protein complexes, where there is a gradual bending in TM6 to provide space for the insertion of the α5 helix. The marked differences in the structural changes of TM6 for family A and family B GPCRs suggest that the energy required to produce the active state of TM6 found in family B GPCRs is higher than for family A GPCRs. To explore the functional impact of the structural differences between these two GPCR families, we compared the β_2 adrenergic receptor (β_2AR), a prototypical family A GPCR, with the glucagon receptor (GCGR), a family B GPCR. The GCGR is an ideal candidate for this comparison because, unlike most other family B GPCRs, it can be purified in a functional state without being complexed with G_s and without thermostabilizing mutations, enabling structure determination as well as biochemical and biophysical studies to characterize G_s activation.

The peptide hormone glucagon, released from the α cells of the pancreas in response to low blood glucose level (*6*), binds and activates the GCGR (*7*, *8*). Upon activation, GCGR predominantly interacts with the adenylyl cyclase stimulatory G protein, G_s, leading to increased cyclic adenosine monophosphate (cAMP) production (*9*). This signaling in turn drives glycogenolysis and gluconeogenesis and an increased hepatic glucose output (*10*). Because of its fundamental role in glucose homeostasis, GCGR represents an important therapeutic target for the treatment of severe hypoglycemia in diabetic patients (*11*). Furthermore, modulation of GCGR signaling has therapeutic potential for obesity and type 2 diabetes therapies (*12*). The sympathetic nervous system is also activated by hypoglycemia, but this leads to activation of hepatocyte β_2ARs and subsequent release of glucose. Previous studies in hepatocytes suggest differences in the kinetics of cAMP accumulation after activation of the GCGR when compared with the kinetics of cAMP accumulation after activation of the β_2AR (*13*). To explore the structural basis of these functional differences, we determined the structure of the GCGR in complex with G_s by means of cryo–electron microscopy (cryo-EM). Comparison of the active-state GCGR structure with previously determined inactive structures, unliganded and bound to a partial agonist, provides insights into ligand efficacy for family B GPCRs. In addition, comparative functional and biophysical studies of the GCGR and the β_2AR revealed marked differences in agonist-induced conformational changes and receptor-mediated G protein activation. Analysis of receptor-mediated G protein dissociation rates of different family A and family B GPCRs suggest that the observed differences in the activation mechanism of GCGR are generalizable to other family B GPCRs.

[1]Department of Molecular and Cellular Physiology, Stanford University School of Medicine, 279 Campus Drive, Stanford, CA 94305, USA. [2]Department of Structural Biology, Stanford University School of Medicine, 279 Campus Drive, Stanford, CA 94305, USA. [3]Zealand Pharma A/S, Sydmarken 11, Søborg 2860, Denmark. [4]Department of Biophysics, Medical College of Wisconsin, Milwaukee, WI 53226, USA. [5]Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai, Miyagi 980-8578, Japan. [6]Department of Photon Science, SLAC National Accelerator Laboratory, Stanford University, Menlo Park, CA 94025, USA.
*These authors contributed equally to this work.
†Present address: Department of Pharmaceutical Chemistry, Philipps-University Marburg, Marbacher Weg 6, Marburg 35037, Germany.
‡Present address: Kobilka Institute of Innovative Drug Discovery, School of Life and Health Sciences, The Chinese University of Hong Kong, Shenzhen, Guangdong Province, 518172, P.R. China.
§Corresponding author. Email: jmm@zealandpharma.com (J.M.M.); yiorgo@stanford.edu (G.S.); kobilka@stanford.edu (B.K.K.)

Results

Development of a soluble glucagon analog, ZP3780

To determine the structure of the GCGR signaling complex, we designed a glucagon analog with improved solubility and stability in aqueous solutions compared with those of the native glucagon peptide that is prone to rapidly form fibrillar aggregates at neutral pH (fig. S1A and tables S1 and S2) (*14*). Building on our previous efforts in designing glucagon analogs (*15*), we retained the native sequence in the N terminus of the peptide, which is important for ligand efficacy (*16*), while improving solubility and stability by iteratively substituting amino acids in the C terminus to change the peptide isoelectric point (pI) (fig. S1B). The resulting glucagon analog, ZP3780, acts as a full agonist and includes four C-terminal mutations that only slightly reduce the affinity while showing similar maximum attainable response (E_{max}) in comparison with that of wild-type (WT) glucagon (fig. S1, C and D, and tables S3 and S4). Receptor activation with ZP3780 enabled the formation of a GCGR-G_s complex that was stable enough for cryo-EM imaging, yielding a final density map at a global nominal resolution of 3.1 Å (Fig. 1A, fig. S2, and table S5).

Cryo-EM structure of ZP3780-bound GCGR–G_s complex

The cryo-EM map shows well-defined density that allowed unambiguous modeling of the secondary structure and side chain orientations of all components of the GCGR-G_s complex, including the full agonist peptide ZP3780 (figs. S2F and S3A). The glucagon analog forms a continuous α-helix bound between the extracellular domain (ECD) and the transmembrane bundle of the receptor (Fig. 1, A and B, and fig. S3). ZP3780 engages GCGR through extensive van der Waals and hydrophobic interactions with residues in the ECD and mostly polar contacts with residues in the transmembrane domain (Fig. 2A and fig. S3, C and D). The overall peptide binding mode and the relative orientation of the ECD

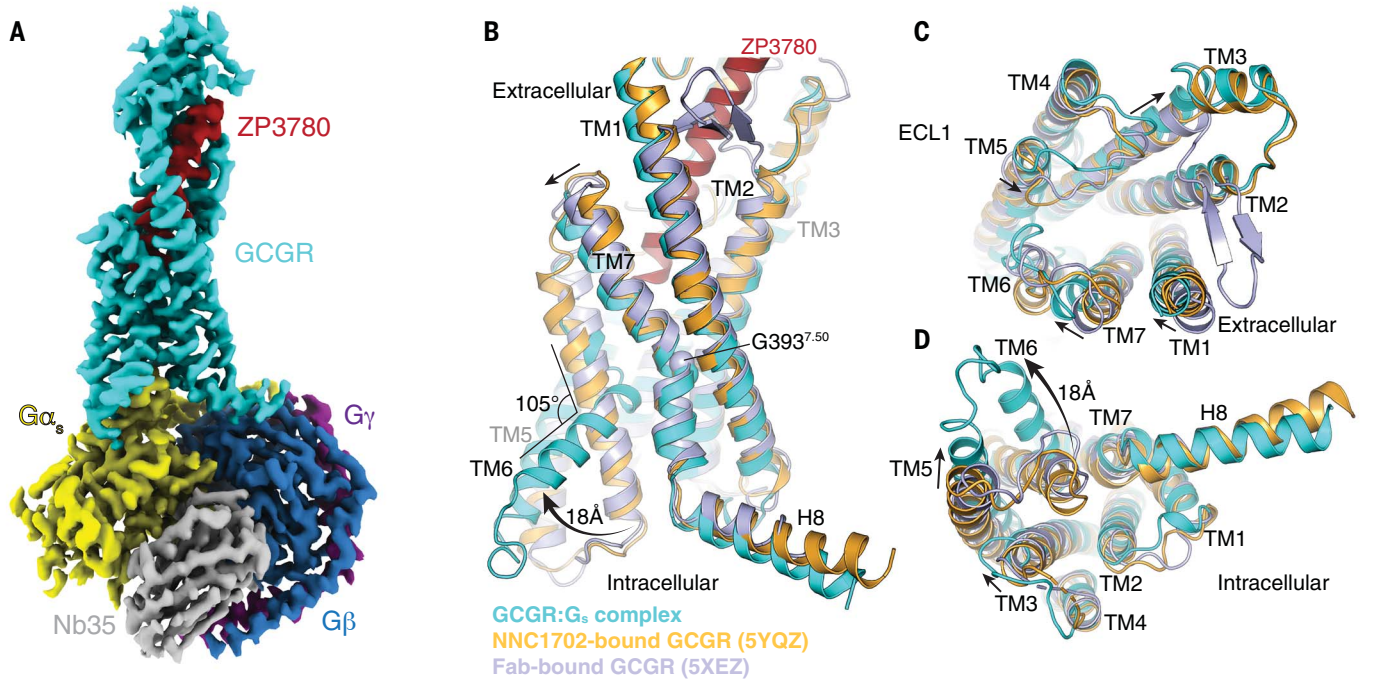


Fig. 1. Cryo-EM structure of the ZP3780-bound GCGR-G_s complex. (**A**) Cryo-EM density map of the ZP3780-bound GCGR-G_s heterotrimeric complex colored by subunit. Cyan, GCGR; red, ZP3780; yellow, $G\alpha_s$; dark blue, Gβ; purple, Gγ; gray, Nb35. (**B** to **D**) Comparison of the inactive, peptide ligand–free state of GCGR (Fab-bound GCGR, blue, PDB 5XEZ) (*19*), partial agonist NNC1702-bound GCGR (orange, PDB 5YQZ) (*17*), and active, full-agonist ZP3780-bound state of GCGR (cyan). Substantial structural change is observed in TM6, which moves outward by 18 Å and partially unwinds to form a kink with an angle of 105°. Additional changes are also observed in TM1, TM3, TM5, and TM7.

with respect to the transmembrane domain is very similar to the recently reported crystal structure of GCGR bound to the partial agonist peptide NNC1702 (fig. S3B) (*17*). Out of the four mutated residues (Q20E, Q24K, M27E, and N28S) that were inserted into the WT glucagon peptide to generate the more soluble derivative ZP3780, only two residues (E20 and E27) interact with the receptor (figs. S1B and S3C). (Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. In the mutants, other amino acids were substituted at certain locations; for example, Q20E indicates that glutamine at position 20 was replaced by glutamic acid.) The glutamate side chain of the Q20E substitution in ZP3780 forms a hydrogen bond (H-bond) with $Q131^{1.29}$ in TM1 similarly to the WT glutamine residue in glucagon, as seen in the NNC1702-bound crystal structure of GCGR (*17*). In contrast to the apolar WT methionine residue in glucagon (M27), which most likely forms a hydrophobic interaction with the nearby side chain of $Y65^{ECD}$ (*18*), the mutated residue M27E in ZP3780 interacts with $Q122^{ECD}$ of GCGR through a H-bond (fig. S3C). Thus, the nonconservative substitution at position 27 might explain the slightly reduced potency of ZP3780 compared with WT glucagon (fig. S1C).

We observed substantial structural rearrangements in the transmembrane region of the activated receptor (Fig. 1, B to D). Comparison of our structure with the previously determined full-length GCGR structures in the inactive, peptide-free (apo), state (*19*) and in complex with the partial agonist NNC1702 (*17*) shows that ZP3780 binding induces conformational changes not only of the ECD and the stalk region but also in the extracellular loop 1 (ECL1) and the extracellular ends of TM1, TM2, TM6, and TM7 (Fig. 1, B and C). Some of these rearrangements are also found in the NNC1702-bound inactive structure of GCGR, suggesting that they are important for peptide binding by expanding the ligand-binding site to allow the penetration of the peptide N terminus deep into the receptor core. However, the GCGR-G_s complex structure also reveals structural changes that are different from the NNC1702-bound GCGR structure and that are related to full activation of the receptor. These include movement of the extracellular half of TM7 that bends further toward TM6 (6.5 Å C_α-C_α distance between $L377^{7.34}$) (receptor residues in superscript are defined by using the family B numbering system) (*20*), facilitated by the conserved $G393^{7.50}$ in the center of the transmembrane domain (Fig. 1, B and C). Furthermore, the entire TM3 shifts up by half a helix turn, and its extracellular end moves away from the receptor core. On the intracellular side, TM5 moves toward TM6 by 6.5 Å, followed by the C-terminal end of TM3 (Fig. 1D). The largest structural rearrangement between the NNC1702-bound structure of GCGR and the fully active receptor conformation in the G protein complex involves the formation of a sharp kink in the middle of TM6 (105.5°, measured between V368-G359-K344) (Fig. 1B). As reported for the previously determined active structures of other family B GPCRs [glucagon-like peptide-1 (GLP-1), calcitonin receptor (CTR), calcitonin gene-related peptide (CGRP), and parathyroid hormone receptor (PTH)], the kink pivots the intracellular half of TM6 to move outward by ~18 Å, while the α-helical structure of the extracellular half partially unwinds (*1*–*5*).

The conformational differences between the NNC1702- and the ZP3780-bound structures are reflected in the distinct efficacy profiles of the two ligands, which likely result from sequence variations in the N-terminal part of the peptides, and the stabilization of the active state by the receptor-engaged heterotrimeric G protein G_s. In the native glucagon peptide, the N-terminal residues histidine 1 (H1) and aspartate 9 (D9) are critical positions for glucagon activity (*21*, *22*). In contrast to ZP3780, which retains these residues, NNC1702 has a

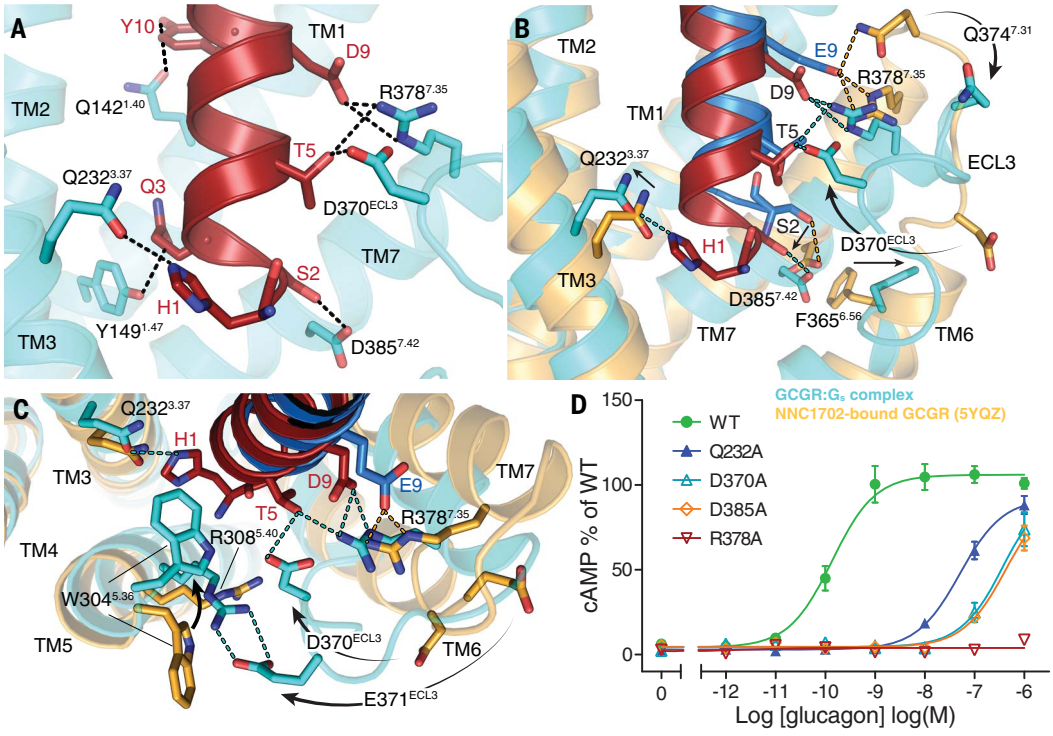


Fig. 2. ZP3780 binding and activation at GCGR. (**A**) The N terminus of ZP3780 (red) is required for full ligand efficacy and penetrates deep in the receptor core to make H-bonds (dotted lines) with residues in TM1, TM3, TM7, and ECL3 (cyan). (**B**) The difference in receptor recognition by full-agonist ZP3780 (red) and partial-agonist NNC1702 (dark blue) that lacks H1 and has a D9E mutation is shown. NNC1702-bound GCGR is shown in orange (PDB 5YQZ) (*17*), and the ZP3780-bound GCGR-G_s complex structure is shown in cyan. The polar interactions are shown as dotted lines colored according to the GCGR structures bound to the two ligands, respectively (ZP3780, cyan; NNC1702, orange). (**C**) Comparison of the structures of the partial agonist NNC1702 (blue)–bound GCGR (orange, PDB 5YQZ) (*17*) and the full-agonist ZP3780 (red)–bound GCGR (cyan) reveal conformational differences in ECL3 and TM5, TM6, and TM7. The presence of H1 in ZP3780 ensures interaction with $Q232^{3.37}$ and may induce rearrangement of residues in TM5. The interaction of D9 seems to stabilize TM7 and ECL3 displacement, which might trigger GCGR activation. (**D**) Mutation of $Q232^{3.37}$, $D370^{ECL3}$, $R378^{7.35}$, and $D385^{7.42}$ to alanine has a large effect on GCGR-mediated cAMP signaling. All mutants were expressed to similar levels as that of the WT receptor. For (D), data represent mean ± SEM from at least three independent experiments, performed in triplicates. Superscripts are Wootten numbering.

deletion of H1 and substitution of D9 by glutamate (D9E). Substitution of H1 or D9 to alanine in glucagon results in a marked reduction in ligand efficacy (*23*). Therefore, a comparison between the interaction patterns of the N termini of the peptides ZP3780 and NNC1702 and the transmembrane bundle (TMs 3, 5, 6, 7, and ECL3) of the receptor can provide a structural framework to explain the molecular basis of ligand efficacy for the glucagon receptor.

Structural insights into ligand efficacy in GCGR

In the full agonist ZP3780, the carboxyl group of D9 forms a salt bridge with residue $R378^{7.35}$ on the N terminus of TM7 that seems to draw TM7 toward TM6 (Fig. 2, A to C). This polar interaction is also present in previously determined complex structures of the family B GPCR GLP-1 with G_s (*2*, *4*). Mutation of $R378^{7.35}$ to alanine completely inhibits GCGR-mediated cAMP signaling (Fig. 2D and table S6), highlighting the importance of this interaction for ligand-dependent activation within receptors of the glucagon family. The altered position of TM7 is accompanied by movement of ECL3 toward the receptor core with a H-bond formed between $D370^{ECL3}$ and T5 of ZP3780 (Fig. 2, B and C). This interaction is part of a polar network that includes $R378^{7.35}$ and D9, which may stabilize the active TM7 conformation together with the H-bond between S2 and $D385^{7.42}$. Mutation of $D370^{ECL3}$ and $D385^{7.42}$ to alanine reduces glucagon potency from 0.14 nM to more than 100 nM (Fig. 2D and table S6). These results suggest that TM7 and ECL3 play an important role in the process of receptor activation, which is in agreement with previous studies on other family B GPCRs (*24*). At the peptide N terminus, H1 forms a H-bond with residue $Q232^{3.37}$ in TM3 on the other side of the ligand binding pocket (Fig. 2A). Formation of this interaction seems to anchor the peptide deep into the ligand-binding pocket and constrains its movement. Alanine substitution of $Q232^{3.37}$ reduces the glucagon potency from 0.14 to 51 nM (Fig. 2D and table S6), indicating that this interaction is crucial for ligand engagement and receptor activation. Furthermore, $W304^{5.36}$ in TM5 flips toward the receptor core, which may lead to a stabilization of the H1-$Q232^{3.37}$ interaction by forming a "cap" that presumably provides a further steric hindrance to peptide movement within the binding pocket (Fig. 2C). The functional importance of $W304^{5.36}$ is shown by a mutation to glutamine that completely abolishes peptide binding (*25*). In addition, the inward movement of $W304^{5.36}$ appears to transmit the rotation through the helix to facilitate a large-scale translation of the cytoplasmic side of TM5 toward TM6 upon GCGR activation (Fig. 1D).

Comparison of the N-terminal positions of the full agonist ZP3780 and the partial agonist NNC1702 suggests that the absence of H1 results in an upward shift of NNC1702 toward TM7 (Fig. 2B). In this position, the partial agonist would clash with TM1 and TM7 in the fully activated receptor conformation (fig. S4, A and B). In addition to the steric restriction of TM1 and TM7 movement, the D9E mutation allows the partial agonist to form additional polar contacts with the extracellular end of TM7 (residues $R378^{7.35}$ and $Q374^{7.31}$) in comparison with ZP3780, which may stabilize it in an inactive state and prevent the inward movement of ECL3 (Fig. 2, B and C). In addition, the missing N-terminal H1 and the lack of rearrangement in ECL3 most likely restrict the inward rotation of $W304^{5.36}$ and, as a consequence, the translational rotation of TM5 that is important for full activation of GCGR (Fig. 2C).

Agonist-induced structural changes in TM6

As noted above, a hallmark of GPCR activation is the outward movement of TM6 that is accompanied in family B GPCRs by the formation of a sharp kink in the middle of the helix at the highly conserved PxxG motif (*26*). Compared with the NNC1702-bound receptor structure showing an outward movement of the extracellular tip of TM6 induced by peptide binding, the ZP3780-bound GCGR-G_s complex

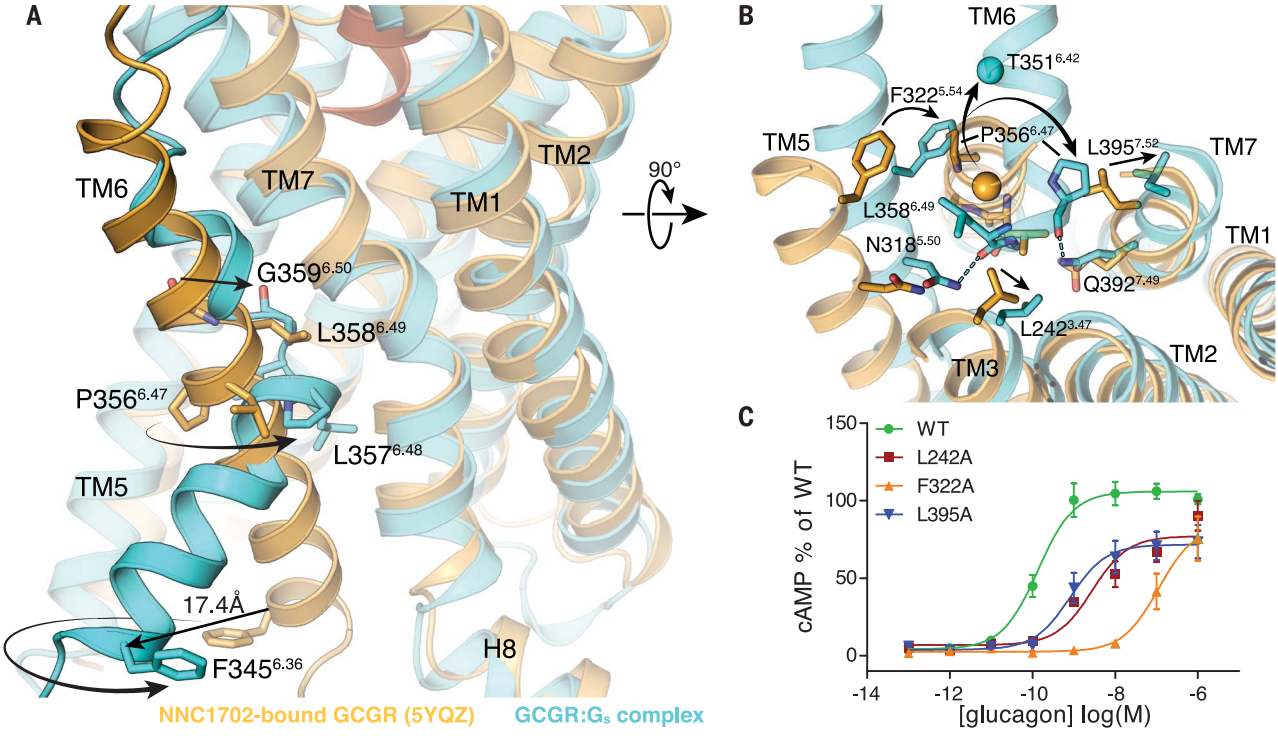


Fig. 3. Structural changes in the PxxG motif of GCGR induced by full-agonist ZP3780 binding and G_s coupling. (**A**) The rearrangement of TM3, TM5, and TM7 allows TM6 outward movement and kink formation at the conserved PxxG motif (P356$^{6.47}$-L357$^{6.48}$-L358$^{6.49}$-G359$^{6.50}$). The ZP3780-bound GCGR-G_s complex (cyan) is superimposed on the NNC1702-bound GCGR structure (orange) to highlight structural changes that result in the TM6 outward movement in the fully active GCGR-G_s complex structure. (**B**) Reorientation of residues in TM3, TM5, and TM7 around the PxxG motif that seems to facilitate kink formation in TM6. The polar interactions are shown as dotted lines colored according to the GCGR structure. The Cα atom of T351$^{6.42}$ is presented as a sphere to highlight the extent of TM6 outward movement. (**C**) Alanine substitution of residues around the PxxG motif significantly reduces GCGR-mediated cAMP signaling. All mutants were expressed to similar levels as that of the WT receptor. For (C), data represent mean ± SEM from at least three independent experiments, performed in triplicates. Superscripts are Wootten numbering.

reveals partial unwinding of helix 6 above the PxxG (P356$^{6.47}$-L357$^{6.48}$-L358$^{6.49}$-G359$^{6.50}$ in GCGR) motif to avoid spatial clashes with both the peptide and the repositioned TM7 (Fig. 3, A and B, and fig. S4, C to E). Although this most likely initiates the destabilization of TM6 helicity, other conformational rearrangements around the PxxG motif may be required for kink formation. Ligand-induced movement of TM5 seems to cause F322$^{5.54}$ to relocate toward TM6 and probably creates a clash between this bulky side chain and P356$^{6.47}$ of the PxxG motif in the helical conformation of TM6, as seen in the NNC1702 bound structure (Fig. 3B) (*17*). As a result, P356$^{6.47}$ rotates away by approximately 90°, which probably causes the PxxG motif to shift inward toward TM3 and TM7 (Fig. 3, A and B). Substitution of F322$^{5.54}$ to alanine results in decreased potency of glucagon from 0.14 nM to 110 nM, indicating that the rearrangement of TM5 and packing of F322$^{5.54}$ against the PxxG motif is important for receptor activation (Fig. 3C and table S6). Furthermore, a slight rotation of TM7 away from TM6, presumably caused by the ligand-induced changes, may help to accommodate the displaced PxxG motif in the active state of the receptor by releasing the steric hindrance between the newly positioned P356$^{6.47}$ and L395$^{7.52}$ in TM7 (Fig. 3B). Another obstacle for kink formation is found in TM3, where the side chain of L242$^{3.47}$ would prevent inward movement of the PxxG motif. The observed upward shift of TM3 may lead to the removal of this restriction by shifting the leucine side chain away from the PxxG motif (Fig. 3B). Individual replacement of the two leucine residues (L242$^{3.47}$ and L395$^{7.52}$) in TM3 and TM7 to the smaller alanine side chain causes a decrease in glucagon potency (from 0.14 to 3 nM and 0.82 nM, respectively) as well as a significant decrease in the maximal glucagon response (Fig. 3C and table S6). This highlights the importance of the residues at these positions to produce conformational rearrangements that seem to be necessary for the inward rotation of the PxxG motif upon receptor activation. The repositioning of the PxxG motif is probably further stabilized by a H-bond with TM7 (P356$^{6.47}$-Q392$^{7.49}$) (Fig. 3B). The functional relevance of these residues for activation of the glucagon receptor family have been previously shown by mutagenesis studies (*20*). As a result of the 90° rotation of TM6 at P356$^{6.47}$, the N terminus of TM6 moves away from TM5, maintaining the 90° rotation (compare the positions of F345$^{6.36}$ in Fig. 3A and of the α carbon of T356$^{6.42}$ Fig. 3B), and dislocates from the receptor core by 18 Å to enable the coupling of the G protein (Figs. 1B and 3A).

Coupling of GCGR to G_s

The conformational rearrangements in the transmembrane bundle of the receptor open up an intracellular cavity that allows engagement of the heterotrimeric G protein, G_s. In the complex structure, the C-terminal α5 helix of $G\alpha_s$ inserts into the receptor core to form extensive hydrophilic interactions with residues in TM5 and TM6 and helix 8 (H8) and hydrophobic contacts with TM3 (Fig. 4, A and B). In general, the GCGR-G_s complex displays high similarity in the receptor–G protein interactions compared with other reported structures of family B GPCRs-G_s complexes (such as GLP-1, CTR, CGRP, and PTH), suggesting a common mechanism for G protein engagement and activation (fig. S5) (*1*–*5*, *27*).

A notable difference between GCGR and β_2AR, a family A GPCR that couples predominantly to G_s, is the extent of interaction

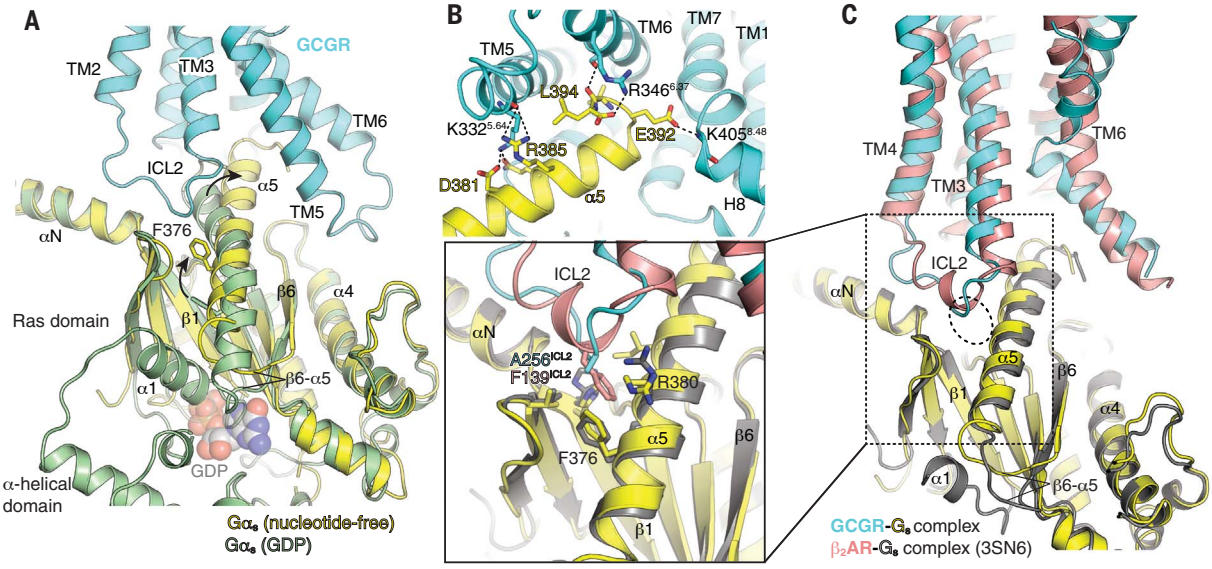


Fig. 4. Structural changes in G_s upon coupling to GCGR. (**A**) Comparison of GDP-bound $G\alpha_s$ (green, PDB 6EG8) (*73*) and nucleotide-free $G\alpha_s$ (yellow) in complex with GCGR (cyan). Major differences between these two structures are the opening and displacement of the α-helical domain, the rotational translation of the α5 helix to engage the receptor core, and conformational rearrangements in the α1 helix and α5-β6 loop. (**B**) The α5 helix (yellow) of $G\alpha_s$ in the nucleotide-free GCGR-G_s complex engages the cytoplasmic core of GCGR (cyan) to form extensive polar interactions (dotted lines) with TM5, TM6, and helix 8 (H8). (**C**) Comparison between the interaction of ICL2 of GCGR (cyan) and β_2AR (pink) (PDB 3SN6) (*50*) and α5 of G_s (yellow and dark yellow, respectively). Close-up view shows that residue $F139^{ICL2}$ of β_2AR engages the hydrophobic pocket lined by residues from the αN-β1 loop, the β2-β3 loop, and the α5 helix. In GCGR, the corresponding residue is $A256^{ICL2}$, which engages the hydrophobic pocket less efficiently.

between the intracellular loop 2 (ICL2) and the hydrophobic pocket in G_s formed by the αN-β1 loop, the β2-β3 loop, and the α5 helix of $G\alpha_s$ (Fig. 4C). In the β_2AR-G_s complex, ICL2 forms a helix that positions F139 to penetrate deeply into this pocket. This interaction has been shown to be functionally important because the F139A mutation dramatically reduces GDP release even though it still forms a complex with the G protein (*28*). By contrast, ICL2 in GCGR is a loop where A256 occupies the position of F139 in β_2AR, having weaker interactions with the hydrophobic pocket (Fig. 4C, close-up view).

Functional impact of structural differences between family B and family A GPCRs

Despite the lack of sequence homology between GPCR families, the activation of family B receptors involves structural rearrangements that are similar to family A receptors, including the outward movements of TM5 and TM6. However, conserved structural motifs and conformational switches that play important roles for the activation of family A GPCRs (such as the DRY/ERY, NPxxY, and PIF motifs) are absent in family B receptors (*26*, *29*). This implies that the mechanisms of activation between family B and family A receptors are substantially different from each other. In family B GPCRs, peptide binding results in conformational changes on the extracellular side that leads to an expansion rather than a contraction of the ligand binding cavity, as seen in family A receptors. Furthermore, in the structures available to date, the active conformation of family B receptors in complex with G_s shows a distinct sharply kinked TM6 when compared with G_s-coupled family A GPCRs such as the β_2AR, where TM6 moves outward without the formation of a helix break (Fig. 5A). We speculated that the energy required to produce the kinked conformation of the active state of TM6 found in family B GPCRs is much higher than the energy required to produce the much smaller change in TM6 observed in family A GPCRs.

To better understand the impact of the TM6 helix kink on receptor activation and signaling of family B GPCRs, we compared the activity of GCGR to β_2AR in in vitro GTP-turnover assays using purified receptors reconstituted into high-density lipoprotein (HDL) particles (also known as nanodiscs) (3:2 molar ratio of POPC: POPG) and purified G_s heterotrimer. The functionality of both receptors in HDL particles was tested by means of radioligand filtration binding to ensure that the specific activity was comparable between the two receptors (fig. S6, A to D, and table S8). GCGR shows significantly lower guanine nucleotide exchange factor (GEF) activity than β_2AR, with a maximal GTP turnover rate of 0.11 GTP min^{-1} G_s^{-1} versus ~8 GTP min^{-1} G_s^{-1} seen for β_2AR (Fig. 5B and table S7). The reduced GCGR-dependent G protein activation might be related to any or all of the following: less efficient G protein association, reduced receptor-induced GDP release, reduced GTP binding, and slower complex dissociation. To distinguish between these possibilities, we performed several biophysical and functional assays. First, we tested the rate of G protein association to the receptors. For this purpose, we measured Förster resonance energy transfer (FRET) between Cy3B-labeled receptors and Sulfo-Cy5-labeled heterotrimeric G protein, G_s, in bulk detergent solutions to avoid nonspecific adsorption of lipidated G proteins to HDL particles. The result shows that the G protein association rate for GCGR with an association rate constant (k_{on}) of 0.0044 s^{-1} is only slightly slower than for β_2AR (k_{on} = 0.0077 s^{-1}) (Fig. 5C and table S7) and hence cannot be the only cause for the substantially slower GTP-turnover rate of GCGR described above. We next tested whether the receptor-induced GDP release is significantly different between the two receptors by using a [^{3}H]-GDP dissociation assay in HDL particles. GCGR catalyzed nucleotide release from [^{3}H]-GDP loaded G_s with a substantially slower rate [dissociation rate constant (k_{off}) = 0.002 s^{-1}] compared with β_2AR (k_{off} = 0.042 s^{-1}) (Fig. 5D and table S7). Together, these results suggest that GCGR forms a more stable intermediate state with GDP-bound G_s and shows lower activity toward the formation of the nucleotide-free complex than that of β_2AR. This might be explained by the weaker interaction between ICL2 of the GCGR and the hydrophobic pocket formed by the αN-β1 loop, the β2-β3 loop, and the α5 helix of $G\alpha_s$ (Fig. 4C).

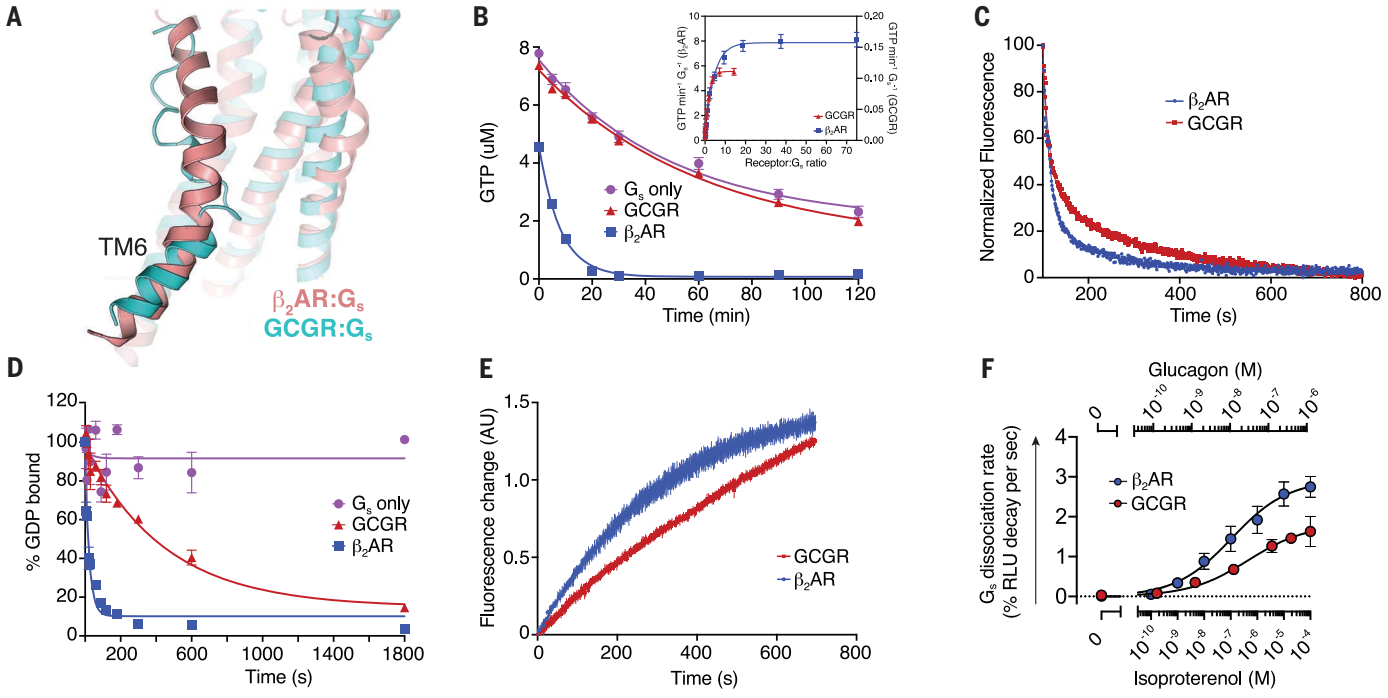


Fig. 5. Engagement and activation of G_s by GCGR and β_2AR. (**A**) Conformation of TM6 in the GCGR-G_s (cyan) and β_2AR-G_s (pink, PDB 3SN6) (*50*) complex structures. (**B**) The GTP-turnover assay shows that β_2AR (blue) activates G_s much faster than GCGR (red), with GCGR inducing a maximal GTP-turnover rate of 0.11 GTP min^{-1} $G_s{}^{-1}$ compared with (inset) 8 GTP $min^{-1}G_s{}^{-1}$ for β_2AR. (**C**) Monitoring of GPCR-G_s association by means of FRET between Cy3B-labeled GCGR (red) or Cy3B-labeled β_2AR (blue) with Sulfo-Cy5-labeled G_s. The decrease in donor fluorescence shows comparable rates of association between the receptors and G_s. (**D**) The rate of receptor-induced [^{3}H]-GDP dissociation from G_s shows faster release for β_2AR (blue; k_{off} = 0.042 s^{-1}) compared with GCGR (red; k_{off} = 0.0022 s^{-1}). (**E**) Bodipy-GTPγS binding to nucleotide-free complex is slower for the GCGR-G_s complex (k_{on} = 0.001 s^{-1}) compared with the β_2AR-G_s complex (k_{on} = 0.003 s^{-1}). (**F**) Concentration-response curves for G_s dissociation rates in HEK293 cells show slower G_s activation by GCGR compared with β_2AR. In (B) and (D), data represent mean ± SEM from at least three independent experiments performed in triplicates. In (C) and (E), data represent mean ± SEM of triplicate measurements. In (F), data represent mean ± SEM of three to seven independent experiments.

Single-molecule FRET studies on β_2AR reveal that one or more nucleotide-bound transient intermediate states play an important role in the process of receptor-mediated G protein activation (*30*). Supported by the inhibitory effect of GDP on the GTP turnover of GCGR (fig. S8A) and the previously reported higher potency of GDP in destabilizing high-affinity glucagon binding of the GCGR in comparison with GTP (*31*), we propose that the longer-lived GDP-bound complex of GCGR is at least in part responsible for the overall lower GTP turnover rate of GCGR in comparison with β_2AR. Another rate-limiting step for G protein activation was found to be the GTP loading of the receptor–G protein complex, as evidenced by Bodipy-GTPγS binding studies on nucleotide-free receptor complexes of GCGR and β_2AR (Fig. 5E and table S7). GCGR shows a slower Bodipy-GTPγS (k_{on} = 0.001 s^{-1}) binding rate compared with the β_2AR (k_{on} = 0.003s^{-1}), which demonstrates that the efficiency of GDP release and GTP binding are significantly reduced compared with those of β_2AR and thus represents the major underlying rate limiting steps for GCGR-mediated G protein activation. In addition, these results also suggest that differences in the conformations of the nucleotide-free receptor–G protein complex of GCGR and β_2AR are responsible for differences in nucleotide binding. This is further supported by the previously observed higher affinity of the GCGR-G_s complex for GDP over GTP (*31*), which is in contrast to the higher GTP affinity found for G_s coupled to the full-agonist-bound β_2AR (*30*, *32*).

To provide evidence for the less efficient G protein activation mediated by GCGR in comparison with β_2AR in intact cells, we performed a NanoBiT complementation-based G protein dissociation assay in human embryonic kidney (HEK) 293 cells. In this in vivo assay, an agonist-dependent decrease in signal was measured because of the receptor-mediated dissociation of the $G\alpha_s$ and $G\beta\gamma$ subunits (*33*). With similar expression levels of both receptors (table S9), concentration-response curves allowed us to determine the maximum rate of G_s activation, which was found to be slower for GCGR (2.0 % relative light unit (RLU) change s^{-1}) compared with the β_2AR (3.24 RLU change s^{-1}) (Fig. 5F and table S9). The observed difference in the G_s dissociation rate in our cell-based assay between GCGR and β_2AR is not as large as in the in vitro assays described above (Fig. 5B). This could be due to several reasons, including differences in the lipid composition of cellular membranes in comparison with that of HDL particles, the presence of accessory proteins such as receptor activity–modifying proteins (RAMPs) (*34*, *35*), or variations in the internalization rates. Despite these differences, the results from whole cells confirm the lower GEF activity of GCGR in comparison with the β_2AR observed in our in vitro studies.

Agonist binding alone does not stabilize an active conformation of the GCGR

Dissimilarities in the GEF activity between GCGR and β_2AR might suggest some differences in the process involved in the formation of the fully active conformation of the receptor. For GCGR, this process includes the kink formation and outward movement of the cytoplasmatic end of TM6, a conformation that is further stabilized by movement of TM5 toward TM6 (Figs. 1D and 3B). To explore the effect of agonists on conformational changes associated with receptor activation of GCGR, we used double electron-electron resonance (DEER) spectroscopy (*36*). We established a minimal cysteine GCGR construct (mC-GCGR) to enable site-specific labeling through thiol-based chemistry. Four cysteine residues in the

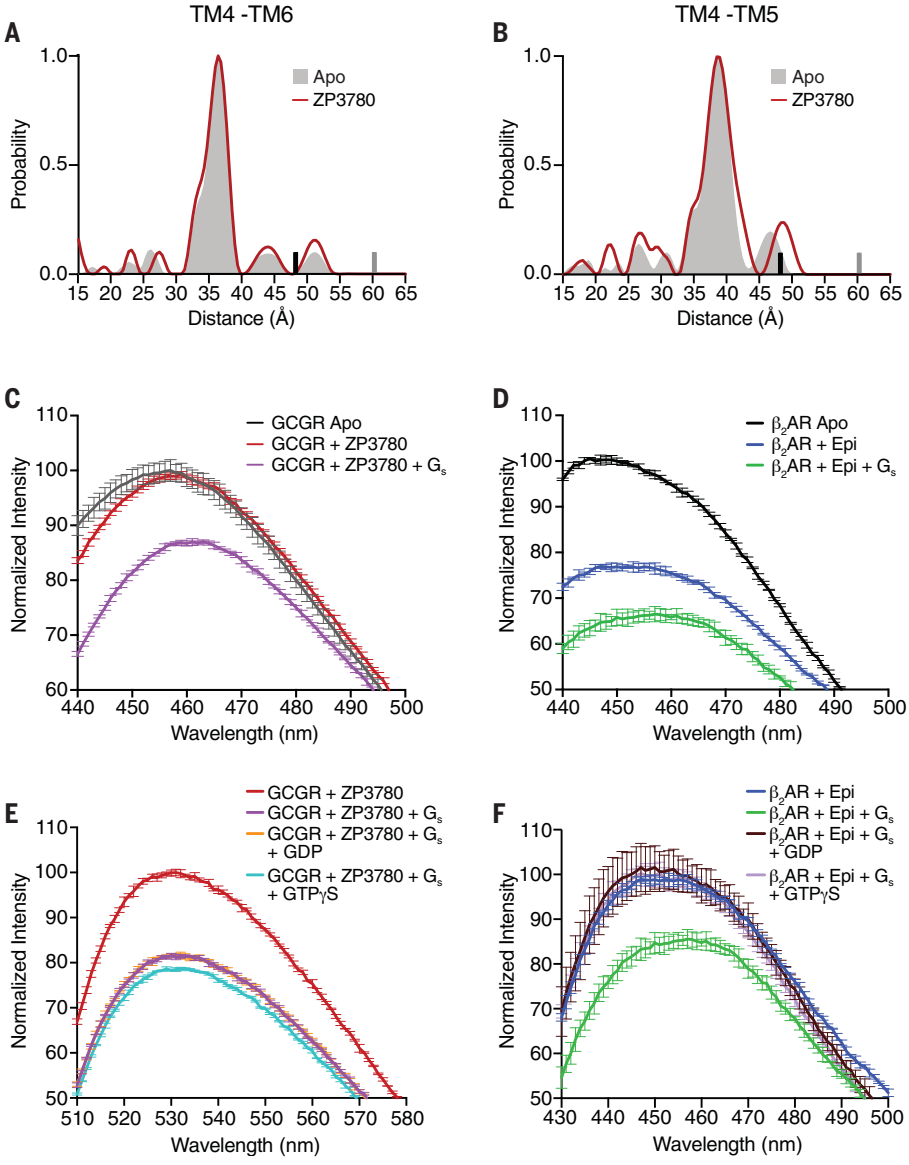


Fig. 6. Conformational changes in GCGR and β_2AR upon ligand binding, G protein coupling, and G protein dissociation. (**A** and **B**) DEER data showing no change in the distance distribution, between Apo (gray shading) and ZP3780 binding (red), of the TM4-TM6 and TM4-TM5 pairs. The upper limit of reliable distance (r) and width (σ) determination are shown as gray and black bars, respectively. (**C**) The apo (gray) spectrum of bimane-labeled GCGR (TM6) does not change upon ZP3780 addition (red). Further addition of G_s (purple) results in a decrease in fluorescence intensity and a 4-nm λ_{max} shift owing to outward movement of TM6. (**D**) β_2AR labeled with bimane in TM6 shows a decrease in fluorescence and a redshift in λ_{max} when agonist, epinephrine (Epi) (blue), is added to apo (black), which changes further upon G_s (green) addition. (**E**) The addition of GDP (orange) or GTPγS (cyan) to NBD-labeled GCGR (TM6) bound to ZP3780 and coupled to G_s (purple) does not result in an increase of the fluorescence intensity to the level of ZP3780-bound GCGR in the absence of G protein (red). (**F**) The addition of GDP (brown) or GTPγS (light purple) to bimane-labeled β_2AR bound to Epi and coupled to G_s (green) leads to an increase in fluorescence intensity and a blueshift in λ_{max} to the same values as that for for Epi-bound β_2AR in the absence of G protein (blue). In (C) to (F), data represent mean ± SEM of triplicate measurements.

TM domains, not involved in disulfide bonding, were mutated (C171T, C240A, C287A, and C401V). These modifications did not substantially affect ligand binding or G protein signaling (fig. S7A). Guided by the GCGR-G_s complex structure and the available inactive structure of GCGR, we engineered cysteine residues, pairwise, at the cytoplasmic ends of TM4 and TM5 (265C to 335C) and TM4 and TM6 (265C to 345C) to probe distance changes upon agonist binding (fig. S7B). Introducing double cysteines into the mC-GCGR background did not substantially alter the pharmacological properties of the receptor (fig. S7A). For the DEER experiments, we labeled the receptor with bis-(2,2,5,5-tetramethyl-3-imidazoline-1-oxyl-4-yl)disulfide (IDSL), which provided reduced background labeling of the mC-GCGR than did other commonly used spin labels. DEER distance distributions on IDSL-labeled TM4-TM5 and TM4-TM6 constructs were collected for receptor both in detergents and in HDL particles. In the apo (unliganded) state, the distance distributions between TM4-TM6 and TM4-TM5 show single conformations in detergent (Fig. 6, A and B, gray shading) as well as HDL particles (fig. S7C, dark gray shading), which are in good agreement with the inactive state of the receptor. Surprisingly, addition of the full agonist ZP3780 at saturating concentration (5x molar excess, 0.5 mM) did not change the distance distribution, indicating that TM5 and TM6 do not undergo ligand-induced conformational changes upon agonist binding. This is in contrast to DEER, single-molecule FRET, and nuclear magnetic resonance (NMR) studies on family A GPCRs (*30*, *37*–*42*). In the case of β_2AR, for example, DEER distance measurements between TM4 and TM6 have shown that agonist binding increases the population of the active state by stabilizing the outward conformation of TM6, as seen in the β_2AR-G_s crystal structure (*38*). That no such DEER distance population for an active conformation was observed for agonist-bound GCGR implies a substantially higher energy level of the fully active conformational state of GCGR in the presence of agonist alone compared with the β_2AR and other characterized family A GPCRs.

For the β_2AR, as well as other family A GPCRs, full stabilization of the active state requires binding of the G protein or a G protein–mimetic nanobody, in addition to the agonist (*29*). Because of the nature of the IDSL spin label used for labeling of mC-GCGR—which gets stripped from the receptor upon engagement of the G protein, as evidenced by a decrease in modulation depth (fig. S7D)—we were unable to use DEER to study conformational changes of the receptor in the presence of the heterotrimer G_s. As an alternative, we used the environmentally sensitive fluorophore monobromobimane (bimane) that has been previously utilized as a conformational reporter of TM6 activation of β_2AR and other GPCRs (*43*–*45*). We labeled the mC-GCGR construct at position 349C in TM6 with bimane and compared it with β_2AR modified at the endogenous cysteine residue 265 in TM6. Using this reporter system, we observed a small but statistically significant difference in the wavelength of maximum emission (λ_{max}) of the fluorescence spectrum of bimane upon addition of ZP3780 to GCGR (Fig. 6C, red, and fig. S8B). This small redshift in λ_{max} suggests that ligand

binding induces some conformational alterations on the intracellular side of the receptor that slightly change the local environment at around the fluorophore attached to TM6. Because the DEER data show no ligand-dependent distance change in TM5 and TM6, we hypothesize that ligand binding induces conformational changes in other regions of the receptor in close vicinity of TM6 that might trigger initial engagement of the GDP-bound G_s to the receptor. By contrast, addition of the agonist epinephrine (Epi) to β_2AR results in a decrease in intensity and a more pronounced redshift in λ_{max}, demonstrating TM6 outward movement (Fig. 6D, blue). These results are consistent with the DEER data that shows that TM6 displacement occurs on agonist binding in β_2AR but not in GCGR (Fig. 6A) (*38*). Adding G_s to the ZP3780-bound GCGR results in a decrease in intensity and a moderate 5-nm redshift in λ_{max} (Fig. 6C, pink). This suggests that GCGR interaction with both agonist and G_s is required to induce conformational changes in TM6 that are necessary for the formation of the nucleotide-free receptor–G protein complex. For β_2AR, addition of G_s to Epi-bound β_2AR leads to an additional reduction of the fluorescence intensity and λ_{max}, indicating that G_s interaction with the receptor further stabilizes the active state of the receptor, as shown previously (*45*). Taken together, the DEER and fluorescence experiments point to a mechanism for GCGR in which agonist binding to the receptor must trigger conformational changes in regions other than TM5 and TM6 that are important for initial G protein coupling. Subsequent full engagement of the G protein is required to induce the kink formation and outward movement of TM6 that are necessary for establishing the nucleotide-free receptor–G protein complex. Although no DEER or fluorescence data are available for other family B GPCRs, the kinked TM6 that pivots away from the core of the receptor is a common feature of all family B GPCR complexes. Hence, the requirement of G_s binding to initiate structural changes in TM6 might be a general mechanism in family B GPCRs.

Persistent active state of GCGR after G protein dissociation

Although molecular dynamics (MD) simulations and fluorescence studies of family A GPCRs have shown fast relaxation of TM6 upon transducer dissociation, less is known about the process of receptor inactivation of family B GPCRs (*30*, *46*). To better understand this process, we used fluorescence spectroscopy to probe conformational changes in TM6 of GCGR upon nucleotide addition to 4-chloro-7-nitrobenz-2-oxa-1,3-diazole (NBD)–labeled mC-GCGR-349C and bimane-labeled β_2AR in complex with G_s. We used the NBD fluorophore instead of bimane for GCGR labeling for this experiment because it showed larger responses to small conformational differences. In parallel, we performed FRET studies to directly detect nucleotide-dependent dissociation of the G protein from both receptors (fig. S8C). To first establish the signaling complexes, we added G_s to ZP3780-bound GCGR (Fig. 6E, purple) or Epi-bound β_2AR (Fig. 6F, green) and followed the decrease in NBD or bimane fluorescence intensity due to TM6 outward movement in the G protein–coupled fluorescently labeled receptors in comparison with the spectra of ligand-bound receptors in the absence of G protein. Once the steady-state level was reached, addition of GDP (Fig. 6F, brown) or GTPγS (Fig. 6F, violet) to β_2AR-G_s, resulted in G_s dissociation (also shown by means of FRET in fig. S8C, brown and violet), and the fluorescence spectrum returned to the wavelength and intensity of the agonist-bound receptor (Fig. 6F, blue). By contrast, although we observed almost complete dissociation of GCGR-G_s complexes by changes in FRET after the addition of GDP or GTPγS, (fig. S8C, orange and teal), no change in NBD fluorescence was observed after 10 min of incubation (Fig. 6E, orange and teal), indicating that TM6 does not return to the inactive state and remains open. But after 1 hour of incubation, the fluorescence intensity increased back to the level of the spectra of receptor bound to ligand only (fig. S8D). This suggests that the active state of GCGR persists for minutes after the dissociation of the G protein before it retracts back to the transmembrane bundle. A minute-scale-long active conformation of TM6 after Gα-α5 dissociation has also been reported previously for the β_2AR, although the duration time of 90 s was markedly shorter than the one observed here for GCGR (*47*). MD simulation suggests that TM6 remains in a more open conformation in the presence of ZP3780 than in the presence of the partial agonist (fig. S8E), suggesting that the efficacy of the ligand might play a role in stabilizing TM6 in the kinked outward conformation.

Receptor-mediated G_s dissociation of family B receptors is slower compared with family A GPCRs

The similarity in all recently solved family B GPCR-G_s complex structures indicate that the mechanism of receptor activation and receptor-mediated G protein signaling is conserved across family B receptors but distinct from family A GPCRs. However, so far no comprehensive studies have been conducted to compare the G protein signaling kinetics of different family B GPCRs among each other and with members of family A. We used the NanoBiT complementation-based G protein dissociation assay to measure the G_s dissociation kinetics of a set of family A and family B GPCRs in order to compare their G protein activation rates. In total, we tested three family B GPCRs, in addition to GCGR—GLP1, PAC1, and PTH—and four family A GPCRs, other than β_2AR, that couple primarily to G_s: 5-hydroxytryptamine receptor 4 (5-HT4), histamine H2 receptor (H2R), prostaglandin D2 receptor (DP), and vasopressin receptor 2 (V2R). As described above, we determined the maximum receptor-mediated G protein dissociation rates by measuring the time-dependent G_s dissociation at increasing ligand concentrations and fitting of the plotted concentration-dependent rates (fig. S9). Comparison of the maximum G_s dissociation rates shows that family B receptors in general catalyze a significantly slower G_s dissociation rate relative to all class B receptors tested (P = 0.019) (Fig. 7A and table S9). The expression levels of all GPCRs were determined for each experiment to ensure similar receptor densities in the cytoplasmic membrane (Fig. 7B). The differences in the G_s dissociation rates between the two GPCR families suggest a conserved mechanism of activation among family B GPCRs that is distinct to the previously described signaling mechanism of family A receptors.

Discussion

Here, we report the structure of the GCGR-G_s signaling complex bound to an engineered glucagon derivative with improved aqueous solubility. The structure provides a framework to explain the wealth of existing structure-activity relationship data for glucagon peptides that has accumulated since the discovery of glucagon by Kimball and Murlin almost 100 years ago (*48*). One hallmark of GPCR activation is the outward movement of TM6. In the GCGR-G_s complex, the conformational change of TM6 is accompanied by a partial unwinding of the secondary structure and formation of a sharp kink in the middle of the helix. This kink represents a characteristic structural feature seen in all recently solved family B GPCR-G_s protein complex structures, but not in family A complexes such as the β_2AR-G_s. Characterization of the functional impact of the structural differences observed between GCGR and β_2AR revealed that GCGR possesses a lower GEF activity compared with that of β_2AR. We show that the two main rate-limiting steps are the GDP release and GTP binding, which are significantly slower for GCGR-G_s than for the β_2AR-G_s complex. This suggests that GCGR induces a specific conformation of the receptor-coupled heterotrimeric G protein with different nucleotide sensitivity compared with G_s bound to β_2AR. Biophysical characterization of the ligand-induced activation of the GCGR indicates that the low GEF activity could be related to the mechanism of TM6 activation. Our

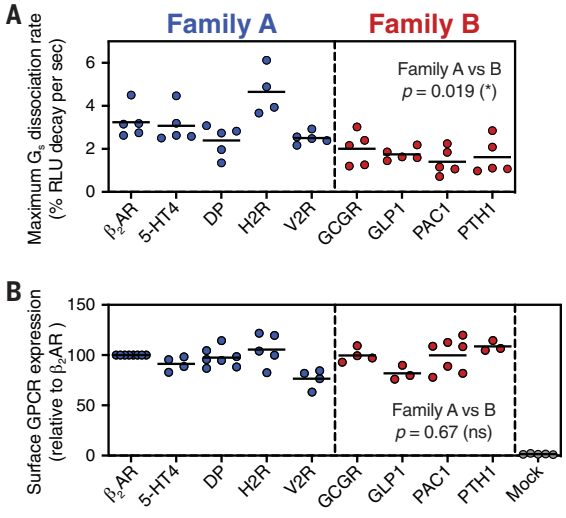


Fig. 7. Comparison of G_s activation kinetics between family A and family B GPCRs. (**A**) G_s dissociation rates among family A and family B GPCRs. GPCR ligand-induced G_s dissociation was measured by monitoring real-time, luminescent signals from NanoBiT-G_s protein and by fitting the plot to a one-phase decay curve. (**B**) Surface expression levels of tested GPCRs. Cell-surface GPCRs were fluorescently labeled with anti-FLAG epitope-tagged antibody and analyzed by means of flow cytometry. Each dot represents data from individual independent experiments. *P* values were obtained by means of two-tailed Student's *t* test with Welch's correction.

data shows a distinct activation mechanism of GCGR in which G_s coupling is necessary to induce kink formation and outward movement of TM6. This is in contrast to the β_2AR, where agonist binding alone promotes some outward movement of TM6 (*38*). Furthermore, once activated, TM6 of GCGR stays in the active kinked conformation longer than TM6 of β_2AR before it eventually relaxes back toward the inactive state.

On the basis of these results, we propose an activation model for GCGR and illustrate the effect of agonist, G protein, and GTP on the equilibrium of three different key functional states: inactive, intermediate, and fully active, with energy landscape diagrams (Fig. 8, A to C). In the unliganded (apo) state, GCGR predominantly exists in the low-energy inactive state, with a large energy barrier toward the intermediate and active states of the receptor, which is in agreement with the low reported basal activity of the receptor (*49*). By contrast, the β_2AR exhibits a moderate level of basal activity, suggesting a higher energy level in the apo state (Fig. 8C). Glucagon binding induces some conformational change on the intracellular side of the receptor without triggering TM6 outward movement. This promotes the formation of an "intermediate state" that we propose plays an important role for the initial engagement with the GDP-bound G protein. This is reflected in the energy landscape where agonist binding results in a decrease in the energy of the intermediate state, with little change in the energy of the fully active receptor conformation with the kinked and unraveled TM6. By contrast, DEER and NMR studies show that in β_2AR, agonist binding stabilizes an intermediate state in which TM6 moves outward and also produces a small proportion of the receptor in a fully active state (Fig. 8C) (*38*). Therefore, agonist binding to β_2AR not only lowers the energy of the intermediate state substantially but also decreases the energy of the fully active state.

G protein engagement and establishment of the nucleotide-free state of the GCGR-G_s complex is required to decrease the energy level of the active state. This leads to kink formation and movement of TM6 away from the receptor core in order to accommodate the C-terminal α5 helix of $G\alpha_s$ (Fig. 8, A and C, and fig. S4E). On the basis of the observation that the nucleotide release and subsequent GTP binding are the most rate-limiting steps in the process of GCGR-mediated G protein activation, we speculate that the intermediate and the fully active states are separated by a higher activation barrier for GCGR than for β_2AR (Fig. 8C). This most likely leads to a slow transition from the GDP-bound receptor complex to the nucleotide-free complex because of the high energy barrier for the unraveling and formation of the helix break in TM6. However, once outward movement of TM6 occurs, we propose that the G protein can engage the receptor more tightly, establishing the functionally important interactions with ICL2 and the intracellular cavity of GCGR that eventually trigger nucleotide release. Subsequent binding of GTP to the nucleotide-free complex, which results in G protein dissociation from the receptor, converts the energy landscape back to the one of the agonist-bound receptor in the absence of the G protein. However, on the basis of the proposed high energy barrier between the active and intermediate receptor conformation, TM6 of GCGR persists in the fully active conformation for an extended time compared with the β_2AR (Fig. 8, A and B). The prolonged active state of TM6 after disengagement of G_s might contribute to the persistent and sustained cAMP accumulation that was previously reported after glucagon addition to perfused rat livers (*13*).

In β_2AR, probably owing to the intermediate state already showing an outward displacement of TM6, full engagement of G_s occurs more rapidly in comparison with GCGR. This proposed rapid tight interaction of G_s with the β_2AR most likely leads to a faster decrease in GDP affinity and thus to a more rapid nucleotide release than for GCGR. As shown previously in DEER and NMR studies, G protein coupling further lowers the energy level of the active state of β_2AR to fully stabilize the receptor in the conformation observed in the x-ray structure of the nucleotide-free β_2AR-G_s complex (*38*, *50*). GTP-induced dissociation of the β_2AR signaling complex fully reverses the effect of the G protein, resulting in a similar energy landscape of the receptor as in the presence of agonist alone. In comparison with GCGR, the relaxation is relatively fast because of the lower energy barrier between the fully active and intermediate states (Fig. 8C).

Together, the observed differences in the signaling behavior of GCGR and β_2AR and the comprehensive comparison between the G protein activation kinetics of a number of family A and family B receptors suggest that the activation mechanism described for GCGR might be conserved in other members of family B GPCRs. Thus, our results provide insight into the distinct mechanism of receptor and G protein activation of family A and family B GPCRs.

Materials and methods

Solubility assay

Stock solutions of 2.0 mg ml^{-1} native glucagon and ZP3780 were prepared by dissolving the compounds in milliQ water adjusted to pH 2.5 by HCl. The concentration of compound was determined by absorbance measurement at 280 nm (Nanodrop 2000c, Thermo Scientific) using a calculated extinction coefficient of 8480 M^{-1} cm^{-1}. Aliquots of 50 μl buffer (table S1) were transferred to a clear bottom, UV compatible 96 wells plate and mixed with 50 μl aliquots of the peptide stock. The final peptide concentrations were 1 and 0.5 mg ml^{-1} and the final buffer concentration 50 mM. ZP3780 solubility was also determined at 5 mg ml^{-1} using a stock solution of 10 mg ml^{-1}. The samples were incubated at room temperature for 15 min and the UV absorbance at 325 nm was measured with an absorbance microplate reader (Spectramax 190). The samples were considered fully dissolved if the absorbance at 325 nm was below 0.02 arbitrary units (a.u.), which corresponds to 5 to 6 times the standard deviation of 8 buffer samples.

Aggregation assay

Samples were prepared according to the solubility assay, with the addition of a final concentration of 40 μM ThT. Samples were loaded

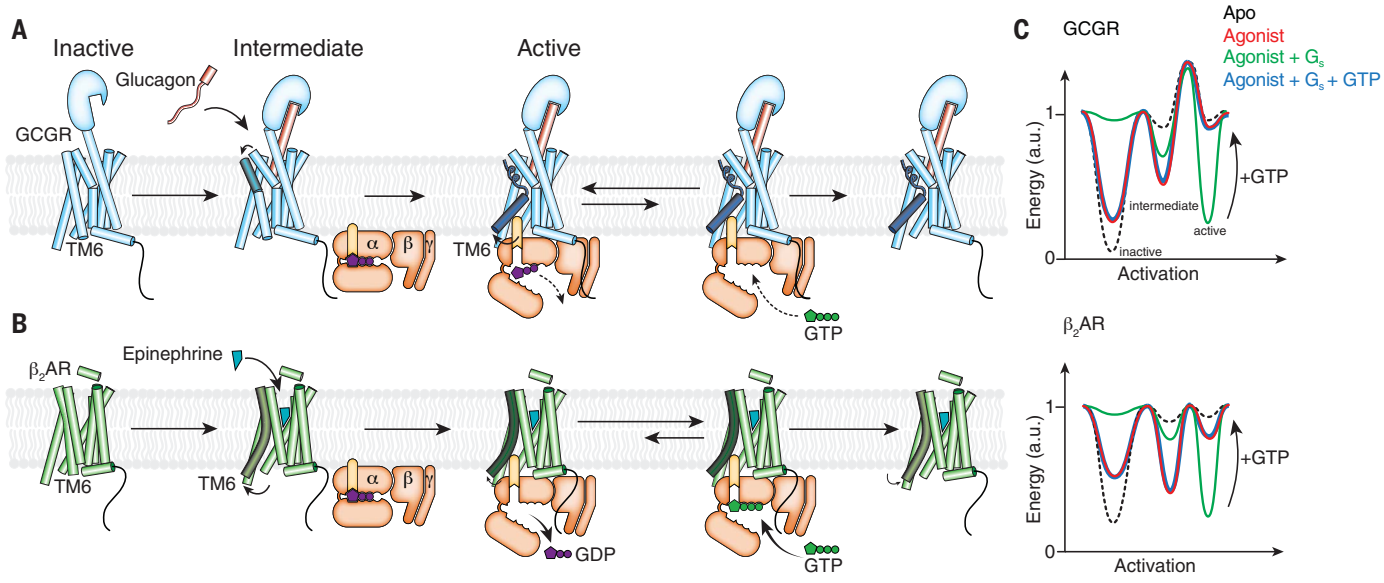


Fig. 8. Proposed model for GCGR activation and signaling in comparison with β_2AR. (**A**) Glucagon binding to GCGR induces conformational change on the extracellular side of the receptor (ECD, TM1, TM2, TM6, and TM7) without inducing outward movement of TM6 on the intracellular side. Coupling of GDP-bound G_s enables TM6 outward movement. The putative high energy required to produce the kinked and outward-moved TM6 may result in slower rates for the receptor-catalyzed nucleotide release of GCGR in comparison with β_2AR. Another rate-limiting step for GCGR-mediated G protein activation is GTP binding to the nucleotide-free G protein that leads to dissociation of the G protein from the receptor. After disengagement of the G protein, relaxation of TM6 to the inactive state is very slow, which might lead to the previously observed prolonged G protein signaling of GCGR in comparison with β_2AR (*13*). (**B**) β_2AR activation by an agonist increases the active state population of the receptor with an outward-moved TM6. G_s coupling to β_2AR fully stabilizes the active state and leads to rapid GDP release. The very transient nucleotide-free complex exhibits a high affinity for GTP that readily binds and dissociates the complex. After disengagement of the G protein, β_2AR relaxes back to the more conformational heterogeneous agonist-bound but G protein–free state. (**C**) Model of the simplified free energy landscapes for GCGR and β_2AR. Shown are the effects of agonist, G protein coupling, and GTP binding to the receptor–G protein complex on the equilibrium between the inactive and active states of the receptors.

in a 96-well black fluorescence plate (clear bottom) in triplicates of 150 µl in each well at ambient temperature and subsequently the plate was covered with UV compatible sealing tape to avoid evaporation. The plates were placed in a Spectramax Gemini XS fluorescence plate reader. The plate reader was programmed to read fluorescence from the top of the plate at 355 nm (excitation 295 nm) and at 485 nm (excitation 450 nm) at fixed intervals of 10 min for 96 hours at 40°C preceded by 300 s of automixing.

[^{125}I]-glucagon competition binding assay

Membranes containing the human GCGR (hGCGR) were prepared as described (*51*) from a HEK293 cell line stably overexpressing the hGCGR. In brief, for generation of the overexpressing cell line, cDNA encoding the hGCGR (P47871) was amplified by means of polymerase chain reaction (PCR) using primers encoding terminal restriction sites for subcloning. The 5′-end primer additionally contained a near Kozak consensus sequence. The fidelity of the DNA encoding the hGCGR was confirmed by DNA sequencing. The PCR product encoding the hGCGR was subcloned into a mammalian expression vector containing a neomycin (G418) resistance marker. The mammalian expression construct was transfected into HEK293 cells by a standard liposome transfection method. 48 hours post-transfection, cells were seeded for limited dilution cloning and selected with 0.5 mg ml^{-1} G418 in the culture medium. After 2 weeks, surviving colonies of hGCGR-expressing cells were picked, propagated and tested for cAMP accumulation. One hGCGR-expressing clone was selected and used for membrane preparation.

For the competition binding experiments 0.06 nM [^{125}I]-glucagon (PerkinElmer, NEX207) was added together with 10 µg hGCGR containing membranes, 0.5 mg SPA-beads (PerkinElmer, RPNQ0001) and varying concentrations of native glucagon or ZP3780 in a buffer consisting of 50 mM HEPES pH 7.4, 5 mM EGTA, 5 mM magnesium chloride, 0.005% Tween-20 and 0.05% casein (Sigma, C4765). After shaking at 600 rpm for 2 hours at room temperature the plate was transferred to a MicroBeta2 scintillation counter (PerkinElmer) and left for 4 hours to allow the SPA-beads to settle before quantifying ^{125}I radioactivity. To determine the respective K_i-values, data were fitted by nonlinear regression to a one-site competition model using GraphPad Prism.

Radioligand binding to receptor reconstituted HDL-particles

The binding capacity of HDL particles reconstituted with the β_2AR or the GCGR were determined by radioligand saturation binding. HDL particles were incubated in a buffer consisting of 20 mM Tris pH 7.5, 100 mM sodium chloride, 1 mM magnesium chloride and 0.013% *n*-dodecyl-β-D-maltopyranoside (DDM) with increasing concentrations of radioligand in absence or presence of 10 µM alprenolol or 50 µM ZP3780, respectively, to detect nonspecific binding, in a final volume of 100 µl in 5 ml falcon PP tubes (Corning #352063). For the β_2AR [^{3}H]-Dihydroalprenolol (DHA) with a specific activity of 105 Cimmol^{-1} (NET720, Perkin Elmer) was used. For the GCGR, [^{125}I]-glucagon with a specific activity of 2200 Cimmol^{-1} (NEX207, Perkin Elmer) was diluted with unlabeled native glucagon to a specific activity of around 27.7 Cimmol^{-1}. G_s was added in 38-fold excess of GCGR to detect the maximal specific binding capacity of GCGR. Reactions were incubated for 3 hours at room temperature. Then, receptor-bound radioligand was captured on 25 mm GF/B Whatman filters (Sigma #WHA1821025) by washing with a total of 7.5 ml icecold above mentioned buffer supplemented with 0.1% DDM. The GF/B filters were preincubated with 0.5% PEI at 4°C for 3 hours and washed twice with 2 ml icecold buffer containing 0.1% DDM prior to the capture on a vacuum manifold. Radioactivity was detected on a Hidex Triathler

scintillation counter. Specific binding was determined by subtraction of nonspecific binding counts from total counts and B_{max} and K_d determined by fitting data to a one-site binding isotherm. B_{max} of GCGR and β_2AR HDL particles was calculated as the ratio of captured sites (i.e., detectable sites) per reconstituted GCGR or β_2AR HDL particle. To determine the amount of G_s protein required to obtain maximal [^{125}I]-glucagon binding, HDL-particles were incubated with fixed amounts of GCGR and [^{125}I]-glucagon (8 nM) and increasing amounts of G_s. Specific binding was determined as described above.

Functional cAMP accumulation assay

The in vitro potencies and E_{max}s of glucagon and ZP3780 were assessed in a cAMP accumulation assay using HEK293 cells transiently transfected with the hGCGR. In brief, HEK293 cells were brought into suspension (0.25×10^6 cells ml^{-1}) by trypsination. For transfection of 1 ml of cells in suspension, a total DNA amount of 1 μg, consisting of 0.2 μg hGCGR (accession no P47871) expression vector DNA and 0.8 μg pcDNA3.1(zeo), in 25 μl OptiMEM was mixed with 3 μl FuGene6 in 57 μl OptiMEM and incubated for 20 min. Hereafter the 85 μl DNA/FuGene6 mix was added to 1 ml cell suspension, mixed and seeded in a clear poly-D-lysine coated 96-well in a volume of 100 μl/well (final cell density of 2.5×10^5 cells/well). The amounts were scaled up to the number of data points needed to generate concentration response curves for glucagon and ZP3780. Twenty-four hours after transfection the assay was performed by washing wells once in 100 μl HBSS buffer [HBSS Gibco 14025, 20 mM HEPES pH 7.5, 1 mM calcium chloride, 1 mM magnesium chloride and 0.1% bovine serum albumin (BSA)]. After the wash, 100 μl/well of HBSS buffer supplemented with IBMX (100 μM final concentration) were added, quickly followed by addition of 50 μl of HBSS buffer containing compound dilutions of glucagon and ZP3780. After incubation at 37°C for 15 min, the reactions were stopped by addition of 160 μl lysis buffer. The accumulated cAMP levels were quantified using the CisBIO cAMP dymanic HTRF kit (Cisbio, 62AM4PEC) according to the manufacturer's instructions and measured on an EnVision plate reader (PerkinElmer). Data were fitted by non-linear regression using GraphPad Prism.

Functional testing of the hGCGR mutations was, in general, conducted as described above. The constructs used for the mutational studies all contained a N-terminal FLAG-tag preceded by a HA-signal peptide and were inserted into the pcDNA3.1(neo) expression vector. The following DNA amounts were transfected to ensure similar surface expression to WT hGCGR: WT GCGR was transfected with 0.2 μg ml^{-1} cell suspension, Q232A and D385A GCGR with 0.05 μg DNA ml^{-1} cell suspension, and finally L242A, F322A, D370A, R378A, L395A, and Y400F GCGR were transfected with 0.035 μgml^{-1} cell suspension. Again, in all cases, the total DNA amount was adjusted to 1 μg ml^{-1} cell suspension by addition of pcDNA3.1(zeo). Twenty-four hours after transfection the assay was performed by washing wells once in 100 μl DPBS + 1 mM calcium chloride (DPBS Gibco 14190). After the wash, 50 μl/well HBSS buffer (HBSS Gibco 14025, 20 mM HEPES pH 7.5, 1 mM calcium chloride, 1 mM magnesium chloride and 0.1% BSA) was added and the plates were incubated for 30 min at 37°C. Hereafter, 50 μl/well glucagon dilutions were added, which were prepared in HBSS buffer supplemented with IBMX to a final concentration of 100 μM. After incubation at 37°C for 15 min, the reactions were stopped and the accumulated cAMP levels were quantified as described above.

Enzyme-linked immunosorbent assay

Surface expression levels of WT and mutated hGCGR variants were tracked using a direct enzyme-linked immunosorbent assay (ELISA) against the N-terminal FLAG tag. In brief, cells were transfected and seeded in a white Poly-D-Lysine-coated 96-well plate as described for the functional cAMP accumulation assay. Twenty-four hours after transfection, cells were washed with 100 μl/well DPBS + 1 mM calcium chloride and fixed with 50 μl/well 4% paraformaldehyde solution for 5 min at room temperature. The wells were washed twice with 100 μl DPBS + 1 mM calcium chloride and blocked with 100 μl/well blocking solution (3% drymilk, 1 mM calcium chloride 50 mM Tris-HCl, pH 7.5) for 30 min at room temperature. Hereafter 75 μl/well HRP-conjugated anti-FLAG antibody (Sigma Aldrich, A8592), diluted 1:2000 in blocking solution, was added and the plate was incubated for 1 hour at room temperature. The plate was washed four times with 100 μl/well blocking solution followed by four washes with 100 μl/well DPBS + 1 mM calcium chloride. The relative amount of present HRP was detected by adding 60 μl/well DPBS + 1 mM calcium chloride and 20 μl/well HRP substrate (Bio-Rad, 170-5060). The plate was incubated for 10 min at room temperature before recording of luminescence on an EnVision plate reader (Perkin Elmer).

Purification of GCGR

Human GCGR (Q27-F477) with N-terminal FLAG and C-terminal octahistidine tag was expressed in *Spodoptera frugiperda Sf9* insect cells using the baculovirus method (Expression Systems) in the presence of the L-168,049 (Tocris Bioscience). Cells were collected 48 hours after infection and stored at -80°C until further use. GCGR was extracted with 1% lauryl maltose neopentyl glycol (L-MNG; Anatrace), 0.1% cholesteryl hemisuccinate (CHS; Sigma) in 20 mM HEPES pH 7.5, 150 mM sodium chloride, 20% glycerol, and protease inhibitors and purified by nickel-chelating sepharose chromatography in the presence of the ligand NNC0640. The eluate from the nickel resin was supplemented with 2 mM calcium chloride and applied to the M1 anti-FLAG immunoaffinity resin and washed with progressively decreasing concentration of NNC0640. The receptor was eluted in a buffer consisting of 20 mM HEPES pH 7.5, 150 mM sodium chloride, 0.05% L-MNG, 0.005% CHS, FLAG peptide and 5 mM EDTA. The final step of purification was size exclusion chromatography on Superdex 200 10/300 gel filtration column (GE Healthcare) in 20 mM HEPES pH 7.5, 150 mM sodium chloride, 0.02% MNG and 0.002% CHS. Finally, apo-GCGR was concentrated to ~400 μM and stored at −80°C.

To perform DEER spectroscopy, human GCGR (Q27-F477) with N-terminal FLAG and C-terminal octahistidine tag was cloned into pcDNA-Zeo-tetO (*52*). Four native cysteine residues were mutated to make the minimal cysteine (mC) construct (C171T, C240A, C287A, C401V). Engineered cysteines were introduced pairwise into TM4/TM5 and TM4/TM6, respectively. Constructs were transfected into Expi293F (Thermo Fisher) cells stably expressing the tetracycline repressor using Expifectamine transfection kit (Thermo Fisher) following the manufacturers recommendations with the following modifications. Two days post-transfection, GCGR expression was induced with doxycycline (4 μg ml^{-1} and 5 mM sodium butyrate) in the presence of 1 μM L-168,049 (Tocris Bioscience). Cells were harvested 30 hours post-induction and stored at -80°C until use. Pellets were thawed and solubilized with 20 mM HEPES pH 7.5, 150 mM sodium chloride, 20% glycerol, 1% lauryl maltose neopentyl glycol (L-MNG, Anatrace), 0.1% CHS (Sigma), and protease inhibitors and purified by anti-FLAG immunoaffinity chromatography. From the FLAG resin, GCGR was eluted with 20 mM HEPES pH 7.5, 150 mM sodium chloride, 0.05% L-MNG, 0.005% CHS, FLAG peptide and 5 mM EDTA and labeled with 20-fold molar excess of bis-(2,2,5,5-tetramethyl-3-imidazoline-1-oxyl-4-yl)disulfide (IDSL) for 2 hours at room temperature. IDSL-labeled GCGR was purified from free spin label and buffer exchanged into 40 mM HEPES pH 7.5, 150 mM sodium chloride, 0.02% MNG and 0.002% CHS made with D_2O (Cambridge Isotopes) using size exclusion chromatography on Superdex 200 10/300 gel filtration column (GE Healthcare).

Expression and purification of heterotrimeric G_s

Heterotrimeric G_s was expressed and purified as previously described (*30*). Briefly, heterotrimeric G_s was expressed in *Trichoplusia ni* (*T. ni*) insect cells using baculoviruses generated by the BestBac method (Expression Systems).

Two separate baculoviruses were used, one encoding the human $G\alpha_s$ short splice variant and the other encoding both the $G\beta_1$ and $G\gamma_2$ subunits, with an histidine tag and HRV 3C protease site inserted at the amino terminus of the β-subunit. *T. ni* cells were infected with the baculoviruses followed by an incubation of 48 hours at 27°C. Cells were harvested by centrifugation and lysed in a buffer comprised of 10 mM Tris, pH 7.5, 100 μM magnesium chloride, 5 mM β-mercaptoethanol (βME), 20 μM GDP and protease inhibitors. The membrane fraction was collected by centrifugation and solubilized with a buffer comprised of 20 mM HEPES, pH 7.5, 100 mM sodium chloride, 1% sodium cholate, 0.05% DDM, 5 mM magnesium chloride, 5 mM βME, 5 mM imidazole, 20 μM GDP and protease inhibitors. The solubilization reaction was incubated for 45 min at 4°C after homogenization with a Dounce homogenizer. After centrifugation, the soluble fraction was loaded onto Ni-chelated sepharose followed by a gradual detergent exchange into 0.1% DDM.

The protein was eluted in buffer supplemented with 200 mM imidazole and dialyzed overnight in 20 mM HEPES, pH 7.5, 100 mM sodium chloride, 0.1% DDM, 1 mM magnesium chloride, 5 mM βME and 20 μM GDP together with HRV 3C protease to cleave off the amino- terminal 6xHis tag. The cleaved 6xHis tag, uncleaved fractions and 3C protease were removed by Ni-chelated sepharose and the G protein was dephosphorylated by lambda protein phosphatase (NEB), calf intestinal phosphatase (NEB), and antarctic phosphatase (NEB) in the presence of 1 mM manganese chloride. Lipidated G_s heterotrimer was isolated using a MonoQ 10/100 GL column (GE Healthcare). After binding of the protein to the column in buffer A [20 mM hydroxyethylpiperazine ethane sulfonic acid (HEPES), pH 7.5, 50 mM sodium chloride, 1 mM magnesium chloride, 0.05% DDM, 100 μM *tris*(2-carboxyethyl)phosphine (TCEP), and 20 μM GDP], the column was washed with buffer A and the G protein heterotrimer was eluted with a linear gradient of 0–50% buffer B (buffer A with 1 M sodium chloride). The main peak containing isoprenylated G protein heterotrimer was collected and the protein was dialyzed into 20 mM HEPES, pH 7.5, 100 mM soidum chloride, 0.02% DDM, 100 μM TCEP and 20 μM GDP. After concentrating the protein to 250 μM, 20% glycerol was added and the protein was flash frozen in liquid nitrogen and stored at −80°C until use.

Purification of Nb35

Nb35 was expressed and purified as described previously (*50*). Briefly, Nb35 was expressed in *Escherichia coli* BL21 cells. After lysis, it was purified on a nickel affinity chromatography and finally subjected to size exclusion chromatography on a Superdex 200 10/300 gel filtration column (GE Healthcare) in 20 mM HEPES pH 7.5, 150 mM sodium chloride. Purified Nb35 was concentrated, flash frozen, and stored at −80°C until further use.

Formation and purification of the GCGR-G_s-Nb35 complex

Purified $G\alpha_s\beta_1\gamma_2$ in 0.02% DDM was incubated with 1% MNG for one hour on ice and, simultaneously, GCGR was incubated with 5-fold molar excess ZP3780 (500 μM) at room temperature. A 1.5-fold molar excess of detergent-exchanged $G\alpha_s\beta_1\gamma_2$ was incubated with ZP3780-bound GCGR at room temperature for two hours, after which a 2-fold molar excess (in terms of $G\alpha_s\beta_1\gamma_2$) of Nb35 was added and incubated on ice for 1.5 hours. To stabilize the nucleotide-free complex, Apyrase (1 unit, NEB) was added and incubated overnight at 4°C. A 4-fold volume of 20 mM HEPES pH 7.5, 100 mM sodium chloride, 0.8% L-MNG/0.08% CHS, 0.27% GDN/0.027% CHS, 1 mM magnesium chloride, 5 μM ZP3780, and 2 mM calcium chloride was added to the complexing reaction and complex was purified by M1 anti-FLAG affinity chromatography. The complex was eluted in 20 mM HEPES pH 7.5, 100 mM sodium chloride, 0.01% MNG/0.001% CHS, 0.0033% GDN (Anatrace)/0.00033% CHS, 5 μM ZP3780, 5 mM EDTA, and FLAG peptide. The eluted complex was supplemented with 100 μM TCEP and subjected to size exclusion chromatography on a Superdex 200 10/300 Increase column in 20 mM HEPES pH 7.5, 100 mM sodium chloride, 5 μM ZP3780, 0.00075% MNG, and 0.00025% GDN. Peak fractions were concentrated to ~16 mg ml^{-1} for electron microscopy studies.

Cryo-EM data acquisition and data processing

An aliquot of 3.5 μl GCGR-G_s-Nb35 complex was applied to glow-discharged 200 mesh grids (Quantifoil R1.2/1.3), at a concentration of 17 mg ml^{-1} and subsequently vitrified using a Vitrobot Mark IV (Thermo Fischer Scientific) at 100% humidity and 4°C. CryoEM images were collected on a Titan Krios operated at 300 kV at a nominal magnification of 130,000× with a Gatan GIF Quantum LS Imaging energy filter using a Gatan K2 Summit direct electron camera in counted mode, corresponding to a pixel size of 1.06 Å. A total number of 3724 movie stacks were obtained with a dose rate of 7 e^-/pix/s and total exposure time of 8 s with 0.2 s per frame, resulting in a total dose of 50 electron per $Å^2$. The defocus range was set to 1.2–2.2 μm.

Dose-fractionated image stacks were subjected to beam-induced motion correction using MotionCor2 (*53*). Contrast transfer function parameters for each micrograph were determined by Gctf v1.06 (*54*). Data processing was performed in RELION3.0 (*55*). A total number of 2,039,910 particles were selected from a template-based auto-picking. A subset of 296,516 particles were selected after one round of two-dimensional (2D) and 3D classification. Particles projections from micrographs with signal better than 3.5 Å were subjected to Bayesian polishing, CTF refinement and 3D reconstruction. The final subset of 266,267 particles was imported to cisTEM for 3D local refinement, resulting in a reconstruction with global resolution of 3.1 Å at FSC 0.143 (*56*). Local resolution was determined using the Bsoft package at cutoff FSC of 0.5 (*57*).

Model building and refinement

The initial template of GCGR was derived from the crystal structure of GCGR (PDB 5YQZ) (*17*). The structural coordinates of PDB 5VAI (*4*) were used as initial models for the G_s and Nb35. Models were docked into the EM density map using *UCSF Chimera* (*58*), followed by iterative manual building in *Coot* (*59*). The final model was subjected to global refinement and minimization in real space using *phenix.real_space_refine* in *Phenix* (*60*). *Molprobity* was used to evaluate model geometry (*61*). FSC curves were calculated between the resulting model and the half map used for refinement as well as between the resulting model and the other half map for cross-validation.

β_2AR purification

β_2AR was expressed in *Sf9* insect cells using the baculovirus method (Expression Systems) and purified as described previously (*50*).

Purification of $G\alpha_s$ and $G_{\beta\gamma}$ for GDP release assay

Human $G\alpha_s$ subunit with an amino-terminal 6x histidine tag followed by an HRV 3C protease site were expressed in Rosetta 2 (DE3) cells (EMD Millipore) using pET28a. Cells were grown in Terrific Broth to OD600 of 0.6, and protein expression was induced by addition of 0.5 mM IPTG. After 15 hours of incubation at room temperature, cells were harvested and resuspended in lysis buffer (50 mM HEPES pH 7.5, 100 mM sodium chloride, 1 mM magnesium chloride, 50 μM GDP, 5 mM βME, 5 mM imidazole, and protease inhibitors). Cells were disrupted by sonification using a 50% duty cycle, 70% power for four times 45 s. Intact cells and cell debris were subsequently removed by centrifugation and the supernatant was incubated with Ni-chelated sepharose for 1.5 hours at 4°C. The Ni-chelated sepharose resin was washed multiple times with lysis buffer in batch and then loaded into a wide-bore glass column, and protein was eluted with lysis buffer containing 200 mM imidazole. The eluted protein was dialyzed overnight in dialysis buffer (20 mM HEPES pH 7.5, 100 mM sodium chloride, 1 mM magnesium chloride, 20 μM GDP, 5 mM βME, and

5 mM imidazole). The amino terminal histidine tag was cleaved by adding 1:1000 w/w 3C protease into the dialysis bag. Uncleaved protein, cleaved histidine tag, and 3C protease were subsequently removed by incubation with Ni-chelated sepharose for 45 min at 4°C. The resin was loaded into a wide-bore glass column and the flow-through containing the Gα subunit was collected. The protein was concentrated and run on a Superdex 200 10/300 GL column in SEC buffer (20 mM HEPES pH 7.5, 100 mM sodium chloride, 1 mM magnesium chloride, 20 μM GDP, and 100 μM TCEP).

The Gβγ heterodimer was expressed in *Tni* cells using a baculovirus generated by the BestBac (Expression systems) method. We used one virus containing the genes for $G\beta_1$ and $G\gamma_2$ subunits. The sequence for the Gβ subunit contains an amino terminal 6x histidine tag followed by an HRV 3C protease sequence, allowing us to cleave off the histidine tag after purification. *T. ni* cells were infected at a density of 3.0×10^6 cells ml^{-1} and incubated at 27°C for 48 hours. After centrifugation, cells were resuspended in lysis buffer (10 mM Tris, pH 7.5, 5 mM βME, and protease inhibitors). The membrane fraction was collected by centrifugation and solubilized by a dounce homogenizer with a buffer comprised of 20 mM HEPES pH 7.5, 100 mM sodium chloride, 1.0% sodium cholate, 0.05% DDM, 5 mM βME, and protease inhibitors. The solubilization reaction was stirred at 4°C for 40 min, and then centrifuged to remove insoluble debris. Ni-chelating sepharose was added to the supernatant and stirred for 1.5 hours at 4°C followed by multiple washes of the Ni-chelating sepharose resin in batch with solubilization buffer. The resin was collected into a wide-bore glass column, and the detergent was gradually exchanged from sodium cholate to 0.1% DDM. The protein was eluted with Ni-chelating sepharose elution buffer (20 mM HEPES pH 7.5, 100 mM sodium chloride, 0.1% DDM, 5 mM βME, 200 mM imidazole and protease inhibitors). To cleave off the amino terminal histidine tag, HRV 3C protease (1:1000 w/w) was added and the sample dialyzed overnight in 20 mM HEPES pH 7.5, 100 mM sodium chloride, 0.1% DDM, 5 mM βME, and 5 mM imidazole. The cleaved histidine tag, uncleaved fractions, and HRV 3C protease were removed by loading of the sample on Ni-chelating sepharose resin and the flow-through containing the Gβγ heterodimer was collected. Lambda protein phosphatase (2000 units, NEB), calf intestinal phosphatase (10 units, NEB), and Antarctic phosphatase (5 units, NEB) were added together with 1 mM manganese chloride, followed by a 1 hour incubation at 4°C. After adjusting the sodium chloride concentration to 50 mM using dilution buffer composed of 20 mM HEPES pH 7.5, 0.02% DDM, and 100 μM TCEP, the sample was passed through a 0.22 μm filter and loaded onto a MonoQ 10/100 GL column (GE Healthcare) equilibrated in buffer A (20 mM HEPES pH 7.5, 50 mM sodium chloride, 0.02% DDM, and 100 μM TCEP). The column was washed with 5 CV of buffer A and bound protein was eluted over 7.5 CV with a linear gradient from 0–25% buffer B (buffer A with 1 M sodium chloride). The fractions containing prenylated heterodimer were pooled and dialyzed in buffer containing 20 mM HEPES pH 7.5, 100 mM sodium chloride, 0.02% DDM, 100 μM TCEP. After concentrating the protein, 20% glycerol was added and protein aliquots were flash frozen in liquid nitrogen.

HDL reconstitution

Reconstitution of GCGR and β_2AR into high-density lipoprotein particles was performed using a 3:2 mixture of POPC/POPG and MSP1E3D1 as belt protein. Receptor:lipid:MSP was mixed at 1:10:60 and incubated for 2 hour at 4°C. Biobeads (Biorad) at a ratio of lipid: beads, 1:8 was added and rotated at 4°C for 2 hours. Further addition of biobeads at ratio of 1:8 of lipid:beads was done and incubated overnight, rotating at 4°C. Reconstituted HDL-GCGR and HDL-β_2AR were separated from empty discs by FLAG-M1 chromatography and further purified by SEC using a Superdex 200 10/300 Increase column into 20 mM HEPES pH 7.5, 100 mM sodium chloride.

GTP turnover assay

The GTP turnover assay was performed by using a modified protocol of the GTPase-Glo™ assay (Promega) as described previously (*30*) with the following modification. The final reaction buffer consisted of 20 mM Hepes pH 7.5, 100 mM sodium chloride, 10 mM magnesium chloride, 100 uM TCEP, 0.02% DDM, 0.1 μM GDP, 10 μM GTP. Before the reaction was started, HDL-reconstituted GCGR (300 nM) or β_2AR (300 nM) was incubated in the presence of 100 μM full agonist (ZP3780 or epinephrine) in buffer containing 20 mM HEPES pH 7.5, 100 mM sodium chloride, 20 μM GTP. After incubation for 60 min at room temperature, G protein (500 nM) was added in buffer containing 20 mM HEPES pH 7.5, 100 mM sodium chloride, 20 mM magnesium chloride, 200 μM TCEP, 0.04% DDM, and 0.2 μM GDP. After incubation for a given time (Fig. 5C), reconstituted GTPase-Glo reagent supplemented with 10 μM adenosine 5′-diphosphate (ADP) was added to the sample and incubated for 30 min at room temperature. Luminescence was measured after the addition of detection reagent and incubation for 10 min at room temperature using a SpectraMax Paradigm plate reader. The GTP concentration in the samples was calculated by using the relative light units (RLUs) of GTP at different concentrations and the data was analyzed using *GrapPad Prism*.

To calculate the maximum rate for the GTP turnover, the GTPase-Glo™ assay (Promega) was performed at different concentrations of receptor reconstituted into HDL particles (GCGR: 0–7.1 uM; β_2AR: 0–7.5 uM). After addition of G_s at concentrations of 0.1 μM and 0.5 μM to β_2AR and GCGR, respectively, reactions proceeded for 10 min (β_2AR) or 90 min (GCGR) before addition of GTPase Glo reagent and ADP, as described in the manufacturer's protocol. Rates were calculated using the RLUs of a GTP standard curve and analyzed using *GraphPad Prism*.

GDP release assay

The single turnover GDP release assay was performed with GCGR and β_2AR reconstituted into HDL particles and purified G protein subunits ($G\alpha_s$ and Gβγ) in detergent. For this purpose, $G\alpha_s$ in 20 mM HEPES pH 7.5, 100 mM sodium chloride, 1 mM magnesium chloride, 100 μM TCEP, and 5 μM GDP was diluted to 0.8 μM in GDP-loading buffer (20 mM HEPES pH 7.5, 150 mM sodium chloride, 1 mM EDTA pH 8.0, 100 μM TCEP) and [^{3}H]-GDP (39.8 $Cimmol^{-1}$, Perkin Elmer) was added to a final concentration of 2.5 μM. After incubation for 60 min at room temperature, Gβγ in 20 mM HEPES pH 7.5, 100 mM sodium chloride, 0.08% DDM, 100 μM TCEP was added with a 1.2-fold molar access and incubated for 30 min to allow formation of the heterotrimer. GCGR and β_2AR in HDL particles were diluted to 10 μM in 20 mM HEPES pH 7.5, 100 mM sodium chloride, and 2 mM GTP and incubated with 100 μM of agonist (ZP3780 or epinephrine) for 1 hour at room temperature. The reaction was started by mixing the ligand-bound receptor with [^{3}H]-GDP loaded heterotrimer to reach final concentrations of 200 nM G_s and 5 μM receptor. At various time points, 500 μl of ice-cold wash buffer (20 mM HEPES pH 7.5, 150 mM sodium chloride, and 20 mM magnesium chloride) was added to 20 μl of the reaction and the mixture was immediately filtered using a microanalysis filter holder (EMD Milipore) and pre-wet mixed cellulose filters (25 mm, 0.22 μm). The filter was washed three times with 500 uL ice-cold wash buffer and dried for at least 1 hour at room temperature. The amount of radioactivity that remained bound to the filters was determined by liquid scintillation spectrometry. GDP release data were analyzed using GraphPad Prism.

DEER

Sample preparation

IDSL-labeled GCGR in detergent or HDL was diluted to 300 nM and incubated with 500 μM ZP3780 for one hour on ice. Labeled-GCGR was then concentrated to ~100 μM and additional 100 μM ZP3780 was added and frozen with 20% D8-glycerol (Camebridge Isotope). Sample was transferred to 1.4/1.7 mm (i.d./o.d.)

borosilicate capillaries (VitroCom, Mountain Lakes, NJ) and flash frozen.

Data Collection

DEER data were acquired on a Bruker ELEXSYS E580 spectrometer equipped with a SuperQ-FT bridge and 10 W AmpQ amplifier operated at Q-band frequency (~33.5 GHz) with an ER 5107D2 resonator (Bruker Biospin). Sample temperature was maintained at 50 K during data collection using a recirculating helium cryo-cooling system (ColdEdge Technologies). A dead-time free four-pulse DEER sequence (*62*) was used with an 8-step suppression of the deuterium nuclear modulation (*63*) and 8-step phase cycle. The pump pulse (rectangular; ~36 ns) was applied at the maximum of the low-field line of the absorbance spectrum. The observe pulses (rectangular; ~20 ns π/2 and ~40 ns π) were applied at a −70 MHz offset from the pump pulse. Optimal pulse lengths for each DEER experiment were determined by an echo nutation experiment. Background correction and model-free fitting analysis of dipolar evolution functions were performed with the program LongDistances933 (developed by Christian Altenbach and available at www.biochemistry.ucla.edu/biochem/Faculty/Hubbell/). The regularization (smoothness) parameter for distance distribution determination was selected using the L-curve criterion (*64*), and the background signal was fitted assuming a three-dimensional distribution of remote spins. The upper limit of reliable distance (r) and width determination (σ) for each mutant in nanometers was calculated using the following equations (*36*):

$$r_{\max,\langle r\rangle} \approx 5^{3}\sqrt{t_{\max}/2\mu s}$$

$$r_{\max,\langle \sigma\rangle} \approx 4^{3}\sqrt{t_{\max}/2\mu s}$$

where t_{max} is the maximum time domain recorded for each sample. All distributions were area-normalized and plotted in GraphPad Prism.

Bimane and NBD Fluorescence

Sample preparation

Receptors (minimal cysteine-GCGR (mC-GCGR) 349C and minimal cysteine-β_2AR (mC-β_2AR) 265C) at 10 μM were incubated with 5-molar excess of bimane or NBD at room temperature for one hour. After concentration the labeled receptors were further purified, to remove excess label, using size exclusion chromatography on a Superdex 200 10/300 Increase column in 20 mM HEPES pH 7.5, 100 mM sodium chloride, 0.01% MNG/0.001% CHS.

Data Collection

Bimane- and NBD-labeled receptor were used at a concentration of 0.1 μM in buffer containing 20 mM HEPES, pH 7.5, 100 mM sodium chloride, 0.01% MNG/0.001% CHS. The concentration of G_s added was 10 μM. Fluorescence data was collected in a 150 μl cuvette with a FluorEssence v3.8 software on a Fluorolog instrument (Horiba) in photon-counting mode. Bimane fluorescence was measured by excitation at 370 nm with excitation and emission bandwidth passes of 4 nm. The emission spectra were recorded from 410 to 510 nm with 1 nm increment and 0.1 s integration time. NBD was measured by excitation at 420 nm with excitation and emission bandwidth passes of 4 nm, and the emission spectra were recorded from 510 nm to 610 nm with 1 nm increment and 0.1 s integration time.

The statistical analysis for the λ_{max} change on ligand and G_s binding, was carried out by independently fitting a single Gaussian between 440 and 485 nm for each of the three measurements of 349C-bimane-labeled GCGR in the Apo state, ZP3780-bound and G_s- bound states. All fluorescence profiles were well-represented by the single Gaussians ($R^2 > 0.98$). Fitted λ_{max} values were taken as the mean of the Gaussian fits and subsequent statistical analyses were performed on the triplicate values for each experimental condition. The λ_{max} values for each condition were compared with unpaired t tests assuming a Gaussian distribution using Welch's correction so as to not assume equal standard deviations.

Ensemble FRET

Sample preparation

Receptors (minCys-GCGR S265C and minCys-β_2AR) at 10 μM were incubated with 5-fold molar excess of Cy3B maleimide (GE Healthcare) at room temperature for 45 min. After quenching with cysteine, size exclusion chromatography on a Superdex 200 10/300 Increase column in 20 mM HEPES pH 7.5, 100 mM sodium chloride, 0.01% MNG/0.001% CHS was performed to remove excess label.

G_s at 10 μM were incubated with 5 molar equivalence of Sulfo-Cy5-NHS ester (Lumiprobe) at room temperature for 45 min before quenching with TRIS-HCl pH 7.5. To remove excess label, size exclusion chromatography was performed on a Superdex 200 10/300 Increase column in 20 mM HEPES pH 7.5, 100 mM sodium chloride, 0.05% DDM/0.001% CHS. 10 μM GDP, 1 mM $MgCl_2$ and 100 μM TCEP.

Data Collection

Cy3B and Sulfo-Cy5 labeled receptors and G_s, respectively, were used to measure the rate of association using FRET. Fluorescence spectra was recorded on a Fluorolog instrument (Horiba) in photon-counting mode. All data were collected with emission (567 nm) and excitation (520 nm) bandpass of 2 nm and integration time of 0.1 s nm^{-1}. The baseline was recorded with Cy3B-labeled receptor (100 nM) till it stabilized (~ 100 s), then 1 μM Sulfo-Cy5-labeled G_s was added and the drop in fluorescence intensity was recorded as a function of time. When the spectra reached steady-state (~800 s), to analyze G_s dissociation GDP or GTPγS was added and the spectra was monitored till 1200 s.

Bodipy-GTPγS binding

Bodipy-GTPγS is a non-hydrolyzable fluorescently labeled GTP analog, which is self-quenched in solution. However, upon binding to G proteins, Bodipy-GTPγS fluorescence increases due to unquenching of the fluorophore. This property was used to investigate the rate of GTP binding to nucleotide-free GPCR–G protein complexes. Nucleotide-free GCGR-G_s and β_2AR-G_s complexes were prepared as described in this manuscript and previously (*50*). For the nucleotide-binding experiment, fluorescence form BODIPY-FL-GTPγS (Thermo Fisher) was recorded in 500 uL quarz cuvettes using the Fluorolog spectrophotometer (HORIBA). The fluorophore was exited at 495 nm and emission was detected at 508 nm at 22°C. All experiments were performed in a buffer comprised of 20 mM HEPES, pH 7.5, 100 mM sodium chloride, 0.01%L-MNG, 0.001% CHS, 10 mM magnesium chloride, and 100 μM TCEP. Typically, kinetics data were collected for 150 nM BODIPY-FL-GTPγS in the absence of receptor-G protein complexes for 100 s to establish the baseline fluorescence intensity. Receptor-G protein complexes were added to 1 μM and rapidly mixed in the fluorescence cuvette without halting data collection. The resulting kinetics spectra were plotted and fit to a one-phase association function using GraphPad Prism.

System setup and protocols for molecular dynamics simulations

We performed all-atom MD simulations for both ZP3780- and NNC1702-bound GCGR. Simulations were initiated from the cryo-EM structure after removing the G protein. The NNC1702-bound structure was obtained by deleting H1 and mutating D9 to E. An initial system containing the peptide bound GCGR and POPC membrane was assembled with Visual Molecular Dynamics 1.9.3 (VMD) (*65*) Membrane Builder module with 0.15 M sodium chloride in the solution. Bilayer dimensions were chosen to maintain at least a 30Å buffer between protein images in the x–y plane and a 20Å buffer between protein images in the z direction. Final system dimensions were approximately 90 × 90 × 130 $Å^3$. There were 5 replicas for each system.

Simulations were performed with the Compute Unified Device Architecture (CUDA) version of NAMD 2.12 (*66*), the CHARMM36m force field (*67*) for proteins, lipids and ions, and TIP3P water. Long-range electrostatic forces were calculated with the Particle Mesh

Ewald method (*68*), and an interaction cutoff of 12 Å was applied within periodic boundary conditions. Van der Waals (vdW) forces were smoothly shifted to zero between 10 and 12 Å. Equations of motion were integrated with a time step of 2 fs. Langevin dynamics with a damping coefficient of 5 ps^{-1} was used to keep the temperature at 310 K. The pressure was maintained at 1 atm using a Nosé–Hoover –Langevin piston (*69*, *70*). Bond lengths were constrained using the SHAKE algorithm (*71*).

Simulation systems were equilibrated in three stages. In stage one, a 1000 step minimization followed by 0.5 ns simulation in which everything except lipid tails was fixed was performed to equilibrate lipid layer. In stage two, another 1000 step minimization followed by a 1 ns equilibration with the protein constrained was performed to permit the environment to relax. In stage three, a 5-ns-long simulation was performed to release harmonic constraints and to further equilibrate the whole system.

After equilibration, each replicate was simulated for 300 ns. Atomic coordinates of all atoms were saved every 2 ps. We visualized and aligned trajectories by using VMD to eliminate all rotational and translational motions.

NanoBiT–G protein dissociation assay

GPCR-induced G_s dissociation was measured by a NanoBiT–G protein dissociation assay (*33*), in which interaction between a Gα subunit and a Gβγ subunit was monitored by a NanoLuc-based enzyme complementation system called NanoBiT (Promega). Specifically, a NanoBiT-G_s protein consisting of $G\alpha_s$ subunit fused with a large fragment (LgBiT) at the alpha helical domain and an N-terminally small fragment (SmBiT)-fused $G\gamma_2$ subunit with a C68S mutation was expressed along with untagged $G\beta_1$ subunit, RIC8B and a test GPCR. HEK293A cells (Thermo Fisher Scientific) were seeded in a 6-well culture plate at a concentration of 2×10^5 cells ml^{-1} [2 ml per well in Dulbecco's modified Eagle's medium (DMEM; Nissui) supplemented with 10% fetal bovine serum (Gibco), glutamine, penicillin, and streptomycin] 1 day before transfection. Transfection solution was prepared by combining 5 μl (per well in a 6-well plate hereafter) of polyethylenimine Max solution (Polysciences; 1 mg ml^{-1}), 200 μl of Opti-MEM (Thermo Fisher Scientific) and a plasmid mixture consisting of 100 ng LgBiT-containing $G\alpha_s$ subunit, 500 ng $G\beta_1$, 500 ng SmBiT-fused $G\gamma_2$ (C68S) 100 ng, RIC8B and a test GPCR that was optimized to match an expression level. After incubation for 1 day, transfected cells were harvested with 0.5 mM EDTA-containing Dulbecco's PBS, centrifuged and suspended in 2 ml of HBSS containing 0.01% bovine serum albumin (BSA; fatty acid–free grade; SERVA) and 5 mM HEPES pH 7.4 (assay buffer). The cell suspension was dispensed in a white 96-well plate at a volume of 80 μl per well and loaded with 20 μl of 50 μM coelenterazine (Carbosynth) diluted in the assay buffer. After 2 hours of incubation at room temperature, the plate was measured for baseline luminescence (Spectramax L, Molecular Devices) and a titrated test ligand (20 μl; 6X of final concentrations) were manually added. Ligands and their sources were as follows: (-)-isoproterenol hydrochloride (Sigma-Aldrich) for β_2AR; serotonin (5-hydroxytryptamine hydrochloride; FUJIFILM Wako Pure Chemical) for 5-HT_4; prostaglandin D_2 (Cayman Chemical) for DP; histamine dihydrochloride (FUJIFILM Wako Pure Chemical) for H_2R; arginine vasopressin (Peptide Institute) for V_2R; glucagon (*Zealand Pharma*) for GCGR; glucagon-like peptide 1 (7-37) (Peptide Institute) for GLP-1; PACAP27 (Peptide Institute) for PAC1; parathyroid hormone (1-34) (Peptide Institute) for PTH. The plate was immediately read at room temperature for the following 3 min at a measurement interval of 7 s with an accumulation time of 0.17 s per read. The kinetics luminescence counts were normalized to the initial count and fold-change signals over vehicle treatment were used to plot G protein dissociation response. G protein dissociation kinetics were calculated by fitting the normalized luminescent data to a one-phase dissociation model built in Prism 8 software (GraphPad Prism). Initial G protein dissociation speed was calculated by a formula of (Plateau – 1)*K where "Plateau" represents a saturated normalized luminescent counts whereas "K" denotes a rate constant in a unit that is reciprocal of time (sec). The resulting dissociation speed data were fitted to a three-parameter sigmoidal concentration-response curve ("Bottom" is fixed to a value of 0), from which an E_{max} value ("Top") was used to represent G protein dissociation for a given experiment. When a sigmoidal curve did not converge due to a lack of saturating data points, a dissociation speed at the highest ligand concentration was used as an E_{max} value.

Flow cytometry analysis

HEK293A cells (Thermo Fisher Scientific) were seeded in a 6-well culture plate and transfection was performed as following the same procedure as described in the "NanoBiT–G protein dissociation assay" section. Our optimization of GPCR plasmid volume to match surface expression level found following GPCR to be used in the study (volume is denoted as per well in the 6-well plate): β_2AR (200 ng), 5-HT_4 (40 ng), DP (200 ng), H_2R (20 ng), V_2R (500 ng), GCGR (40 ng), GLP-1R (40 ng), PAC1R (500 ng), PTH1R (40 ng). Except for PAC1R, all of the eight GPCRs (β_2AR, 5-HT_4, DP, H_2R, V_2R, GCGR, GLP-1R, PTH1R) were N-terminally fused to a hemagglutinin signal sequence (KTIIALSYIFCLVFA) and a FLAG epitope (DYKDDDDK) tag with or without a linker sequence between the FLAG epitope tag and a GPCR. For construction of the Class A GPCRs, a full-length coding sequence except for the initial methionine was used. For the N-terminal modification, a signal sequence of the Class B GPCRs was removed as following: GCGR (amino acid positions 1-24), GLP-1R (1-23) and PTH1R (1-23). A PAC1R construct contained the FLAG epitope tag at the amino acid position between 23 and 24 of the native PAC1R sequence. One day after transfection, the cells were collected by adding 200 μl of 0.53 mM EDTA-containing Dulbecco's PBS (D-PBS), followed by 200 μl of 5 mM HEPES pH 7.4-containing Hank's Balanced Salt Solution (HBSS). The cell suspension was transferred to a 96-well V-bottom plate in duplicate and fluorescently labeled with an anti-FLAG epitope (DYKDDDDK) tag monoclonal antibody (Clone 1E6, FujiFilm Wako Pure Chemicals; 10 μg ml^{-1} diluted in 2% goat serum- and 2 mM EDTA-containing D-PBS (blocking buffer)) and a goat anti-mouse IgG secondary antibody conjugated with Alexa Fluor 488 (Thermo Fisher Scientific, 10 μg ml^{-1} diluted in the blocking buffer). After washing with D-PBS, the cells were resuspended in 200 μl of 2 mM EDTA-containing-D-PBS and filtered through a 40-μm filter. The fluorescent intensity of single cells was quantified by an EC800 flow cytometer equipped with a 488 nm laser (Sony). The fluorescent signal derived from Alexa Fluor 488 was recorded in an FL1 channel, and the flow cytometry data were analyzed with the FlowJo software (FlowJo). Live cells were gated with a forward scatter (FS-Peak-Lin) cutoff at the 390 setting, with a gain value of 1.7. Values of mean fluorescence intensity (MFI) from approximately 20,000 cells per sample were used for analysis.

Figure preparation

Figures were created using the *PyMOL* Molecular Graphics System, Version 2.20 Schrödinger, (http://pymol.org), and the UCSF *Chimera X* package (*72*). In vitro graphs were created using *GraphPad Prism*.

REFERENCES AND NOTES

1. Y.-L. Liang *et al*., Cryo-EM structure of the active, G_s-protein complexed, human CGRP receptor. *Nature* **561**, 492–497 (2018). doi: 10.1038/s41586-018-0535-y; pmid: 30209400
2. Y.-L. Liang *et al*., Phase-plate cryo-EM structure of a biased agonist-bound human GLP-1 receptor-Gs complex. *Nature* **555**, 121–125 (2018). doi: 10.1038/nature25773; pmid: 29466332
3. Y.-L. Liang *et al*., Phase-plate cryo-EM structure of a class B GPCR-G-protein complex. *Nature* **546**, 118–123 (2017). doi: 10.1038/nature22327; pmid: 28437792
4. Y. Zhang *et al*., Cryo-EM structure of the activated GLP-1 receptor in complex with a G protein. *Nature* **546**, 248–253 (2017). doi: 10.1038/nature22394; pmid: 28538729
5. L.-H. Zhao *et al*., Structure and dynamics of the active human parathyroid hormone receptor-1. *Science* **364**, 148–153 (2019). doi: 10.1126/science.aav7942; pmid: 30975883

6. G. Jiang, B. B. Zhang, Glucagon and regulation of glucose metabolism. *Am. J. Physiol. Endocrinol. Metab.* **284**, E671–E678 (2003). doi: 10.1152/ajpendo.00492.2002; pmid: 12626323
7. K. Hollenstein *et al*., Insights into the structure of class B GPCRs. *Trends Pharmacol. Sci.* **35**, 12–22 (2014). doi: 10.1016/j.tips.2013.11.001; pmid: 24359917
8. L. J. Jelinek *et al*., Expression cloning and signaling properties of the rat glucagon receptor. *Science* **259**, 1614–1616 (1993). doi: 10.1126/science.8384375; pmid: 8384375
9. K. E. Mayo *et al*., International Union of Pharmacology. XXXV. The glucagon receptor family. *Pharmacol. Rev.* **55**, 167–194 (2003). doi: 10.1124/pr.55.1.6; pmid: 12615957
10. T. D. Müller, B. Finan, C. Clemmensen, R. D. DiMarchi, M. H. Tschöp, The New Biology and Pharmacology of Glucagon. *Physiol. Rev.* **97**, 721–766 (2017). doi: 10.1152/physrev.00025.2016; pmid: 28275047
11. P. E. Cryer, S. N. Davis, H. Shamoon, Hypoglycemia in diabetes. *Diabetes Care* **26**, 1902–1912 (2003). doi: 10.2337/diacare.26.6.1902; pmid: 12766131
12. Y. M. Cho, C. E. Merchant, T. J. Kieffer, Targeting the glucagon receptor family for diabetes and obesity therapy. *Pharmacol. Ther.* **135**, 247–278 (2012). doi: 10.1016/j.pharmthera.2012.05.009; pmid: 22659620
13. J. H. Exton, G. A. Robison, E. W. Sutherland, C. R. Park, Studies on the role of adenosine 3′,5′-monophosphate in the hepatic actions of glucagon and catecholamines. *J. Biol. Chem.* **246**, 6166–6177 (1971). pmid: 4331382
14. J. S. Pedersen *et al*., The changing face of glucagon fibrillation: Structural polymorphism and conformational imprinting. *J. Mol. Biol.* **355**, 501–523 (2006). doi: 10.1016/j.jmb.2005.09.100; pmid: 16321400
15. U. Hövelmann *et al*., Pharmacokinetic and pharmacodynamic characteristics of dasiglucagon, a novel soluble and stable glucagon analog. *Diabetes Care* **41**, 531–537 (2018). doi: 10.2337/dc17-1402; pmid: 29273578
16. C. G. Unson, Molecular determinants of glucagon receptor signaling. *Biopolymers* **66**, 218–235 (2002). doi: 10.1002/bip.10259; pmid: 12491536
17. H. Zhang *et al*., Structure of the glucagon receptor in complex with a glucagon analogue. *Nature* **553**, 106–110 (2018). doi: 10.1038/nature25153; pmid: 29300013
18. C. M. Koth *et al*., Molecular basis for negative regulation of the glucagon receptor. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14393–14398 (2012). doi: 10.1073/pnas.1206734109; pmid: 22908259
19. H. Zhang *et al*., Structure of the full-length glucagon class B G-protein-coupled receptor. *Nature* **546**, 259–264 (2017). doi: 10.1038/nature22363; pmid: 28514451
20. D. Wootten, J. Simms, L. J. Miller, A. Christopoulos, P. M. Sexton, Polar transmembrane interactions drive formation of ligand-specific and signal pathway-biased family B G protein-coupled receptor conformations. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5211–5216 (2013). doi: 10.1073/pnas.1221585110; pmid: 23479653
21. C. G. Unson, D. Andreu, E. M. Gurzenda, R. B. Merrifield, Synthetic peptide antagonists of glucagon. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4083–4087 (1987). doi: 10.1073/pnas.84.12.4083; pmid: 3035568
22. C. G. Unson, E. M. Gurzenda, R. B. Merrifield, Biological activities of des-His1[Glu9]glucagon amide, a glucagon antagonist. *Peptides* **10**, 1171–1177 (1989). doi: 10.1016/0196-9781(89)90010-7; pmid: 2560175
23. C. G. Unson, D. Macdonald, K. Ray, T. L. Durrah, R. B. Merrifield, Position 9 replacement analogs of glucagon uncouple biological activity and receptor binding. *J. Biol. Chem.* **266**, 2763–2766 (1991). pmid: 1847133
24. E. Dal Maso *et al*., Extracellular loops 2 and 3 of the calcitonin receptor selectively modify agonist binding and efficacy. *Biochem. Pharmacol.* **150**, 214–244 (2018). doi: 10.1016/j.bcp.2018.02.005; pmid: 29454620
25. F. Y. Siu *et al*., Structure of the human glucagon class B G-protein-coupled receptor. *Nature* **499**, 444–449 (2013). doi: 10.1038/nature12393; pmid: 23863937
26. C. de Graaf *et al*., Extending the structural view of class B GPCRs. *Trends Biochem. Sci.* **42**, 946–960 (2017). doi: 10.1016/j.tibs.2017.10.003; pmid: 29132948
27. A. Glukhova *et al*., Rules of engagement: GPCRs and G proteins. *ACS Pharmacol. Transl. Sci.* **1**, 73–83 (2018). doi: 10.1021/acsptsci.8b00026; pmid: 32219204
28. Y. Du *et al*., Assembly of a GPCR-G protein complex. *Cell* **177**, 1232–1242.e11 (2019). doi: 10.1016/j.cell.2019.04.022; pmid: 31080064
29. A. Manglik, A. C. Kruse, Structural Basis for G Protein-Coupled Receptor Activation. *Biochemistry* **56**, 5628–5634 (2017). doi: 10.1021/acs.biochem.7b00747; pmid: 28967738
30. G. G. Gregorio *et al*., Single-molecule analysis of ligand efficacy in β_2AR-G-protein activation. *Nature* **547**, 68–73 (2017). doi: 10.1038/nature22354; pmid: 28607487
31. F. J. Rojas, L. Birnbaumer, Regulation of glucagon receptor binding. Lack of effect of Mg and preferential role for GDP. *J. Biol. Chem.* **260**, 7829–7835 (1985). pmid: 2989262
32. R. Seifert, U. Gether, K. Wenzel-Seifert, B. K. Kobilka, Effects of guanine, inosine, and xanthine nucleotides on β_2-adrenergic receptor/G_s interactions: Evidence for multiple receptor conformations. *Mol. Pharmacol.* **56**, 348–358 (1999). doi: 10.1124/mol.56.2.348; pmid: 10419554
33. A. Inoue *et al*., Illuminating G-protein-coupling selectivity of GPCRs. *Cell* **177**, 1933–1947.e25 (2019). doi: 10.1016/j.cell.2019.04.044; pmid: 31160049
34. D. S. Serafin, N. R. Harris, N. R. Nielsen, D. I. Mackie, K. M. Caron, Dawn of a New RAMPage. *Trends Pharmacol. Sci.* **41**, 249–265 (2020). doi: 10.1016/j.tips.2020.01.009; pmid: 32115276
35. E. Lorenzen *et al*., Multiplexed analysis of the secretin-like GPCR-RAMP interactome. *Sci. Adv.* **5**, eaaw2778 (2019). doi: 10.1126/sciadv.aaw2778; pmid: 31555726
36. G. Jeschke, DEER distance measurements on proteins. *Annu. Rev. Phys. Chem.* **63**, 419–446 (2012). doi: 10.1146/annurev-physchem-032511-143716; pmid: 22404592
37. L. Sušac, M. T. Eddy, T. Didenko, R. C. Stevens, K. Wüthrich, A_{2A} adenosine receptor functional states characterized by ^{19}F-NMR. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12733–12738 (2018). doi: 10.1073/pnas.1813649115; pmid: 30463958
38. A. Manglik *et al*., Structural insights into the dynamic process of β2-adrenergic receptor signaling. *Cell* **161**, 1101–1111 (2015). doi: 10.1016/j.cell.2015.04.043; pmid: 25981665
39. J. Okude *et al*., Identification of a conformational equilibrium that determines the efficacy and functional selectivity of the μ-opioid receptor. *Angew. Chem. Int. Ed.* **54**, 15771–15776 (2015). doi: 10.1002/anie.201508794; pmid: 26568421
40. R. Sounier *et al*., Propagation of conformational changes during μ-opioid receptor activation. *Nature* **524**, 375–378 (2015). doi: 10.1038/nature14680; pmid: 26245377
41. L. M. Wingler *et al*., Angiotensin analogs with divergent bias stabilize distinct receptor conformations. *Cell* **176**, 468–478.e11 (2019). doi: 10.1016/j.cell.2018.12.005; pmid: 30639099
42. L. Ye, N. Van Eps, M. Zimmer, O. P. Ernst, R. S. Prosser, Activation of the A2A adenosine G-protein-coupled receptor by conformational selection. *Nature* **533**, 265–268 (2016). doi: 10.1038/nature17668; pmid: 27144352
43. J. F. Fay, D. L. Farrens, Purification of functional CB_1 and analysis by site-directed fluorescence labeling methods. *Methods Enzymol.* **593**, 343–370 (2017). doi: 10.1016/bs.mie.2017.06.026; pmid: 28750810
44. C. T. Schafer, J. F. Fay, J. M. Janz, D. L. Farrens, Decay of an active GPCR: Conformational dynamics govern agonist rebinding and persistence of an active, yet empty, receptor state. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11961–11966 (2016). doi: 10.1073/pnas.1606347113; pmid: 27702898
45. X. J. Yao *et al*., The effect of ligand efficacy on the formation and stability of a GPCR-G protein complex. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9501–9506 (2009). doi: 10.1073/pnas.0811437106; pmid: 19470481
46. R. O. Dror *et al*., Activation mechanism of the β2-adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18684–18689 (2011). doi: 10.1073/pnas.1110499108; pmid: 22031696
47. T. M. Gupte, M. Ritt, M. Dysthe, R. U. Malik, S. Sivaramakrishnan, Minute-scale persistence of a GPCR conformation state triggered by non-cognate G protein interactions primes signaling. *Nat. Commun.* **10**, 4836 (2019). doi: 10.1038/s41467-019-12755-9; pmid: 31645561
48. C. P. Kimball, J. R. Murlin, Aqueous extracts of pancreas III. Some precipitation reactions of insulin. *J. Biol. Chem.* **58**, 337–346 (1923).
49. Y. Yin *et al*., Rearrangement of a polar core provides a conserved mechanism for constitutive activation of class B G protein–coupled receptors. *J. Biol. Chem.* **292**, 9865–9881 (2017). doi: 10.1074/jbc.M117.782987; pmid: 28356352
50. S. G. F. Rasmussen *et al*., Crystal structure of the β2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549–555 (2011). doi: 10.1038/nature10361; pmid: 21772288

ACKNOWLEDGMENTS

We thank S. Reedtz-Runge (Novo Nordisk A/S) for sharing the ligand NNC0640. We thank K. Sato, S. Nakano, and A. Inoue (Tohoku University) for their assistance of plasmid preparation and cell-based GPCR assays. We thank P. Aschauer for initial characterization of glucagon analogs. **Funding:** D.H. was supported by the German Academic Exchange Service (DAAD). K.K.K. was supported by the American Diabetes Association (ADA) Postdoctoral Fellowship. A.I. was funded by the PRIME 18 g5910013 from the Japan Agency for Medical Research and Development (AMED) and the Japan Society for the Promotion of Science (JSPS) KAKENHI 17K08264. The work is supported by NIH grants R01GM083118 and R01NS028471 to B.K.K.; B.K.K. is a Chan Zuckerberg Biohub investigator. **Author contributions:** D.H. and K.K.K. developed the procedure for and prepared the GCGR-G_s complex; modeled the structure in the cryo-EM map; performed HDL reconstitutions; prepared DEER samples; and performed the GTP-turnover assay, GDP release assay, Bodipy GTPγS assay, NBD fluorescence, and ensemble FRET. H.H. froze grids, obtained cryo-EM data, processed data, and obtained the cryo-EM map. M.F.P. and J.M.M. performed the radioligand binding assays, ELISA, and functional cAMP assays. E.S.O. performed bimane spectroscopy and helped with DEER sample preperation. L.G. and J.M.M. identified the soluble glucagon analog and performed the solubility and aggregation assays. M.L. supervised and performed, with C.J., DEER experiments and data acquisition and analysis. A.I. performed the NanoBiT–G protein dissociation assay and analyzed the surface expression of receptors by means of flow cytometry analysis. G.E. performed and analyzed molecular dynamics simulations. D.H., K.K.K., J.M.M., G.S. and B.K.K. wrote the manuscript with input from all the authors. **Competing interests:** B.K.K. is a cofounder of and consultant for ConfometRx. J.M.M. and L.G. are employees of Zealand Pharma. M.F.P. is a former employee of Zealand Pharma. **Data and materials availability:** The cryo-EM density map has been deposited in the Electron Microscopy Data Bank (EMDB) under accession code EMD-21866, and model coordinates have been deposited in the Protein Data Bank (PDB) under accession no. 6PWP. All other data are available in the manuscript or supplementary materials. Materials described in this study are available upon request sent to the corresponding authors.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/eaba3373/suppl/DC1
Figs. S1 to S9
Tables S1 to S9
References (*51*–*73*)

View/request a protocol for this paper from *Bio-protocol*.

27 November 2019; accepted 17 June 2020
10.1126/science.aba3373

RESEARCH ARTICLE

BIOCHEMISTRY

ANGEL2 is a member of the CCR4 family of deadenylases with 2′,3′-cyclic phosphatase activity

Paola H. Pinto[1], Alena Kroupova[2], Alexander Schleiffer[3], Karl Mechtler[4,5], Martin Jinek[2], Stefan Weitzer[1]*, Javier Martinez[1]*

RNA molecules are frequently modified with a terminal 2′,3′-cyclic phosphate group as a result of endonuclease cleavage, exonuclease trimming, or de novo synthesis. During pre-transfer RNA (tRNA) and unconventional messenger RNA (mRNA) splicing, 2′,3′-cyclic phosphates are substrates of the tRNA ligase complex, and their removal is critical for recycling of tRNAs upon ribosome stalling. We identified the predicted deadenylase angel homolog 2 (ANGEL2) as a human phosphatase that converts 2′,3′-cyclic phosphates into 2′,3′-OH nucleotides. We analyzed ANGEL2's substrate preference, structure, and reaction mechanism. Perturbing ANGEL2 expression affected the efficiency of pre-tRNA processing, X-box–binding protein 1 (*XBP1*) mRNA splicing during the unfolded protein response, and tRNA nucleotidyltransferase 1 (TRNT1)–mediated CCA addition onto tRNAs. Our results indicate that ANGEL2 is involved in RNA pathways that rely on the ligation or hydrolysis of 2′,3′-cyclic phosphates.

Cellular RNAs are heavily modified during synthesis and posttranscriptionally in response to internal and external stimuli. Enzymes that sculpt the sequence and chemistry of RNA termini include 5′ capping and decapping enzymes, 5′ kinases, endonucleases, exonucleases, poly(A) and poly(U) polymerases, and deadenylases. Independent of its identity, a nucleotide at the 3′ end of an RNA molecule displays chemical groups including: 2′,3′-OH (hydroxyl), 2′-P (phosphate)/3′-OH, 2′-OH/3′-P, or 2′,3′>P (2′,3′-cyclic phosphate). In particular, 2′,3′>Ps are generated through endonucleolytic cleavage (*1*, *2*), exonucleolytic trimming (*3*, *4*) or de novo by the RNA 3′-terminal phosphate cyclase (RTCA) (*5*).

In mammalian cells, RNA molecules ending with 2′,3′>P are substrates for the tRNA ligase complex during pre-tRNA processing (*6*) and X-box–binding protein 1 (*XBP1*) mRNA splicing in the unfolded protein response (UPR) (*7*–*9*). Additionally, 2′,3′>P termini are generated during the ribosome-associated quality control (RQC) pathway upon ribosome stalling (*10*) and during stress-induced tRNA cleavage by angiogenin (*11*). Hydrolysis of 2′,3′>P yields 2′-P/3′-OH (*12*) or 2′-OH/3′-P (*13*). Enzymes known to convert 2′,3′>P into 2′,3′-OH include bacteriophage T4 polynucleotide kinase (PNK) (*14*) and other RNA repair enzymes from the DxDxT acyl phosphatase superfamily, such as baculoviral RNA ligase 1 (Rnl1) (*15*). Examples from bacteria include the histidine-aspartate (HD)–Pnk RNA healing enzymes (*16*) and Pnkp from the binuclear metallophosphoesterase superfamily (*17*). However, such activity has not been detected in eukaryotic cells.

[1]Max Perutz Labs, Medical University of Vienna, Vienna BioCenter, 1030 Vienna, Austria. [2]Department of Biochemistry, University of Zurich, Zurich, Switzerland. [3]IMP/IMBA Bioinformatics Core Facility, Research Institute of Molecular Pathology, Vienna BioCenter, 1030 Vienna, Austria. [4]Research Institute of Molecular Pathology, Vienna BioCenter, 1030 Vienna, Austria. [5]Institute of Molecular Biotechnology (IMBA), Austrian Academy of Sciences, Vienna BioCenter, 1030 Vienna, Austria.
*Corresponding author. Email: stefan.weitzer@meduniwien.ac.at (S.W.); javier.martinez@meduniwien.ac.at (J.M.)

ANGEL2 is a mammalian 2′,3′-cyclic phosphatase

To search for an RNA 2′,3′>Pase activity in human cells, we established an assay using a radiolabeled, short double-stranded RNA (dsRNA) substrate. The labeled strand, with 5′-P and 2′,3′>P termini, was annealed to a complementary strand containing a 5′-amino linker moiety (L) to block interstrand ligation by the human tRNA ligase complex (*6*) (Fig. 1A, left). A reaction in HeLa cell extracts yielded a slower-migrating band on a denaturing gel, whose mobility was similar to a band identified from treatment of the dsRNA substrate with T4 PNK (Fig. 1A, right), suggesting RNA 2′,3′>Pase activity. We followed the activity through six purification steps (Fig. 1B) and subjected the activity peak and surrounding fractions of a glycerol gradient centrifugation (Fig. 1C) to tandem mass spectrometry. Candidate genes were repressed by RNA interference (RNAi) in HeLa cells. Knockdown of ANGEL2 led to a large decrease in 2′,3′>Pase activity (table S1 and fig. S1).

Human ANGEL2 (UniProtKB Q5VTE6), a 544–amino acid, 62-kDa protein, is a homolog of yeast carbon catabolite repression 4 (CCR4), a member of the CCR4-NOT (negative on TATA-less) deadenylase complex (*18*). Of the five human CCR4 homologs only two, CNOT6 (CCR4a) and CNOT6L (CCR4b), are alternatively assembled into the CCR4-NOT complex (*19*). Nocturnin (CCR4c), ANGEL2 (CCR4d), and ANGEL1 (CCR4e) were considered to be deadenylases (*20*) because they contain the conserved magnesium (Mg^{2+})–dependent nuclease domain shared by members of the exonuclease, endonuclease, and phosphatase (EEP) superfamily and characteristic of the CCR4 family. However, their N-terminal regions lack the leucine-rich repeat region required for incorporation into the CCR4-NOT complex (*18*, *19*). Purified ANGEL2, expressed as a FLAG-hemagglutinin (FLAG-HA)–tagged protein in human cells, was unable to remove poly(A) tails (fig. S2, A to C). In contrast to ANGEL2, RNAi-mediated knockdown of ANGEL1 and CNOT6 did not abrogate 2′,3′>Pase activity (Fig. 1D and fig. S3).

Wild-type ANGEL2 expressed as a glutathione *S*-transferase (GST) fusion protein in *Escherichia coli* showed intrinsic 2′,3′>Pase activity, indicating that ANGEL2 alone performs the reaction (fig. S4, A and B). Reaction products were ligated to a 5′-P–ended RNA by T4 RNA ligase, a result only possible in the presence of a 3′-OH–containing RNA (fig. S5, A and B). The substrate specificity of ANGEL2 was assessed using RNA substrates with 2′,3′-OH, 2′-P/3′-OH, 2′-OH/3′-P, or 2′,3′>P termini. ANGEL2 reacted only with 2′-P and 2′,3′>P RNAs to yield 2′,3′-OH products, which suggests that 2′-P is a reaction intermediate (Fig. 1E). Because the N-terminal region of ANGEL2 is predicted to be unstructured, and because the full-length protein has poor expression yield and solubility, we also expressed and purified an N-terminally truncated (ΔN) version of ANGEL2 ($ANGEL2^{119-544}$) (fig. S6, A and B). Michaelis-Menten kinetics analysis performed with ANGEL2 ΔN and a 2′,3′>P RNA oligonucleotide resulted in Michaelis constant (K_M), maximal velocity (V_{max}), and catalytic constant (k_{cat}) values of 0.3282 μM, 5.026 nM/s, and 2.513 s^{-1}, respectively (Fig. 1F and fig. S6, C and D).

Crystal structure of ANGEL2

To analyze the 2′,3′>Pase activity of the conserved EEP domain in ANGEL2 (residues 169 to 544) at the structural level, we crystallized ANGEL2 ΔN in its apo and substrate-bound forms. The structure of the apo enzyme was determined by molecular replacement and refined to a resolution of 1.45 Å (table S2). We cocrystallized ANGEL2 ΔN in complex with Mg^{2+} and the substrate analog adenosine-2′,3′-vanadate (AV), which mimics the pentavalent transition state assumed upon 2′,3′>P hydrolysis, resulting in a structure at 2.10 Å resolution (table S2).

ANGEL2 adopts a conserved two-layer α/β sandwich fold typical of EEP family proteins (Fig. 2A) (*21*–*23*). Structural alignment of ANGEL2 with CCR4b (CNOT6L; PDB ID 3ngo), phosphodiesterase 12 (PDE12; PDB ID 4z0v), and apurinic/apyrimidinic endonuclease (APE1; PDB ID 1de9) yielded superpositions with root mean square deviations (RMSD) of 2.2 Å (over 337 Cα atoms), 2.2 Å (over 434 Cα atoms), and

3.0 Å (over 276 Cα atoms), respectively (fig. S7A). The core of the EEP domain is formed by two conserved β sheets and surrounded by two outer layers of α helices and loops. The active site of apo ANGEL2 ΔN contains one Mg^{2+} ion coordinated by the carboxyl group of the invariant Glu^{220} and five water molecules, two of which form hydrogen bonds with Asp^{532} and one with Asn^{173} (Fig. 2B).

In the transition-state mimic complex, the adenine base of the AV ligand forms π-stacking interactions with Tyr^{313}. The 2′,3′-cyclic vanadate moiety, a mimic of 2′,3′>P, is bound by direct coordination of the 2′-O and O3V atoms to Mg^{2+}, as well as by hydrogen bonding contacts from the 3′-O atom to Asn^{353} and His^{310}, and between O3V and His^{533} (Fig. 2C and fig. S7B). The residues involved in the coordination of the Mg^{2+} ion and substrate binding are evolutionarily invariant in EEP family enzymes, except for Tyr^{313}, which is replaced by phenylalanine in some enzymes (fig. S7C). Analysis of electrostatic surface potential revealed a conserved, positively charged groove lined with residues Arg^{246}, Arg^{249}, Lys^{250}, Asn^{285}, Arg^{283}, and Asn^{314}, which may have a role in binding the ribose-phosphate backbone of RNA substrates (fig. S7D).

High conservation of the overall fold and active site of the EEP family enzymes contrasts with their distinct biochemical activities. Unlike CNOT6L, which hydrolyzes the phosphodiester group linking a 3′-terminal adenosine to the penultimate nucleotide in an RNA substrate, ANGEL2 acts on a terminal phosphate. Superposition of the ANGEL2 and CNOT6L active sites reveals a loop in ANGEL2, spanning residues 374 to 390, inserted in the substrate-binding cleft (fig. S7, E and F). Here, Gln^{382} of ANGEL2 occupies the position corresponding to the 3′-terminal nucleotide-binding site in CNOT6L, thus rendering it inaccessible in ANGEL2. In addition, an aromatic residue in CNOT6L (Phe^{484}) that mediates a π-stacking interaction with the adenine base of the 3′-terminal nucleotide is not conserved in ANGEL2. As a result, the ANGEL2 active-site pocket accommodates the 3′-terminal nucleotide of an RNA substrate in the −1 position (upstream of the catalytic site), whereby the active-site residues are positioned to catalyze the hydrolysis of the terminal 2′,3′>P. A similar substrate-binding mode is observed in other 3′-terminal phosphatases of the EEP family, including human inositol polyphosphate-5-phosphatase

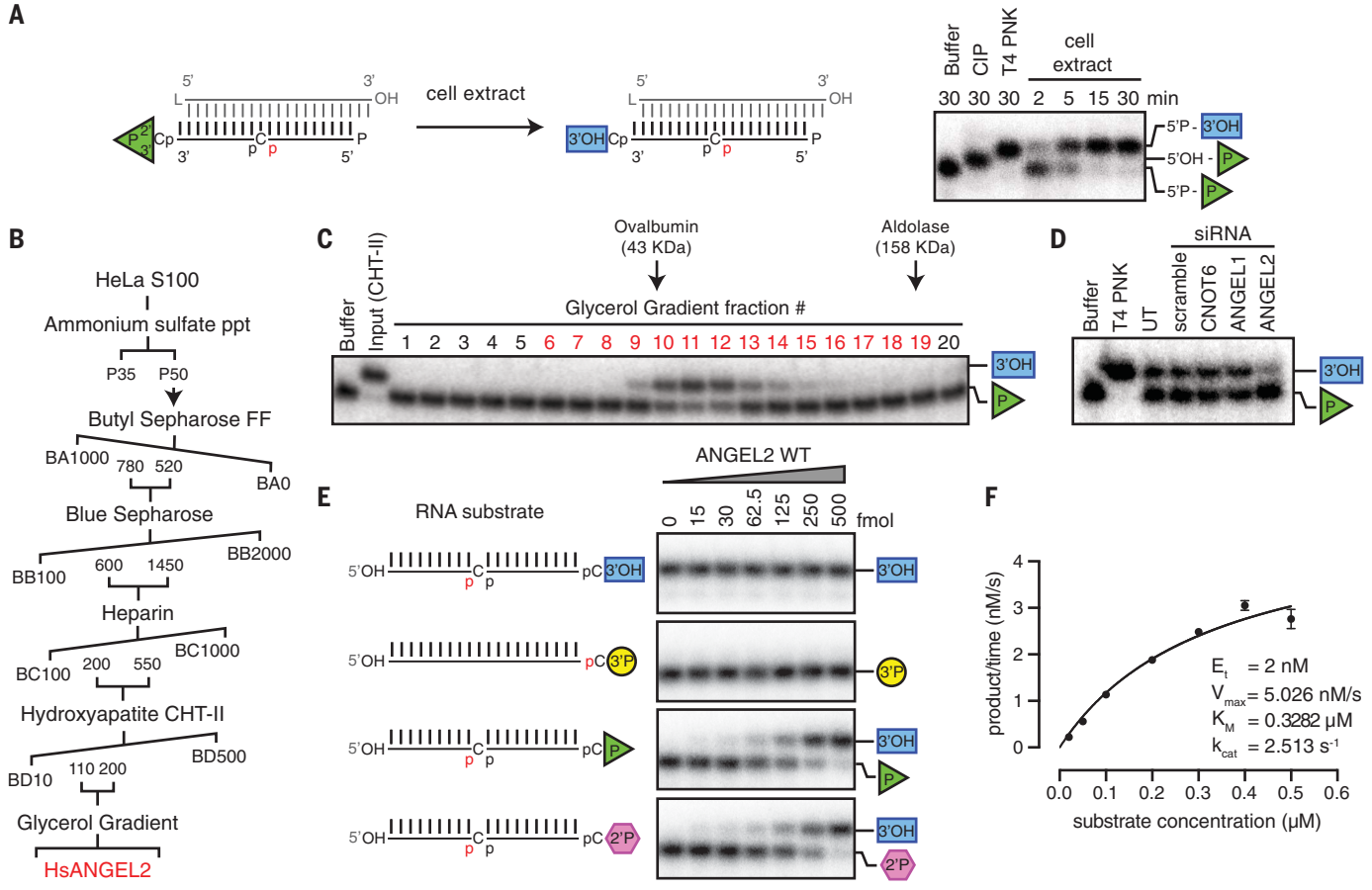


Fig. 1. Identification of ANGEL2 as a human 2′,3′-cyclic phosphatase. (**A**) Biochemical assay to measure 2′,3′>Pase activity. Incubation of a 5′-P/2′,3′>P short, internally radiolabeled RNA duplex with HeLa cell extracts (left; red denotes radioactive ^{32}P) resulted in a migration shift when resolved by denaturing gel electrophoresis (right). As controls, the substrate RNA was incubated with buffer or treated with calf intestinal phosphatase (CIP, which dephosphorylates terminal phosphates except 2′,3′>P) or T4 PNK (which phosphorylates 5′-OH and dephosphorylates any phosphate at the 3′ end). L, amino linker. (**B**) Fractionation scheme for partial purification of a 2′,3′>Pase activity from HeLa S100 cytoplasmic extracts. FF, fast flow; ppt, precipitation; CHT-II, ceramic hydroxyapatite type II. (**C**) Activity assay after separation through glycerol gradient centrifugation. The input, combined active fractions from the CHT-II column, was used as a positive control. Fractions in red were analyzed by mass spectrometry. (**D**) Extracts from HeLa cells untransfected (UT) or transfected with control small interfering RNA (siRNA; scramble) or siRNAs complementary to ANGEL2 or homologous proteins CNOT6 and ANGEL1 were assayed for 2′,3′>Pase activity. (**E**) Radiolabeled RNA substrates (red denotes ^{32}P) with 2′,3′-OH, 2′-OH/3′-P, 2′,3′>P, or 2′-P/3′-OH termini were incubated with recombinant ANGEL2 and analyzed by denaturing gel electrophoresis. (**F**) Michaelis-Menten kinetics of ANGEL2 ΔN. Reaction rates were estimated as in fig. S6. Michaelis-Menten parameters were obtained by nonlinear regression. K_{cat} values were obtained by dividing the V_{max} value by the concentration of ANGEL2 (E_t).

B (INPP5B; fig. S7G) (*24*). Overall, the ANGEL2 structure confirms that 3′-terminal phosphatases of the EEP family contain residues that occlude the +1 nucleotide-binding position, which distinguishes them from the endo-/exonucleases and facilitates correct substrate positioning with respect to the catalytic residues and the divalent cation.

ANGEL2 generates a 2′-P reaction intermediate

To dissect the reaction mechanism of ANGEL2, we performed structure-guided mutational analysis of selected conserved residues in

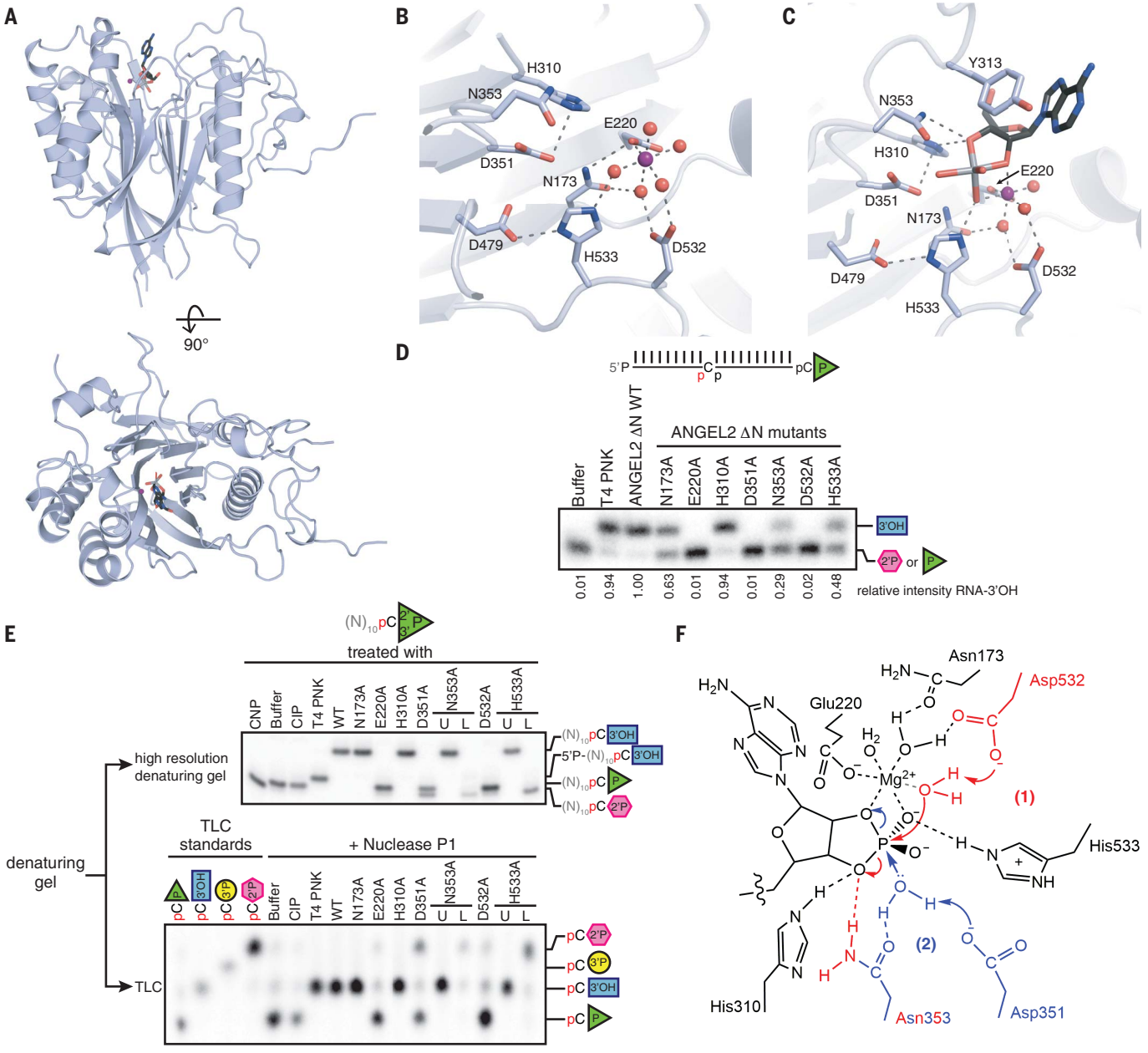


Fig. 2. Crystal structure of human ANGEL2 ΔN and reaction mechanism. (**A**) Overall fold of the ANGEL2 ΔN-AV structure. The AV is shown as black sticks, the Mg^{2+} ion as a purple sphere. (**B**) Detail of apo-ANGEL2 ΔN active site with bound Mg^{2+} (purple sphere). Catalytic residues are shown as sticks; water molecules coordinating Mg^{2+} are shown as red spheres. Gray dashed lines show hydrogen bonds and metal coordination bonds. (**C**) Detail of the AV substrate mimic (black sticks) bound in the ANGEL2 active site. Catalytic residues are shown as blue sticks; water molecules and Mg^{2+} are shown as red and purple spheres, respectively. Hydrogen bonds and metal coordination bonds are shown as gray dashed lines. (**D**) 2′,3′>Pase activity of recombinant ANGEL2 ΔN WT and mutants (10 nM) incubated for 5 min with an internally radiolabeled, 21-nucleotide RNA substrate displaying a terminal 2′,3′>P; red denotes ^{32}P. Numbers below lanes indicate relative amounts of reaction product (RNA-3′OH). (**E**) An 11-nucleotide RNA oligonucleotide terminating with [^{32}P]cytidine-2′,3′>P was incubated with recombinant active CNP (fig. S8C), buffer, CIP, T4 PNK, and ANGEL2 ΔN WT or mutants (0.8 μM) for 30 min to ensure reaction completeness. Individual bands detected on a short denaturing gel (fig. S8B) were purified, and one-half of each was reanalyzed by high-resolution long denaturing gel electrophoresis (top; U, upper band corresponding to 2′,3′-OH; L, lower band corresponding to 2′-P or 2′,3′>P; note RNA 5′ phosphorylation upon T4 PNK treatment due to presence of ATP in the enzyme storage buffer). The remaining half of each was treated with nuclease P1, and the resulting cytidine mono- and bisphosphates were visualized by TLC (bottom); red denotes ^{32}P. (**F**) Proposed reaction mechanism for the two-step 2′,3′>Pase reaction catalyzed by ANGEL2. The catalytic residues involved in the first step are colored red; residues catalyzing the second step are colored blue.

ANGEL2 ΔN (fig. S8A). Alanine substitutions of Glu^{220}, Asp^{351}, and Asp^{532} abolished 2′,3′>Pase activity, whereas mutating Asn^{173}, Asn^{353}, and His^{533} into alanine resulted in decreased activity; the His^{310} → Ala mutant remained active (Fig. 2D).

To capture reaction intermediates, we treated an RNA substrate bearing a terminal [^{32}P]cytidine-2′,3′>P with recombinant wild-type ANGEL2 ΔN (ANGEL2 ΔN WT) or mutants. Individual bands detected by denaturing gel electrophoresis (fig. S8B) were extracted and the isolated RNA was analyzed on a high-resolution gel (Fig. 2E, top) or digested with nuclease P1 (NP1), which hydrolyzes P-O3′ bonds but does not cleave P-O2′ and 2′,3′>P-O bonds in RNA oligo- and mononucleotides (*5*, *25*). Analysis of NP1 digests by thin-layer chromatography (TLC) showed that ANGEL2 ΔN WT yielded ^{32}P-labeled cytidine-2′,3′-OH, confirming the lack of phosphate in the final reaction product (Fig. 2E, bottom). The RNA substrate treated with recombinant 2′,3′-cyclic nucleotide 3′ phosphodiesterase [CNP (*26*); fig. S8C] and digested with NP1 served as control for 2′-P. Mutations in Asp^{351} and His^{533} generated a 2′-P (Fig. 2E and fig. S8D), indicating their involvement in catalyzing the second step in the 2′,3′>P hydrolysis reaction.

On the basis of these observations and the ANGEL2 ΔN-AV structure, we propose a two-step catalytic mechanism (Fig. 2F). In the first step, Asp^{532} acts as a general base to deprotonate a water molecule in the Mg^{2+} coordination shell for an in-line nucleophilic attack on the 2′,3′>P, resulting in the breakage of the P-O3′ bond, while His^{533} may contribute to transition-state stabilization. The leaving group in the transition state is stabilized by Asn^{353}. The resultant 2′-P intermediate is further hydrolyzed in the second step, whereby Asp^{351} acts as a general base to deprotonate another water molecule that carries out a nucleophilic attack to break the P-O2′ bond, and Asn^{353} stabilizes such attack. Overall, the mechanism requires a single Mg^{2+} ion, which coordinates the 2′,3′>P moiety and the water nucleophile in the first step and stabilizes the transition state.

ANGEL1 exhibits weak 2′,3′>Pase activity

To assess ANGEL2 activity and the contribution of ANGEL1 to the overall 2′,3′>Pase activity in human cells, we generated ANGEL2 knockout (KO) human embryonic kidney (HEK293) cells. Unlike parental HEK293 WT cells, extracts from ANGEL2 KO clonal cell lines failed to convert 2′,3′>P into 2′,3′-OH (Fig. 3A). The lack of 2′,3′>Pase activity in ANGEL2 KO cells could be rescued by overexpressing c-Myc–tagged ANGEL2 ΔN WT, but not mutant ANGEL2 ΔN E220A (Glu^{220} → Ala; Fig. 3B and fig. S9A). Highly concentrated extracts from ANGEL2 KO cells revealed traces of 2′,3′>Pase activity that were abolished by RNAi-mediated silencing of ANGEL1 (Fig. 3C and fig. S9B). ANGEL1 2′,3′>Pase activity was detected in c-Myc immunoprecipitation (IP) eluates from both ANGEL2 ΔN WT and ANGEL2 KO cells overexpressing c-Myc–ANGEL1 ΔN ($ANGEL1^{207-670}$), indicating that ANGEL1 is also capable of removing 2′,3′>P (Fig. 3D and fig. S9, C and D). The 2′,3′>Pase activity of recombinant ANGEL1 ΔN was comparable to that of ANGEL2 ΔN when used at concentrations that were higher by factors of ~3000 to 4000 (Fig. 3E and fig. S9E). Taken together, these data indicate that ANGEL2 is the primary 2′,3′>Pase in human cells and that ANGEL1 may be its functional paralog.

ANGEL2 antagonizes pre-tRNA and *XBP1* mRNA splicing

Given the 2′,3′>Pase activity of ANGEL2, we investigated its roles in pathways involving 2′,3′>P RNAs. tRNAs harboring intronic sequences within their anticodon loop must undergo splicing to form mature tRNAs. In mammalian cells, this reaction is catalyzed by the tRNA splicing endonuclease (TSEN) complex (*27*), generating two exon halves that are joined by the tRNA ligase complex (*6*). After cleavage, the 5′-exon contains a 2′,3′>P group required for the tRNA ligase reaction. We found that overexpressing ANGEL2 ΔN WT, but not a mutant version (D532A_H533A), in HEK293 cells impaired tRNA ligation in an in vitro pre-tRNA splicing reaction, as indicated by the accumulation of unligated 3′-exon and dephosphorylated 5′-exon halves (Fig. 4A and fig. S10,

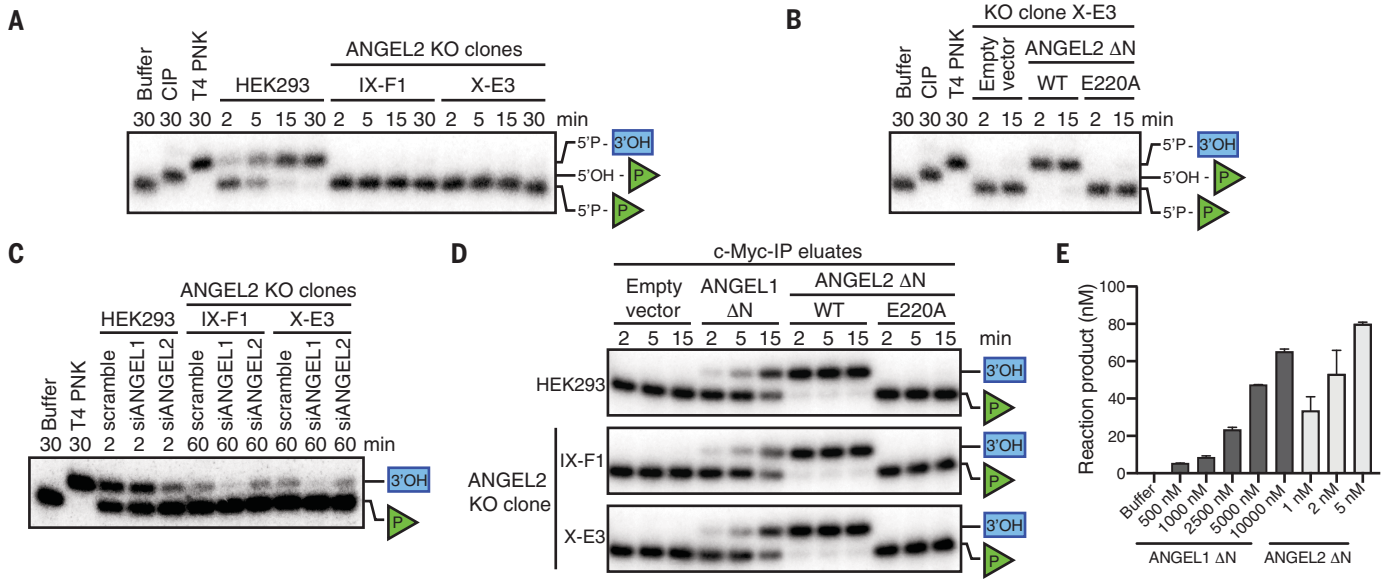


Fig. 3. ANGEL1 displays a weak 2′,3′-cyclic phosphatase activity relative to ANGEL2. (**A**) HEK293 WT and ANGEL2 KO cells (clones IX-F1 and X-E3) were assayed for 2′,3′>Pase activity. (**B**) ANGEL2 KO clone X-E3 was transfected with an empty vector or an expression vector encoding c-Myc–tagged ANGEL2 ΔN WT or mutant (E220A) proteins. Extracts were assayed for 2′,3′>Pase activity. (**C**) HEK293 WT and ANGEL2 KO cells (clones IX-F1 and X-E3) were transfected with control siRNA (scramble) or siRNAs complementary to ANGEL1 or ANGEL2. Extracts were assayed for 2′,3′>Pase activity using concentrations 10 times those used in standard assays. (**D**) HEK293 WT and ANGEL2 KO cells (clones IX-F1 and X-E3) were transfected with an empty vector or an expression vector encoding c-Myc–tagged ANGEL1 ΔN or c-Myc–tagged ANGEL2 ΔN WT or mutant (E220A) proteins. c-Myc IP eluates were assayed for 2′,3′>Pase activity. (**E**) Quantification of 2′,3′>Pase activity of recombinant ANGEL1 ΔN ($ANGEL1^{207-670}$, 500 to 10,000 nM) compared to that of recombinant ANGEL2 ΔN ($ANGEL2^{119-544}$, 1 to 5 nM). Data are means ± SEM (n = 2; denaturing gels are shown in fig. S9E).

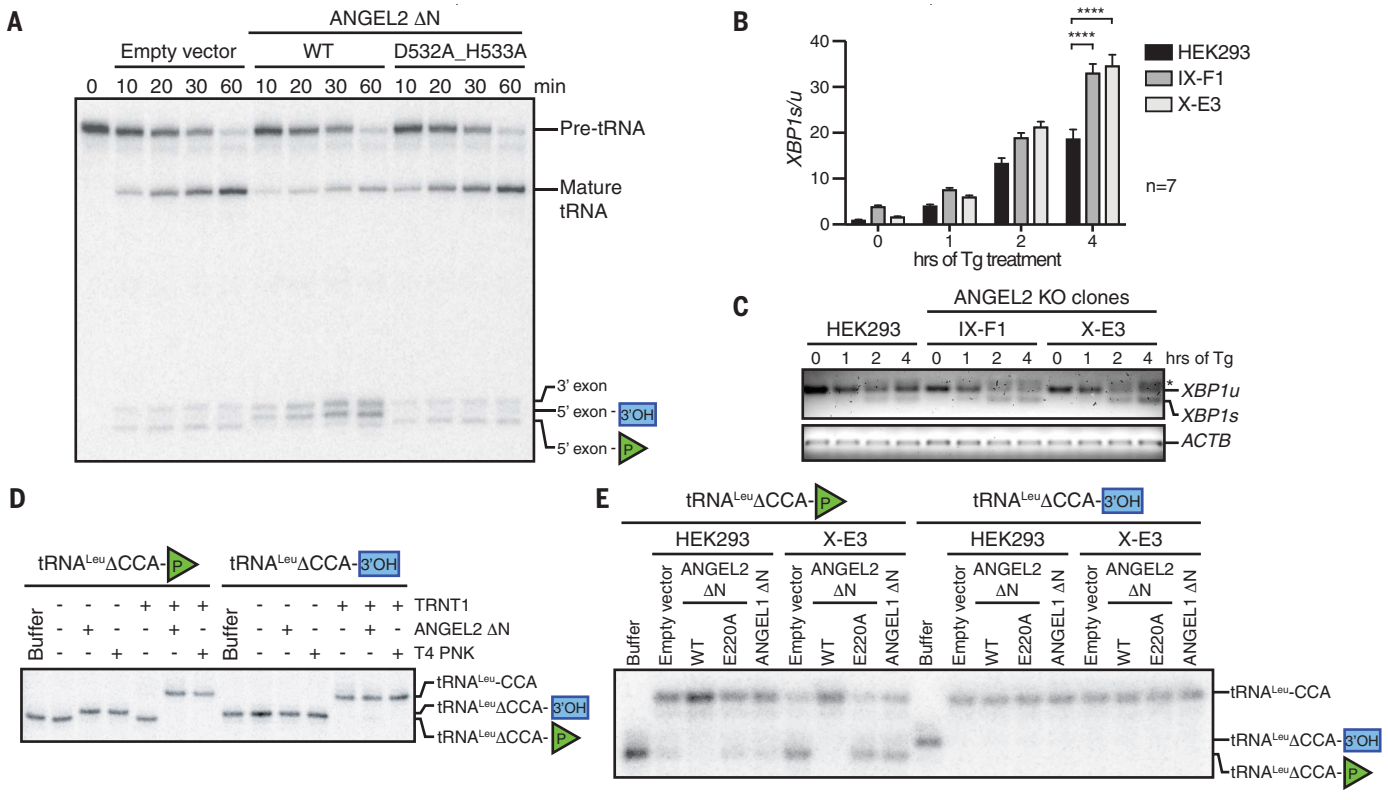


Fig. 4. Misexpression of ANGEL2 affects tRNA ligation, *XBP1* mRNA splicing, and TRNT1-mediated CCA addition. (**A**) HEK293 cell extracts from c-Myc–tagged ANGEL2 ΔN WT or mutant (D532A_H533A) transfected cells were assayed for pre-tRNA splicing by incubation with an internally radiolabeled, intron-containing yeast pre-tRNA^{Phe} for the indicated time periods. Pre-tRNA processing was monitored by denaturing gel electrophoresis. (**B**) HEK293 WT and ANGEL2 KO cells (clones IX-F1 and X-E3) were treated with 500 nM Tg for the indicated time periods. Relative mRNA levels of *XBP1s* and unspliced *XBP1* (*XBP1u*) were measured by quantitative polymerase chain reaction (qPCR) and shown as a ratio of *XBP1s* to *XBP1u*. Expression levels were normalized to *ACTB* mRNA levels and to the untreated HEK293 WT sample. Values represent means ± SEM from seven independent experiments. ****P < 0.0001 (two-way analysis of variance test). (**C**) Reverse transcription PCR analysis of *XBP1* mRNA splicing in samples described in (B). *ACTB* was used as a loading control. The asterisk indicates the position of a hybrid amplicon. A representative result of seven independent experiments is shown. (**D**) Internally radiolabeled tRNA^{Leu}ΔCCA-2′,3′>P or ΔCCA-3′OH were incubated with rNTPs and recombinant TRNT1 in the presence of recombinant ANGEL2 or T4 PNK. Control reactions with buffer alone were also performed. We assessed 3′-terminal CCA addition (tRNA^{Leu}-CCA) by denaturing gel electrophoresis. (**E**) Internally radiolabeled tRNA^{Leu}ΔCCA-2′,3′>P or ΔCCA-3′OH were incubated with extracts of HEK293 WT and ANGEL2 KO (clone X-E3) cells transfected with c-Myc–tagged ANGEL2 ΔN WT or mutant (E220A), ANGEL1 ΔN, or an empty vector control; 3′-terminal CCA addition (tRNA^{Leu}-CCA) was monitored by denaturing gel electrophoresis.

A and B). Extracts of cells overexpressing ANGEL2 ΔN also rapidly dephosphorylated a radiolabeled 2′,3′>P strand in dsRNA, preventing interstrand ligation to the complementary 5′-OH–containing strand by the tRNA ligase complex (*6*) (fig. S10, C to E). In contrast, overexpression of ANGEL1 ΔN caused only a mild impairment in pre-tRNA splicing and interstrand ligation, consistent with its weaker 2′,3′>Pase activity (fig. S11, A to F). Thus, ANGEL2 competes with the tRNA ligase complex for suitable substrates and inhibits the formation of mature tRNAs in vitro by dephosphorylation of 5′ tRNA exon halves.

XBP1 mRNA splicing during the UPR (*28*) involves a two-step, noncanonical splicing reaction, similar to pre-tRNA splicing (*2*). In the first step, cleavage by inositol-requiring enzyme 1 (IRE1), a protein kinase/endoribonuclease, generates mRNA exons with 2′,3′>P and 5′-OH (*29*) that are joined by the tRNA ligase complex in the second step (*7*–*9*). To assess whether ANGEL2 plays an inhibitory role in *XBP1* mRNA splicing, we compared UPR induction using thapsigargin (Tg), an inhibitor of endoplasmic reticulum Ca^{2+} adenosine triphosphatases (ATPases), in HEK293 WT and ANGEL2 KO cells. Tg treatment resulted in higher levels of spliced *XBP1* (*XBP1s*) mRNA in ANGEL2 KO cells (Fig. 4, B and C). Treatment with tunicamycin (Tm), an inhibitor of N-linked glycosylation and UPR trigger, showed a similar effect (fig. S12, A and B). Overexpressing ANGEL2 ΔN in HEK293 WT cells had a significant, albeit small, detrimental effect on *XBP1* mRNA splicing upon induction with Tm (fig. S12, C to E). Together with the increase in *XBP1* mRNA splicing in the absence of ANGEL2, these findings lead us to conclude that ANGEL2 can interfere with *XBP1* mRNA splicing, most likely by attacking 2′,3′>P generated by IRE1, modulating ligation by the tRNA ligase complex, and thus affecting the efficiency of UPR activation.

ANGEL2 allows TRNT1-mediated CCA addition on damaged tRNAs

After the activation of the RQC pathway in response to stalled ribosomes, nascent protein chains are ubiquitinated and released from the tRNA by the peptidyl-tRNA hydrolase ANKZF1. This cleavage event also removes the tRNA 3′ terminal CCA sequence, generating a 2′,3′>P (*10*, *30*–*33*). As a prerequisite for tRNAs to become aminoacylated after ribosome stalling, cleaved tRNAs are recycled via a two-step process, comprising removal of the terminal 2′,3′>P and readdition of CCA by the template-independent tRNA nucleotidyltransferase 1 enzyme (TRNT1) (*10*). To investigate whether ANGEL2 could generate a tRNA 3′-OH needed for CCA addition, we transcribed a tRNA^{Leu} lacking the 3′ CCA and fused it to a 3′ hepatitis

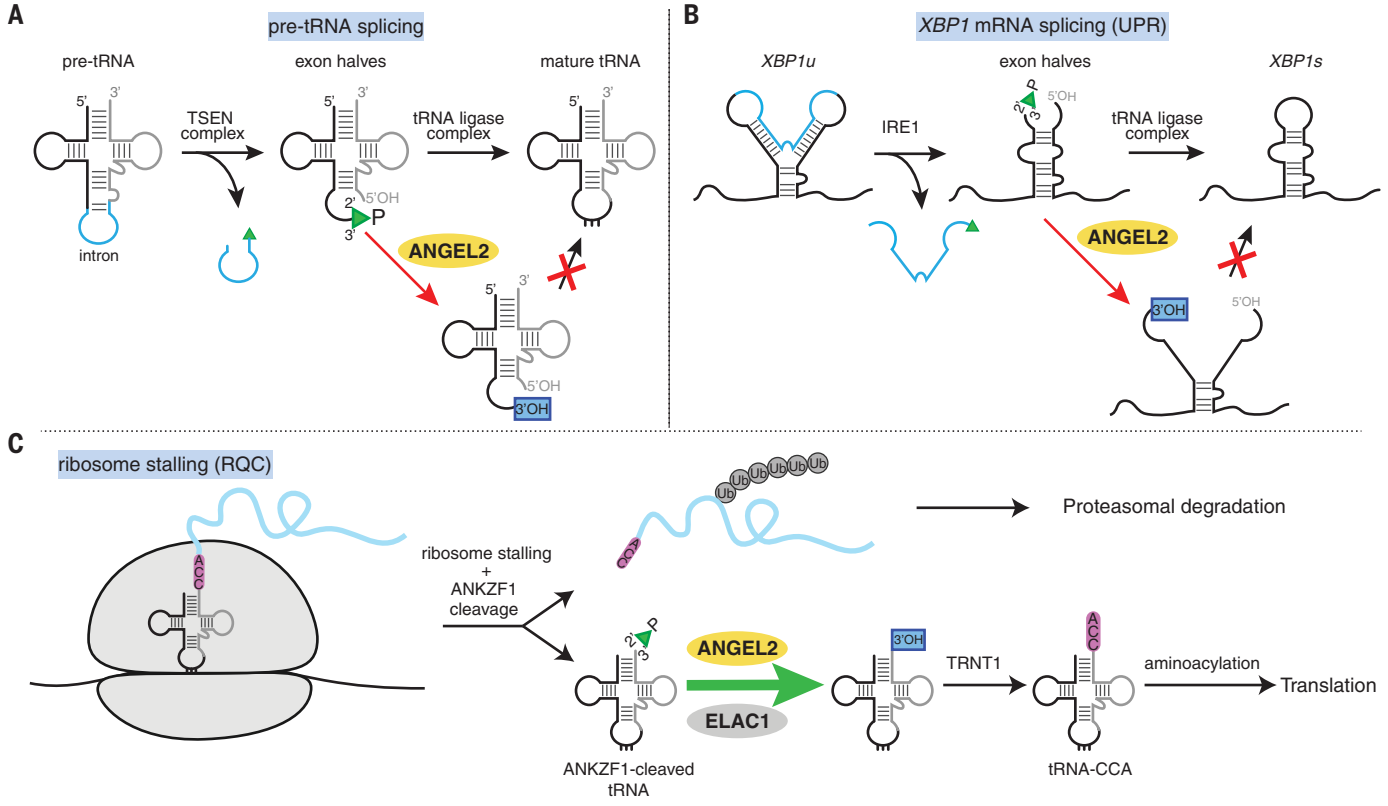


Fig. 5. Summary of the effects of ANGEL2 on pathways requiring ligation or hydrolysis of 2′,3′-cyclic phosphates. (**A**) Pre-tRNA splicing. Intron-containing tRNAs are cleaved at exon-intron borders by the TSEN complex; this generates exon halves displaying 2′,3′>P and 5′-OH that are subsequently joined by the tRNA ligase complex. ANGEL2 modulates levels of mature tRNA during pre-tRNA processing by competing with the tRNA ligase complex for 2′,3′>P substrates. (**B**) *XBP1* mRNA splicing. During UPR activation, IRE1 cleaves a 26-nucleotide intron from the *XBP1* mRNA, generating two exon halves with 2′,3′>P and 5′-OH termini that are also substrates for ligation by the tRNA ligase complex. ANGEL2 can affect the efficiency of *XBP1* mRNA splicing upon induction of the UPR by limiting the availability of suitable 2′,3′>P substrates for the tRNA ligase complex. (**C**) Ribosome stalling. Upon ribosome stalling, ANKZF1 cleaves the 3′ CCA of the peptidyl-tRNA, releasing the ubiquitinated (Ub) nascent polypeptide (light blue) attached to the CCA from the 60*S* ribosome–nascent chain complex, and generates a tRNA fragment with a terminal 2′,3′>P. ANGEL2 aids in recycling of tRNAs to be reused for translation by removing the 2′,3′>P, thus permitting CCA addition by TRNT1 and subsequent aminoacylation.

delta virus (HDV) ribozyme sequence. Upon self-cleavage, a homogeneous 2′,3′>P was generated ($tRNA^{Leu}$ΔCCA-2′,3′>P; fig. S13, A and B), resembling a CCA-less tRNA after ANKZF1 cleavage (*10*). When incubating with recombinant TRNT1 (fig. S13C) and ribonucleoside triphosphates (rNTPs), CCA addition occurred only in the presence of recombinant ANGEL2 ΔN or, as a control, T4 PNK (Fig. 4D). Both enzymes were dispensable for TRNT1 activity when the $tRNA^{Leu}$ΔCCA-2′,3′>P substrate was pretreated with recombinant ANGEL2 ΔN ($tRNA^{Leu}$ΔCCA-3′OH).

To assess whether ANGEL2 is required for CCA addition in cell extracts, we incubated $tRNA^{Leu}$ΔCCA-2′,3′>P with lysates from HEK293 WT and ANGEL2 KO cells. Inefficient CCA addition observed in ANGEL2 KO extracts could be rescued by overexpression of c-Myc–tagged ANGEL2 ΔN WT, but not mutant ANGEL2 ΔN E220A or ANGEL1 ΔN (Fig. 4E and fig. S14, A and B). Likewise, incubation with a $tRNA^{Leu}$ΔCCA-3′OH substrate rescued the CCA addition defect, indicating that TRNT1 activity is not compromised in ANGEL2 KO cell extracts. Although CCA addition was not fully abolished in ANGEL2 KO cell extracts (Fig. 4E and fig. S14B), no further 2′,3′>Pase activity was detected by additionally including a double-stranded 2′,3′>P RNA substrate in the assay (fig. S15). We conclude that ANGEL2 contributes to resolving 2′,3′>P for efficient CCA addition by TRNT1.

Discussion

Our study shows that ANGEL2 modulates levels of mature tRNAs and *XBP1* mRNA during pre-tRNA processing in vitro and in the UPR, respectively (Fig. 5, A and B). Two opposing enzymatic activities have been reported (*34*) to fine-tune the splicing efficiency of *XBP1* mRNA by acting on its exon termini: (i) RTCA, which converts RNA terminal 3′-P, and less efficiently 2′-P, into 2′,3′>P; and (ii) CNP (*26*), which generates 2′-P/3′-OH termini from 2′,3′>P. The ubiquitously expressed ANGEL2 (*35*) confers an additional regulatory mechanism to globally modulate UPR.

Furthermore, we have shown that ANGEL2 contributes to the removal of 2′,3′>Ps prior to 3′ CCA readdition by TRNT1 (Fig. 5). Similarly, after exposure to oxidative stress, angiogenin generates tRNA fragments with 2′,3′>P termini by cleaving mature tRNAs at the anticodon loop (*11*, *36*) and at the 3′ CCA sequence (*37*), thereby leading to a rapid shutdown of translation. ANGEL2 might thus function in tRNA recycling and return to homeostasis upon stress relief, providing suitable tRNA 3′-OH ends for CCA readdition by TRNT1. ELAC1, an isoform of the tRNAse Z family generally known for removing 3′-trailer elements during tRNA biogenesis, was also recently reported to display 2′,3′>Pase activity on ANKZF1-cleaved tRNAs (*38*), but presumably not on short RNA duplexes with terminal 2′,3′>P (fig. S15); this suggests a restricted specificity toward damaged tRNAs.

We found that relative to ANGEL2, ANGEL1 displays a similar but significantly weaker 2′,3′>Pase activity, which could be relevant through its association with the cap binding protein eIF4E (*39*). Preserving both ANGEL paralogs through evolution might indicate that the two 2′,3′>Pase activities execute complementary cellular functions.

Our work paves the way for studying the role of ANGEL proteins on various RNA species with 2′,3′>P termini, mostly derived from tRNAs, mRNAs, and rRNAs (*40*).

REFERENCES AND NOTES

1. C. L. Peebles, P. Gegenheimer, J. Abelson, *Cell* **32**, 525–536 (1983).
2. C. Sidrauski, P. Walter, *Cell* **90**, 1031–1039 (1997).
3. S. Mroczek *et al*., *Genes Dev.* **26**, 1911–1925 (2012).
4. V. Shchepachev, H. Wischnewski, E. Missiaglia, C. Soneson, C. M. Azzalin, *Cell Rep.* **2**, 855–865 (2012).
5. P. Genschik, E. Billy, M. Swianiewicz, W. Filipowicz, *EMBO J.* **16**, 2955–2967 (1997).
6. J. Popow *et al*., *Science* **331**, 760–764 (2011).
7. J. Jurkin *et al*., *EMBO J.* **33**, 2922–2936 (2014).
8. S. G. Kosmaczewski *et al*., *EMBO Rep.* **15**, 1278–1285 (2014).
9. Y. Lu, F. X. Liang, X. Wang, *Mol. Cell* **55**, 758–770 (2014).
10. M. C. J. Yip *et al*., *Nat. Struct. Mol. Biol.* **26**, 343–349 (2019).
11. S. Yamasaki, P. Ivanov, G. F. Hu, P. Anderson, *J. Cell Biol.* **185**, 35–42 (2009).
12. P. R. Whitfeld, L. A. Heppel, R. Markham, *Biochem. J.* **60**, 15–19 (1955).
13. A. K. Chakravarty, S. Shuman, *Nucleic Acids Res.* **40**, 8558–8567 (2012).
14. U. Das, S. Shuman, *Nucleic Acids Res.* **41**, 355–365 (2013).
15. A. Martins, S. Shuman, *J. Biol. Chem.* **279**, 18220–18231 (2004).
16. A. Munir, S. Shuman, *J. Bacteriol.* **199**, e00739-16 (2017).
17. N. Keppetipola, S. Shuman, *Nucleic Acids Res.* **35**, 7721–7732 (2007).
18. A. R. Godwin, S. Kojima, C. B. Green, J. Wilusz, *Biochim. Biophys. Acta* **1829**, 571–579 (2013).
19. E. Wahle, G. S. Winkler, *Biochim. Biophys. Acta* **1829**, 561–570 (2013).
20. A. Dupressoir *et al*., *BMC Genomics* **2**, 9 (2001).
21. C. D. Mol, T. Izumi, S. Mitra, J. A. Tainer, *Nature* **403**, 451–456 (2000).
22. H. Wang *et al*., *EMBO J.* **29**, 2566–2576 (2010).
23. E. R. Wood *et al*., *J. Biol. Chem.* **290**, 19681–19696 (2015).
24. L. Trésaugues *et al*., *Structure* **22**, 744–755 (2014).
25. M. Fujimoto, A. Kuninaka, H. Yoshino, *Agric. Biol. Chem.* **38**, 1555–1561 (1974).
26. L. Bernier *et al*., *J. Neurosci.* **7**, 2703–2710 (1987).
27. S. V. Paushkin, M. Patel, B. S. Furia, S. W. Peltz, C. R. Trotta, *Cell* **117**, 311–321 (2004).
28. P. Walter, D. Ron, *Science* **334**, 1081–1086 (2011).
29. H. Yoshida, T. Matsui, A. Yamamoto, T. Okada, K. Mori, *Cell* **107**, 881–891 (2001).
30. C. A. P. Joazeiro, *Nat. Rev. Mol. Cell Biol.* **20**, 368–383 (2019).
31. K. Kuroha, A. Zinoviev, C. U. T. Hellen, T. V. Pestova, *Mol. Cell* **72**, 286–302.e8 (2018).
32. R. Verma *et al*., *Nature* **557**, 446–451 (2018).
33. O. Zurita Rendón *et al*., *Nat. Commun.* **9**, 2197 (2018).
34. I. Unlu, Y. Lu, X. Wang, *J. Biol. Chem.* **293**, 19365–19376 (2018).
35. X. Yi *et al*., *J. Biol. Chem.* **287**, 21045–21057 (2012).
36. H. Fu *et al*., *FEBS Lett.* **583**, 437–442 (2009).
37. A. Czech, S. Wende, M. Mörl, T. Pan, Z. Ignatova, *PLOS Genet.* **9**, e1003767 (2013).
38. M. C. J. Yip, S. Savickas, S. P. Gygi, S. Shao, *Cell Rep.* **30**, 2106–2114.e5 (2020).
39. P. Gosselin *et al*., *Nucleic Acids Res.* **41**, 7783–7792 (2013).
40. M. Shigematsu, K. Morichika, T. Kawamura, S. Honda, Y. Kirino, *PLOS Genet.* **15**, e1008469 (2019).

ACKNOWLEDGMENTS

We thank D. Koneswarakantha, A. Castello, and A. Jarvelin for help, advice, and reagents; R. Imre, O. Hudecz, I. Steinmacher, and S. Opravil for mass spectrometry analysis; B. Blattmann at the Protein Crystallization Center (University of Zurich) for assistance with initial crystallization experiments; and M. Wang, V. Olieric, and T. Tomizaki at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) for assistance with x-ray diffraction measurements. **Funding:** Work at the Martinez lab was funded by the Medical University of Vienna and the Fonds zur Förderung der wissenschaftlichen Forschung (FWF) as Stand-Alone Projects (P29888 and P32011) and through the RNA Biology Doctoral Program. A.K. is funded by the Boehringer Ingelheim Fonds PhD Fellowship. M.J. is an International Research Scholar of the Howard Hughes Medical Institute, Vallee Scholar of the Bert L. & N. Kuggie Vallee Foundation, and a member of the National Competence Center for Research (NCCR) RNA & Disease. Mass Spectrometry analysis was funded by FWF (SFB F3402; P24685, TRP 308-N15, and I 3686) and by Horizon 2020 EPIC-XS (P823839). **Author contributions:** P.H.P. designed and carried out most of the experiments, A.K. designed and carried out crystallization and structural experiments, A.S. performed bioinformatics analysis, K.M. performed mass spectrometry analysis, M.J. designed crystallization and structural experiments, S.W. designed and carried out experiments, and J.M. designed experiments. P.H.P., A.K., M.J., S.W., and J.M. wrote the manuscript. **Competing interests:** No competing interests. **Data and materials availability:** Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession numbers 6RW0 (apo ANGEL2 ΔN) and 6RVZ (ANGEL2 ΔN-AV). All other data are available in the main text or the supplementary materials. Materials requests should be addressed to S.W.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/524/suppl/DC1
Materials and Methods
Tables S1 and S2
Figs. S1 to S15
References (*41*–*61*)

View/request a protocol for this paper from *Bio-protocol*.

20 January 2020; accepted 2 June 2020
10.1126/science.aba9763

IMMUNOGENOMICS

Human fetal microglia acquire homeostatic immune-sensing properties early in development

L. Kracht[1]*, M. Borggrewe[1]*, S. Eskandar[1,2]*, N. Brouwer[1], S. M. Chuva de Sousa Lopes[3,4], J. D. Laman[1], S. A. Scherjon[2], J. R. Prins[2], S. M. Kooistra[1]†‡, B. J. L. Eggen[1]†‡

Microglia, immune cells of the central nervous system (CNS), are important for tissue development and maintenance and are implicated in CNS disease, but we lack understanding of human fetal microglia development. Single-cell gene expression and bulk chromatin profiles of microglia at 9 to 18 gestational weeks (GWs) of human fetal development were generated. Microglia were heterogeneous at all studied GWs. Microglia start to mature during this developmental period and increasingly resemble adult microglia with CNS-surveilling properties. Chromatin accessibility increases during development with associated transcriptional networks reflective of adult microglia. Thus, during early fetal development, microglia progress toward a more mature, immune-sensing competent phenotype, and this might render the developing human CNS vulnerable to environmental perturbations during early pregnancy.

Microglia are the resident myeloid cells of the central nervous system (CNS) and contribute to tissue homeostasis and pathology. Under homeostasis, microglia survey the CNS parenchyma and express receptors involved in monitoring and immune-sensing functions, termed the sensome (*1*). Microglia colonize the brain prior to neurogenesis, myelination, and blood-brain barrier formation and are a largely self-maintaining and long-lived population in healthy CNS (*2*, *3*). Environmental perturbations during pregnancy affect microglia functioning (*4*, *5*) and can be associated with epigenetic reprogramming (*6*, *7*). Microglia are important for CNS development, and in view of their longevity and epigenetic memory, early perturbations in microglia might have long-lasting effects that could affect CNS development and function (*8*, *9*).

In mice, microglia emerge from early erythromyeloid progenitors in the extraembryonic yolk sac at embryonic day 7.5 (E7.5) and subsequently colonize the developing brain rudiment at E9.5 (*10*). Mouse microglia development is accompanied by changes in gene expression and epigenetic profiles (*5*). Microglia proliferate and differentiate in early embryonic stages and support neuronal development (synaptic pruning) in late embryonic and early postnatal stages. They acquire their homeostatic and immune surveillance profile at late postnatal and adult stages (*4*, *5*). During embryonic development and early postnatal stages, mouse microglia are heterogeneous (*11*–*13*).

In human embryos, shortly after the closure of the neural tube, amoeboid ionized calcium-binding adaptor protein 1–positive ($IBA1^{pos}$) microglia appear at gestational week (GW) 4.5 in the leptomeninges, the ventricular edge, and the choroid plexus of the brain (*14*). From these sites, microglia colonize the telencephalon and diencephalon. During this process, microglia transform to ramified cells, which are recognizable as early as GW12 (*15*, *16*).

To date, limited bulk RNA sequencing (RNA-seq) data of human mid-gestation fetal microglia indicate that mouse and human microglia share developmental gene expression signatures (*4*). However, an in-depth understanding of cellular development and heterogeneity of microglia during human fetal development is lacking. Here, 15,782 microglia were analyzed with single-cell RNA sequencing (scRNA-seq), and underlying gene regulatory mechanisms were observed by assay for transposase-accessible chromatin sequencing (ATAC-seq) in 23 human fetuses ranging from GW9 to -18.

Isolation and characterization of microglia from human fetal CNS tissue

Single-cell transcriptional profiles (n = 20 fetuses) and chromatin accessibility (n = 8 fetuses) were generated from viable [4′,6-diamidino-2-phenylindole–negative, DRAQ5-positive ($DAPI^{neg}DRAQ5^{pos}$)] human fetal microglia ($CD11B^{pos}CD45^{int}$) that were isolated by fluorescence-activated cell sorting (FACS) of CNS tissue from 23 early to mid-gestation (GW9 to -18) fetuses after elective pregnancy termination (Fig. 1, A and B, and table S1). Using a modified Smart-seq2 scRNA-seq protocol, we sequenced 15,782 microglia and 781 nonmicroglia CNS cells, with a median number of 22,330 unique molecular identifiers (UMIs) and 977 unique genes per cell after filtering. The number of UMIs and percentages of ribosomal and mitochondrial RNA were similar across all samples and ages (fig. S1, A and B). To verify that the $CD11B^{pos}CD45^{int}$ population was microglia, we compared the transcription profile to CNS cell type–specific gene sets from two independent human datasets (*17*, *18*) (table S2). Expression of microglia genes was enriched in $CD11B^{pos}CD45^{int}$ microglia, but not in other ($CD11B^{neg}CD45^{neg}$) CNS cells (Fig. 1C). IBA1 immunoreactivity was detected in fetal CNS tissue at all gestational ages, and a typical ramified microglia morphology was observed in most tissues, confirming the presence of microglia (Fig. 1D).

Microglia are heterogeneous during human fetal development and exhibit an activated, phagocytic gene expression profile

To determine transcriptional microglia heterogeneity during development, we performed an unsupervised clustering analysis, resulting in 16 distinct clusters (Fig. 2A) present at all GWs (Fig. 2B). All clusters exhibited similar amounts of mitochondrial and ribosomal RNA and UMIs per cell (fig. S1C). Thus, cluster formation was not caused by differences in cell quality. Variation between samples was minor, and sample bias did not affect clustering (fig. S2A). Although sex did not affect cluster distribution (fig. S2B), male/female gene expression differences were difficult to assess, owing to unequal male and female sample numbers per GW (table S1). Clusters were annotated on the basis of cluster-enriched genes (fig. S2, C and D, and table S3), which were determined by differential gene expression analysis of one cluster compared with all other clusters.

Cells in myeloid clusters 9 and 10 expressed the nonmicroglia myeloid cell markers *LYVE1* and *S100A9*, respectively (fig. S2, C and D). These clusters also expressed genes of the *MS4A* family, which have been associated with a microglia-brain border macrophage intermediate during mouse development (*13*). Microglia clusters 11 to 16 were characterized on the basis of unique expression of cluster-enriched genes *MRPL23*, *PARP4*, *MTX1*, *HBA/HBG*, *ZP3*, and *NAMPT* (fig. S2, C and D).

All microglia clusters expressed canonical microglia markers *CSF1R* and *CX3CR1* (Fig. 2C). Expression of homeostatic microglia genes *P2RY12* (Fig. 2C), *P2RY13*, and *TMEM119* (fig. S3A) was less frequent but present and validated in situ for TMEM119 (fig. S3, B and C). Microglia of all clusters expressed multiple genes previously associated with an activated, phagocytic microglia phenotype identified in aging and neurodegenerative diseases [also called the DAMs/MGnD (disease-associated microglia and microglial neurodegenerative) phenotype] (*19*–*21*) (Fig. 2C and fig. S3A). AXL, APOE, and CD68, markers for activated,

[1]Department of Biomedical Sciences of Cells and Systems, Section Molecular Neurobiology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands. [2]Department of Obstetrics and Gynecology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands. [3]Department of Anatomy and Embryology, Leiden University Medical Center, Leiden, Netherlands. [4]Department for Reproductive Medicine, Ghent University Hospital, Ghent, Belgium.
*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. Email: s.m.kooistra@umcg.nl (S.M.K.); b.j.l.eggen@umcg.nl (B.J.L.E.)

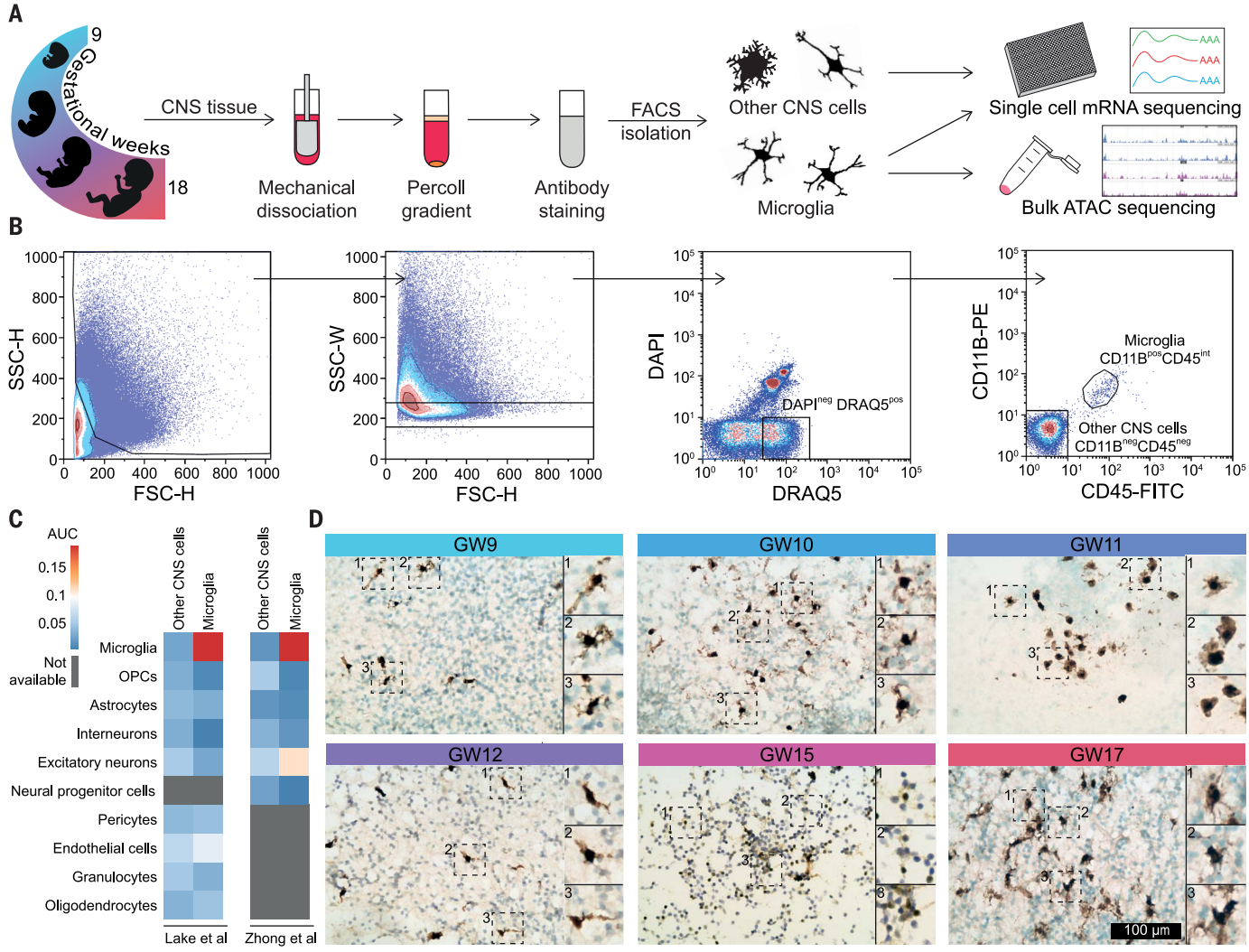


Fig. 1. Study design and isolation of microglia from fetal CNS tissue. (**A**) CNS tissue from fetuses at GW9 to -18 was processed on ice with mechanical dissociation and Percoll gradient centrifugation (for GWs >12). Individual microglia and other CNS cells were sorted into 384-well plates for scRNA-seq (n = 20). Bulk microglia were collected for ATAC-seq (n = 8). (**B**) Representative FACS plots of the gating strategy for isolation of microglia and other CNS cells (n = 23). FSC-H, forward scatter height; SSC-W, side scatter width; PE, phycoerythrin; FITC, fluorescein isothiocyanate. (**C**) Heatmap depicting expression of CNS cell type–specific gene sets from two different datasets (*17*, *18*) (table S2) in FACS-isolated microglia and other CNS cells, reported as AUC values. (**D**) IBA1 immunoreactivity in CNS tissue of GW9 to -17 fetuses. Insets depict 2× magnifications of the indicated areas. OPC, oligodendrocyte progenitor cell.

phagocytic microglia, colocalized with IBA1 in situ, confirming their expression by microglia (Fig. 2, D and E, and fig. S3D).

Microglia clusters are associated with GW and exhibit distinct functional profiles

In view of the GW-dependent distribution of cells on the uniform manifold approximation and projection (UMAP) (Fig. 3A) and the differential contribution to clusters (Fig. 2B), we characterized GW-associated clusters.

The percentage of cells in cluster 5 increased from GW9 to -18 and showed distinct expression of immediate-early genes (IEGs) (Fig. 3B). cJUN protein was also detected in IBA1^{pos} microglia in situ (Fig. 3C), making artifactual induction of IEGs exclusively by the microglia isolation procedure unlikely. Gene ontology (GO) terms associated with the cluster-enriched genes were mainly immune- and inflammation-related (Fig. 3D and table S4).

Microglia from all gestational ages contributed to cluster 6, but it was enriched for cells from GW10 to -13 (Fig. 3E). Cells of cluster 6 uniquely expressed the cell cycle genes *MKI67* and *SPC24* (Fig. 3E), were annotated to G_2/M and S phase on the basis of conserved cell cycle genes (fig. S2B), and were associated with GO term cell division (Fig. 3D and table S4). Immunostaining confirmed expression of MKI67 in IBA1^{pos} microglia across multiple GWs (Fig. 3F), indicating the presence of proliferating microglia.

Clusters 7 and 8 had a strong bias for GW9 to -10 microglia, whereas these clusters had little contribution of cells from GWs >15 (Fig. 3G). Cluster-enriched genes included neuronal genes and phagocytosis genes (Fig. 3, G and H). Phagocytosis of other cells or debris may explain the presence of these neuronal transcripts.

More than 40% of microglia in clusters 2 and 3 were derived from GW11 to -12 samples (Fig. 3H). Although microglia of all clusters expressed genes associated with microglia activation (Fig. 2, C, D, and E), some of these genes were more enriched in GW11 to -12 microglia (Fig. 3H). GO terms associated with GW11 to -12 microglia cluster–enriched genes were related to phagocytosis and brain development (Fig. 3D and table S4).

Clusters 1 and 4 exhibited higher expression of homeostatic microglia markers and primarily contained cells from GW15 to -17 (Fig. 3I). Homeostatic markers CX3CR1 and VISTA

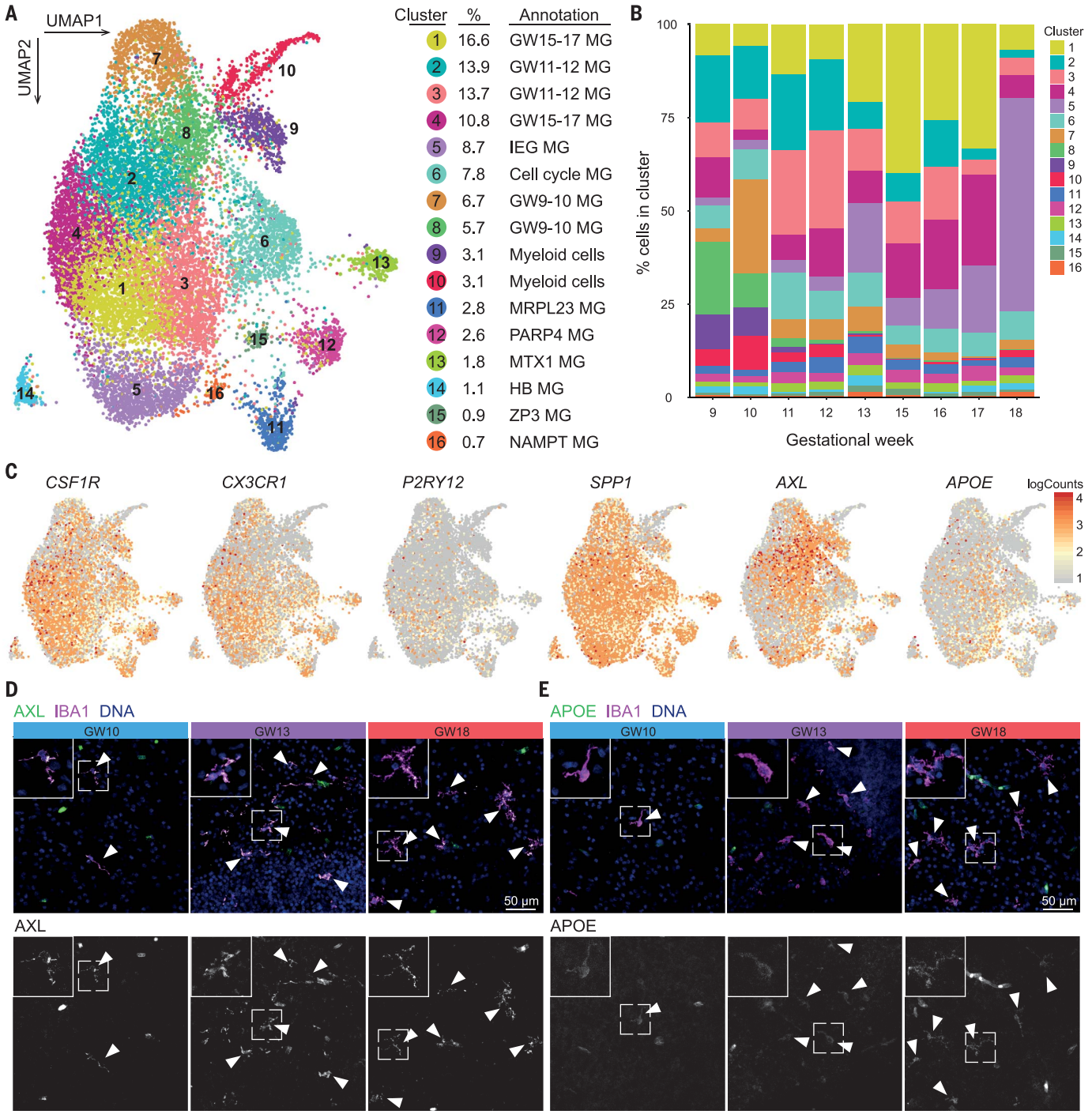


Fig. 2. Human fetal microglia are heterogeneous and exhibit an activated, phagocytic profile. (**A**) UMAP of 15,782 cells and 16 clusters depicted with size (percentage of total cells in a specific cluster) and annotation (20 fetal samples; *n* = 1 to 4 per GW). (**B**) Bar plots depicting the percentage of cells in each cluster at different GWs. (**C**) UMAPs depicting expression of microglia markers in log counts. (**D** and **E**) AXL and IBA1 (D) and APOE and IBA1 (E) coexpression in CNS tissue of GW10 to -18 fetuses. Lower images show single channels. White arrowheads indicate colocalization. Insets depict 2× magnifications of the indicated areas. MG, microglia.

colocalized with IBA1 in situ (fig. S3, E and F), demonstrating their expression by microglia (*22*). Enriched genes of GW15 to -17 microglia were annotated with GO terms for cytokine and immune system processes (Fig. 3D and table S4).

Microglia undergo developmental transition toward adult, homeostatic microglia

To further delineate the observed developmental progression (Fig. 3), we performed pseudotime analysis (*23*). Assigned pseudotimes (Fig. 4A) corresponded with the gestational ages (Fig. 4B), suggesting a GW-dependent trajectory. This trajectory contained seven segments, or states, which were separated by branching points (Fig. 4, A and C), and differential gene expression analysis was performed to identify state-enriched genes (table S5). State

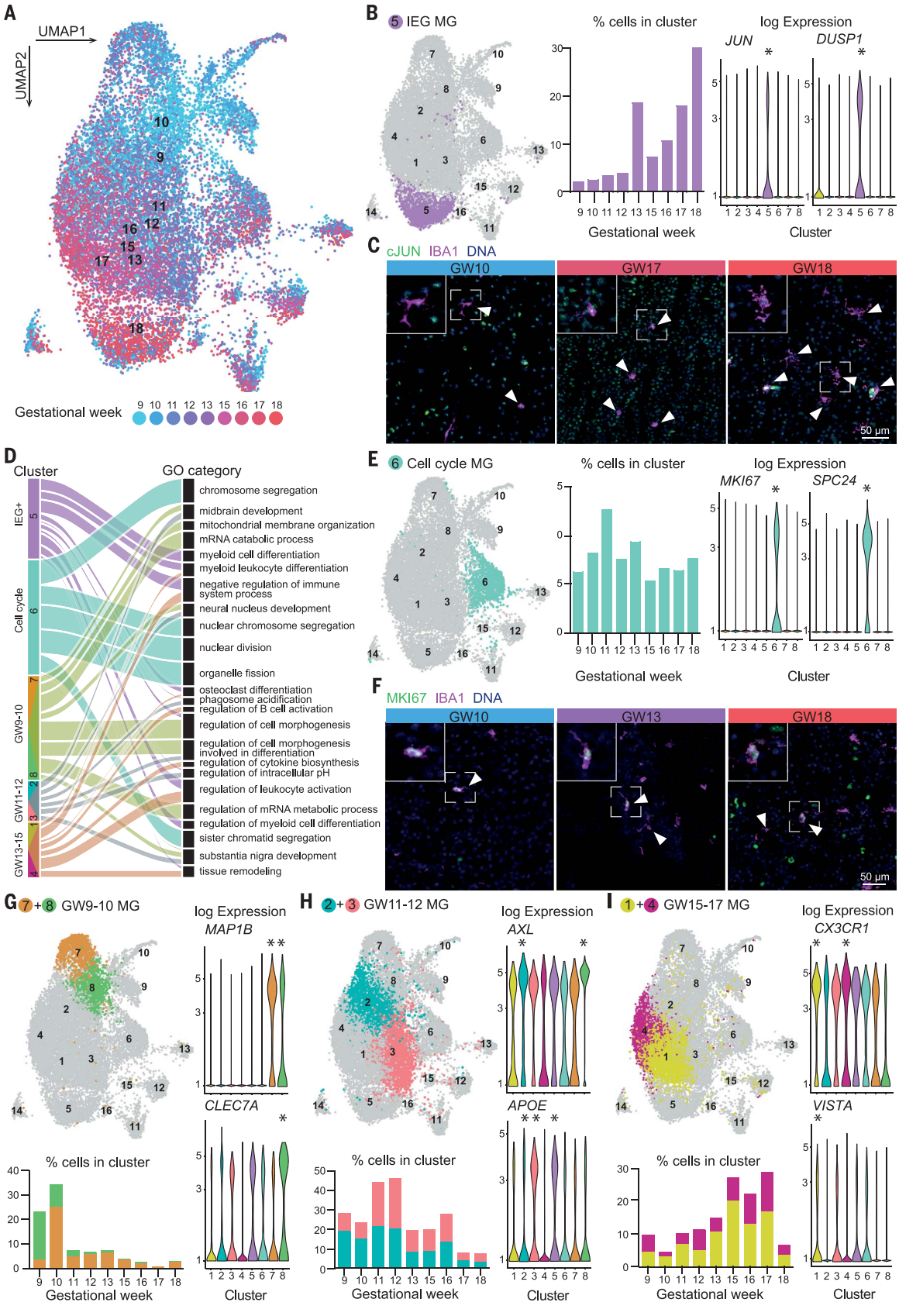


Fig. 3. GW-associated microglia clusters have diverse functional profiles. (**A**) UMAP depicting GWs (20 fetal samples; n = 1 to 4 per GW). (**B**) UMAP highlighting specific cluster, bar plots indicating the percentage of cells in cluster across GWs, and violin plots depicting log expression of cluster-enriched genes for IEG MG. (**C**) cJUN and IBA1 coexpression in CNS tissue of GW10 to -18 fetuses. (**D**) Alluvial plot depicting the top five GO terms per cluster, plotted for all clusters. Ribbon thickness reflects the number of genes in GO terms. (**E**) UMAP highlighting specific cluster, bar plots indicating the percentage of cells in cluster across GWs, and violin plots depicting log expression of cluster-enriched genes for cell cycle MG. (**F**) MKI67 and IBA1 (F) coexpression in CNS tissue of GW10 to -18 fetuses. (**G** to **I**) UMAPs highlighting specific clusters, bar plots indicating the percentage of cells in respective clusters across GWs, and violin plots depicting log expression of cluster-enriched genes for GW9 to -10 MG (G), GW11 to -12 MG (H), and GW15 to -17 MG (I). Significantly enriched genes per cluster, compared with all other clusters, are indicated. *P(adjusted) < 0.05, model-based analysis of single-cell transcriptomics (MAST). White arrowheads indicate colocalization. Insets depict 2× magnifications of the indicated areas. MG, microglia.

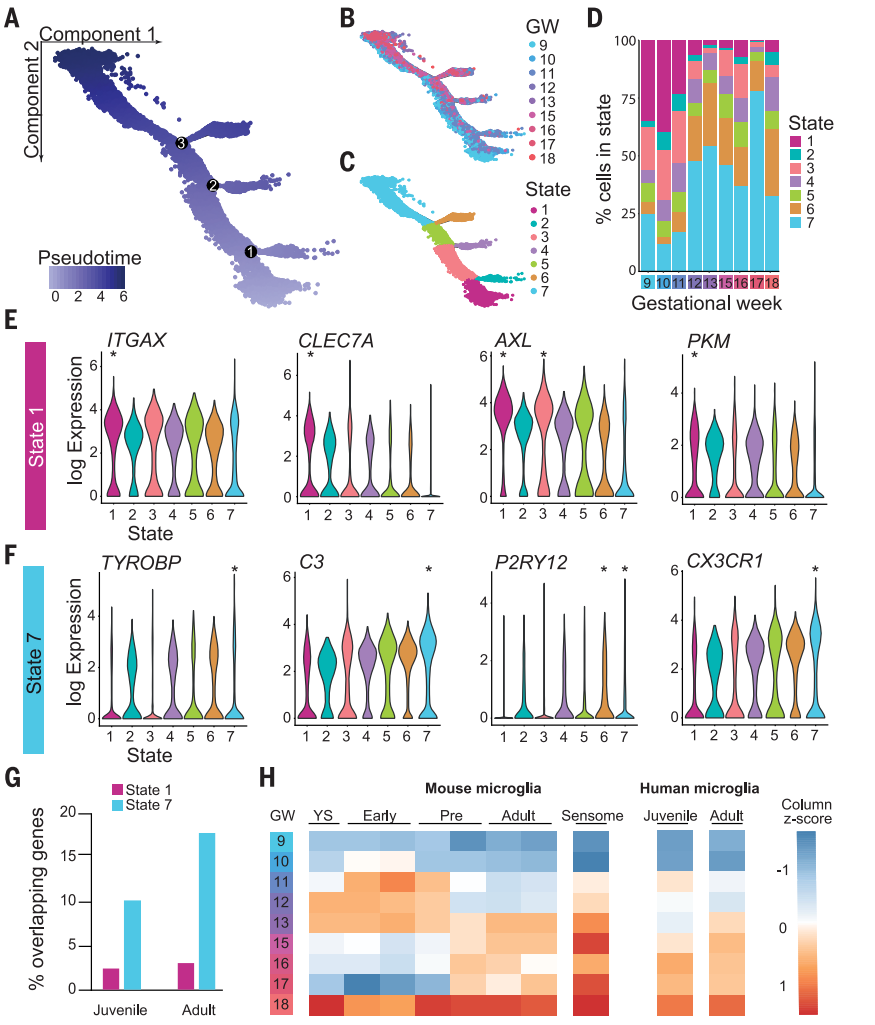


Fig. 4. Developmental progression toward homeostatic microglia. (**A** to **C**) Trajectory plots depicting pseudotime (numbers indicate branching points) (A), GW (B), and state (C). (**D**) Bar plots depicting the percentage of microglia in each state at different GWs. (**E** and **F**) Violin plots of log expression of state-enriched genes for state 1 (GW9 to -11) (E) and state 7 (GW12 to -18) (F). Significantly enriched genes per state were compared to all other states. *P(adjusted) < 0.05, MAST test. (**G**) Bar plot depicting the percentages of overlapping genes between state 1 and 7 genes (table S6) with the top 500 human juvenile (*26*) and adult (*27*) microglia genes (table S2). (**H**) Heatmap depicting enrichment of mouse developmental stage genes (*5*), mouse sensome (*1*), and human juvenile (*26*) and adult microglia (*27*) genes (table S2) across GWs visualized as column z-scores of AUC values. YS, yolk sac.

formation was not driven by individual samples (fig. S4A).

State-enriched genes of side branches (states 2, 4, and 6) were associated with oxidative phosphorylation, cell cycle, and the immediate-early response (table S5). Whether these side branches represent distinct cell fates or transient transcriptional divergences from the main trajectory is not clear.

State 1 contained mainly GW9 to -11 microglia, whereas state 7 largely consisted of GW12 to -18 microglia (Fig. 4D), which is consistent with the GW-associated clusters (fig. S4B). State 1–enriched genes were associated with microglia development (*24*), microglia activation and phagocytosis (*19*, *20*), and glycolysis (*13*, *25*) (Fig. 4E and fig. S4C). Over the pseudotime trajectory, expression of homeostatic sensome markers and a different set of microglia activation markers increased, and this was most pronounced in state 7 (Fig. 4F and fig. S4D).

To confirm the developmental progression of fetal microglia, we compared their transcriptional profiles with profiles of mouse developmental microglia (*5*), the mouse sensome (*1*), and juvenile (*26*) and adult human microglia (*27*) (Fig. 4, G and H, and table S2). The overlap with juvenile and adult microglia genes increased from ≈2 to 3% of state 1 genes to 10 to 18% of state 7 genes (Fig. 4G), suggesting developmental progression of human microglia from GW9 to -11 (state 1) to GW12 to -18 (state 7). Progressive enrichment of mouse and human microglia gene sets in GW9 to -18 fetal microglia was determined using area under the curve (AUC) analysis (Fig. 4H). Microglia younger than GW13 were enriched for genes from yolk sac and early mouse microglia (E10 to -16.5), and microglia older than GW13 were enriched for genes from pre- and adult microglia (P3 to -56), with the exception of GW18 microglia, which were enriched for genes at all developmental stages (Fig. 4H). An increasing overlap with mouse sensome genes was observed with progressing fetal GW (Fig. 4H). Also, enrichment of human juvenile and adult microglia genes gradually increased in fetal microglia with increasing GW (Fig. 4H). Together, these data suggest that microglia progress toward a homeostatic state during early human fetal development.

Distinct gene regulatory networks are active in early and mid-gestational microglia

Mouse microglia development is orchestrated by distinct transcription factors (TFs) (*5*). To unravel putative gene regulatory mechanisms underlying the transcriptional changes in developing human fetal microglia, single-cell regulatory network inference and clustering (SCENIC) (*28*) was used. SCENIC analyzes co-expression and TF motifs in scRNA-seq data.

Sixty-six gene regulatory networks were identified that segregated into two main hubs after unsupervised clustering associated with older and younger GWs (Fig. 5A). Gene regulatory networks associated with younger GWs were enriched for diverse general cellular functions, including cell cycle (E2F2), morphogenesis (SOX4/11), and differentiation and chromatin remodeling (SP1) (*29*) (Fig. 5, A and B). Gene regulatory networks associated with older GWs were more microglia specific, including many ETS TF family members, such as ETS1-2, ELF1, ELK3, and SPI1 (PU.1) (Fig. 5, A and C), which are crucial for microglia development and function (*30*, *31*). Furthermore, gene regulatory networks associated with IEGs were identified in older GWs (Fig. 5A), and JUN activity overlapped with the IEG microglia cluster (Fig. 5C). These data show that gene regulatory networks progress from more general cellular functions in young GWs to microglia-specific properties in older GWs.

To verify scRNA-seq data–derived differences in gene regulatory network activity between microglia from young and old GWs, we assessed chromatin accessibility with ATAC-seq (Fig. 5, D and E, and fig. S5). The DNA sequences underlying peaks were analyzed for enrichment of putative TF-binding motifs

with HOMER (*32*). In the combined peaks of all samples, putative binding motifs of essential microglia TFs were enriched (table S7). Differential peak analysis indicated that more peaks were associated with older (GW >13; 1338 peaks) than with younger developmental stages (GW <13; 33 peaks) (Fig. 5D and table S8). In peaks that were enriched in microglia at GW <13, the SP1 TF motif was present, which was also detected by SCENIC (Fig. 5, A, B, and E). DNA sequences underlying peaks enriched in microglia at GW >13 contained many ETS TF family motifs, including PU.1 (Fig. 5E), which was validated at the protein level (Fig. 5F), in agreement with the SCENIC results (Fig. 5, A and C). These findings indicate that increased chromatin accessibility during microglia development is accompanied by the activation of gene networks driving microglia-specific functions.

Discussion

Microglia are the CNS-resident myeloid cells and are critical for brain development and later tissue homeostasis (*4*, *5*). After seeding the developing CNS from the yolk sac, and also closure of the blood-brain barrier, microglia form a self-sustained population with highly variable turnover and with negligible contribution from peripheral immune cells (*33*). Studies in mice have revealed that during development, different gene regulatory networks drive microglia proliferation, differentiation, and maturation and that perturbances have long-lasting functional consequences (*4*, *5*).

Here, the transcriptomic profile and chromatin organization of human fetal microglia during early to mid-gestation (GW9 to -18) development are presented. In contrast to microglia in the healthy adult CNS (*34*), we found that fetal microglia are highly heterogeneous. They progressively mature from GW13 onward, which is regulated by the activity of increasingly complex gene regulatory networks, and they already display functional properties characteristic of mature human microglia at mid-gestation.

Human fetal microglia share extensive transcriptional similarities with microglia during mouse development, with analogous proliferative, glycolytic, and activated, phagocytic capacities (*4*, *12*, *13*). The presence of glycolysis-related genes in microglia at GW9 to -10 underscores their undifferentiated and activated state, since immune-activated mouse and human microglia (*25*, *35*) as well as undifferentiated cells such as stem cells (*36*) and embryonic cells (*13*) use glycolysis as an energy source.

Early to mid-gestation microglia share transcriptional features with a phagocytic microglia population that is transiently present during postnatal mouse development and associated with myelinating brain regions (*12*, *13*). In mice, microglia support myelination, neurogenesis (*37*), and oligodendrogenesis (*38*). Human microglia may play similar roles, as the gestational period GW9 to -18 coincides with oligodendrocyte (*39*) and neuronal (*40*) development. These human fetal phagocytic microglia express genes detected in DAM/MGnD microglia that are observed in neurodegenerative mouse models (*19*–*21*). The similarities between developmental human microglia

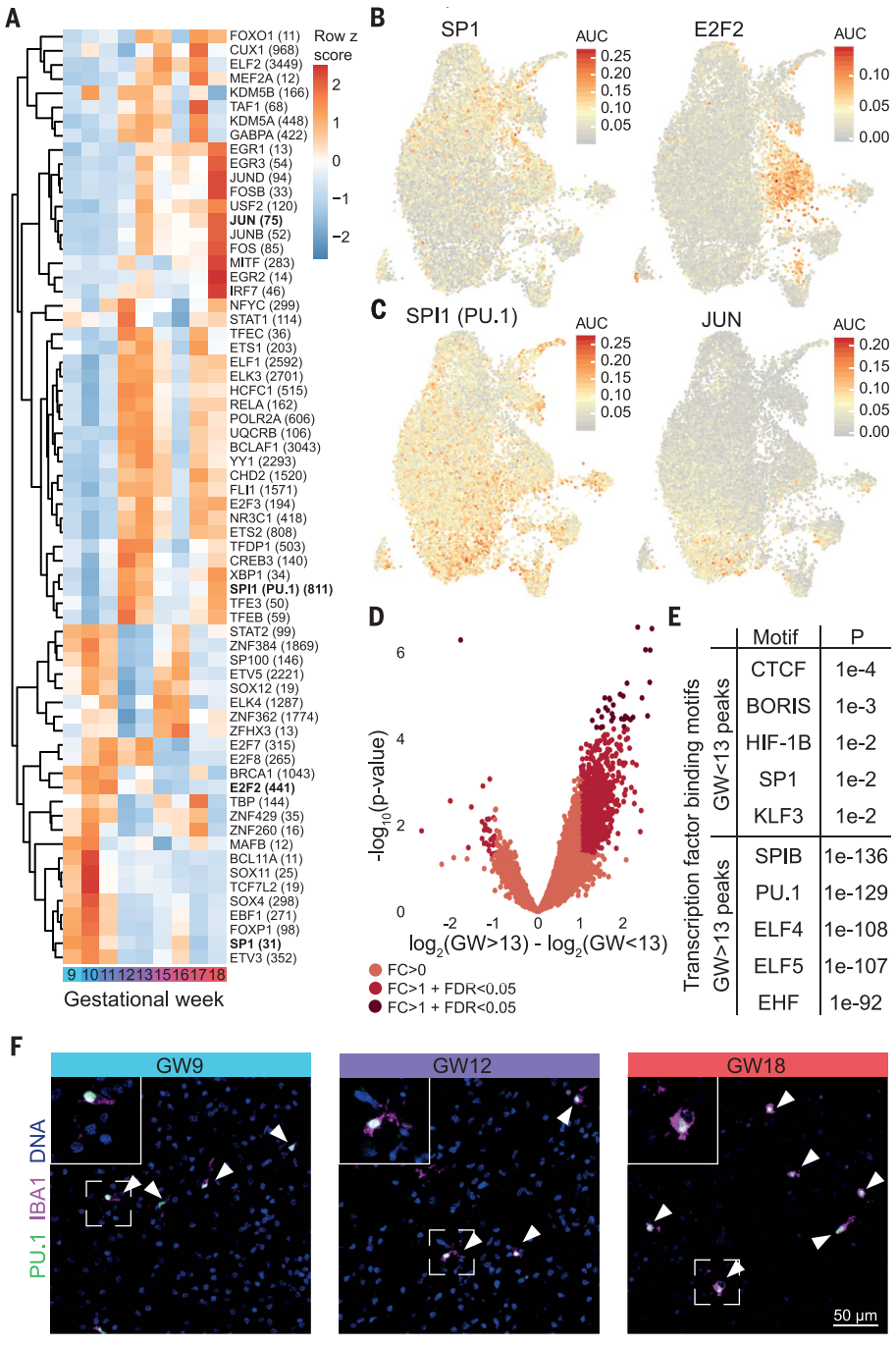


Fig. 5. Transition of gene regulatory networks and chromatin landscape in microglia during development. (**A**) Heatmap with unsupervised clustering of gene regulatory network activity in GW9 to -18 microglia visualized as row *z*-scores of mean AUC values per GW (20 fetal samples; n = 1 to 4 per GW). (**B** and **C**) UMAPs of gene regulatory network activity visualized as AUC values. (**D**) Volcano plot of differential ATAC peaks between GW>13 and GW<13 microglia (table S8) (n = 4 in each group). (**E**) TF motif enrichment of sequences underlying differential ATAC peaks. (**F**) PU.1 and IBA1 coexpression in fetal CNS tissue. White arrowheads indicate colocalization. Insets depict 2× magnifications of the indicated areas. FC, fold change; FDR, false discovery rate.

subtypes and DAM/MGnD microglia in mice suggest that developmental transcriptional programs are reactivated in neurodegenerative diseases (*12*, *13*) in which microglia are increasingly implicated.

At later GWs, the frequency of IEG-expressing microglia increases. Expression of IEGs in microglia was previously attributed to ex vivo activation (*6*, *12*). However, we detected cJUN in fetal CNS tissue prior to microglia isolation, indicating that IEGs were not (exclusively) induced by the isolation procedure or the collection and experimental methods used here. IEG expression might reflect a necessary responsiveness of microglia to local environmental cues of the developing CNS.

In all investigated GWs, small microglia clusters were present (clusters 11 to 16) for which we could not assign specific functional properties because of a low number of cluster-enriched genes. In situ validation of these minor clusters was further hampered by the limited number of cells in these respective clusters and by the small size of early fetal brain tissues.

With increasing GWs, human microglia acquire a more homeostatic phenotype, reflected by an increasing overlap between genes expressed in fetal (especially beyond GW13) and juvenile and adult human microglia. This overlap was likely underestimated, because juvenile and adult human microglia were analyzed by bulk sequencing and may have included less abundantly expressed genes not detected in single-cell sequencing. Moreover, expression of microglia sensome genes (*1*), encoding receptors important in environmental sensing, increased at later gestational ages, pointing to the emergence of immune-sensing microglia during early fetal development. The detected increase in chromatin accessibility and associated gene regulatory networks in fetal microglia from older GWs allows for the activation of more complex gene programs that are required for the immune sensing, synaptic pruning, phagocytic, and tissue-supportive functions of microglia.

The emergence of immune-sensing microglia is highly relevant in view of environmental perturbations during pregnancy that disturb mouse microglia development (*4*, *5*) and impair CNS functions in adult mice (*41*). Perturbed microglial development has been linked to human neurodevelopmental and psychiatric disorders (*8*, *9*). A higher risk for the development of autism is associated with feverish infections, particularly in the second trimester (*42*, *43*). Because microglia express receptors involved in environmental sensing at this gestational period, they may contribute to fetal CNS sensitivity to the environment during early pregnancy.

Together, our findings demonstrate that microglia are highly heterogeneous during early human fetal development and mature by mid-gestation. This explains the vulnerability of the developing human CNS to environmental perturbations at this developmental period during pregnancy, with potentially long-lasting consequences.

REFERENCES AND NOTES

1. S. E. Hickman *et al.*, *Nat. Neurosci.* **16**, 1896–1905 (2013).
2. P. Füger *et al.*, *Nat. Neurosci.* **20**, 1371–1376 (2017).
3. T. L. Tay *et al.*, *Nat. Neurosci.* **20**, 793–803 (2017).
4. M. S. Thion *et al.*, *Cell* **172**, 500–516.e16 (2018).
5. O. Matcovitch-Natan *et al.*, *Science* **353**, aad8670 (2016).
6. P. Ayata *et al.*, *Nat. Neurosci.* **21**, 1049–1060 (2018).
7. A. C. Wendeln *et al.*, *Nature* **556**, 332–338 (2018).
8. T. L. Tay *et al.*, *Front. Mol. Neurosci.* **10**, 421 (2018).
9. M. Prinz, J. Priller, *Nat. Rev. Neurosci.* **15**, 300–312 (2014).
10. G. Hoeffel *et al.*, *Immunity* **42**, 665–678 (2015).
11. T. Masuda *et al.*, *Nature* **566**, 388–392 (2019).
12. Q. Li *et al.*, *Neuron* **101**, 207–223.e10 (2019).
13. T. R. Hammond *et al.*, *Immunity* **50**, 253–271.e6 (2019).
14. D. A. Menassa, D. Gomez-Nicola, *Front. Immunol.* **9**, 1014 (2018).
15. A. Monier *et al.*, *J. Neuropathol. Exp. Neurol.* **66**, 372–382 (2007).
16. A. Monier, P. Evrard, P. Gressens, C. Verney, *J. Comp. Neurol.* **499**, 565–582 (2006).
17. B. B. Lake *et al.*, *Nat. Biotechnol.* **36**, 70–80 (2018).
18. S. Zhong *et al.*, *Nature* **555**, 524–528 (2018).
19. H. Keren-Shaul *et al.*, *Cell* **169**, 1276–1290.e17 (2017).
20. S. Krasemann *et al.*, *Immunity* **47**, 566–581.e9 (2017).
21. I. R. Holtman *et al.*, *Acta Neuropathol. Commun.* **3**, 31 (2015).
22. M. Borggrewe *et al.*, *Glia* **66**, 2645–2658 (2018).
23. X. Qiu *et al.*, *Nat. Methods* **14**, 979–982 (2017).
24. R. Parakalan *et al.*, *BMC Neurosci.* **13**, 64 (2012).
25. S. Ghosh, E. Castillo, E. S. Frias, R. A. Swanson, *Glia* **66**, 1200–1212 (2018).
26. D. Gosselin *et al.*, *Science* **356**, eaal3222 (2017).
27. T. F. Galatro *et al.*, *Nat. Neurosci.* **20**, 1162–1171 (2017).
28. S. Aibar *et al.*, *Nat. Methods* **14**, 1083–1086 (2017).
29. M. Marin, A. Karis, P. Visser, F. Grosveld, S. Philipsen, *Cell* **89**, 619–628 (1997).
30. A. M. Smith *et al.*, *Glia* **61**, 929–942 (2013).
31. K. Kierdorf *et al.*, *Nat. Neurosci.* **16**, 273–280 (2013).
32. S. Heinz *et al.*, *Mol. Cell* **38**, 576–589 (2010).
33. P. Réu *et al.*, *Cell Rep.* **20**, 779–784 (2017).
34. T. Masuda, R. Sankowski, O. Staszewski, M. Prinz, *Cell Rep.* **30**, 1271–1281 (2020).
35. M. van der Poel *et al.*, *Nat. Commun.* **10**, 1139 (2019).
36. A. Moussaieff *et al.*, *Cell Metab.* **21**, 392–402 (2015).
37. A. Wlodarczyk *et al.*, *EMBO J.* **36**, 3292–3308 (2017).
38. N. Hagemeyer *et al.*, *Acta Neuropathol.* **134**, 441–458 (2017).
39. I. Jakovcevski, R. Filipovic, Z. Mo, S. Rakic, N. Zecevic, *Front. Neuroanat.* **3**, 5 (2009).
40. I. Kostović, I. Ž. Išasegi, Ž. Krsnik, *J. Anat.* **235**, 481–506 (2019).
41. D. Mattei *et al.*, *Transl. Psychiatry* **7**, e1120 (2017).
42. M. Hornig *et al.*, *Mol. Psychiatry* **23**, 759–766 (2018).
43. L. A. Croen *et al.*, *Autism Res.* **12**, 1551–1561 (2019).
44. L. Kracht *et al.*, Molecular Neurobiology/Human fetal microglia acquire homeostatic immune-sensing properties early in development, Version v1.1, Zenodo (2020); https://doi.org/10.5281/zenodo.3835875.

ACKNOWLEDGMENTS

We thank Stichting Stimezo Groningen and Gynaikon Clinic Rotterdam for providing fetal tissue samples. We thank M. Nieveen for assistance with sample collections, A. Alsema and M. Meijer for assistance in imaging and data analysis, and the UMCG Central Flow Cytometry Unit for aiding in cell sorting. We thank P. van der Vlies, D. Brandenburg, N. Festen, and W. Uniken Venema for assistance in the implementation of the modified Smart-seq2 protocol, D. Spierings and K. Hoekstra-Wakker for sequencing-related advice, and D. Smith for scientific editing. **Funding:** L.K. is funded by a fellowship from the Graduate School of Medical Sciences, University of Groningen. M.B. and S.M.K. are supported by the Dutch MS Research Foundation (13-833, 16-947). S.E. holds a scholarship from the Junior Scientific Masterclass, University of Groningen. J.R.P. received a Mandema fellowship, University of Groningen. The Zabawas Foundation (Den Haag, The Netherlands) is thanked for financial support of sequencing. **Author contributions:** B.J.L.E., L.K., M.B., and S.M.K. contributed to conceptualization. L.K., M.B., N.B., and S.E. designed and performed the experiments. L.K. and M.B. analyzed and visualized the data. B.J.L.E., J.D.L., and S.M.K. acquired funding. B.J.L.E., J.R.P, S.A.S., and S.M.K. contributed to supervision. L.K., M.B., S.E., and S.M.C.d.S.L. collected and isolated tissue samples. M.B., L.K., and B.J.L.E. wrote the manuscript. All authors revised and edited the manuscript. **Competing interests:** The authors have no competing interests to disclose. **Data and materials availability:** All next-generation sequencing data can be viewed at NCBI GEO under accession number GSE141862. Code used for data analysis is deposited at Zenodo (*44*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/530/suppl/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S10
References (*45*–*57*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

16 December 2019; accepted 5 June 2020
10.1126/science.aba5906

REPORT

MOLECULAR DIFFUSION

Boosted molecular mobility during common chemical reactions

Huan Wang[1], Myeonggon Park[1,2], Ruoyu Dong[1], Junyoung Kim[1,3], Yoon-Kyoung Cho[1,3], Tsvi Tlusty[1,2], Steve Granick[1,4]*

Mobility of reactants and nearby solvent is more rapid than Brownian diffusion during several common chemical reactions when the energy release rate exceeds a threshold. Screening a family of 15 organic chemical reactions, we demonstrate the largest boost for catalyzed bimolecular reactions, click chemistry, ring-opening metathesis polymerization, and Sonogashira coupling. Boosted diffusion is also observed but to lesser extent for the uncatalyzed Diels-Alder reaction, but not for substitution reactions S_N1 and S_N2 within instrumental resolution. Diffusion coefficient increases as measured by pulsed-field gradient nuclear magnetic resonance, whereas in microfluidics experiments, molecules in reaction gradients migrate "uphill" in the direction of lesser diffusivity. This microscopic consumption of energy by chemical reactions transduced into mechanical motion presents a form of active matter.

How to transduce chemical activity into motion is of mounting interdisciplinary interest in chemistry, biology, and physics (*1*–*17*). However, the known examples of how to do so on the molecular level were produced by nature to achieve biomolecular functions with molecular motors (*2*, *3*) and catalytic enzymes (*4*–*6*) or in specially designed synthetic chemical systems (*7*–*9*), making their generality difficult to assess.

Our experiments show that common chemical reactions can produce mobility more rapid than Brownian diffusion. On the basis of the use of two independent techniques, pulsed field–gradient nuclear magnetic resonance (NMR) spectroscopy and gradient microfluidics, the findings cannot be attributed to gas production or convection and instead are qualitatively consistent with findings reported for catalytic enzymes (*4*–*6*, *15*, *16*), whose interpretation excites debate (*17*). Our interpretation of these findings suggests that a chemical reaction in liquids can be a nonlocal phenomenon with substantial mechanical perturbation of nearby solvent molecules over distances tens to hundreds of solvent molecules. This finding raises fundamental questions about how energy is funneled from reactants to products in the solvent environment.

To measure molecular diffusion in solution, we used pulsed field–gradient NMR spectroscopy, which provides chemical specificity regarding both the abundance and the mobility of multiple species in the same sample (*18*). Experimental precautions that need to be taken against solution currents from convection are known (*19*, *20*). Briefly, we labeled molecules spatially using a magnetic gradient along the NMR tube. Radio-frequency pulses with equal duration and opposite magnitude were applied as echo, with a set diffusion time (Δ) between them so that nondiffusing spins were unaffected but displacement of the diffusing ones was encoded as intensity attenuation (*18*). A time limitation of the latter is that the time for each scan must exceed the spin-lattice relaxation time (T_1) of nuclear spins, a number that can be hundreds of milliseconds for catalysts that contain a paramagnetic metal. Another limitation is that the method's limited sensitivity demands numerous averaging scans to reach the needed signal-to-noise ratio.

Thus, we studied chemical reaction conditions with rates sufficiently slow to permit 3 to 5 min of signal averaging per data point and validity of the steady-state assumption that underlies the data analysis yet is sufficiently high to give adequate signal regarding mobility changes. Reactions were typically completed in ~2 hours. We confirmed the successful suppression of convection in a reaction known (*20*) to produce bulk convection (fig. S1) and used the same convection suppression pulses routinely (figs. S2 to S6). The successful suppression of convection artifacts lends confidence to the present measurements.

As summarized in table S1, we began by screening 15 organic reactions that could be run under mild conditions—for example, room temperature and standard pressure. The selected reactions spanned a wide range of thermodynamic driving force (reaction free energy) and reaction rate, and we considered both catalyzed and uncatalyzed reactions. Some reactions with large thermodynamic driving force were, given the above-mentioned technical constraints, too rapid to follow, such as acid-base reactions, ketone-hydroxylamine condensations, and strain-promoted copper-free click reactions. Some were too slow to follow, such as unimolecular elimination. Some produce gas by-products, such as Grubbs ring-closing metathesis, tetrazine-strained alkene addition, and Staudinger ligation. From this screening, we identified four reactions—copper-catalyzed click, Grubbs ring-opening metathesis polymerization (ROMP), Sonogashira coupling, and Diels-Alder—during which molecular mobility was demonstrably more rapid than Brownian diffusion, whereas two other reactions (substitution reactions S_N1 and S_N2) showed null results. The parameters of the NMR experiments are listed in table S2. Although making comparisons between these different reactions in the same solvent would have been optimal, different solvents were needed to obtain the needed reaction rates.

In an example of a click reaction (*21*), the aqueous-phase copper (*I*) catalyzed 1,3-dipolar cycloaddition of an alkyne and an azide (Fig. 1A); a schematic depiction of the catalytic cycle is shown in Fig. 1B. Chemical shifts of different proton nuclei were monitored during the reaction, with raw data illustrated in fig. S2A, and from the peak heights as a function of time elapsed, we quantified the reaction kinetics. We monitored the solvent in which the catalyst was dispersed by using D_2O solvent containing H_2O tracer and found no dependence on tracer concentration in the range 0.4 to 2%. Raw data in pulsed-field gradients (figs. S3 and S4) gave the apparent translational diffusion coefficient D_{app} of different proton nuclei as a function of elapsed time, deduced from a log-linear plot of the signal intensity against b, a function of gradient field (*G*) squared, $I/I_0 = \exp(-D_{app}b)$, whose slope is D_{app} (figs. S5 and S6, and fits in tables S3 and S4). The high linearity we found experimentally (correlation coefficient > 0.9999) (Fig. 1C and table S3) validated the applicability of this method to analyze the data. Varying the diffusion time Δ by an order of magnitude, 10 to 100 ms, made no difference (fig. S7). We confirmed known Brownian diffusion coefficients—for example, 1.9×10^{-9} m^2 s^{-1} before and after reaction for H_2O in D_2O from the chemical shift at 4.7 parts per million (ppm), and 1.0×10^{-9} m^2 s^{-1} for ascorbate before reaction from chemical shifts at 4.8 and 3.7 ppm (for ascorbate, meaningful comparison was not possible after reaction given its complexation with copper ion). These control experiments lent credibility to the measurement of increased D_{app} during chemical reaction. In Fig. 1, D and E, we plot $\Delta D_{app}/D_o$, the relative diffusion increase over the Brownian diffusion coefficient of the same molecules.

[1]Center for Soft and Living Matter, Institute for Basic Science (IBS), Ulsan 44919, South Korea. [2]Department of Physics, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, South Korea. [3]Department of Biomedical Engineering, UNIST, Ulsan 44919, South Korea. [4]Department of Chemistry, UNIST, Ulsan 44919, South Korea.
*Corresponding author. Email: sgranick@gmail.com

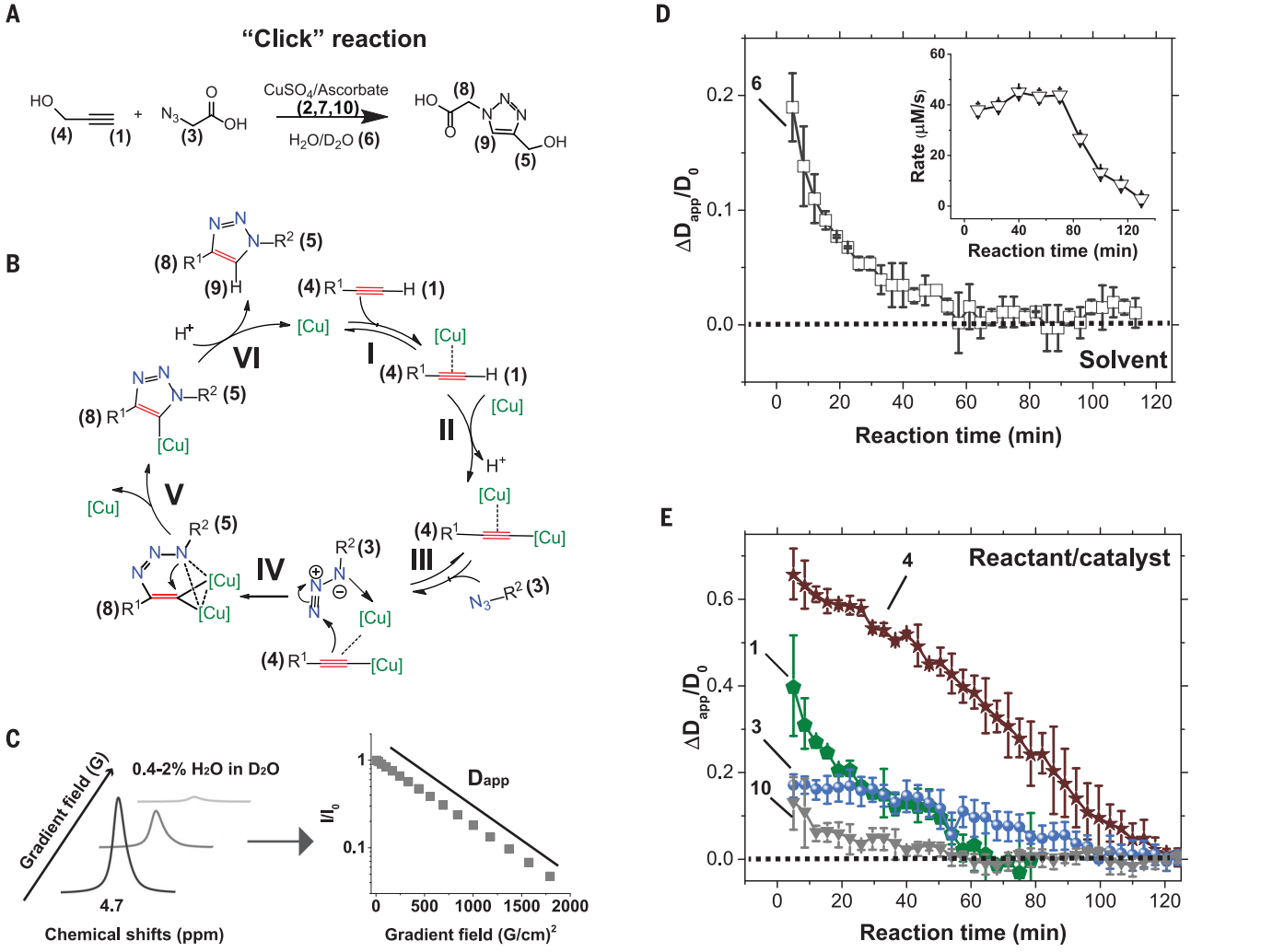


Fig. 1. Increased diffusion during click chemical reaction. (**A**) Reaction equation. The numbers correspond to the order of NMR shifts whose time dependence we track in (B) to (E) (fig. S2A). (**B**) Schematic depiction of the click reaction catalytic cycle, adapted from (*29*). Double arrows indicate reversible steps, and single arrows indicate irreversible steps. Intermediate reactions are identified by roman numerals. (**C**) Representative diffusion NMR spectra showing that a magnetic gradient produces attenuation I/I_0 in magnetic gradient squared and linearity on the log-linear scale from which apparent diffusion coefficient is calculated. (**D**) Increased diffusion plotted against reaction time for solvent, with reaction rate plotted in the inset. (**E**) Increased diffusion plotted against reaction time for the reactant and catalyst. Reaction conditions: 0.25 M reactant, 20 mM catalyst, in 0.4 to 2% H_2O-doped D_2O.

The earliest time point measurements were ~5 min after mixing the reactants, the time to mount the NMR tube and align it. The diffusion of water increased similarly as much (Fig. 1D and fig. S2A, peak 6) as for the catalyst Cu-ascorbate complex, which had a molar mass an order of magnitude greater, 240 g mol^{-1} (Fig. 1E, gray curve, and fig. S2A, peak 10). During subsequent reaction, the diffusion increase lessened but more slowly than would be consistent with a single exponential process. Comparing with reaction rate, which is the rate of reactant consumption (Fig. 1D, inset), the diffusion rate increased, whereas chemical reaction rate remained constant. In other experiments, we changed the reactant concentration. For higher reactant concentrations, increased diffusion persisted longer (fig. S8), but the maximum amount of increase was unaffected. Below, we discuss the possible dependence on catalyst concentration.

The click reaction is exothermic, but exothermicity alone did not explain the observed increase of up to 20%. First, temperature rise was <1.5°C (fig. S9), which is too small by an order of magnitude because diffusion coefficient is usually proportional to absolute temperature and inversely proportional to viscosity. Physical considerations show that local temperature pulses dissipated too rapidly to be relevant (*5*). Pulses of heat from local reaction events might produce convection on nanometer and micrometer scales. We conducted simulations of heat transfer (supplementary text) and did not find support for this hypothesis (fig. S10).

Olefin metathesis through the ruthenium-catalyzed (Grubbs catalyst) ROMP of norbornene to polynorbornene (Fig. 2, A and B) (*22*) showed similar patterns. The solvent was benzene, a nonpolar molecule, so its consistency with findings for the aqueous click reaction showed that the effects we report were neither solvent nor metal-catalyst specific. Diffusion of both solvent (Fig. 2C) and catalyst (Fig. 2D) increased while the reaction took place, and we tracked the reaction kinetics (Fig. 2C, inset). We monitored the catalyst diffusion coefficient from the ligand chemical shift (figs. S2B, S11, and S12 and fits in table S5) and observed an initial increase of ~15%, followed by decay, which was consistently higher than solvent diffusion increase throughout the reaction time. These trends correspond to the Grubbs catalyst

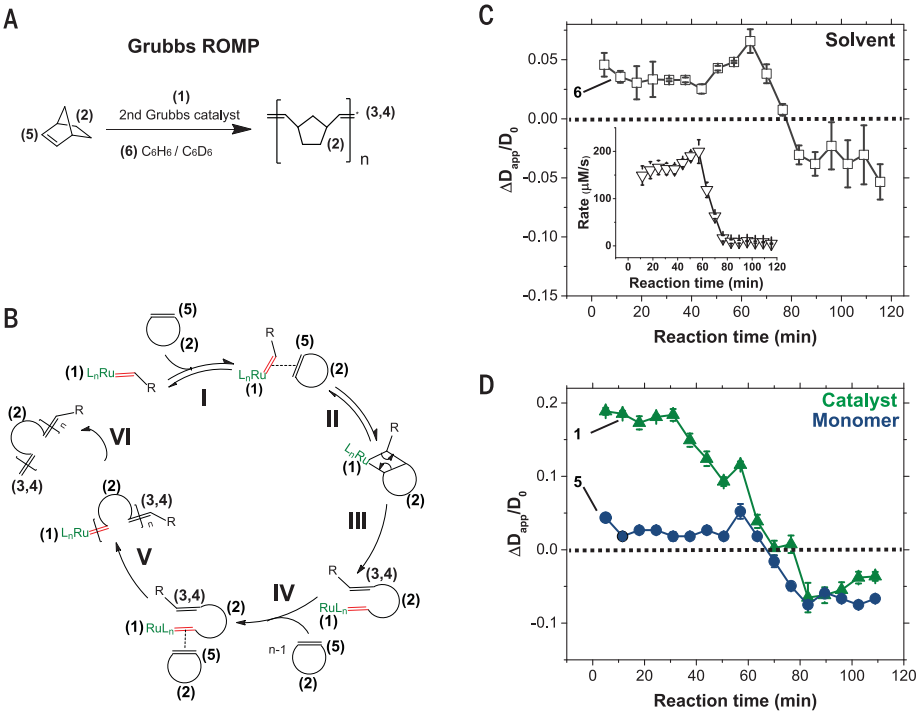


Fig. 2. Increased diffusion during Grubbs-catalyzed ROMP. (**A**) Reaction equation. The numbers correspond to the order of NMR shifts whose time dependence we track in (B) to (D) (fig. S2B). (**B**) Schematic depiction of the catalytic cycle, adapted from (*22*). Double arrows indicate reversible steps, and single arrows indicate irreversible steps. Intermediate reactions are identified by roman numerals. (**C**) Increased diffusion plotted against reaction time for solvent with reaction rate plotted in the inset. (**D**) Increased diffusion plotted against reaction time for the monomer and catalyst. Reaction conditions are 1 M reactant, 0.12 mM catalyst, in 2 to 10% C_6H_6-doped C_6D_6.

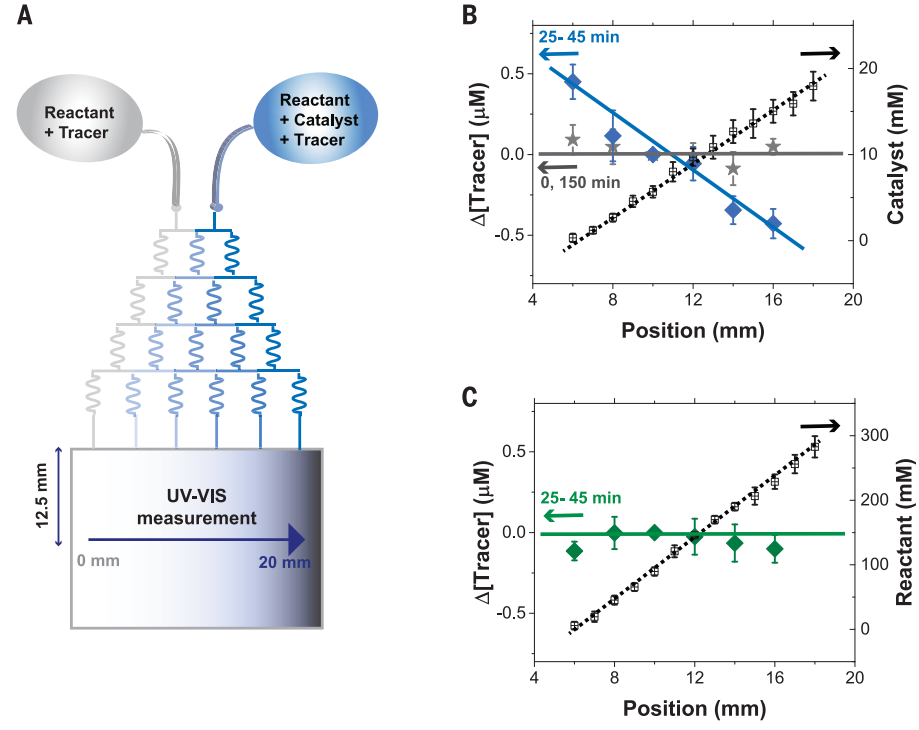


Fig. 3. Increased diffusion induces antichemotaxis. (**A**) Microfluidic gradient chip and inverse relation of reaction gradient and tracer gradient for click reaction. Condition is the same as in Fig. 1. Reaction gradient is produced by the catalyst gradient across the chip as mixture containing catalyst is infused from one sole inlet, with the same tracer concentration in both inlets. Tracer molecules close to catalysts are most strongly influenced by the momentum of local hydrodynamic flow. (**B**) Reaction produces dynamic tracer concentration profile across the channel, determined with spatially resolved UV-vis absorption measurement. Change of tracer concentration (left ordinate) and catalyst concentration (right ordinate) are plotted against position normal to flow during reaction (25 to 45 min) and after reaction is completed (150 min). (**C**) Control experiment testing for antichemotaxis in a large reactant concentration gradient gives a null result in the absence of catalyst.

requiring an initiation time (Fig. 2B, intermediate steps I, II, and III) before reaching maximum catalytic efficiency. Because the catalyst attaches to the end of propagating polymer chain (Fig. 2B, intermediate steps IV and V) during ROMP, the mass of the diffusing molecule should increase by the 92 g mol^{-1} per catalysis cycle (*23*). We confirmed that the increase of viscosity in the system created by the growing polymer chains slowed diffusion of reactant molecules (Fig. 2D). Thus, by the end of reaction, the catalyst diffusivity was ~5% slower than before reaction started, unlike the click reaction, which after completion had produced no permanent diffusivity changes and hence, by inference, no viscosity change.

Sonogashira coupling is a more complex system because it uses cocatalysts. This reaction forms a carbon-carbon bond between terminal alkyne and aryl halide through a main palladium catalyst and a copper cocatalyst (figs. S13 to S15 and fits in table S6). Our fourth to sixth examples were noncatalyzed reactions, Diels-Alder (figs. S16 to S18 and fits in table S7), S_N2 (figs. S19 to S21 and fits in table S8) and S_N1 (figs. S22 to S24 and fits in table S9). The results for the S_N1 and S_N2 reactions gave

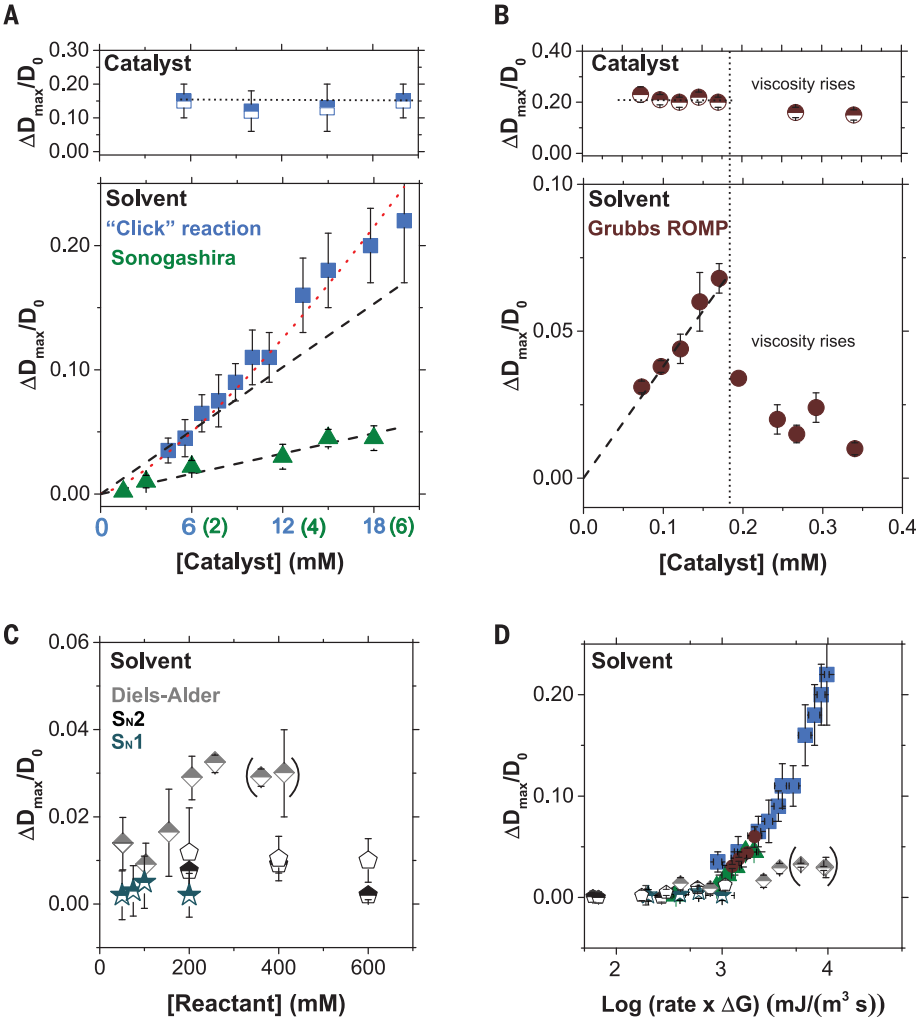


Fig. 4. Summary of five chemical systems. (**A** to **D**) Click, Grubbs ROMP, Sonogashira reaction, and noncatalyzed reactions: Diels-Alder (rhomboids), S_N1 (stars), S_N2 at (open pentagons) and above (half-filled pentagons) room temperature as specified in the supplementary materials. Ordinates are maximal normalized increased diffusion plotted against catalyst concentration regarding [(A) and (B)] catalyst diffusion and [(A), (B), and (C)] solvent diffusion. Lines with slope 1 (black dashed) and power 4/3 (red dotted) are guides to the eye. (D) Maximal normalized increased diffusion is plotted against free-energy release rate by using the thermodynamic parameters in table S10. Points in parentheses are believed less reliable because the rate of this second-order reaction is highest in the beginning, so its maximum was likely missed in these measurements averaged over 10 to 15 min per datum.

putative findings of up to 1% diffusion increase, which we consider to be a null result given the experimental uncertainty. The Diels-Alder addition of maleimide and furfurylamine showed 1 to 3% diffusion increases, which was greater than the experimental uncertainty.

Our last example used an independent experimental method. Microfluidics-based experiments allowed us to set up catalyst gradients to test the prediction that when chemical stimuli form a gradient, migration of catalyst, reactant, and nearby solvent up (or down) this gradient should display spatially dependent D_{app} values of the kind observed long ago for the different problem of bacteria chemotaxis (*24*) and more recently for catalytic enzymes (*6*). To implement this approach to follow solvent diffusion, we added a dye of high absorbance to water for the click reaction and measured its spatially resolved ultraviolet-visible (UV-vis) absorbance to gauge local water diffusivity, near the channel middle to avoid possible wall boundary effects downstream from the channel entry (12.5 mm downstream with laminar flow of 100 μm s^{-1}). This stratagem was needed because for technical reasons, it was not feasible to directly spatially resolve molecules that participated in the chemical reactions. Fluorescence detection, which could be more sensitive, suffers from quenching from the azide reagent of click chemistry.

A glass microfluidic chip adapted from earlier studies (*6*, *16*) was used to produce uniform reactant and dye concentrations accompanied by linear gradients of catalyst concentration across the width of the microfluidic channel (Fig. 3A). Plotting change in local dye concentration (left ordinate) and local catalyst concentration (right ordinate) against position across width of the channel (Fig. 3B), we observed that the initially uniform dye concentration (0 min) adopted a concentration gradient with slope opposite to that of the catalyst gradient (25 to 45 min). After the reaction ended, the dye concentration reverted to homogeneity (150 min).

Ruling out origin from the reactant gradient alone, we performed a control experiment without catalyst but with a large gradient of reactant concentration and observed a null result (Fig. 3C), which was as expected because the gradient produces negligible concentration difference across the nanometer size of these molecules. The conceptual issue—how it is possible on physical grounds to have a stable, spatially dependent concentration gradient—was sorted out in the context of bacteria chemotaxis (*24*, *25*). Applying those ideas to this situation, conserving constant fluid density in the face of increased diffusion of molecules in the vicinity of catalysis events requires a counterbalanced backflow of mostly solvent molecules. As the concentration gradients stabilized toward steady state, the net currents of catalyst and solvent, including the backflow, vanish (*24*, *25*). We interpret these findings to confirm those from diffusion NMR.

What governs magnitude? The concentration dependence in all of these systems is summarized in Fig. 4, A to C. We surveyed and summarized relevant thermodynamic properties, noting that values computed from density functional theory (DFT) can have a large scatter. The free energy change ΔG, enthalpy change ΔH, activation energy barrier, activation volume ΔV, maximum reaction rate, and boosted diffusion for all of the reactions reported are summarized in table S10. Taken together, the data for catalyzed reactions over a range of catalyst and reactant concentrations, 0.1 to 20 mM for catalysts and 50 mM to 1 M for reactants, fall empirically on a master curve so that the mobility increase depends on the energy release rate, the product of ΔG times rate (Fig. 4D), with increase observed only beyond a threshold. As $\Delta G \approx \Delta H$, the data correlate similarly with ΔH. The results for the S_N1 and S_N2 reactions are below the threshold, likely suggesting that null results for these reactions (within experimental uncertainty) signifies that threshold energy release rate was not reached. For intramolecular alkyne cyclization, from the maximum energy

release rate one can anticipate increased diffusion of 2 to 3% according to this correlation, which is within the standard deviation of data reported in an earlier study (*20*, *26*) and consistent with the null result reported (*20*).

At energy release rates beyond the threshold, a diffusion increase was observed for Diels-Alder cycloaddition, but the magnitude was less than for the catalyzed reactions. This difference may indicate the prominent contribution of kinetic intermediates because catalytic reactions have broad time scales, broader than single-step reactions, and also multiple elemental reactions during each catalytic cycle. It is likely that multiple factors—including activation energy barrier, rate-limiting step, kinetic reversibility of individual elemental steps, and characteristic time scales of intermediates (supplementary text)—could all contribute to the magnitude of diffusion increase.

That catalyst diffusion is unaffected by catalyst concentration (Fig. 4, A and B) signifies that this is a single-particle property. At the same time, overall solvent mobility scales much more strongly with catalyst concentration (Fig. 4, A to C). Millimolar catalyst concentrations boosted the overall solvent mobility by 2 to 20%. We interpret this to signify that moving reaction centers coupled to nearby solvent hydrodynamically (*1*, *11*, *12*, *14*). As for the concentration dependence, data over limited concentration spans were consistent with the linearity drawn as guide to the eye (Fig. 4, A and B), complicated in the case of ROMP by viscosity rise. For the click reaction (Fig. 4A), it was possible to achieve the most extended concentration range, giving data more consistent with the power 4/3. Viewing the catalysts as units of "active matter" (*10*–*14*), they are predicted to induce velocity disturbance that decays as the inverse distance squared $v \sim 1/r^2$, boosting the diffusion by a factor that increases as 4/3 power of catalyst concentration, $\Delta D_{max}/D_0 \sim v^2 \sim 1/r^4 \sim c^{4/3}$, as evidenced by the click reaction. For the cases of more limited concentration spans, linear and 4/3 power laws are practically indistinguishable (supplementary text and fig. S25). Taken together, from all of these data, boosted solvent diffusion appears to a nonlocal phenomenon, with the reacting centers producing substantial mechanical perturbation of surrounding solvent.

The largest activation energy step often correlates to the rate-determining step in a chemical reaction, so transition states likely account for some of the differences we report. For example, catalysis speeds up reaction by lowering the activation barrier. Specifically, activation volume is larger for 1,3 dipolar addition (~–65 Å^3/molecule) (*27*) and Diels-Alder (~–65 to –50 Å^3/molecule) reactions than for substitution reactions (<–30 Å^3/molecule) (*28*). Unfortunately, values have not to our knowledge been reported for click (copper-catalyzed 1,3 dipolar addition), Grubbs-ROMP, or Sonogashira reactions, which possess multiple, incompletely elucidated transition states during their catalytic cycles, with the extra complication of precatalysis initiation.

Full understanding of these effects is likely to require consideration of not only barrier crossing rates but also ramifications of electronic structure changes during chemical reaction. These rapid electronic changes, manifested as electric field changes, will require solvated solvent molecules to reorganize—a long-range readjustment process in the environment that is slower than actual changes of the electronic structure. On physical grounds, we conjecture the likely mechanism to be that this mode of energy release produces transient translational motion of reacting centers at the nanoscale, accompanied by their random reorientations to produce a random walk over distances of tens to hundreds of solvent molecule dimensions. Conformational changes likely contribute (*5*, *15*) but cannot fully explain the results because the common S_N2 reaction (figs. S19 to S21) produced null results despite the Walden conformation inversions that they are known to experience during reactions over the same regime of reaction rate as for the click reaction. Our findings suggest a breakdown of the classical chemistry paradigm that chemical reactions and diffusivity are uncoupled. The empirical correlations presented to summarize the magnitude of boosted diffusion, although imperfect, may be useful to estimate the effect in as-yet untested reactions.

REFERENCES AND NOTES

1. K. Kanazawa, T. G. Sano, A. Cairoli, A. Baule, *Nature* **579**, 364–367 (2020).
2. R. D. Vale, R. A. Milligan, *Science* **288**, 88–95 (2000).
3. M. J. Shelley, *Annu. Rev. Fluid Mech.* **48**, 487–506 (2016).
4. S. Sengupta *et al*., *J. Am. Chem. Soc.* **135**, 1406–1414 (2013).
5. R. Golestanian, *Phys. Rev. Lett.* **115**, 108102 (2015).
6. A.-Y. Jee, S. Dutta, Y.-K. Cho, T. Tlusty, S. Granick, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 14–18 (2018).
7. J. Li, B. Esteban-Fernández de Ávila, W. Gao, L. Zhang, J. Wang, *Sci. Robot.* **2**, eaam6431 (2017).
8. V. García-López *et al*., *Nature* **548**, 567–572 (2017).
9. A. Coskun, M. Banaszak, R. D. Astumian, J. F. Stoddart, B. A. Grzybowski, *Chem. Soc. Rev.* **41**, 19–30 (2012).
10. B. Liebchen, H. Löwen, *Acc. Chem. Res.* **51**, 2982–2990 (2018).
11. S. Ramaswamy, *J. Stat. Mech. Theory Exp.* **2017**, 054002 (2017).
12. R. Kapral, A. S. Mikhailov, *Physica D* **318-319**, 100–104 (2016).
13. C. Bechinger *et al*., *Rev. Mod. Phys.* **88**, 045006 (2016).
14. M. C. Marchetti *et al*., *Rev. Mod. Phys.* **85**, 1143–1189 (2013).
15. J. Agudo-Canalejo, P. Illien, R. Golestanian, *Nano Lett.* **18**, 2711–2717 (2018).
16. A.-Y. Jee, Y.-K. Cho, S. Granick, T. Tlusty, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E10812–E10821 (2018).
17. M. Feng, M. K. Gilson, *Annu. Rev. Biophys.* **49**, 87–105 (2020).
18. E. O. Stejskal, J. E. Tanner, *J. Chem. Phys.* **42**, 288–292 (1965).
19. I. Swan *et al*., *J. Magn. Reson.* **252**, 120–129 (2015).
20. T. S. C. MacDonald, W. S. Price, R. D. Astumian, J. E. Beves, *Angew. Chem. Int. Ed.* **131**, 19040–19043 (2019).
21. H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **40**, 2004–2021 (2001).
22. O. M. Ogba, N. C. Warner, D. J. O'Leary, R. H. Grubbs, *Chem. Soc. Rev.* **47**, 4510–4544 (2018).
23. M. Yasir, P. Liu, I. K. Tennie, A. F. M. Kilbinger, *Nat. Chem.* **11**, 488–494 (2019).
24. M. J. Schnitzer, S. M. Block, H. C. Berg, E. M. Purcell, *Symp. Soc. Gen. Microbiol.* **46**, 15–34 (1990).
25. P. F. Tupper, X. Yang, *Math. Phys.* **468**, 3864–3881 (2012).
26. T. S. C. MacDonald, W. S. Price, J. E. Beves, *ChemPhysChem* **20**, 926–930 (2019).
27. B.-B. Ni *et al*., *Chem. Commun.* **49**, 10130–10132 (2013).
28. T. Asano, W. J. Le Noble, *Chem. Rev.* **78**, 407–489 (1978).
29. B. T. Worrell, J. A. Malik, V. V. Fokin, *Science* **340**, 457–460 (2013).

ACKNOWLEDGMENTS

We thank D. Frenkel, B. Grzybowski, and K. Schweizer for discussions. **Funding:** This work was supported by the taxpayers of South Korea through the Institute for Basic Science, project code IBS-R020-D1. **Author contributions:** H.W. and S.G. conceived and designed the project and wrote the paper. H.W. did experiments. M.P. contributed to experiments. R.D. analyzed heat effects. J.K. and Y.-K.C. designed the microfluidic chip. T.T. did physical analysis. All coauthors reviewed and approved the final manuscript. **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/369/6503/537/suppl/DC1
Materials and Methods
Figs. S1 to S25
Tables S1 to S10
References (*30*–*53*)

9 January 2020; accepted 2 June 2020
10.1126/science.aba8425

MATERIALS SCIENCE

Exceptional plasticity in the bulk single-crystalline van der Waals semiconductor InSe

Tian-Ran Wei[1,2]*, Min Jin[3]*, Yuecun Wang[4]*, Hongyi Chen[2], Zhiqiang Gao[2], Kunpeng Zhao[1,2], Pengfei Qiu[2], Zhiwei Shan[4], Jun Jiang[5], Rongbin Li[3], Lidong Chen[2]†, Jian He[6]†, Xun Shi[1,2]†

Inorganic semiconductors are vital for a number of critical applications but are almost universally brittle. Here, we report the superplastic deformability of indium selenide (InSe). Bulk single-crystalline InSe can be compressed by orders of magnitude and morphed into a Möbius strip or a simple origami at room temperature. The exceptional plasticity of this two-dimensional van der Waals inorganic semiconductor is attributed to the interlayer gliding and cross-layer dislocation slip that are mediated by the long-range In-Se Coulomb interaction across the van der Waals gap and soft intralayer In-Se bonding. We propose a combinatory deformability indicator (Ξ) to prescreen candidate bulk semiconductors for use in next-generation deformable or flexible electronics.

Inorganic semiconductors are versatile and have applications in information, sensor, catalysis, and energy technologies (*1*). Most bulk inorganic semiconductors are intrinsically brittle at ambient condition (*2*). The lack of intrinsic plasticity and deformability restricts the scope of application for bulk inorganic semiconductors. While doping or alloying may alter the electrical, magnetic, optical, and thermal properties of an inorganic semiconductor by orders of magnitude, doping hardly bestows any plasticity and deformability. A handful of examples, however, suggest that this is not always the case. Bulk single-crystalline ZnS is brittle under normal lighting but becomes plastically deformable in complete darkness at ambient temperature, as plastic deformation is inhibited by the trapping of photoexcited charge carriers and the electrostatic interactions between dislocation cores (*3*). Brittle materials in the bulk form can also exhibit plasticity as nanowires (*4*). The discovery of intrinsic plastic deformability in bulk Ag_2S (*5*) has enabled the fabrication of a fully inorganic Ag_2S-based flexible thermoelectric power generation module (*6*). The plasticity of bulk Ag_2S is attributed to multicentered chemical bonding and solid linkage among easy slip planes (*5*) or to easy motion of Ag-S octagons without breaking Ag-S bonds in the presence of metallic Ag-Ag bonding (*7*).

Aside from Ag_2S, evidence for a wide range of bulk inorganic semiconductors being plastically deformable at ambient condition is not yet available. Note that two-dimensional van der Waals (2D vdW) inorganic materials gain pliability upon exfoliation, after thinning down to a monolayer or a few layers (*8*, *9*). This observation follows the general rule that the bending modulus is proportional to the cube of the thickness (*10*), and thus the stiffness of a bulk material can be lowered by reducing its cross-sectional area (*11*–*13*). Hence, monolayer or few-layer 2D vdW inorganic semiconductors become promising candidates for nanoelectronics (*14*, *15*).

Still, plastic deformability is not expected for a bulk 2D vdW inorganic semiconductor at room temperature. The thickness-induced flexibility does not change the brittle nature of the material, and large stress during deformation can still rupture a material of finite thickness, especially for those high–elastic modulus materials such as graphene or graphite (*16*). In a typical 2D vdW bulk inorganic semiconductor, the intralayer bonding is often covalent or ionic-covalent, creating intralayer rigidity. The presence of a vdW gap makes the material easy to cleave, thus making it difficult for the material to retain its structural integrity through plastic deformation.

In this work, we discovered that indium selenide (InSe), a 2D vdW group IIIA metal chalcogenide, exhibits exceptional plasticity in the bulk single-crystalline form but not in the bulk polycrystalline form. We characterized and pinpointed the atomic origin of deformability and plasticity and proposed a deformability factor (Ξ) to prescreen materials that can be morphed into various shapes for specific applications.

Despite the potential applications in electronics and optoelectronics (*17*–*19*), InSe is underutilized compared with other III-V and II-VI semiconductors. InSe adopts a layered structure in which each layer is formed by the In-Se honeycomb lattice in the *ab* plane, and the sequence of atoms in the *c* direction is Se-In-In-Se, with a vdW gap between Se and Se (Fig. 1A). These layers are stacked along the *c* axis with a period of ~0.8 nm (*c*/2, where *c* is the lattice constant along the *c* axis). Different layer-stacking sequences lead to various polytypes (*20*–*22*). The as-grown InSe crystalline ingot (*23*) and its cleaved facets have a metallic luster (Fig. 1, B and C). The measurements of packing density, x-ray diffraction (XRD), and transmission electron microscope–selected-area electron diffraction (TEM-SAED) (fig. S1) corroborate that the InSe samples studied here are β-phase single crystals with a space group of $P6_3/mmc$ (*20*). The unit cell contains two Se-In-In-Se layers, with the In atoms in one layer vertically aligned with Se atoms in the adjacent layer (Fig. 1A).

The as-grown InSe single crystal is a nondegenerate, native n-type semiconductor with a measured bandgap around 1.2 eV (fig. S2), consistent with previous reports (*18*, *24*). The sheets peeled off from an InSe single crystal can be bent, wound, twisted into a Möbius strip, or folded into a simple origami (Fig. 1, D to F). Large deformation induces large local stress, as reflected by the few wrinkles and microcleavages on the compression side of an ~300-μm-thick sheet (~10^5 unit cells along the *c* axis) that is folded into a "U" shape with an ~400-μm radius semicircle without visible fractures (Fig. 1G). Rather than extending through the sample, these microcleavages and wrinkles are local, and most can be healed by gentle pressing after unfolding (fig. S3). Cyclic bending deformation is found to have a trivial effect on the photoluminescence and Raman spectra and a limited effect on the electrical resistance (fig. S4). Such exceptional deformability and plasticity are not observed in other 2D materials with similar thickness or in inorganic semiconductors such as SnSe and Bi_2Te_3 (fig. S5). In fact, 2D materials such as graphene can only be folded and unfolded when they are one or a few layers thick (*25*) or by artificial microstructural manipulations (*26*).

As shown in Fig. 1, H and I, InSe crystals exhibit ~80% compression strain along and perpendicular to the *c* axis. Because of the equipment limit, the ultimate strain is not reached. The serrations in the stress-strain curve indicate the dislocation nucleation, crosslink, and break off. Large plastic deformability is also confirmed in the uniaxial tensile tests (fig. S6A) and three-point bending tests (fig. S6B). Progressive roller pressing can compress a 660-μm-thick crystal slab to the equipment limit of ~15-μm thickness (fig. S7), corresponding

[1]State Key Laboratory of Metal Matrix Composites, School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China. [2]State Key Laboratory of High Performance Ceramics and Superfine Microstructure, Shanghai Institute of Ceramics, Chinese Academy of Sciences, Shanghai 200050, China. [3]College of Materials, Shanghai Dianji University, Shanghai 201306, China. [4]Center for Advancing Materials Performance from the Nanoscale (CAMP-Nano) and Hysitron Applied Research Center in China (HARCC), State Key Laboratory for Mechanical Behavior of Materials, Xi'an Jiaotong University, Xi'an 710049, China. [5]Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo 315201, China. [6]Department of Physics and Astronomy, Clemson University, Clemson, SC 29634-0978, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: cld@mail.sic.ac.cn (L.C.); jianhe@g.clemson.edu (J.H.); xshi@sjtu.edu.cn and xshi@mail.sic.ac.cn (X.S.)

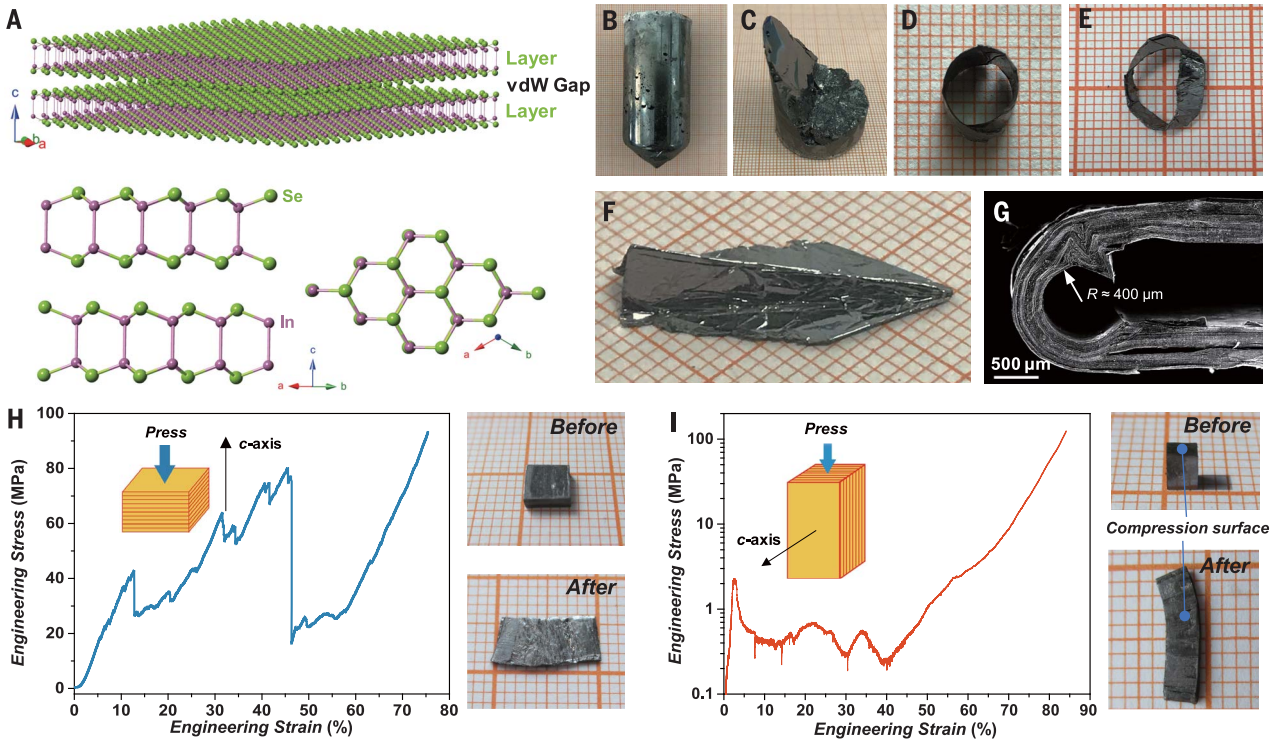


Fig. 1. Deformability of InSe single crystals. (**A**) Crystal structure of β-InSe and the projection on the (110) and (001) planes. (**B**) As-grown crystal and (**C**) cleavage surface. (**D** to **F**) InSe single crystal is morphed into various shapes without breaking. (**G**) Scanning electron microscopy (SEM) image of a folded crystal slab. *R*, radius. Compression engineering stress-strain curves along (**H**) and perpendicular to (**I**) the *c* axis. The smallest grid denotes 1 mm in all photographs.

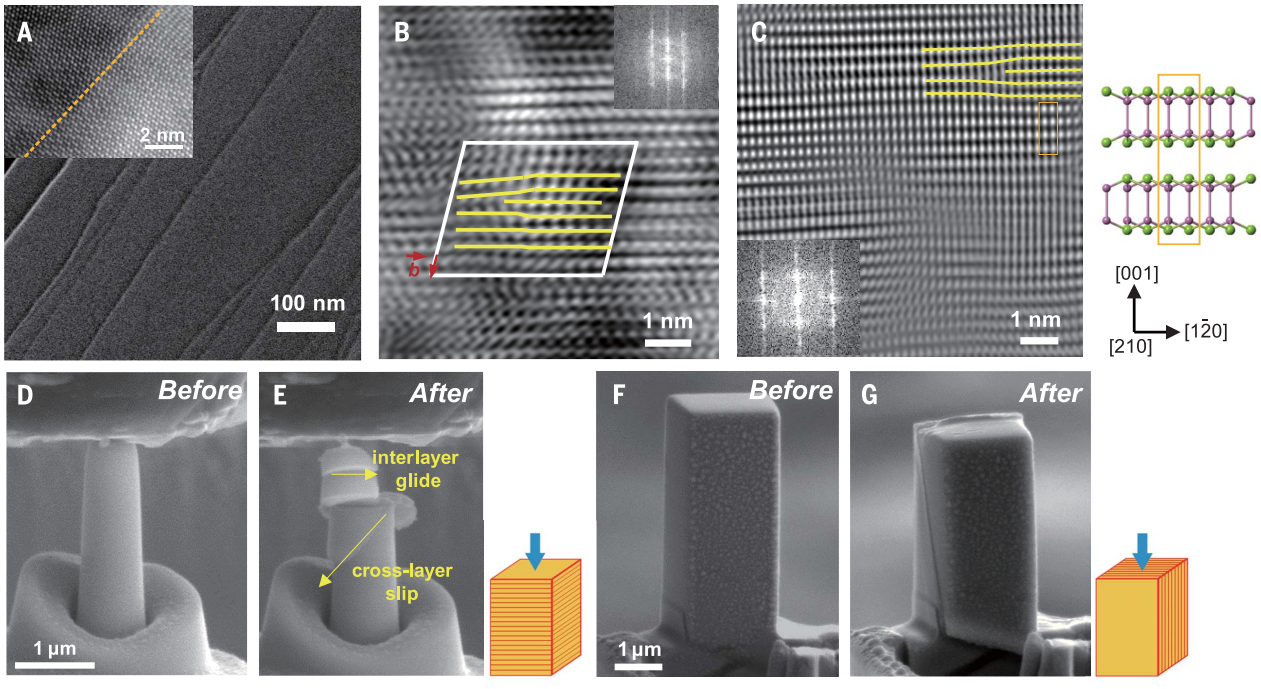


Fig. 2. Microscopic characterization and in situ SEM compression tests revealing the plasticity mechanisms of InSe. (**A** to **C**) TEM-STEM analyses of deformed InSe crystal; the zone axis is along (A) and perpendicular to [(B) and (C)] the *c* axis. (A) The slip steps are the result of easy interlayer gliding. The inset is the inverse Fourier transform from dark-field STEM (IFT-DF-STEM) image of the area near the slip step (dashed orange line). [(B) and (C)] IFT-DF-STEM images of edge dislocations. Yellow lines highlight the edge dislocation. (**D** to **G**) In situ SEM compression tests on InSe micropillars along [(D) and (E)] and perpendicular to [(F) and (G)] the *c* axis, showing interlayer gliding and cross-layer slip.

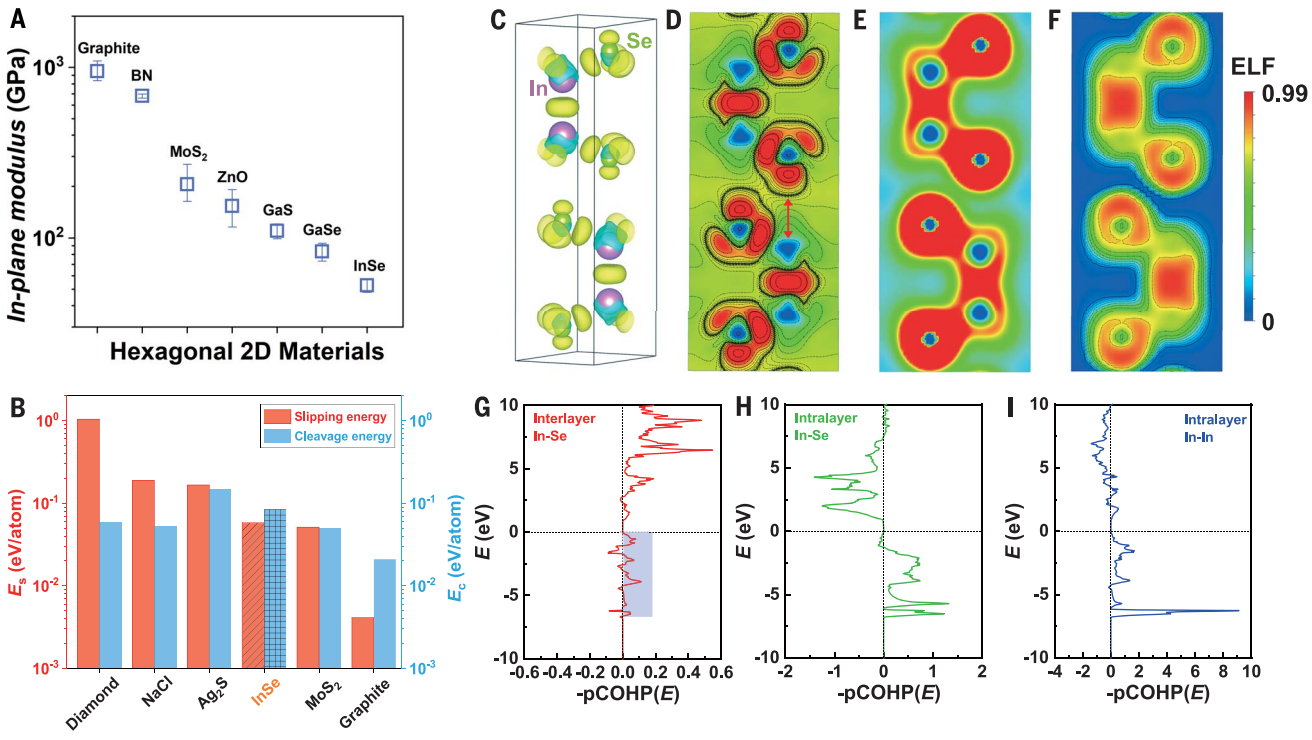


Fig. 3. Bond characters of InSe. (**A**) Intralayer modulus of InSe and several other hexagonal 2D inorganic semiconductors (table S1). (**B**) Calculated slipping energy (E_s, red, left axis) and cleavage energy (E_c, blue, right axis) for several representative materials. (**C**) Differential charge density and (**D**) projection on the (110) plane. (**E**) Charge density. (**F**) Electron localization function (ELF). (**G** to **I**) COHPs for interlayer In-Se bonding (G), intralayer In-Se bonding (H), and intralayer In-In bonding (I).

to a total compression ratio of >4000%. Brittle materials are shattered through a similar compression process. Scanning transmission electron microscopy (STEM) and TEM characterization of the deformed InSe single crystal provide atomic-scale details of plastic deformation and hints of the underlying mechanism. As shown in Fig. 2A and fig. S1, the intralayer lattice is nearly perfect, with very few defects; the slip steps are the sign of interlayer gliding. Meanwhile, the cross-layer lattice contains abundant defects, such as stacking faults, consistent with the streaked SAED spots (fig. S8A). Figure S8B displays the continuous, smooth, and fracture-free bending over many layers in a collective fashion that underlies the macroscopic bendability. Edge dislocations are clearly seen in Fig. 2, B and C, with the extra half-plane of atoms parallel to the basal *ab* plane. The cross-layer dislocation slip thus enables shear deformation under compression.

We also conducted in situ SEM uniaxial compression tests on the InSe single-crystalline micropillars along and perpendicular to the *c* axis (Fig. 2, D to G, and fig. S9). Consistent with the results of macroscopic compression tests (Fig. 1, H and I), the stress-strain curve of microscopic compression tests clearly shows slip-induced strain bursts (fig. S9), similar to those of metals (*27*), and the morphology after testing is free of discernible cleavage, confirming the exceptional plasticity. Through the process of compression, interlayer gliding occurs first, followed by cross-layer dislocation slip, as evidenced by the cross-layer shear band [Fig. 2, D to G, and movies S1 and S2 (*23*)]. The shear band angle (Fig. 2E) and the XRD patterns of roller-pressed samples (fig. S10) suggest the (10L) slip planes. Hence, it is the interlayer gliding and the cross-layer dislocation slip that enable the exceptional plasticity of bulk InSe single crystal.

To elucidate the atomic origin of the exceptional plasticity of bulk InSe single crystal, we inspected the intra- and interlayer interactions. For intralayer mechanical properties, we used the elastic stiffness constants (c_{ij}) data to deduce the intralayer Young's modulus E_{in}, which is isotropic within the *ab* plane of a hexagonal crystal lattice (*23*). The E_{in} value essentially reflects the intralayer bond rigidity, and the contribution from interlayer interactions is small. The E_{in} of InSe is ~53 GPa (Fig. 3A), among the lowest for the hexagonal-structured semiconductors, and explains the pliability of mono- or few-layer InSe.

Given the difficulty of experimentally characterizing the interlayer interactions and relative glide process, we invoked density functional theory calculations. On the basis of our TEM observations, we assume that the (001) planes slip along the [100] direction in InSe. For convenience, we divide each slipping period into 12 steps and calculate the energy variation versus the interplanar distance for each step (*23*). The slipping energy (E_s) barrier of InSe is as low as 0.058 eV per atom (Fig. 3B). In contrast, the cleavage energy (E_c) is 0.084 eV per atom, which is much higher than that of 2D MoS_2 and graphene. The E_c of InSe is also substantially higher than that of well-known brittle materials such as NaCl and diamond. The small slipping energy E_s points toward an easy interlayer gliding, while the relatively large cleavage energy E_c value favors a strong interlayer integrity during slip. Facilitated by the interlayer gliding and cross-layer slip (compare Fig. 2, D to G), plastic deformation takes place. The interlayer interaction, along with the pliable layers, makes large plastic deformation possible. The plasticity of InSe is strongly anisotropic owing to the limited slip systems.

The intralayer and interlayer interactions have a common origin in chemical bonding. We calculated the charge density map and the differential charge density map. One Se atom is bonded with three In atoms in one layer, leaving lone-pair electron lobes pointing to the interlayer van der Waals gap (Fig. 3, C and D). The electron localization function (ELF) of the

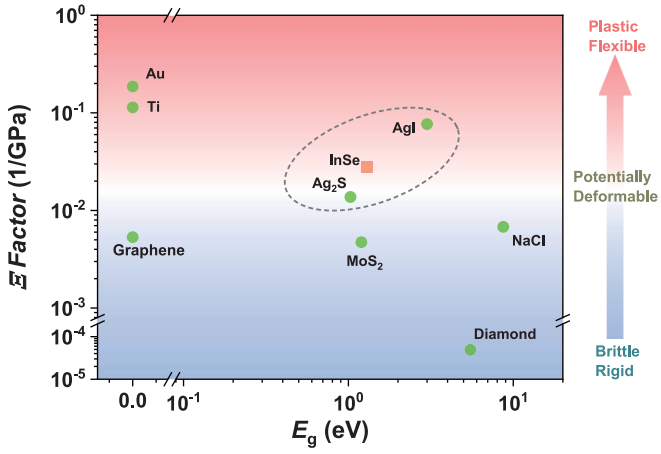


Fig. 4. Deformability factor Ξ = (E_c/E_s)(1/E_{in}) as a function of bandgap (E_g). E_c and E_s are the cleavage energy and slipping energy of the slip system, respectively; E_{in} is the Young's modulus along the slip direction.

intralayer In-Se bond is as low as 0.34 to 0.5 (Fig. 3F), which is comparable to that of metals (*28*), suggesting a considerable degree of delocalization, a low bond rigidity, and thus a low intralayer modulus (Fig. 3A). Similarly, one In atom is bonded to three Se atoms and another In atom. The In-In bond in the slab has a high ELF of 0.94 and a strong, sharp peak in the crystal orbital Hamilton population (COHP) (Fig. 3I), indicating a localized covalent feature.

For the interlayer interactions, we clearly see the Se-Se interaction from the distortion of the electron cloud in the differential charge density maps (Fig. 3, C and D). The vdW-like force is the prevalent feature of 2D materials. Of special interest is the presence of a long-range Coulombic interaction reflected in the charge density distortion between In and Se atoms across the layer (Fig. 3D, red arrows), spreading from 0 to −7 eV in the calculated COHP (Fig. 3G, blue rectangle). This long-range Coulombic In-Se interlayer interaction explains the peak-splitting in the Raman and infrared spectra of InSe (*29*) and GaSe (*30*). Here, the long-range coulombic In-Se interlayer interaction and the Se-Se vdW-like force work in tandem with a sufficiently large E_c to enable the interlayer gliding and cross-layer dislocation slip while retaining the structural integrity through plastic deformation.

More important than the specific criteria for deformability in InSe and Ag_2S is the development of a more general rule to find other candidate bulk materials. Any criteria should apply to both Ag_2S and InSe. As the deformability and bendability of 2D materials require large cleavage energy (E_c) and small slipping energy (E_s) to facilitate plastic deformation as well as pliable layers with low modulus to allow elastic bending, we propose that a deformability factor Ξ = (E_c/E_s)(1/E_{in}) (in units of inverse gigapascals) might be useful to prescreen bendable and deformable inorganic semiconductors. The E_c/E_s ratio quantifies the plasticity of the material that is in line with the criterion proposed by Rice *et al.* (*31*, *32*), while E_{in} is the in-plane Young's modulus along the slip direction. We summarize Ξ as a function of the bandgap (E_g) for a number of representative materials (Fig. 4). Metals are in the top-left corner with extremely large plasticity and deformability but zero bandgap. The wide-bandgap inorganic materials such as NaCl and diamond are in the bottom-right corner with brittleness and sizable bandgap. Notably, deformable semiconductor Ag_2S and InSe have close Ξ values, and they are located in the middle area with medium bandgaps. The diagram also shows that β-AgI with a large bandgap (~2.9 eV) should be deformable. This seems to be the case according to early literature (*33*). The Ξ-E_g diagram can thus be used as a guide to prescreen and develop deformable and plastic bulk inorganic semiconductors, especially in the vicinity of InSe and Ag_2S.

In this work, we discovered exceptional deformability and plasticity in bulk single-crystalline InSe, the first of its kind in van der Waals inorganic semiconductors. We envisage applications in deformable and flexible thermoelectric devices, sensors, and photodetectors, to name a few. The results of comprehensive experimental and theoretical study corroborate that such extraordinary mechanical properties stem from both the intralayer pliability and the interlayer long-range In-Se coulombic interaction across the van der Waals gap, a trade-off and synergy between the retention of substructures and multiple easy slip pathways. We believe that our proposed deformability factor will help guide materials development for discovering other bulk semiconductors useful in next-generation deformable or flexible electronics.

REFERENCES AND NOTES

1. K. Seeger, *Semiconductor Physics: An Introduction* (Springer, ed. 9, 2004).
2. K. T. Faber, K. J. Malloy, *The Mechanical Properties of Semiconductors*, vol. 37 of *Semiconductors and Semimetals* (Academic Press, 1992).
3. Y. Oshima, A. Nakamura, K. Matsunaga, *Science* **360**, 772–774 (2018).
4. X. D. Han *et al.*, *Adv. Mater.* **19**, 2112–2118 (2007).
5. X. Shi *et al.*, *Nat. Mater.* **17**, 421–426 (2018).
6. J. Liang *et al.*, *Energy Environ. Sci.* **12**, 2983–2990 (2019).
7. G. Li *et al.*, *npj Comput. Mater.* **4**, 44 (2018).
8. S. Das, J. A. Robinson, M. Dubey, H. Terrones, M. Terrones, *Annu. Rev. Mater. Res.* **45**, 1–27 (2015).
9. S. J. Kim, K. Choi, B. Lee, Y. Kim, B. H. Hong, *Annu. Rev. Mater. Res.* **45**, 63–84 (2015).
10. S. P. Timoshenko, S. Woinowsky-Krieger, *Theory of Plates and Shells* (McGraw-Hill, ed. 2, 1959).
11. J. A. Rogers, T. Someya, Y. Huang, *Science* **327**, 1603–1607 (2010).
12. H. Li, Z. Tang, Z. Liu, C. Zhi, *Joule* **3**, 613–619 (2019).
13. J. Peng, G. J. Snyder, *Science* **366**, 690–691 (2019).
14. G. R. Bhimanapati *et al.*, *ACS Nano* **9**, 11509–11539 (2015).
15. G. Fiori *et al.*, *Nat. Nanotechnol.* **9**, 768–779 (2014).
16. C. Lee, X. Wei, J. W. Kysar, J. Hone, *Science* **321**, 385–388 (2008).
17. T. Shen, J.-C. Ren, X. Liu, S. Li, W. Liu, *J. Am. Chem. Soc.* **141**, 3110–3115 (2019).
18. D. A. Bandurin *et al.*, *Nat. Nanotechnol.* **12**, 223–227 (2017).
19. S. R. Tamalampudi *et al.*, *Nano Lett.* **14**, 2800–2806 (2014).
20. S. Popović, A. Tonejc, B. Gržeta-Plenković, B. Čelustka, R. Trojko, *J. Appl. Cryst.* **12**, 416–420 (1979).
21. J. Rigoult, A. Rimsky, A. Kuhn, *Acta Crystallogr. B* **36**, 916–918 (1980).
22. G. Han, Z. G. Chen, J. Drennan, J. Zou, *Small* **10**, 2747–2765 (2014).
23. Materials and methods are available as supplementary materials.
24. S. Lei *et al.*, *ACS Nano* **8**, 1263–1272 (2014).
25. H. Chen *et al.*, *Science* **365**, 1036–1040 (2019).
26. L. Peng *et al.*, *Adv. Mater.* **29**, 1700589 (2017).
27. Z.-J. Wang *et al.*, *Appl. Phys. Lett.* **100**, 071906 (2012).
28. A. Ormeci, H. Rosner, F. R. Wagner, M. Kohout, Y. Grin, *J. Phys. Chem. A* **110**, 1100–1105 (2006).
29. C. Carlone, S. Jandl, H. R. Shanks, *Phys. Status Solidi B* **103**, 123–130 (1981).
30. T. Wieting, J. Verble, *Phys. Rev. B* **5**, 1473–1479 (1972).
31. J. R. Rice, *J. Mech. Phys. Solids* **40**, 239–271 (1992).
32. J. R. Rice, R. Thomson, *Philos. Mag.* **29**, 73–97 (1974).
33. S.-I. Pyun, J.-S. Bae, *J. Power Sources* **63**, 109–113 (1996).

ACKNOWLEDGMENTS

We thank F. Xu and C. Zhu at the Shanghai Institute of Ceramics, Chinese Academy of Sciences, for assisting with the TEM-STEM experiments and analyses, and C. Guo at Xi'an Jiaotong University for assistance with the micropillar compression tests. **Funding:** This work is supported by the National Key Research and Development Program of China (2018YFB0703600 and 2017YFB0702001), the National Natural Science Foundation of China (51625205, 91963208, 51961135106, and 51902249), and the Nature Science Foundation of Shanghai (19ZR1419900). M.J. and R.L. acknowledge the support of the Shanghai Engineering Research Center of Hot Manufacturing (18DZ2253400). **Author contributions:** T.-R.W., M.J., J.J., and R.L. prepared the samples and measured the physical properties. T.-R.W. and Z.G. measured the mechanical properties. H.C. performed ab initio calculations. Y.W. and Z.S. performed in situ SEM mechanical tests. T.-R.W., Y.W., H.C., Z.G., K.Z., and P.Q. collected the data and provided explanation and model under the guidance of X.S., J.H., and L.C. T.-R.W., M.J., Y.W., L.C., J.H., and X.S. wrote and edited the manuscript. All authors contributed helpful discussions to this work. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript and the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/542/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
Tables S1 and S2
References (*34*–*52*)
Movies S1 and S2

21 January 2020; accepted 8 June 2020
10.1126/science.aba9778

NEURODEVELOPMENT

Human-specific *ARHGAP11B* increases size and folding of primate neocortex in the fetal marmoset

Michael Heide[1]*, Christiane Haffner[1], Ayako Murayama[2,3], Yoko Kurotaki[4], Haruka Shinohara[4], Hideyuki Okano[2,3], Erika Sasaki[4], Wieland B. Huttner[1]*

The neocortex has expanded during mammalian evolution. Overexpression studies in developing mouse and ferret neocortex have implicated the human-specific gene *ARHGAP11B* in neocortical expansion, but the relevance for primate evolution has been unclear. Here, we provide functional evidence that *ARHGAP11B* causes expansion of the primate neocortex. *ARHGAP11B* expressed in fetal neocortex of the common marmoset under control of the gene's own (human) promoter increased the numbers of basal radial glia progenitors in the marmoset outer subventricular zone, increased the numbers of upper-layer neurons, enlarged the neocortex, and induced its folding. Thus, the human-specific ARHGAP11B drives changes in development in the nonhuman primate marmoset that reflect the changes in evolution that characterize human neocortical development.

Evolutionary expansion of the human neocortex is linked to our cognitive abilities (*1*–*6*). The human-specific gene *ARHGAP11B* (*7*, *8*) is implicated in this neocortical expansion because it is expressed in the human progenitor cells giving rise to neocortical neurons, and when overexpressed in developing mouse and ferret neocortex, two evolutionarily distant mammals, can induce features associated with neocortical expansion (*9*, *10*). *ARHGAP11B* arose ≈5 million years ago by partial duplication of ubiquitous *ARHGAP11A*, which encodes a Rho-GAP exhibiting nuclear localization (*7*–*9*, *11*). However, because of a point mutation that presumably occurred after the partial gene duplication event and leads to a human-specific change in protein sequence, ARHGAP11B lacks Rho-GAP activity in vivo and is localized in mitochondria. This promotes the proliferation of basal progenitors, which are implicated in neocortical expansion through glutaminolysis (*11*, *12*). Here, we tested ARHGAP11B's relevance for neocortical expansion in a nonhuman primate by expressing *ARHGAP11B* under the control of its own (human) promoter in transgenic fetal marmosets.

To express human-specific *ARHGAP11B* (*7*, *8*) (fig. S1A) in the common marmoset, we constructed a lentiviral vector. In this functionally verified vector (fig. S1, B and C), an ≈2.7-kb human genomic segment containing the *ARHGAP11B* promoter drives expression of an enhanced green fluorescent protein (EGFP) reporter, followed by the complete ARHGAP11B protein-coding sequence. The two proteins become separate polypeptides after translation because of the presence of a T2A self-cleaving sequence (fig. S1B). This expression vector was used to generate pregnant marmosets carrying *ARHGAP11B*-transgenic fetuses by following a previously established protocol (*13*). This protocol involves microinjection into fertilized marmoset oocytes and transfer of in vitro–developed embryos into foster mothers 3 to 5 days after ovulation (with the day of transfer being defined as day 0 of pregnancy; fig. S1D and table S1).

We confined our analyses to marmoset fetuses because we anticipated that expression of this human-specific gene would affect neocortex development in this animal. In light of potential unforeseeable consequences with regard to postnatal brain function, we considered it a prerequisite—and mandatory from an ethical point of view—to first determine the effects of *ARHGAP11B* expression on the development of fetal marmoset neocortex. To this end, we collected fetuses after Caesarian section at day 101 of the ≈150-day gestation (fig. S1D), a stage when neocortical development shows both progenitor cell division and production of neurons (destined mostly to the upper layers) and which corresponds to fetal human neocortical development at ≈16 weeks after conception. Of the seven *EGFP*- plus *ARHGAP11B*-transgenic marmoset fetuses obtained (table S1), five expressed both EGFP and ARHGAP11B in fetal neocortex, whereas two expressed neither (Fig. 1). In the five transgenic fetuses exhibiting EGFP and ARHGAP11B expression in neocortex, we found three or four lentivirus integration events at random genomic positions

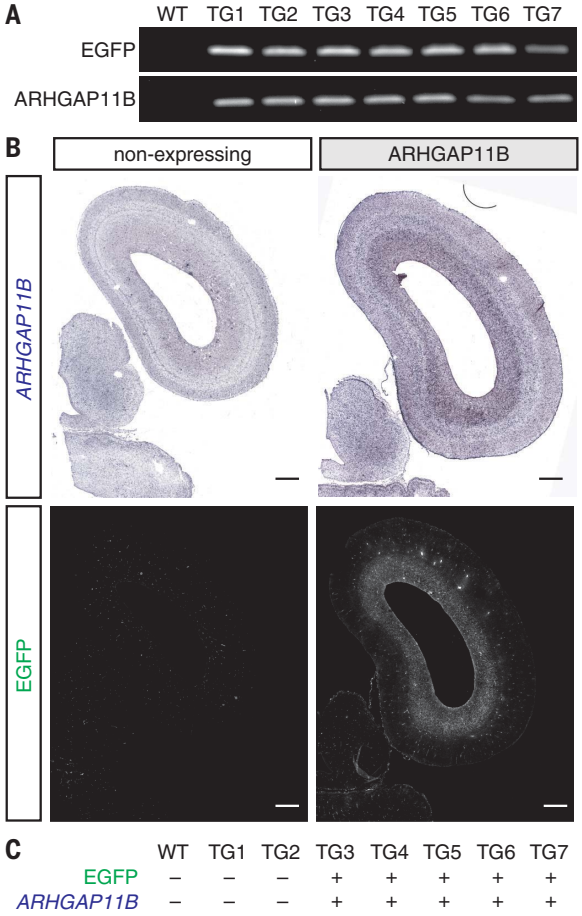


Fig. 1. *ARHGAP11B* and EGFP expression in *ARHGAP11B*-transgenic marmoset 101-day fetuses. Shown are the results of genomic PCR for *EGFP* and *ARHGAP11B* using somatic cells (**A**) and the absence (–) or presence (+) of EGFP protein and *ARHGAP11B* mRNA expression in neocortex (**B** and **C**) of one WT and seven *ARHGAP11B*-transgenic marmoset fetuses. (B) *ARHGAP11B* mRNA in situ hybridization (top) and EGFP immunohistochemistry (bottom) of *ARHGAP11B*-non-expressing (TG2) and *ARHGAP11B*-expressing (TG6) neocortex of marmoset fetuses. Scale bars, 500 μm.

[1]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany. [2]Department of Physiology, Keio University School of Medicine, Tokyo 160-8582, Japan. [3]Laboratory for Marmoset Neural Architecture, RIKEN Center for Brain Science, Wako City, Saitama 351-0198, Japan. [4]Department of Marmoset Biology and Medicine, Central Institute for Experimental Animals, Kawasaki 210-0821, Japan.
*Corresponding author. Email: heide@mpi-cbg.de (M.H.); huttner@mpi-cbg.de (W.B.H.)

per animal (table S2). *ARHGAP11B* mRNA expression in the marmoset neocortical wall resembled that in fetal human neocortex, with similar intensity and occurring preferentially in the germinal zones [i.e., the ventricular zone (VZ), inner subventricular zone (iSVZ), and outer subventricular zone (oSVZ) (*14*)] (*9*, *15*), like EGFP, as revealed by in situ hybridization (fig. S1E) and reverse transcription quantitative polymerase chain reaction (PCR) (fig. S1F).

The *ARHGAP11B*-expressing marmoset neocortex was larger and its cortical plate (CP) thicker than that in normal marmoset neocortex (Figs. 1B and 2A and fig. S1E) and, in contrast to the smooth surface of the normal marmoset brain, exhibited surface folds (Fig. 2A). Quantification of fetal marmoset neocortex as a whole indicated no statistically significant difference in width but a significant increase in length of *ARHGAP11B*-expressing neocortex compared with wild-type (WT) and *ARHGAP11B*-non-expressing neocortex (Fig. 2B). To quantify cortical folding, we analyzed coronal sections of fetal marmoset neocortex along the rostrocaudal axis (Fig. 2C) to obtain the gyrification index (GI) (fig. S2A), which is the ratio of tracing the de facto length of the (unfolded or folded) cortical surface (Fig. 2E, green) over a hypothetical minimal length, i.e., smooth, tracing of the cortical surface (Fig. 2E, magenta) (*16*, *17*). Applying this tracing to the entire dorsoventral dimension of the coronal sections analyzed, WT and *ARHGAP11B*-non-expressing neocortex exhibited a GI of nearly 1.0 (Fig. 2C), consistent with the essentially unfolded, near-lissencephalic nature of the marmoset neocortex (*18*, *19*). The GI of *ARHGAP11B*-expressing neocortex increased rostrally (Fig. 2C) and reached nearly 1.1 when the tracing was confined to the portion of the cortical surface where gyrus-like structures emerged (Fig. 2D and fig. S2B). These structures did not arise by folding of a CP of equal thickness, but rather reflected local CP thickening (fig. S2, C to E), which in turn reflected a specific increase in upper-layer neurons as revealed by immunostaining for markers of specific neuron populations (fig. S2, D and F).

We then quantified CP thickness in WT, *ARHGAP11B*-non-expressing, and *ARHGAP11B*-expressing marmoset neocortex, taking into consideration only regions where no gyrus-like structures emerged in the *ARHGAP11B*-expressing neocortex. This revealed increased CP thickness for *ARHGAP11B*-expressing neocortex compared with WT and *ARHGAP11B*-non-expressing neocortex (Fig. 3, A and B, and figs. S3 and S4A).

To understand the basis of this increase in CP thickness, we quantified CP nuclei that were positive for Tbr1 and Ctip2, two markers of deep-layer neurons, and CP nuclei that were positive for Satb2 and Brn2, which are expressed by upper-layer neurons (*20*, *21*) (Fig. 3A and fig. S4B). We observed a nearly 40 and 50% increase in Satb2$^+$ neurons and Brn2$^+$ neurons, respectively, but not in Tbr1$^+$ and Ctip2$^+$ neurons, in the CP of *ARHGAP11B*-expressing marmoset neocortex compared with WT and *ARHGAP11B*-non-expressing neocortex (Fig. 3C and fig. S4, C and D).

Consistent with the developmental stage of our analyses (fig. S1D), we noted that a substantial proportion of the Satb2$^+$ and Brn2$^+$ neurons observed in the cortical wall were found in the subplate (fig. S5, A and B), consistent with these neurons migrating to the CP (*22*, *23*). Accordingly, the numbers specifically of Satb2$^+$ and Brn2$^+$ neurons in the subplate

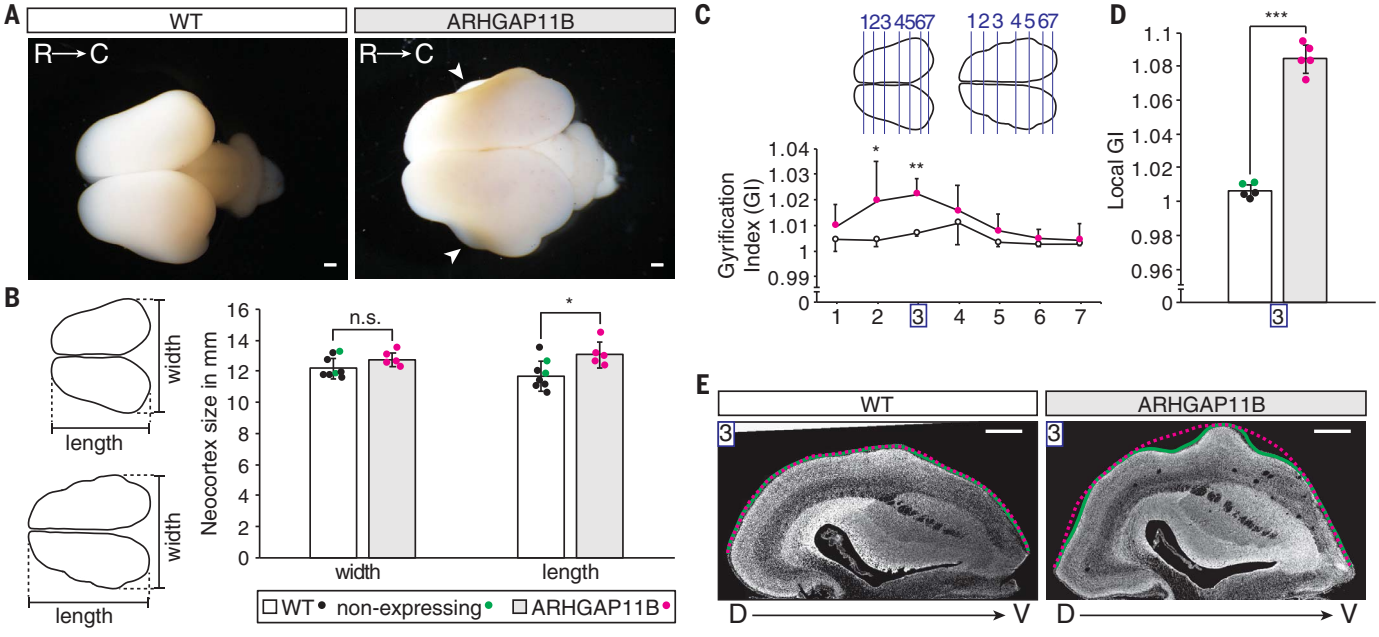


Fig. 2. Size and GI of WT and *ARHGAP11B*-non-expressing versus *ARHGAP11B*-expressing 101-day fetal marmoset neocortex. (**A**) WT brain and brain expressing *ARHGAP11B* in neocortex (TG3). Arrowheads indicate cortical folds. R, rostral; C, caudal. Scale bars, 1 mm. (**B**) Width and length (see diagrams) of six WT (black dots, white columns) plus two *ARHGAP11B*-non-expressing neocortices (green dots, white columns) versus five *ARHGAP11B*-expressing neocortices (magenta dots, gray columns). Data are shown as mean ± SD; n.s., not significant; *P < 0.05 (two-tailed t test). (**C**) GI [see (E) and fig. S2A] of three WT plus two *ARHGAP11B*-non-expressing neocortices (white circles) versus five *ARHGAP11B*-expressing neocortices (magenta circles) at seven positions along the rostrocaudal axis (see diagrams). Data are shown as mean ± SD; *P < 0.05; **P < 0.01 (one-tailed t test). (**D**) Local GI (see fig. S2B) of three WT neocortices (black dots, white column) plus two *ARHGAP11B*-non-expressing neocortices (green dots, white column) versus five *ARHGAP11B*-expressing neocortices (magenta dots, gray column). Data are shown as mean ± SD; ***P < 0.001 (two-tailed t test). (**E**) 4′,6-diamidino-2-phenylindole (DAPI)–stained coronal section of WT and *ARHGAP11B*-expressing (TG4) neocortex at position 3 [see (C)]. D, dorsal; V, ventral. Green line indicates the de facto length of the cortical surface; magenta line indicates the hypothetical minimal length of the cortical surface. Scale bars, 500 μm.

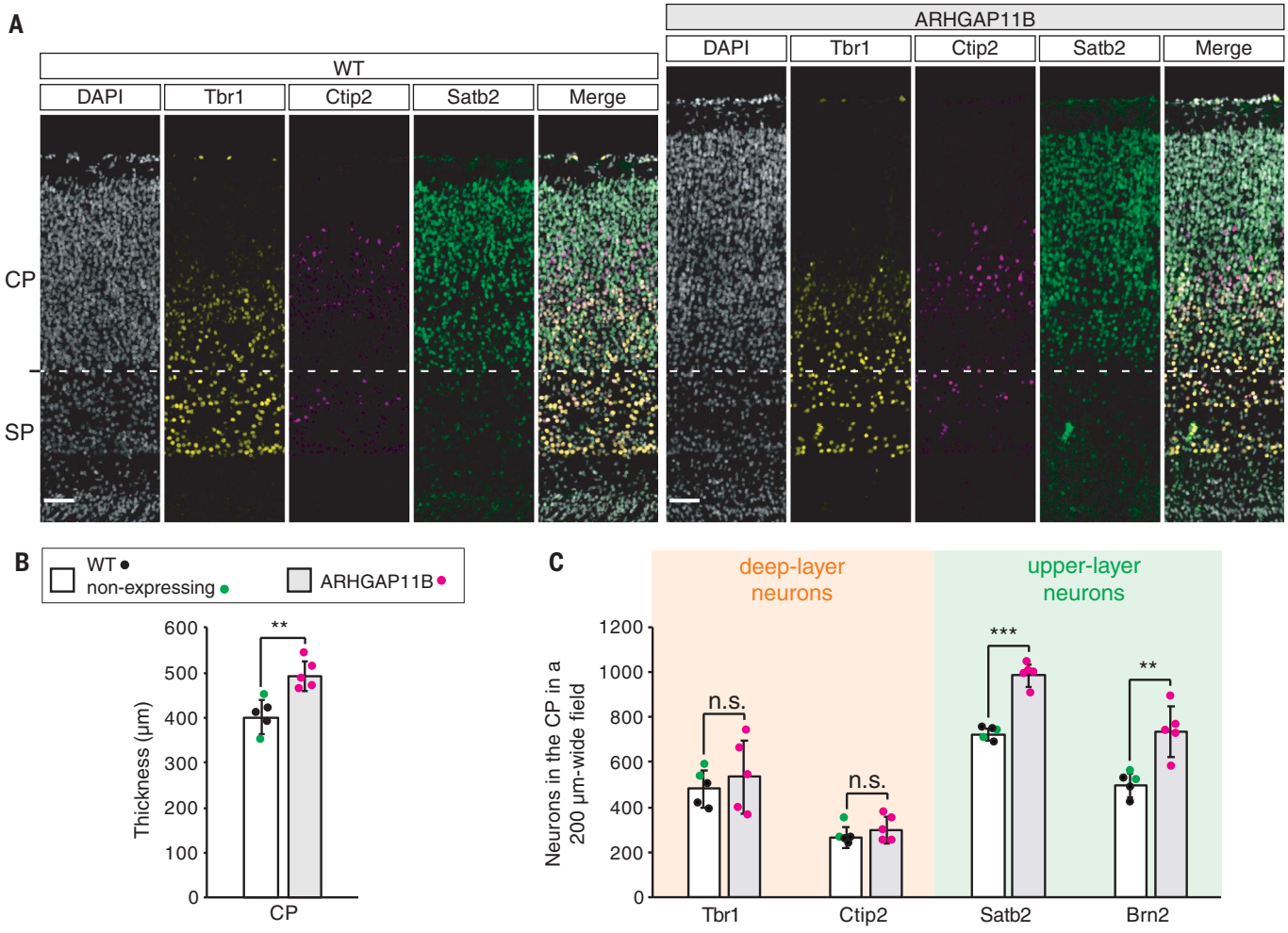


Fig. 3. *ARHGAP11B*-expressing 101-day fetal marmoset neocortex shows increased CP thickness and elevated numbers specifically of upper-layer neurons. (**A**) Triple immunofluorescence for Tbr1 (yellow), Ctip2 (magenta), and Satb2 (green) combined with DAPI staining (white), of WT (left) and an *ARHGAP11B*-expressing (TG3, right) neocortex (occipital lobe). Scale bars, 50 µm. (**B** and **C**) CP thickness (B) and Tbr1$^+$, Ctip2$^+$ Satb2$^+$, and Brn2$^+$ neuron number in CP in a 200-µm-wide field (C) of three WT neocortices (black dots, white columns) plus two *ARHGAP11B*-non-expressing neocortices (green dots, white columns) versus five *ARHGAP11B*-expressing neocortices (magenta dots, gray columns). For *ARHGAP11B*-expressing neocortex, quantification excluded gyrus. Data are shown as mean ± SD; $^{**}P < 0.01$; $^{***}P < 0.001$ (two-tailed *t* test).

were greater, but the subplate thickness was equal, for *ARHGAP11B*-expressing marmoset neocortex compared with WT and *ARHGAP11B*-non-expressing neocortex (fig. S5, C to F).

These data were consistent with an ongoing production of cortical neurons, mostly upper-layer neurons, at the developmental stage of our analyses (fig. S1D). We examined the germinal zones (VZ, iSVZ, and oSVZ) and progenitors therein for WT, *ARHGAP11B*-non-expressing, and *ARHGAP11B*-expressing marmoset neocortex (Fig. 4A). Analysis of the germinal zones showed increased oSVZ thickness for *ARHGAP11B*-expressing neocortex compared with WT and *ARHGAP11B*-non-expressing neocortex (Fig. 4B and fig. S7A). We observed an increase in mitotic basal progenitors that overall was ≈2-fold in the iSVZ and ≈3-fold in the oSVZ, but observed no difference in mitotic apical progenitors in the VZ (Fig. 4C and figs. S6 and S7, B and C).

At least half of the mitotic basal progenitors in the oSVZ of *ARHGAP11B*-expressing neocortex exhibited a basal process and thus were basal (or outer) radial glia (*24*–*27*), whereas this proportion was less (≤40%) for WT and *ARHGAP11B*-non-expressing neocortex (Fig. 4D and figs. S8, A to D), which is consistent with previous data (*28*, *29*). *ARHGAP11B* expression increased mitotic basal radial glia ≈3-fold (Fig. 4E and fig. S8E). A significant increase in basal radial glia caused by *ARHGAP11B* expression was also observed when these cells were quantified in interphase using the marker Hopx (*6*) (fig. S9). More than 99% of the mitotic basal radial glia in oSVZ were Sox2$^+$ (fig. S8F) and about half lacked expression of Tbr2 (Fig. 4, D and E, and fig. S8G). Therefore, the cells amplified upon *ARHGAP11B* expression in fetal marmoset neocortex exhibited a marker signature consistent with the identity of basal radial glia (*5*, *6*, *9*).

In this study, we examined physiologically relevant expression of human-specific *ARHGAP11B* (*7*, *8*) in the fetal neocortex of a nonhuman primate, the common marmoset, by using the human *ARHGAP11B* promoter, in contrast to previous studies using a strong constitutive promoter (*9*, *10*). This expression increased fetal neocortex size, CP thickness, upper-layer neurons, oSVZ size, and basal progenitors—including basal radial glia, the progenitor type that is thought to drive development of the mammalian neocortex (*2*–*6*, *14*, *30*). Our results suggest that the human-specific *ARHGAP11B* gene may have caused neocortex expansion in the course of human evolution.

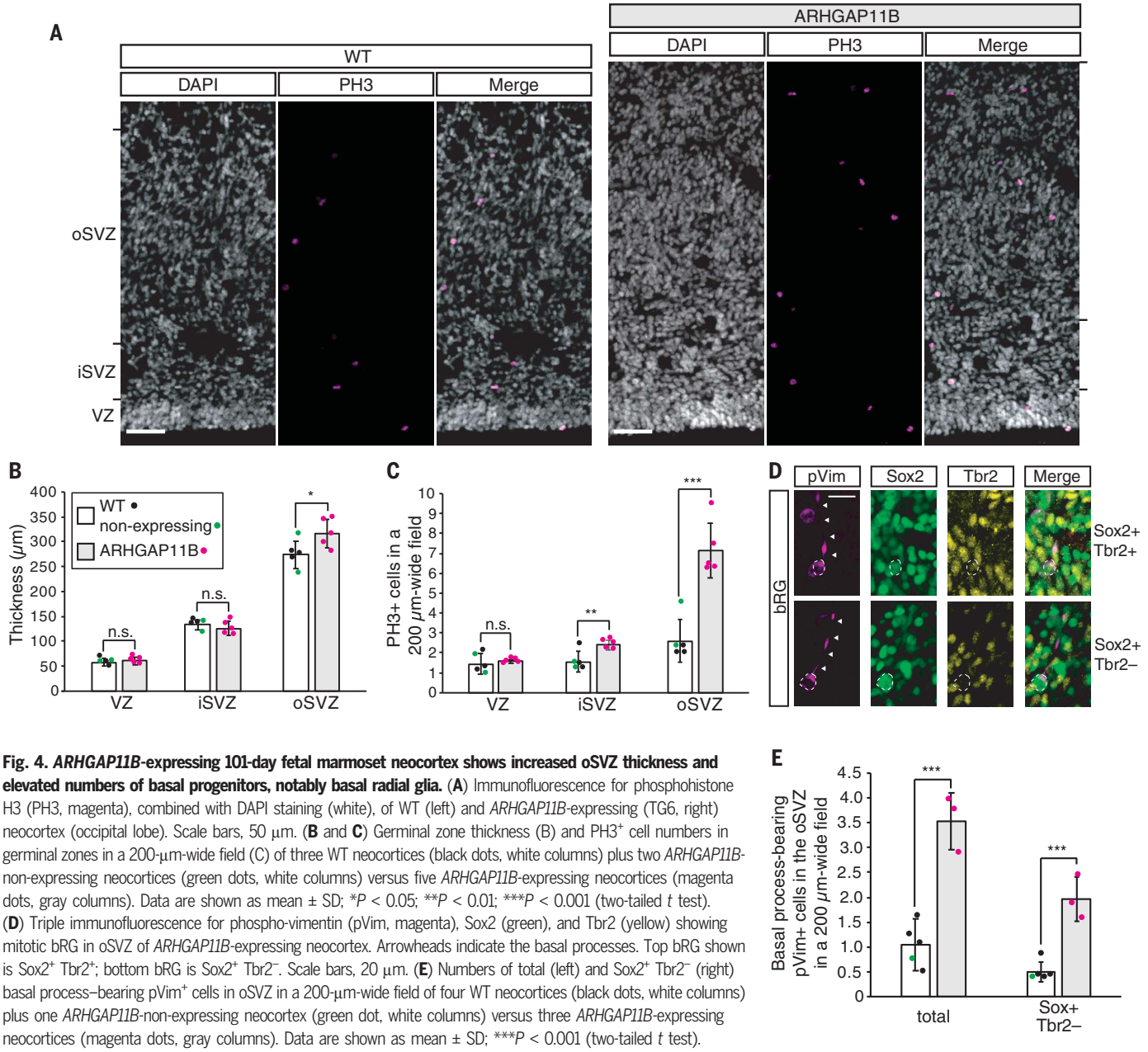


Fig. 4. *ARHGAP11B*-expressing 101-day fetal marmoset neocortex shows increased oSVZ thickness and elevated numbers of basal progenitors, notably basal radial glia. (**A**) Immunofluorescence for phosphohistone H3 (PH3, magenta), combined with DAPI staining (white), of WT (left) and *ARHGAP11B*-expressing (TG6, right) neocortex (occipital lobe). Scale bars, 50 μm. (**B** and **C**) Germinal zone thickness (B) and PH3$^+$ cell numbers in germinal zones in a 200-μm-wide field (C) of three WT neocortices (black dots, white columns) plus two *ARHGAP11B*-non-expressing neocortices (green dots, white columns) versus five *ARHGAP11B*-expressing neocortices (magenta dots, gray columns). Data are shown as mean ± SD; *P < 0.05; **P < 0.01; ***P < 0.001 (two-tailed t test). (**D**) Triple immunofluorescence for phospho-vimentin (pVim, magenta), Sox2 (green), and Tbr2 (yellow) showing mitotic bRG in oSVZ of *ARHGAP11B*-expressing neocortex. Arrowheads indicate the basal processes. Top bRG shown is Sox2$^+$ Tbr2$^+$; bottom bRG is Sox2$^+$ Tbr2$^-$. Scale bars, 20 μm. (**E**) Numbers of total (left) and Sox2$^+$ Tbr2$^-$ (right) basal process–bearing pVim$^+$ cells in oSVZ in a 200-μm-wide field of four WT neocortices (black dots, white columns) plus one *ARHGAP11B*-non-expressing neocortex (green dot, white columns) versus three *ARHGAP11B*-expressing neocortices (magenta dots, gray columns). Data are shown as mean ± SD; ***P < 0.001 (two-tailed t test).

REFERENCES AND NOTES

1. P. Rakic, *Nat. Rev. Neurosci.* **10**, 724–735 (2009).
2. J. H. Lui, D. V. Hansen, A. R. Kriegstein, *Cell* **146**, 18–36 (2011).
3. M. Florio, W. B. Huttner, *Development* **141**, 2182–2194 (2014).
4. C. Dehay, H. Kennedy, K. S. Kosik, *Neuron* **85**, 683–694 (2015).
5. Z. Molnár *et al.*, *J. Anat.* **235**, 432–451 (2019).
6. D. L. Silver *et al.*, in *The Neocortex*, W. Singer, T. J. Sejnowski, P. Rakic, Eds. (MIT Press, 2019), vol. 27 of Strüngmann Forum Reports, pp. 61–109.
7. P. H. Sudmant *et al.*, *Science* **330**, 641–646 (2010).
8. M. Y. Dennis *et al.*, *Nat. Ecol. Evol.* **1**, 0069 (2017).
9. M. Florio *et al.*, *Science* **347**, 1465–1470 (2015).
10. N. Kalebic *et al.*, *eLife* **7**, e41241 (2018).
11. T. Namba *et al.*, *Neuron* **105**, 867–881.e9 (2020).
12. M. Florio, T. Namba, S. Pääbo, M. Hiller, W. B. Huttner, *Sci. Adv.* **2**, e1601941 (2016).
13. E. Sasaki *et al.*, *Nature* **459**, 523–527 (2009).
14. I. H. Smart, C. Dehay, P. Giroud, M. Berland, H. Kennedy, *Cereb. Cortex* **12**, 37–53 (2002).
15. M. Florio *et al.*, *eLife* **7**, e32332 (2018).
16. K. Zilles, E. Armstrong, A. Schleicher, H. J. Kretschmann, *Anat. Embryol. (Berl.)* **179**, 173–179 (1988).
17. K. R. Long *et al.*, *Neuron* **99**, 702–719.e6 (2018).
18. K. Zilles, E. Armstrong, K. H. Moser, A. Schleicher, H. Stephan, *Brain Behav. Evol.* **34**, 143–150 (1989).
19. K. Sawada *et al.*, *Neuroscience* **257**, 158–174 (2014).
20. B. J. Molyneaux, P. Arlotta, J. R. Menezes, J. D. Macklis, *Nat. Rev. Neurosci.* **8**, 427–437 (2007).
21. S. Lodato, P. Arlotta, *Annu. Rev. Cell Dev. Biol.* **31**, 699–720 (2015).
22. K. Y. Kwan, N. Sestan, E. S. Anton, *Development* **139**, 1535–1546 (2012).
23. O. Britanova *et al.*, *Neuron* **57**, 378–392 (2008).
24. D. V. Hansen, J. H. Lui, P. R. Parker, A. R. Kriegstein, *Nature* **464**, 554–561 (2010).
25. S. A. Fietz *et al.*, *Nat. Neurosci.* **13**, 690–699 (2010).
26. I. Reillo, C. de Juan Romero, M. A. García-Cabezas, V. Borrell, *Cereb. Cortex* **21**, 1674–1694 (2011).
27. M. Betizeau *et al.*, *Neuron* **80**, 442–457 (2013).
28. I. Kelava *et al.*, *Cereb. Cortex* **22**, 469–481 (2012).
29. F. García-Moreno, N. A. Vasistha, N. Trevia, J. A. Bourne, Z. Molnár, *Cereb. Cortex* **22**, 482–492 (2012).
30. V. Borrell, M. Götz, *Curr. Opin. Neurobiol.* **27**, 39–46 (2014).

ACKNOWLEDGMENTS

We apologize to all researchers whose work could not be cited due to space limitations. We thank J. Okahara (Okano Lab in RIKEN CBS) for providing WT marmosets; J. Hata (Okano Lab in RIKEN CBS) for MR imaging of fetal marmoset brains; R. Behr (German Primate Center) for providing primary marmoset fibroblasts for initial construct and virus tests; D. Gerrelli, S. Lisgo, and their teams at the HDBR for the invaluable support from this resource; the Light Microscopy Facility, a Core Facility of the CMCB Technology Platform at TU Dresden, for scanning of cryosections; the Histology Facility of the CMCB Technology Platform at TU Dresden for cryosectioning; J. Peychl and his team of the Light Microscopy Facility at MPI-CBG for help with microscopy; G. Nadar of

the Scientific Computing Facility at MPI-CBG for help with data management; F. Friedrich and K. Margitudis for taking overview brain images; and T. Namba for providing the mouse monoclonal anti-ARHGAP11B IgG1 3758-A37-5. **Funding:** H.O. was supported by Brain/MINDS (JP20dm0207001) from AMED. E.S. was supported by the Strategic Research Program for Brain Science (JP17dm0107051) and Brain/MINDS (JP20dm0207065) from AMED. W.B.H. was supported by central funds of the Max Planck Society and by grants from the Deutsche Forschungsgemeinschaft (SFB 655, A2), the European Research Council (Advanced Grant 250197), and ERA-NET NEURON (MicroKin). **Author contributions:** Conceptualization: M.H., H.O., W.B.H.; Funding acquisition: H.O., E.S., W.B.H.; Investigation: M.H., C.H., A.M., Y.K., H.S.; Resources: H.O., E.S.; Writing – original draft: M.H., W.B.H.; Supervision: H.O., E.S., W.B.H.; Writing – review & editing: M.H., A.M., H.O., E.S., W.B.H. **Competing interests:** The authors declare no competing interests. H.O. and E.S. are inventors on the following patent applications: US 8592643 2013/11/26, Europe 2246423 2016/03/23, China ZL200880128383.3 (2014/11/26), Japan 5374389 (2013/9/27), Singapore 163739 (W02009/096101) 2013/09/30, and Korea 10-1588474 (2016/01/19) held by Keio University School of Medicine that covers "Method for introducing foreign gene into early embryo of primate animal" and "Method for production of transgenic primate animal comprising the introduction method." **Data and materials availability:** All data are available in the main text or the supplementary materials. Materials are available from M.H. or W.B.H. under a material transfer agreement and CITES permission.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/546/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S4
Data S1
References (*31*–*33*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

10 February 2020; accepted 1 June 2020
Published online 18 June 2020
10.1126/science.abb2401

ULTRACOLD MATTER

Cooling and entangling ultracold atoms in optical lattices

Bing Yang[1,2,3], Hui Sun[1,2,3], Chun-Jiong Huang[1,3], Han-Yi Wang[1,2,3], Youjin Deng[1,3], Han-Ning Dai[1,2,3], Zhen-Sheng Yuan[1,2,3]*, Jian-Wei Pan[1,2,3]*

Scalable, coherent many-body systems can enable the realization of previously unexplored quantum phases and have the potential to exponentially speed up information processing. Thermal fluctuations are negligible and quantum effects govern the behavior of such systems with extremely low temperature. We report the cooling of a quantum simulator with 10,000 atoms and mass production of high-fidelity entangled pairs. In a two-dimensional plane, we cool Mott insulator samples by immersing them into removable superfluid reservoirs, achieving an entropy per particle of $1.9^{+1.7}_{-0.4} \times 10^{-3} k_B$. The atoms are then rearranged into a two-dimensional lattice free of defects. We further demonstrate a two-qubit gate with a fidelity of 0.993 ± 0.001 for entangling 1250 atom pairs. Our results offer a setting for exploring low-energy many-body phases and may enable the creation of large-scale entanglement.

A paradigmatic quantum simulator is composed of ultracold neutral atoms in an optical lattice, which provides a clean and controllable platform for studying complex many-body problems (*1*, *2*). However, the thermal entropy arising from the intrinsic heating (*3*, *4*), non-adiabaticity (*5*, *6*), and inefficient thermalization (*7*) hinders the exploration of exotic phases of matter (*8*) and the ability to scale up entangled states (*9*, *10*). Numerous methods have been applied to cooling of quantum gases (*11*–*16*); some theoretical schemes suggest immersing lattice-trapped atoms into superfluid reservoirs that could eventually carry away the thermal entropy (*17*, *18*). With the help of a microscope, cooling of an atomic sample with a removable surrounding reservoir was demonstrated, leading to the realization of a Fermi-Hubbard antiferromagnet (*5*, *19*). However, further cooling of atomic samples would require either mixing two distinct quantum phases for efficient thermalization or realizing sublattice addressability to remove the reservoir.

We met these experimental challenges by designing a staggered lattice to immerse samples into superfluid reservoirs for sufficient thermalization, and by using a sublattice-addressing technique to remove all the high-entropy reservoirs and achieve a low-entropy regime. The cooling concept is sketched in Fig. 1A, where a superfluid that holds a large density of states serves as a reservoir for storing the entropy of the joint system. In our two-dimensional (2D) bosonic system with a bichromatic superlattice along the x direction, the competition between the kinetic energy J and interaction energy U leads to the formation of superfluid and Mott insulator phases. A spatially alternating appearance of the insulator and superfluid is realized by adjusting the local chemical potentials. On the basis of the local density variations, we developed a thermometry method to characterize the cooling performance. After we isolated and removed the particles in the reservoirs, the atoms of the samples (Fig. 1B) were redistributed within the superlattice to achieve a unity filling state with more than 10^4 sites.

The defect-free platform after cooling is a good starting point for quantum information processing. Owing to the parallel controllability of atoms in optical lattices, multipartite entanglement can be created by coupling the atomic qubits with nearest-neighbor interactions (*20*, *21*). Although both single- and two-qubit control (*22*–*26*) have been demonstrated in this system, the current challenge toward achieving large-scale entanglement is to develop a high-fidelity entangling gate (*27*, *28*). Here, we enhanced the superexchange interaction by engineering the Hamiltonian in a three-level scheme (Fig. 1C). The dynamical evolution is driven by the coherent competition of superexchange and atom tunneling. Relative to an earlier experiment (*26*), we markedly reduced the gate operation time and thereby improved the operation fidelity, thereby exceeding the threshold of some fault-tolerant models for quantum computing (*29*–*31*).

The experiment starts with a ^{87}Rb Bose-Einstein condensate of ~8.6×10^4 atoms confined in a single well of a pancake-shaped standing wave (*26*). The 2D cloud is adiabatically loaded into a square lattice in the x-y plane with a period of $\lambda_s/2$, where $\lambda_s = 767$ nm denotes the wavelength of the "short" lattices. Another "long" lattice with wavelength $\lambda_l = 2\lambda_s$ is used to construct a superlattice that divides the quantum gas into odd and even subsystems. Their lattice depths are ramped exponentially in 60 ms with a time constant of 20 ms up to $V_s = 26.1 \pm 0.2\, E_r$ and $V_l = 10.0 \pm 0.1\, E_r$, where $E_r = \hbar^2/(2m\lambda_s^2)$ is the recoil energy, $\hbar$ is the Planck constant, and m is the atomic mass. Meanwhile, the transverse trap is adjusted to set the atomic densities to $\bar{n} \approx 1.75$ in the central area. The energy offset Δ between odd and

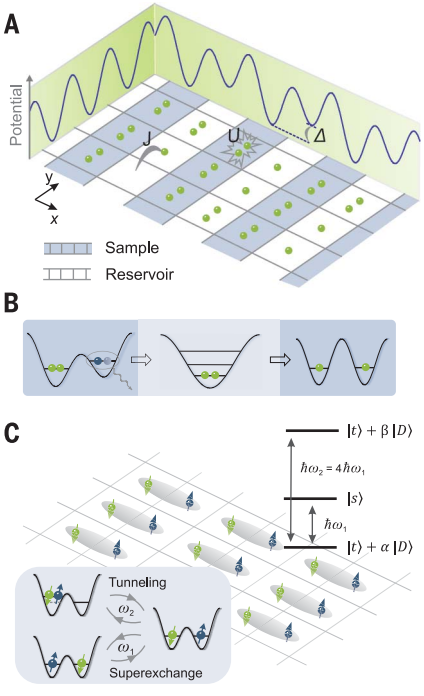


Fig. 1. Cooling and entangling atoms. (**A**) Schematic of staggered-immersion cooling. In an optical superlattice, gapped Mott insulators (samples) are immersed in the gapless superfluid reservoirs with a staggered geometry. J, U, and Δ denote kinetic energy, interaction energy, and energy offset, respectively. (**B**) Redistribution of the atoms into a unity filling state by engineering the lattice potential. (**C**) Entangling thousands of atom pairs in parallel. The atoms rendered in green are in the state of $|\downarrow\rangle$; atoms in blue are in the state of $|\uparrow\rangle$. The blue and light blue atoms in (B) represent an average occupation of ~1.5 atoms in the state of $|\uparrow\rangle$. When the initial state in double wells is $|\downarrow,\uparrow\rangle$, the Hilbert space contains three eigenstates; the inset shows the energy level diagram. The triplet/singlet states are $|t/s\rangle = (|\downarrow,\uparrow\rangle \pm |\uparrow,\downarrow\rangle)/\sqrt{2}$, and the double-occupied state is $|D\rangle = (|\uparrow\downarrow,0\rangle + |0,\uparrow\downarrow\rangle)/\sqrt{2}$; the coefficients α and β depend on J and U. At $J/U = \sqrt{3}/4$, the transition frequencies satisfy the relation $\omega_2 = 4\omega_1$. A two-qubit $\sqrt{\text{SWAP}}$ gate is realized after the evolution time of $\pi/(2\omega_1)$.

[1]Hefei National Laboratory for Physical Sciences at Microscale and Department of Modern Physics, University of Science and Technology of China, Hefei, Anhui 230026, China. [2]Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, 69120 Heidelberg, Germany. [3]CAS Centre for Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei, Anhui 230026, China.
*Corresponding author. Email: yuanzs@ustc.edu.cn (Z.-S.Y.); pan@ustc.edu.cn (J.-W.P.)

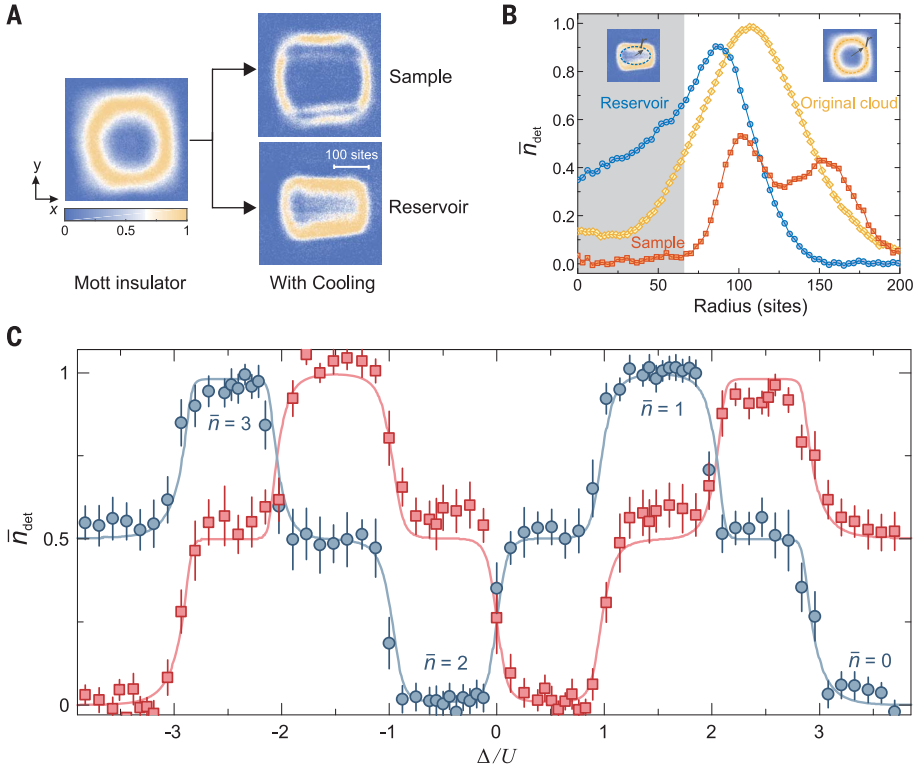


Fig. 2. Cooling a quantum gas. (**A**) Representative images of the atomic densities averaged over 50 repetitions. The original Mott insulator is created in the short lattice without cooling. After applying the cooling sequence, we detect the odd (sample) and even (reservoir) subsystems separately. With a parity-projection measurement on the $\bar{n}$ = 2 Mott insulator, less remaining density in the cloud center indicates lower entropy. (**B**) The density profiles are azimuthal averages over the corresponding images in (A), as sketched in the insets. For calculating the mean densities in the samples (red) and the reservoirs (blue), radius r refers to the geometric mean of the elliptical radii (the aspect ratio for this ellipse is 0.6). The entropy reduction is pronounced in the shaded area containing more than 10^4 lattice sites. (**C**) Density partitions between the odd (blue) and even (red) subsystems in the central region as a function of the energy offset between them. The superfluid and Mott insulator phases are coexistent and emerge alternately. Their densities change from 3.5 atoms per site to 0 and from 0 to 3.5 atoms per site, respectively, as reflected in the parity-projection measurement. The staircase-like changing of the densities is indicative of the low temperature. The solid curves are isentropic simulations at $k_B T_f/U$ = 0.05 (fig. S4) (*32*). The subtle features—such as the larger statistical errors of the superfluid states, higher temperatures at larger $|\Delta|$, and unequal width of the Mott plateaus—are revealed by our data and simulations. Each of the error bars represents SD over 20 images.

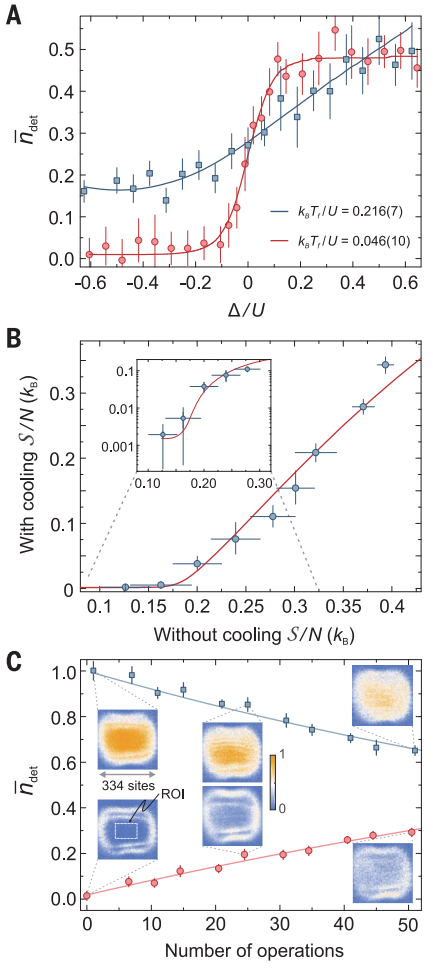


Fig. 3. Thermometry of the quantum gases and atom redistribution. (**A**) Density variations at 60 ms (red circles) and 500 ms (blue squares) lattice ramping. The corresponding curves are isentropic fitting results. (**B**) Cooling efficiency and critical behavior. The red curve is the theoretical prediction based on Monte Carlo simulations (fig. S4) (*32*). The inset shows the entropy of the system with cooling on a logarithmic scale. The horizontal error bars denote SD; the vertical ones represent 1σ of fitting errors (*32*). (**C**) Calibrating the fidelity of atom redistribution. We repeat the operation from $|2,0\rangle$ to $|1,1\rangle$ and its reversal in cycles, then count the atoms on 10^4 lattice sites within the region of interest (ROI). Red circles correspond to parity-projection measurements after even times of the redistribution operation; blue squares correspond to parity-projection measurements after odd times of the operation. Each data point in (A) and (C) is averaged over 20 measurements; error bars denote SD. The red and blue curves are power-law fitting results.

even subsystems is controlled by shifting the relative phase of the superlattice lasers (*26*). As the tunneling strength decreases across a critical value, gapped Mott insulators with $\bar{n}$ = 2 start to separate from the adjacent superfluid states with $\bar{n} \approx 1.5$.

After the atoms enter the deep lattices, coherent tunneling is negligible and defects in Mott insulators are induced only by thermal fluctuations (*13*). The on-site number fluctuations are probed via photoassociation collisions, which reduce the occupation of a site to its odd-even parity. The odd (sample) and even (reservoir) subsystems are distinguished by turning on a spin-dependent superlattice to separate their hyperfine transition by 28 kHz (*26*). Then, we perform a Landau-Zener sweep to transfer the atoms of the even rows from the initial state $|\downarrow\rangle \equiv |F = 1, m_F = -1\rangle$ to $|\uparrow\rangle \equiv |F = 2, m_F = -2\rangle$, with an efficiency of 99.5 ± 0.3% (*32*). The in situ atomic densities $\bar{n}_{det}$ of the odd and even subsystems (Fig. 2A) are successively recorded by absorption imaging with an optical resolution of ~1.0 μm. Figure 2B shows the density profiles, where the number fluctuations of the sample decrease markedly relative to that of the original Mott insulator without applying the staggered-immersion cooling. Nonetheless, the efficiency of the parity projection, 98.7 ± 0.3%, leaves undesired residual atoms, which limits the lowest entropy that can be detected to 0.011 ± 0.002 k_B. To overcome this imperfection for determining the ultralow temperature T_f of the target samples ($\Delta = -U/2$), we applied thermometry on the samples by probing the density redistribution in the odd and even subsystems, as described below.

The partitions of atomic densities are measured at different offsets of the local chemical

potential. Figure 2C shows the densities at different values of Δ/U, where the Mott insulators emerge associated with superfluid states in the other subsystem. Changing Δ represents a small perturbation to the lattice; thus, the size of the whole cloud is maintained and only local transport is allowed. Eight plateaus emerge around the half-integers of Δ/U, which are analogous to the concentric shell structures of the Mott insulator in a harmonic trap (*13*). Intriguingly, the compressible superfluid states exhibit distinctive half-integer plateaus owing to the local particle conservation. In the absence of global mass flow, the entropy is most likely to be conserved locally and the non-adiabaticities of the lattice loading are minimized (*6*). In addition, we measure the thermalization rate by monitoring the quench dynamics (following a reversal of the sign of Δ at certain times during the lattice ramping), and we find that the system can equilibrate on the time scale of h/J (fig. S5) (*32*).

The temperature of the samples is fitted on the basis of particle and entropy conservation at different values of Δ/U. Figure 3A shows $\bar{n}_{\mathrm{det}}$ of the odd subsystem as a function of Δ/U at two experimental settings. The phase transition at the edge of the plateaus leads to sudden changes of the occupancies, which allows the thermometer to access ultralow temperatures. In these samples, the lowest temperature achieved is $k_{\mathrm{B}}T_{\mathrm{f}} = 0.046 \pm 0.010\ U$. To characterize the cooling performance, we control the entropy by changing the adiabaticity or duration of the lattice loading (*3*, *6*). For each superlattice loading sequence, a Mott insulator with central density $\bar{n} = 2$ is prepared by setting $\Delta = 0$ to disable the cooling effect. The entropy in the absence of cooling is thereby deduced from the parity measurement. As shown in Fig. 3B, if the entropy without cooling achieves a critical value $\mathcal{S}/N \sim 0.17k_{\mathrm{B}}$, where $\mathcal{S}$ is von Neumann entropy and N is the number of atoms, the cooling power becomes strong and the entropy of the samples reduces considerably. This corresponds to a phase transition of the reservoir from a normal fluid to a superfluid (*4*), achieving an entropy per particle of less than $k_{\mathrm{B}} \ln 2/(2\bar{n}) = 0.20k_{\mathrm{B}}$ as the system equilibrates at nonzero tunneling strength (*32*). The experimental results are in excellent agreement with quantum Monte Carlo simulations (*32*). In addition, the temperatures of smaller samples can be determined from the density profiles and are consistent with the thermometry method above. Losing less than half of the total atoms in the reservoir (~1.5/3.5), we lower the entropy per particle in the large samples to $\mathcal{S}/N = 1.9^{+1.7}_{-0.4} \times 10^{-3}k_{\mathrm{B}}$, corresponding to a factor of 65 reduction of the original entropy. Besides the entropy inherent in the 2D gas, the intrinsic heating induced by the light-induced scattering during the lattice loading process is 0.025 ± 0.002 k_{B} per particle (*32*), indicating that the cooling power has exceeded the heating rate of the optical lattices.

We implement state engineering across 10^4 sites to turn the low-entropy samples into a unity filling state. This is achieved by manipulating the double wells (DWs) in parallel after removing the particles in the high-entropy reservoirs. Figure 1B shows the sequence from the number state $|2,0\rangle$ to $|1,1\rangle$ in an isolated DW unit. To avoid level crossing, which would cause band excitation, we shift the centroid of atom pairs to match the antinodes of the long lattice before transferring the atoms adiabatically into its ground band. Next, we ramp up the short lattice to split the atom pairs into a product state with one atom per site. The atoms favor occupying different sites in the presence of the interaction blockade U. The achieved fidelity of 99.3 ± 0.1% is calibrated by applying multiple operations to the samples—that is, driving the atoms between $|2,0\rangle$ and $|1,1\rangle$ back and forth as in Fig. 3C. Together with the entropy from the samples, the overall fidelity for preparing a unity filling state across 10^4 lattice sites is 99.2 ± 0.1%.

The defect-free platform provides massive physical qubits for quantum information processing (*26*–*28*). We implement the entangling gate within all the DWs using a three-level scheme. For probing the spin dynamics, the state initialization and detection have the same procedure as our previous experiment (*26*). Here, the potentials are set to 9.76 ± 0.01 E_{r} for the short lattice and 10.0 ± 0.1 E_{r} for the long lattice. In Fig. 4B, we observe an interference between two processes for 6.5 superexchange periods, which is equivalent to the application of 26 gate operations. The lifetime

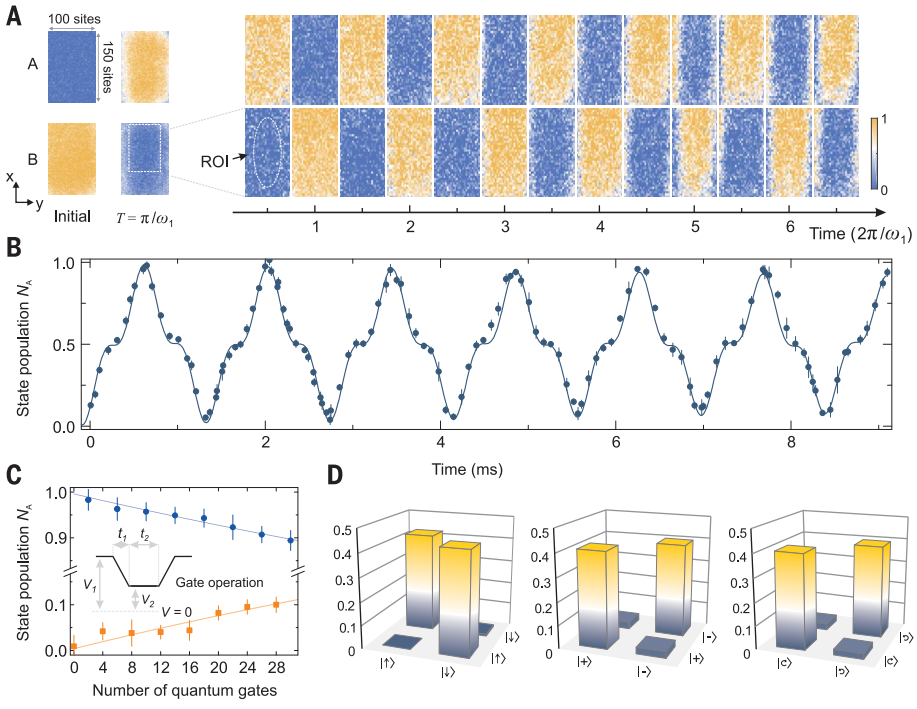


Fig. 4. Two-qubit quantum gate. (**A**) Spin dynamics. We measure the observables $N_{\mathrm{A}} = \langle n_{\uparrow,\mathrm{odd}} + n_{\downarrow,\mathrm{even}}\rangle/2$ on image A and $N_{\mathrm{B}} = \langle n_{\downarrow,\mathrm{odd}} + n_{\uparrow,\mathrm{even}}\rangle/2$ on image B, successively. The representative images show the initial state and the state after one-half of a superexchange period. The elliptical region of interest (ROI) containing 2500 lattice sites is chosen for counting signals. The high contrast of each group of images indicates excellent coherence of the system. The spatial inhomogeneity of the interactions among the sample starts to emerge after multiple periods. (**B**) Time-resolved observation of spin dynamics. The state evolution shows a coherent interference between two processes. Every data point is obtained by spatially averaging N_{A} across the ROI and incorporating five experimental repetitions. The solid curve is a fitting result by considering the coherent evolution and a phenomenological exponential decay, which gives $\omega_1/2\pi = 709 \pm 3$ Hz and a lifetime of 57 ± 16 ms. (**C**) Gate operation and state coherence. The quantum gate is implemented by controlling the depth of the short lattice, where $V_1 = 31.2 \pm 0.2\ E_{\mathrm{r}}$, $V_2 = 9.72 \pm 0.07\ E_{\mathrm{r}}$, $t_1 = 250$ μs, and $t_2 = 270$ μs. We apply multiple two-qubit gates and then record the state populations. Each data point is averaged over 20 measurements; error bars denote SD. The orange and blue symbols correspond to the measured N_{A} after applying $4j$ and $4j + 2$ gates, respectively, with $j = 0, 1, 2, \ldots, 7$. The curves are power-law fitting results. (**D**) Spin correlations. The spin state after 29 gate operations is measured on the $|\downarrow\rangle/|\uparrow\rangle$, $|+\rangle/|-\rangle$, and $|\circlearrowright\rangle/|\circlearrowleft\rangle$ basis.

of the spin dynamics (Fig. 4B), 57 ± 16 ms, is mainly caused by the spatial inhomogeneity of the interactions and fluctuations of the superlattice phase (*26*). We further studied the decoherence caused by light scattering and achieved a coherent oscillation with a lifetime of 646 ± 32 ms (fig. S9) (*32*).

The gate fidelity is characterized by measuring the target Bell state $|t\rangle$, which is acquired by performing a two-qubit gate followed by a phase rotation (*32*). The fidelity of the experimentally produced density matrix is

$$F = \langle t|\rho^{\mathrm{exp}}|t\rangle = \frac{1}{2}(P_{|\uparrow,\downarrow\rangle} + P_{|\downarrow,\uparrow\rangle}) + \frac{1}{4}(\langle\sigma_x\sigma_x\rangle + \langle\sigma_y\sigma_y\rangle) \quad (1)$$

where $\sigma_{x,y}$ are Pauli spin operators. The bases for the σ_x and σ_y operators are $|+/-\rangle = (|\uparrow\rangle \pm |\downarrow\rangle)/\sqrt{2}$ and $|\circlearrowright/\circlearrowleft\rangle = (|\uparrow\rangle \pm i|\downarrow\rangle)/\sqrt{2}$, respectively. The incoherent error of the gate operator accumulates and the fidelity scales as F^{N_g} when multiple gates N_g are applied. From the measurements of the state populations (Fig. 4C), we observe a reduction of the state coherence after implementing multiple quantum gates. To avoid the influence of the detection error, we measure the fidelity of the Bell state after 29 consecutive gate operations and then deduce the single-gate fidelity. Under the $|\uparrow/\downarrow\rangle$ basis, the probabilities of the states are $P_{|\uparrow,\downarrow\rangle}$ = 0.44 ± 0.02, $P_{|\downarrow,\uparrow\rangle}$ = 0.42 ± 0.01, $P_{|\uparrow,\uparrow\rangle}$ = 0.01 ± 0.02, and $P_{|\downarrow,\downarrow\rangle}$ = 0.01 ± 0.03. The states $\{|\uparrow\downarrow,0\rangle, |0,\uparrow\downarrow\rangle\}$ resulting from the operation error are included in the normalization. For the basis of $|+/-\rangle$ and $|\circlearrowright/\circlearrowleft\rangle$, a microwave $\pi/2$ pulse is applied to the atoms and then the populations of the spin states are measured. The σ_y operator given by the pulse has a relative $\pi/2$ phase shift with the pulse of the σ_x operator. The expectation values of the spin correlations are $\langle\sigma_x\sigma_x\rangle$ = 0.75 ± 0.03 and $\langle\sigma_y\sigma_y\rangle$ = 0.75 ± 0.04. The fidelity of this state after 29 gates is 0.80 ± 0.02, from which the fidelity of an individual two-qubit gate is deduced as F = 99.3 ± 0.1%. The infidelity of this two-qubit gate is below the error threshold required by some fault-tolerant models (*29*–*31*).

The cooling method we have demonstrated can markedly reduce the entropy of a quantum simulator. The achieved entropy enables studies of low-energy physics. Furthermore, we have realized a high-fidelity two-qubit gate and prepared 1250 Bell states of neutral atoms in parallel. The defect-free atom arrays and entangling gate are essential building blocks for creating multiparticle entanglement toward measurement-based quantum computing (*28*, *33*, *34*). For building a practical logic qubit toward fault-tolerant quantum computation (*29*–*31*), our system meets the requirements concerning both the entangling gate and the number of physical qubits. Our technique is generic and practical for cooling insulating states of other species and systems in other geometries (*18*). For the gapless fermionic phases, our method could strongly improve the efficiency of entropy transport (*12*, *19*), giving access to lower-temperature regimes for observing the d-wave superfluid phase (*35*). The defect-free platform and entangled pairs are suitable for exploring the bosonic antiferromagnetic order (*36*) and resonating valence bond states (*37*).

REFERENCES AND NOTES

1. I. Bloch, J. Dalibard, W. Zwerger, *Rev. Mod. Phys.* **80**, 885–964 (2008).
2. M. Greiner, O. Mandel, T. Esslinger, T. W. Hänsch, I. Bloch, *Nature* **415**, 39–44 (2002).
3. H. Pichler, A. J. Daley, P. Zoller, *Phys. Rev. A* **82**, 063605 (2010).
4. S. Trotzky *et al*., *Nat. Phys.* **6**, 998–1004 (2010).
5. C. S. Chiu, G. Ji, A. Mazurenko, D. Greif, M. Greiner, *Phys. Rev. Lett.* **120**, 243201 (2018).
6. M. Dolfi, A. Kantian, B. Bauer, M. Troyer, *Phys. Rev. A* **91**, 033407 (2015).
7. C.-L. Hung, X. Zhang, N. Gemelke, C. Chin, *Phys. Rev. Lett.* **104**, 160403 (2010).
8. D. C. McKay, B. DeMarco, *Rep. Prog. Phys.* **74**, 054401 (2011).
9. M. Nielsen, I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2010).
10. P. W. Shor, in *37th Symposium on Foundations of Computing* (IEEE Computer Society Press, 1996), pp. 56–65.
11. J. Catani *et al*., *Phys. Rev. Lett.* **103**, 140401 (2009).
12. T.-L. Ho, Q. Zhou, arXiv 0911.5506 [cond-mat.quant-gas] (29 November 2009).
13. W. S. Bakr *et al*., *Science* **329**, 547–550 (2010).
14. P. Medley, D. M. Weld, H. Miyake, D. E. Pritchard, W. Ketterle, *Phys. Rev. Lett.* **106**, 195301 (2011).
15. S. Taie, R. Yamazaki, S. Sugawa, Y. Takahashi, *Nat. Phys.* **8**, 825–830 (2012).
16. D. Greif, T. Uehlinger, G. Jotzu, L. Tarruell, T. Esslinger, *Science* **340**, 1307–1310 (2013).
17. A. Griessner, A. J. Daley, S. R. Clark, D. Jaksch, P. Zoller, *Phys. Rev. Lett.* **97**, 220403 (2006).
18. A. Kantian, S. Langer, A. J. Daley, *Phys. Rev. Lett.* **120**, 060401 (2018).
19. A. Mazurenko *et al*., *Nature* **545**, 462–466 (2017).
20. D. Jaksch, H.-J. Briegel, J. I. Cirac, C. W. Gardiner, P. Zoller, *Phys. Rev. Lett.* **82**, 1975–1978 (1999).
21. L. M. Duan, E. Demler, M. D. Lukin, *Phys. Rev. Lett.* **91**, 090402 (2003).
22. C. Weitenberg *et al*., *Nature* **471**, 319–324 (2011).
23. Y. Wang, A. Kumar, T.-Y. Wu, D. S. Weiss, *Science* **352**, 1562–1565 (2016).
24. M. Anderlini *et al*., *Nature* **448**, 452–456 (2007).
25. S. Trotzky *et al*., *Science* **319**, 295–299 (2008).
26. H.-N. Dai *et al*., *Nat. Phys.* **12**, 783–787 (2016).
27. T. D. Ladd *et al*., *Nature* **464**, 45–53 (2010).
28. R. Raussendorf, H. J. Briegel, *Phys. Rev. Lett.* **86**, 5188–5191 (2001).
29. E. Knill, *Nature* **434**, 39–44 (2005).
30. R. Raussendorf, J. Harrington, *Phys. Rev. Lett.* **98**, 190504 (2007).
31. A. G. Fowler, M. Mariantoni, J. M. Martinis, A. N. Cleland, *Phys. Rev. A* **86**, 032324 (2012).
32. See supplementary materials.
33. R. Raussendorf, D. E. Browne, H. J. Briegel, *Phys. Rev. A* **68**, 022312 (2003).
34. B. Vaucher, A. Nunnenkamp, D. Jaksch, *New J. Phys.* **10**, 023005 (2008).
35. W. Hofstetter, J. I. Cirac, P. Zoller, E. Demler, M. D. Lukin, *Phys. Rev. Lett.* **89**, 220407 (2002).
36. A. S. Sørensen *et al*., *Phys. Rev. A* **81**, 061603 (2010).
37. S. Trebst, U. Schollwöck, M. Troyer, P. Zoller, *Phys. Rev. Lett.* **96**, 250402 (2006).
38. B. Yang, Harvard Dataverse, doi.org/10.7910/DVN/XSFBAB.

ACKNOWLEDGMENTS

We thank A. Kantian for discussions. **Funding:** Supported by National Key R&D Program of China grant 2016YFA0301603, NNSFC grant 11874341, the Fundamental Research Funds for the Central Universities (special funds for promoting the construction of world-class universities and disciplines), the Anhui Initiative in Quantum Information Technologies, and the Chinese Academy of Sciences. **Author contributions:** Z.-S.Y. and J.-W.P. conceived the research and designed the experiment; B.Y. developed the cooling method; B.Y., H.S., and H.-N.D. set up the experiment; B.Y., H.S., and H.-Y.W. performed the measurement and analyzed the data; and C.-J.H. and Y.D. did the quantum Monte Carlo simulation. All authors contributed to manuscript preparation. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All experimental data and code are available in the database (*38*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/550/suppl/DC1
Supplementary Text
Figs. S1 to S10
Table S1
References (*39*–*65*)

30 September 2019; accepted 5 June 2020
Published online 18 June 2020
10.1126/science.aaz6801

STRUCTURAL BIOLOGY

In-cell architecture of an actively transcribing-translating expressome

Francis J. O'Reilly[1]*, Liang Xue[2,3]*, Andrea Graziadei[1]*, Ludwig Sinn[1], Swantje Lenz[1], Dimitry Tegunov[4], Cedric Blötz[5], Neil Singh[5], Wim J. H. Hagen[2], Patrick Cramer[4], Jörg Stülke[5], Julia Mahamid[2]†, Juri Rappsilber[1,6]†

Structural biology studies performed inside cells can capture molecular machines in action within their native context. In this work, we developed an integrative in-cell structural approach using the genome-reduced human pathogen *Mycoplasma pneumoniae*. We combined whole-cell cross-linking mass spectrometry, cellular cryo–electron tomography, and integrative modeling to determine an in-cell architecture of a transcribing and translating expressome at subnanometer resolution. The expressome comprises RNA polymerase (RNAP), the ribosome, and the transcription elongation factors NusG and NusA. We pinpointed NusA at the interface between a NusG-bound elongating RNAP and the ribosome and propose that it can mediate transcription-translation coupling. Translation inhibition dissociated the expressome, whereas transcription inhibition stalled and rearranged it. Thus, the active expressome architecture requires both translation and transcription elongation within the cell.

The two fundamental processes of gene expression—transcription and translation—are functionally coupled in bacteria. While the transcribing RNA polymerase (RNAP) produces a nascent mRNA chain that can be directly translated by ribosomes (*1*–*3*), translation has been shown to influence the overall transcription rate in *Escherichia coli* (*4*, *5*), implying a physical link between the two processes. Accordingly, an in vitro–reconstituted *E. coli* RNAP-ribosome supercomplex structure was elucidated and termed an "expressome" (*6*). Additional in vitro reconstitutions, some including the essential factor NusG (*7*) that is proposed to link RNAP and the ribosome (*8*, *9*), reveal different structural arrangements of the supercomplex (*10*–*12*). These raise questions as to the mechanisms of coupling that could be used inside the cell, in the context of all regulatory factors.

To structurally analyze transcription-translation coupling inside cells, we combined in-cell cross-linking mass spectrometry (CLMS) (*13*) and cellular cryo–electron tomography (cryo-ET) (*14*). We used the small, genome-reduced bacterium *Mycoplasma pneumoniae*, which is an ideal cell model for system-wide structural studies (*15*). Although *M. pneumoniae* has undergone substantial genome reduction during its evolution as a human pathogen, it has retained the core transcription and translation machineries (*16*–*18*).

To assess the topology of a putative RNAP-ribosome supercomplex and its associated regulatory factors, we performed whole-cell CLMS of intact *M. pneumoniae* cells (*13*, *19*) (fig. S1 and table S1). We identified 10,552 cross-links involving the same protein (self-links) and 1957 heteromeric cross-links with a 5% residue-pair false discovery rate (FDR). These represented 577 distinct protein-protein interactions (PPIs) at a 5% PPI FDR (Fig. 1A and supplementary text). Identified cross-links covered 83% of the detectable proteome (table S2 and fig. S1), including PPIs of membrane proteins (41% of PPIs), 76 uncharacterized proteins, the ribosome, RNAP, and associated factors (Fig. 1B and figs. S1 to S3).

The *M. pneumoniae* RNAP core, which consists of the conserved subunits α, β, and β′, was found to interact with the known auxiliary factors SigA, GreA, NusG, NusA, SpxA, and RpoE (the firmicute-specific RNAP δ subunit) (*20*) (Fig. 1B). Additionally, two uncharacterized essential proteins, MPN555 and MPN530 (*21*), were found, and the interaction of MPN530 with the β and β′ subunits was independently validated by a bacterial two-hybrid screen (fig. S4). Despite these interactions, no direct cross-links between the RNAP core and the ribosome were identified. Interaction between NusG and the ribosomal protein S10, previously reported in *E. coli*, was also not detected (*8*, *9*). Instead, NusA, an essential transcription factor involved in elongation, termination, and antitermination (*21*–*23*), was found to interact with RNAP via its N-terminal domain (NTD) and with the mRNA entry site of the ribosome via its C-terminal region (Fig. 1B). In-cell CLMS thus indicated an unexpected architecture in which RNAP and the ribosome are linked by NusA.

To investigate the structure of this potential association, we acquired cryo-ET data on unperturbed, frozen-hydrated *M. pneumoniae* cells (*19*) (Fig. 2A and fig. S5). We extracted 108,501 ribosome subtomograms and subjected them to classification and refinement (figs. S6 and S7). These subtomograms exhibited marked structural heterogeneity and were first sorted into classes representing the 50*S* subunit (30.3%), 70*S* ribosomes (53.3%), and 70*S* ribosomes in the closely assembled polysome configuration (*24*) (16.4%) (Fig. 2A and fig. S7). Subjecting 73,858 70*S* ribosomes to a newly established subtomogram analysis workflow (fig. S6) (*19*) resulted in a 5.6-Å ribosome density (Fig. 2B and fig. S8). A ribosome homology model [based on Protein Data Bank (PDB) ID 3J9W] was fitted, and most *M. pneumoniae* ribosomal

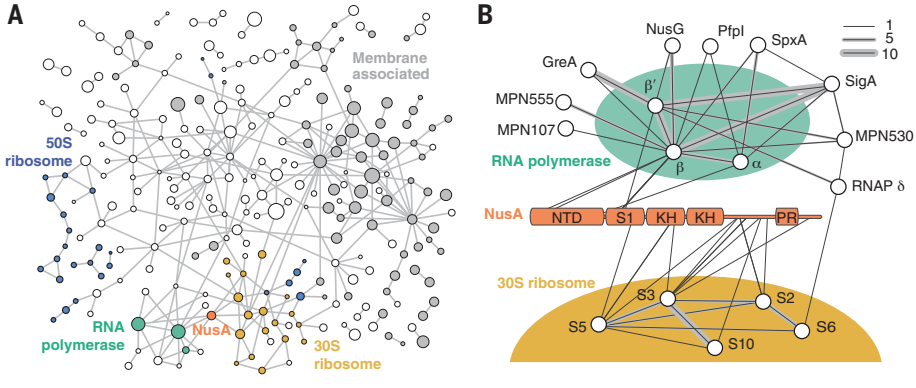


Fig. 1. Cross-link–based protein interaction map of the *M. pneumoniae* proteome. (**A**) Five hundred seventy-seven distinct PPIs identified at 5% PPI-level FDR (interactions with eight abundant glycolytic enzymes and chaperones have been removed for clarity). Circle diameter indicates relative protein size. Each edge represents one or more cross-links. (**B**) Interactors of RNAP and NusA. The NusA NTD, S1, KH domains, and proline-rich region (PR) are annotated. Line thickness represents the number of identified cross-links.

[1]Bioanalytics Unit, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany. [2]Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany. [3]Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences, 69120 Heidelberg, Germany. [4]Department of Molecular Biology, Max-Planck-Institute for Biophysical Chemistry, 37077 Göttingen, Germany. [5]Department of General Microbiology, Institute of Microbiology and Genetics, GZMB, Georg-August-University Göttingen, 37077 Göttingen, Germany. [6]Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh, EH9 3BF, UK.
*These authors contributed equally to this work.
†Corresponding author. Email: julia.mahamid@embl.de (J.M.); juri.rappsilber@tu-berlin.de (J.R.)

proteins could be mapped (fig. S9). Helical densities at the C termini of L22 and L29 (Fig. 2B, inset) were not explained by the homology model and were assigned to two C-terminal extensions that are specific to *M. pneumoniae* and its close relatives (figs. S10 and S11). L23, which also contains a C-terminal extension, was predicted to be unstructured and did not produce any discernible density in the map (fig. S12). Therefore, in conjunction with CLMS, the attained high-resolution map enabled de novo assignment of secondary structures in cellulo.

Focused classification of the 70*S* ribosome on the mRNA entry site allowed us to identify a ribosome class in complex with RNAP (70*S*+RNAP; Fig. 2A and fig. S7). Refinement thereof provided a 9.2-Å map (fig. S13) into which the ribosome and RNAP models fitted unambiguously (Fig. 2C). Consistent with the CLMS data, the map contained additional density at the interface between the two complexes (Fig. 2C, arrowheads), which was further resolved by multibody refinement (fig. S14). The path of the DNA and RNA-DNA hybrid duplex showed that RNAP is in an elongating state (Fig. 3A and fig. S15C). The existence of the elongating RNAP and the 70*S* ribosome demonstrated that the supercomplex represents an actively elongating expressome with a high degree of structural flexibility (movie S1).

Both CLMS and cryo-EM results showed binding of NusG to its conserved site (figs. S15 and S16) (*25*). *M. pneumoniae* NusG contains large inserts of unknown structure but retains the residues involved in the NusG-S10 interaction (*8*). However, the arrangement of RNAP relative to the ribosome placed NusG away from S10, indicating that this interaction does not occur in the elongating expressome (fig. S16). All other proteins found interacting with RNAP by CLMS did not fit in the elongating expressome density (fig. S17). The remaining density between RNAP and the ribosome was therefore consistent with NusA (Fig. 1B and fig. S18).

The CLMS and cryo-EM data were used to derive an integrative model of the elongating expressome (*19*, *26*) (fig. S19, table S4 and S5). *M. pneumoniae* NusA contains a disordered proline-rich C-terminal region that is not found in *E. coli* or *Bacillus subtilis* (fig. S18), which we established to be essential by mutation experiments (fig. S20). This region, which was found to be cross-linked to multiple 30*S* ribosomal proteins (Fig. 1B), was coarse-grained and not fitted into the density. The best-scoring solutions (fig. S21) showed that the NusA NTD and the S1 domain bind RNAP similarly to the *E. coli* paused-elongation complex (*23*),

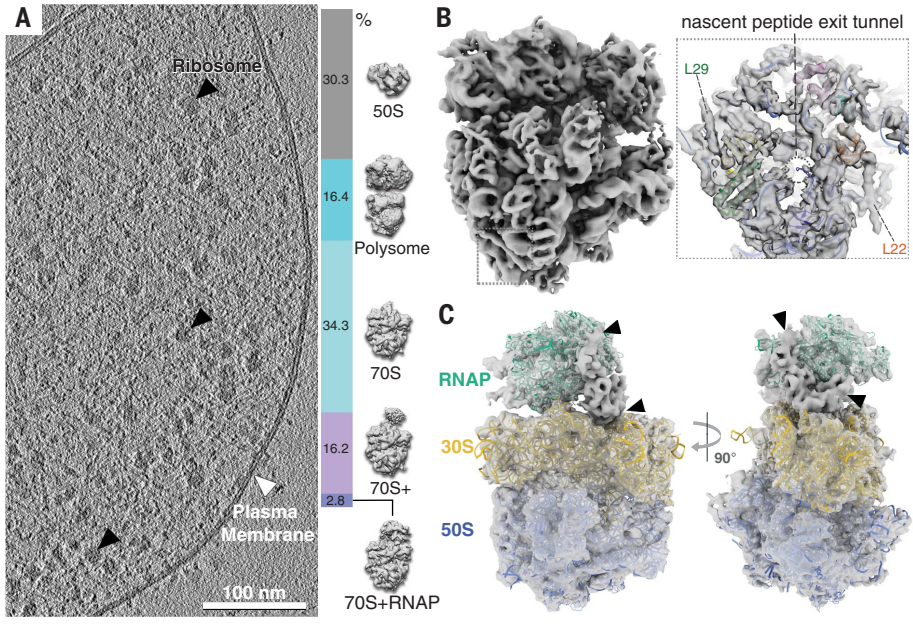


Fig. 2. In-cell cryo-ET reveals the presence of an RNAP-ribosome supercomplex. (**A**) (Left) Tomographic slice of a *M. pneumoniae* cell. (Right) Classification of 108,501 ribosome subtomograms from *M. pneumoniae cells*. (**B**) (Left) In-cell 70*S* ribosome density of 5.6-Å resolution. (Inset) Density near the peptide exit tunnel (dashed circle) shows two helices not accounted for by the fitted homology model (L22 and L29). (**C**) In-cell cryo-EM density (9.2 Å) of RNAP-ribosome supercomplexes [2.8% in (A)], fitted with homology models. Arrowheads indicate remaining unassigned density.

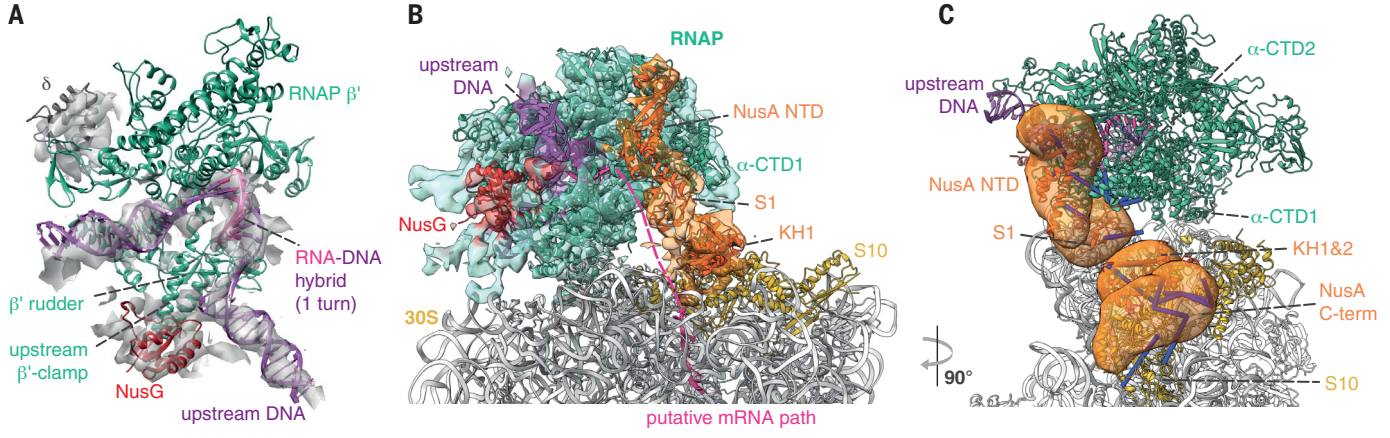


Fig. 3. Integrative model of the *M. pneumoniae* elongating expressome. (**A**) Cryo-EM density of the RNAP corresponding to the DNA, RNA-DNA hybrid, upstream β′ clamp, δ subunit, and NusG accommodates one turn of the RNA-DNA hybrid, consistent with an elongating RNAP. (**B**) Integrative model with cryo-EM density of the RNAP-NusG-NusA-ribosome elongating expressome. Structured regions are represented as colored cartoons. The electron density colors represent the subunits occupying the corresponding volumes. Coarse-grained regions are not shown. The schematic of the putative mRNA path denotes the shortest distance between mRNA exit and entry sites. (**C**) The mRNA exit tunnel face of RNAP is covered by NusA. Localization probability densities for NusA domains are shown in orange. Cross-links between NusA and other proteins are indicated. Satisfied cross-links (<35 Å) are in blue; violated cross-links are in red.

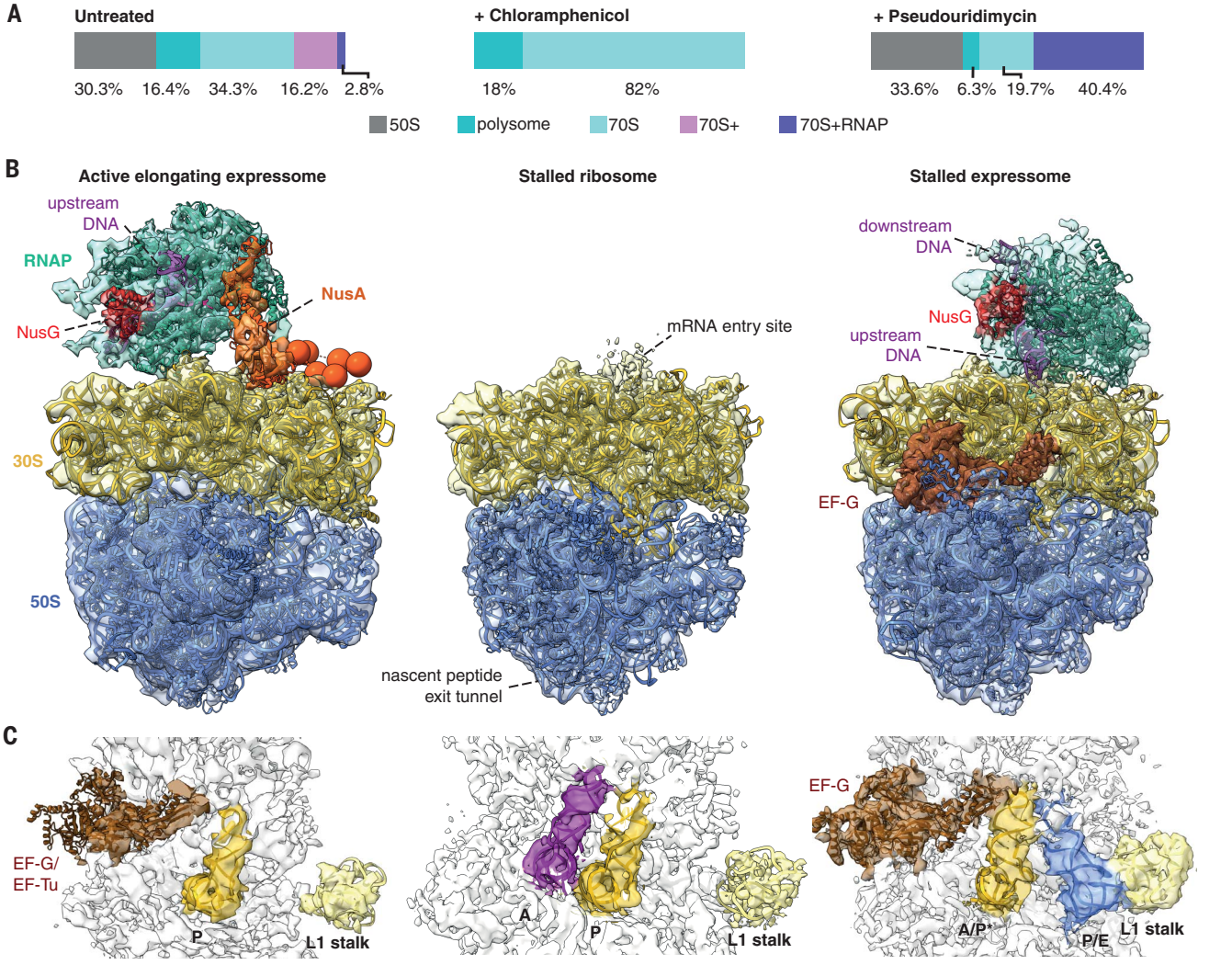


Fig. 4. Stalling translation or transcription alters the expressome architecture in cells. (**A**) Classification of subtomograms in untreated and Cm- and PUM-treated *M. pneumoniae* cells revealed shifts in ribosome populations after perturbations. (**B**) Models of RNAP-ribosome supercomplexes and the Cm-stalled ribosome. (Left) In untreated cells, the expressome compromises an actively elongating RNAP and ribosome. (Center) Cm decoupled the ribosome and RNAP. (Right) In PUM-treated cells, the ribosome encounters the stalled RNAP. (**C**) Ribosome tRNA occupancy states. In untreated cells, densities for P-site tRNA and elongation factors densities were visible, indicating a translating ribosome. Upon addition of Cm, A- and P-site tRNAs were observed, indicating a stalled ribosome. In PUM-treated cells, the presence of EF-G and hybrid A/P*- and P/E-site tRNAs suggested a pretranslocation stalled state.

with the S1 domain near the RNAP mRNA exit tunnel (Fig. 3, B and C, and fig. S22). The two KH domains were positioned near S3, S4, and S5 at the ribosome mRNA entry site. The orientation of the KH domains retained their RNA-binding interface in a position that can interact with the nascent mRNA (*27*) (fig. S23). The C-terminal domains (CTDs) of RNAP α subunits were found to be in a wide range of conformations localized between the NusA NTD and the second KH domain, with one α-CTD fitting a region between the second NusA KH domain and the RNAP core (fig. S23 and supplementary text). Additionally, the firmicute-specific RNAP δ subunit was positioned below RNAP β′ CTD (fig. S24), consistent with its suggested role in regulating RNAP-DNA interactions (*20*).

The integrative model demonstrated that NusA bridges the elongating RNAP and ribosome in the active expressome. To determine whether this architecture requires active translation elongation, we collected cryo-ET data on cells treated with the translation inhibitor chloramphenicol (Cm). The percentage of 70*S* ribosomes increased substantially compared with that of untreated cells (Fig. 4A and fig. S25). The resulting 6.5-Å ribosome density (Fig. 4B) had well-resolved A- and P-site tRNAs, similar to a previous ribosome-Cm structure (*28*) (Fig. 4C), but did not contain any RNAP density near the ribosome mRNA entry site. Thus, stalling ribosomes led to dissociation of the expressome.

The dependence on active transcription was probed by treating cells with the specific RNAP inhibitor pseudouridimycin (PUM) (*29*), which substantially increased the percentage of well-resolved expressomes (Fig. 4A). The PUM-induced expressome was refined to 7.1 Å (figs. S26 and S27 and movie S2), revealing direct interaction between the NusG-bound RNAP and the ribosome and excluding density for NusA (Fig. 4B and fig. S28). This stalled expressome closely resembled the architecture of the *E. coli* expressome solved in vitro (*6*) (fig. S28). Notably, tRNAs in the ribosome were found in hybrid A/P* and P/E states, and density corresponding to EF-G was well resolved (Fig. 4C). This suggested that the ribosome was trapped in a pretranslocation state (fig. S28) (*30*) and unable to complete the translocation step, presumably owing to physical obstruction by the stalled RNAP.

In summary, we have determined the native architecture of the expressome in *M. pneumoniae* and have shown that it requires active transcription and translation elongation. At the RNAP-ribosome interface, we unexpectedly found NusA, which followed the path of nascent mRNA at the nexus of transcription-translation coupling. NusA may act as a sensor of RNAP that detects an approaching ribosome and modulates transcription elongation. However, it remains to be seen whether the involvement of NusA in the *M. pneumoniae* active expressome represents a feature that is conserved across bacteria. Our data highlight the potential of integrative in-cell structural biology methods for elucidating dynamic machineries within their native functional context.

REFERENCES AND NOTES

1. T. J. Santangelo, I. Artsimovitch, *Nat. Rev. Microbiol.* **9**, 319–329 (2011).
2. K. McGary, E. Nudler, *Curr. Opin. Microbiol.* **16**, 112–117 (2013).
3. A. B. Conn, S. Diggs, T. K. Tam, G. M. Blaha, *Int. J. Mol. Sci.* **20**, 2595 (2019).
4. R. Landick, J. Carey, C. Yanofsky, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4663–4667 (1985).
5. S. Proshkin, A. R. Rahmouni, A. Mironov, E. Nudler, *Science* **328**, 504–508 (2010).
6. R. Kohler, R. A. Mooney, D. J. Mills, R. Landick, P. Cramer, *Science* **356**, 194–197 (2017).
7. R. A. Mooney *et al*., *Mol. Cell* **33**, 97–108 (2009).
8. B. M. Burmann *et al*., *Science* **328**, 501–504 (2010).
9. R. S. Washburn, P. K. Zuber, M. Sun, Y. Hashem, B. Shen, bioRxiv 871962 [Preprint]. 11 December 2019. www.biorxiv.org/content/10.1101/871962v1.
10. C. Wang, V. Molodtsov, E. Firlar, J. Kaelber, G. Blaha, bioRxiv 972380 [Preprint]. 2 March 2020. www.biorxiv.org/content/10.1101/2020.03.01.972380v1.
11. M. W. Webster, M. Takacs, C. Zhu, V. Vidmar, A. D. Eduljee, bioRxiv 971028 [Preprint]. 2 March 2020. www.biorxiv.org/content/10.1101/2020.03.01.971028v1.
12. G. Demo *et al*., *eLife* **6**, e28560 (2017).
13. F. J. O'Reilly, J. Rappsilber, *Nat. Struct. Mol. Biol.* **25**, 1000–1008 (2018).
14. M. Beck, W. Baumeister, *Trends Cell Biol.* **26**, 825–837 (2016).
15. S. Kühner *et al*., *Science* **326**, 1235–1240 (2009).
16. M. Güell *et al*., *Science* **326**, 1268–1271 (2009).
17. H. Grosjean *et al*., *PLOS Genet.* **10**, e1004363 (2014).
18. E. Yus *et al*., *Cell Syst.* **9**, 143–158.e13 (2019).
19. See supplementary materials.
20. A. Weiss, L. N. Shaw, *FEMS Microbiol. Rev.* **39**, 541–554 (2015).
21. M. Lluch-Senar *et al*., *Mol. Syst. Biol.* **11**, 780 (2015).
22. X. Yang *et al*., *EMBO Rep.* **10**, 997–1002 (2009).
23. X. Guo *et al*., *Mol. Cell* **69**, 816–827.e4 (2018).
24. F. Brandt *et al*., *Cell* **136**, 261–271 (2009).
25. M. Turtola, G. A. Belogurov, *eLife* **5**, e18096 (2016).
26. D. Russel *et al*., *PLOS Biol.* **10**, e1001244 (2012).
27. B. Beuth, S. Pennell, K. B. Arnvig, S. R. Martin, I. A. Taylor, *EMBO J.* **24**, 3576–3587 (2005).
28. M. S. Svetlov *et al*., *RNA* **25**, 600–606 (2019).
29. S. I. Maffioli *et al*., *Cell* **169**, 1240–1248.e23 (2017).
30. A. F. Brilot, A. A. Korostelev, D. N. Ermolenko, N. Grigorieff, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20994–20999 (2013).
31. F. O'Reilly *et al*., Integrative modeling results of in-cell architecture of an actively transcribing-translating expressome, Version 2.0.0, Zenodo (2020); https://doi.org/10.5281/zenodo.3837625.

ACKNOWLEDGMENTS

We are grateful to M. Chaabou, H. Barysz, Z. A. Chen, L. Fischer, C. Combe, M. Graham, and I. Zagoriy for assistance in data acquisition and processing. We thank J. Kosinski, C. Müller, and M. Beck for critical reading of the manuscript. **Funding:** This project received funding from the Bundesministerium für Bildung und Forschung Grant FKZ 031L0012 to J.S.; Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy – EXC 2008/1 – 390540038, project no. 426290502 and 392923329, the Wellcome Trust Senior Research Fellowship (103139) to J.R.; and EMBL and the European Research Council to J.M. (760067) and P.C. (693023). The Wellcome Centre for Cell Biology is supported by core funding from the Wellcome Trust (203149). **Author contributions:** F.J.O., L.X., J.M., and J.R. designed the experiments; F.J.O., L.S., and S.L. collected and processed CLMS data; C.B. and J.S. performed the bacterial two-hybrid work; N.S. performed mutation experiments; L.X. and J.M. collected and processed cryo-EM data; W.J.H.H. supported cryo-EM data collection; D.T. and P.C. contributed new cryo-EM processing software; A.G. and S.L performed integrative modeling; and F.J.O., L.X., A.G., J.M., and J.R. prepared the figures and wrote the manuscript, with input from all authors. **Competing interests:** The authors have no competing interests. **Data and materials availability:** The supplementary materials contain additional data. EM densities have been deposited in the EM Data Bank with accession numbers EMD-10677, -10678, -10679, -10680, -10681, -10682, -10683, -10684, -10685, -10686, and -10687. CLMS data are available via ProteomeXchange with identifiers PXD017711 (DSSO) and PXD017695 (DSS). The integrative model is available in PDB-Dev with accession number PDBDEV_00000049. Homology models are available at the ModelArchive with accession codes ma-mrryl, ma-7ov95, ma-eeo9f, and ma-8tn6v. Integrative modeling code and raw files are available in Zenodo (*31*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/554/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S28
Tables S1 to S7
References (*32*–*92*)
MDAR Reproducibility Checklist
Movies S1 and S2

View/request a protocol for this paper from *Bio-protocol*.

28 February 2020; accepted 1 June 2020
10.1126/science.abb3758

ORGANIC CHEMISTRY

Direct reversible decarboxylation from stable organic acids in dimethylformamide solution

Duanyang Kong*, Patrick J. Moon*, Erica K. J. Lui, Odey Bsharat, Rylan J. Lundgren†

Many classical and emerging methodologies in organic chemistry rely on carbon dioxide (CO_2) extrusion to generate reactive intermediates for bond-forming events. Synthetic reactions that involve the microscopic reverse—the carboxylation of reactive intermediates—have conventionally been undertaken using very different conditions. We report that chemically stable C(sp^3) carboxylates, such as arylacetic acids and malonate half-esters, undergo uncatalyzed reversible decarboxylation in dimethylformamide solution. Decarboxylation-carboxylation occurs with substrates resistant to protodecarboxylation by Brønsted acids under otherwise identical conditions. Isotopically labeled carboxylic acids can be prepared in high chemical and isotopic yield by simply supplying an atmosphere of $^{13}CO_2$ to carboxylate salts in polar aprotic solvents. An understanding of carboxylate reactivity in solution enables conditions for the trapping of aldehydes, ketones, and α,β-unsaturated esters.

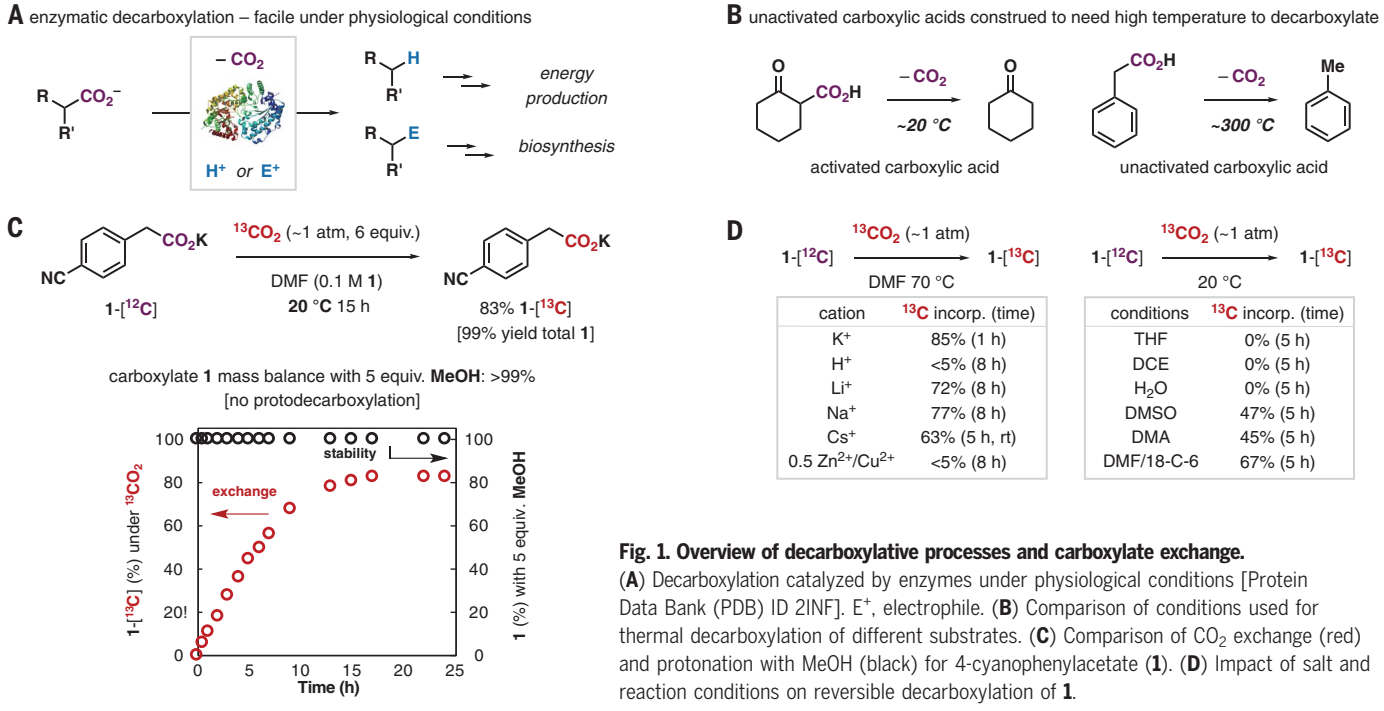
Decarboxylation is a fundamental step in biochemical processes and synthetic organic chemistry. Fermentation, respiration, and the biosynthesis of many secondary metabolites involve the loss of CO_2 from organic acids (*1*, *2*). Decarboxylase enzymes accelerate these reactions by stabilizing developing intermediates (typically carbanions) and promoting CO_2 diffusion from the active site, thereby enabling otherwise unfeasible decarboxylations to occur under physiological conditions (Fig. 1A) (*3*, *4*). Acid substrates lacking strong anion-stabilizing groups adjacent to the reactive carbon center have been construed to be inert toward spontaneous decarboxylation without resorting to thermolysis conditions (Fig. 1B) (*5*). As a result, synthetic reactions driven by decarboxylation are often carried out using high reaction temperatures (*6*), added oxidizing agents (*7*, *8*), or prior stoichiometric chemical modification of the carboxylate unit (*9*–*12*).

Carboxylation reactions, the microscopic reverse of decarboxylations, are equally valuable processes in biology and synthetic chemistry. Despite the possibility of a shared reaction pathway, the biochemical machinery that promotes carboxylation in CO_2 fixation pathways operates with a distinct set of substrates and enzymes from those that promote decarboxylation in all but a few cases (*13*–*17*). Similarly, synthetic techniques that generate carboxylic acid derivatives from CO_2 have tended to apply strongly nucleophilic organometallics and/or in situ stoichiometric (electro)chemical substrate reduction (*18*, *19*).

The potential for the reversibility of decarboxylation-carboxylation mechanisms is largely ignored in reports of synthetic methodologies that rely on these elementary steps. Reports of direct nonenzymatic reversible CO_2 exchange of carboxylic acids are restricted to specialized substrate-mediator pairs (*20*, *21*). Exchange of carboxylate groups in simple aliphatic acids with CO_2 has been documented but requires heating of neat substrates at 280° to 400°C (*22*, *23*). Nonetheless, in the course of our studies on catalytic decarboxylative cross-coupling reactions (*24*, *25*), we questioned whether the apparent stability of organic carboxylates could arise from reversible decarboxylation-carboxylation events in solution. Supporting this hypothesis, we observed that certain simple organic acids that are stable toward protodecarboxylation in solution undergo spontaneous incorporation of $^{13}CO_2$ when this heavier isotope is supplied at atmospheric pressure (Fig. 1C).

The potassium salt of arylacetic acid **1** exemplifies the reversible decarboxylation-carboxylation behavior of otherwise chemically stable carboxylic acids. A 0.1 M solution of **1** in dimethylformamide (DMF) at 20°C underwent CO_2 exchange when placed under an atmosphere of $^{13}CO_2$. In a reaction where about 6 equiv of $^{13}CO_2$ were supplied (13 ml of CO_2 at ~1 atm, dissolved [CO_2] = 0.20 M), equilibrium between ^{12}C and ^{13}C was achieved in 15 hours (Fig. 1C, red trace).

Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada.
*These authors contributed equally to this work.
†Corresponding author. Email: rylan.lundgren@ualberta.ca



cation	^{13}C incorp. (time)
K^+	85% (1 h)
H^+	<5% (8 h)
Li^+	72% (8 h)
Na^+	77% (8 h)
Cs^+	63% (5 h, rt)
0.5 Zn^{2+}/Cu^{2+}	<5% (8 h)

conditions	^{13}C incorp. (time)
THF	0% (5 h)
DCE	0% (5 h)
H_2O	0% (5 h)
DMSO	47% (5 h)
DMA	45% (5 h)
DMF/18-C-6	67% (5 h)

Fig. 1. Overview of decarboxylative processes and carboxylate exchange. (**A**) Decarboxylation catalyzed by enzymes under physiological conditions [Protein Data Bank (PDB) ID 2INF]. E^+, electrophile. (**B**) Comparison of conditions used for thermal decarboxylation of different substrates. (**C**) Comparison of CO_2 exchange (red) and protonation with MeOH (black) for 4-cyanophenylacetate (**1**). (**D**) Impact of salt and reaction conditions on reversible decarboxylation of **1**.

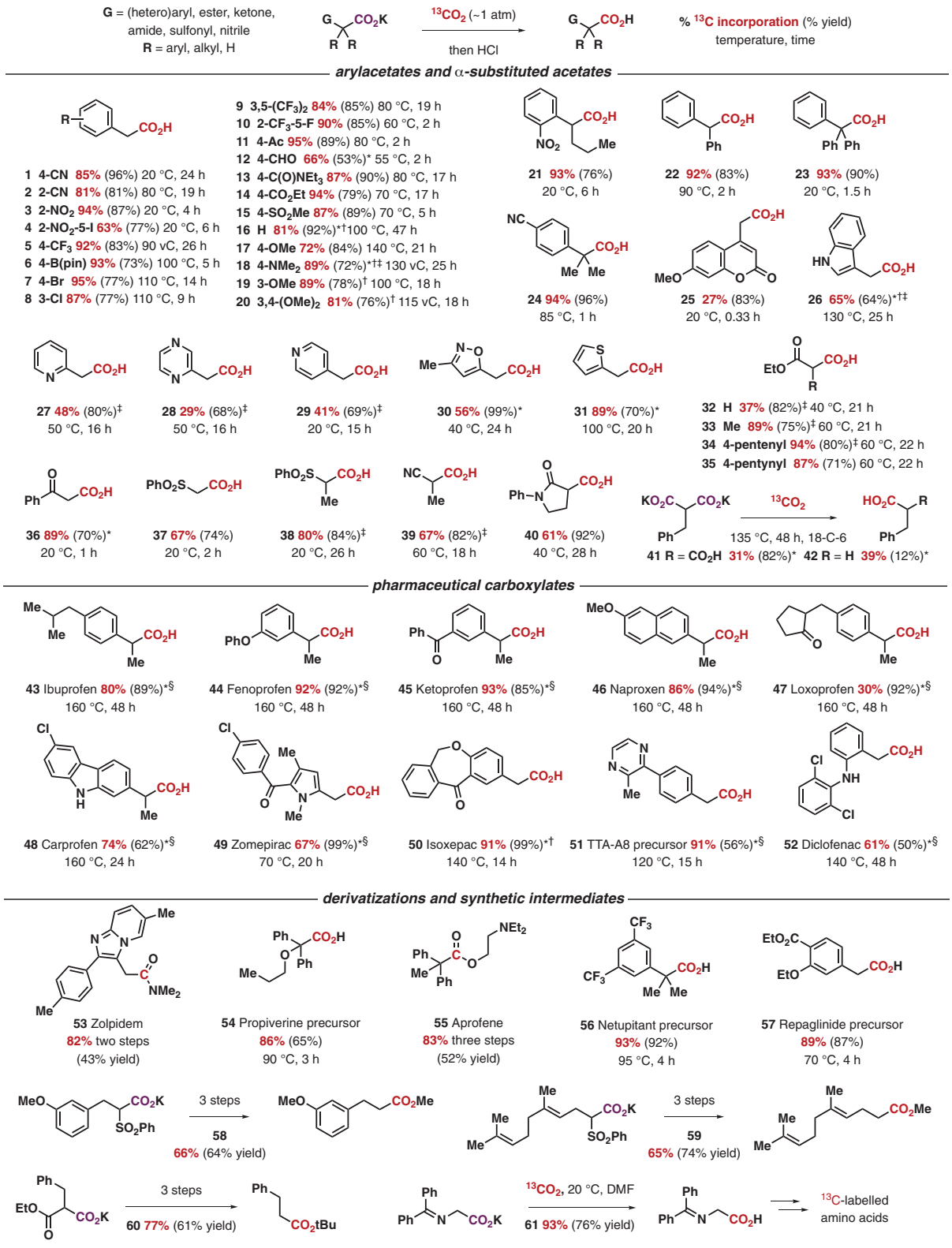


Fig. 2. Carboxylate exchange scope and application. Unless noted otherwise, yields are of isolated material. *Calibrated ^{1}H NMR spectroscopy yield. †One equivalent of 18-C-6 added. ‡Percent ^{13}C incorporation and yield determined by analysis of the corresponding methyl or benzyl ester. §DMSO used instead of DMF. See the supplementary materials for complete details. Ac, acetyl; Et, ethyl; Ph, phenyl.

Quantitative recovery of carboxylate **1** with 83% ^{13}C-enrichment was possible by acid-base extractive workup. Under similar conditions at 20°C with 5 equiv of a weak Brønsted acid [methanol (MeOH)], no protodecarboxylation of **1** was observed (Fig. 1C, black trace). These results demonstrate that capture of the putative nucleophilic intermediate generated from **1** with dissolved CO_2 is considerably more favorable than protonation. The process tolerates up to 0.01 M H_2O and does not require rigorous exclusion of air (see fig. S2). During the review of this work, isotopic exchange of carboxylate groups in cesium arylacetic acid salts with ^{14}C-, ^{13}C-, and ^{11}C-labeled CO_2 in dimethyl sulfoxide (DMSO) at 80° to 190°C was reported (*26*).

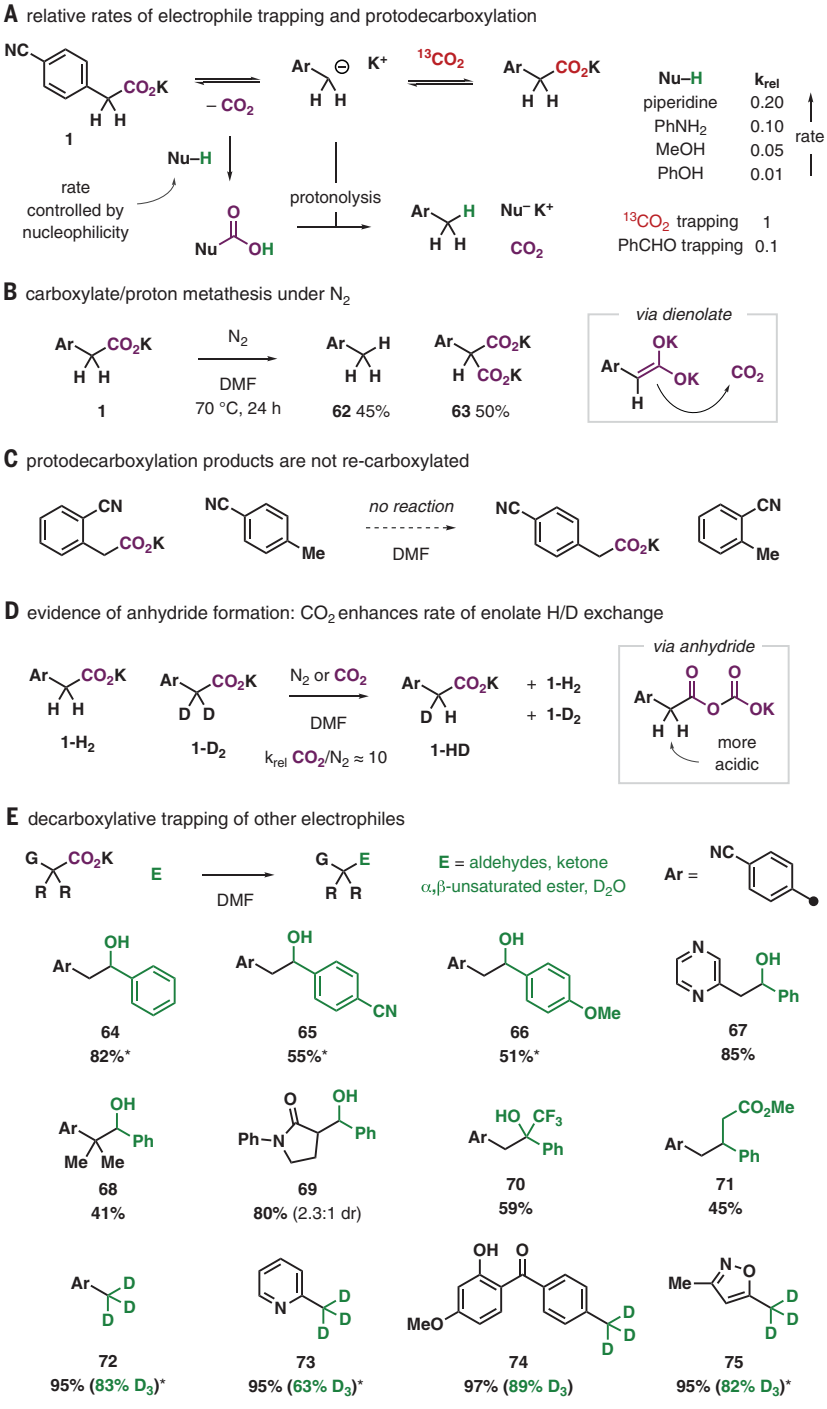


Fig. 3. Mechanistic control experiments and electrophile trapping. (**A**) Relative rates for arylacetic acid salt protodecarboxylation by Brønsted acids and decarboxylative trapping by benzaldehyde. (**B**) Net carboxylate/proton metathesis of **1** in the absence of additional CO_2. (**C**) CO_2 transfer between arylacetic acid salt and alkylarene does not occur. (**D**) The presence of CO_2 accelerates the rate of methylene C–H/D exchange in **1**. (**E**) Scope examples for the decarboxylative trapping of alternative classes of electrophiles. See the supplementary materials for complete details and full reaction conditions. *Yield determined by ^{1}H NMR spectroscopy. Ar, (4-CN)C_6H_4; dr, diastereomeric ratio.

Reversible decarboxylation occurred for an array of carboxylate-containing molecules that contain adjacent aryl, carbonyl, cyano, or sulfonyl groups, including valuable synthetic precursors, drug molecules, and amino acid derivatives (Fig. 2). The incorporation of $^{13}CO_2$ and product recovery remained high (>80%) across several substrate classes. The degree of incorporation is largely a function of the amount of $^{13}CO_2$ supplied: Greater than 95% enrichment can be obtained when ~50 equiv is provided (see fig. S5). For successful cases, this carboxylate exchange process compares favorably in terms of operational simplicity to current state-of-the-art methods to prepare C(sp^3)–$^{13/14}CO_2$–labeled carboxylic acid-derivatives sought after in (pre)clinical absorption, distribution, metabolism, and excretion (ADME) studies (*28*). Reported approaches require either chemical activation-decarboxylation-metalation-carboxylation sequences mediated by transition metals (*29*–*31*), indirect nucleophilic substitution reactions with labeled cyanide followed by hydrolysis (*32*), or introduction of labeled carbon monoxide in place of CO_2 (*33*, *34*). The direct exchange of C(sp^2)-carboxylate groups catalyzed by transition metals has been demonstrated; however, reactivity is restricted to nitro- or sulfonyl-containing arenes or 2-heteroatom substituted electron-rich heterocycles at high temperature (≥150°C) (*35*).

(Hetero)arylacetic acid salts with anion-stabilizing groups underwent exchange at moderate temperatures (Fig. 2, **1** to **4**, **9**, **10**,

and **11** to **15** at 20° to 80°C), whereas arylacetates with strongly electron-donating OMe or NMe_2 groups (Me, methyl) required higher temperatures (**17** to **20** at 100° to 130°C) and benefitted from the addition of 18-C-6. The simplicity of the process enabled broad functional group compatibility, including tolerance to boronic esters (**6**), aryl halides (I, Br, Cl, F; **4**, **7**, **8**, and **10**), ketones (**11**), aldehydes (**12**), amides (**13**), esters (**14**), sulfonyls (**15**), and potentially reactive heterocycles [chromenone (**25**), NH-indole (**26**), pyridines (**27** and **29**), pyrimidine (**28**), isoxazole (**30**), and thiophene (**31**)]. Alkyl and aryl substitution adjacent to the carboxylate was tolerated, including examples of trisubstituted, nonenolizable arylacetic acid salts (**23** and **24**). Other classes of potassium carboxylates that underwent reversible decarboxylation include malonate half-esters (**32** to **35**), keto acids (**36**), β-carboxysulfonyls (**37** and **38**), cyanoacetates (**39**), and carboxylactams (**40**). Alkene and terminal alkyne functional groups did not interfere with the process (**34** and **35**). Potassium malonates underwent CO_2 exchange at higher temperature (135°C) to give a mixture of mono- and doubly labeled product along with ^{13}C-enriched monoacid (**41** and **42**).

Carboxylate exchange could also be used to directly prepare isotopically labeled drug molecules, including arylacetic acid salts and propionate nonsteroidal anti-inflammatory drugs (NSAIDs) of varying complexity (**43** and **52**; Fig. 2). Pharmaceuticals featuring amide or ester groups were obtained via derivatization of the acid group [zolpidem (**53**), aprofene (**55**)] or could be prepared according to established literature protocols [propiverine (**54**), netupitant (**56**), repaglinide (**57**)]. Consistent with the generation of a carbanion, racemization of enantiopure naproxen (**46**) was observed (fig. S6). Reversible decarboxylation may explain reports of aryl propionate racemization required for kinetic resolution manufacturing processes (*36*, *37*). Simple alkyl carboxylates did not undergo CO_2 exchange; however, isotopically labeled products of this class can be readily obtained by carboxylate exchange-desulfonylation reactions of β-sulfonyl acids or exchange-decarboxylation sequences of malonic acids in three steps (**58** to **60**). The facile generation of ^{13}C-diphenylmethylidene glycine at room temperature (**61**, 93% incorporation, 76% yield) serves as a starting point for the synthesis of other labeled amino acids (*38*).

The reversible CO_2 exchange process likely involves the formation of a carbon nucleophile either from direct decarboxylation or, potentially in the case of enolizable substrates, through an enolate intermediate. The reaction rates and required temperatures for $^{12}CO_2$-$^{13}CO_2$ interconversion correlate with the substrate's capacity to stabilize negative charge and not with oxidation potential (compare **1**, **14**, **16**, and **17**). The addition of radical inhibitors [(2,2,6,6-tetramethylpiperidin-1-yl)oxyl (TEMPO), butylated hydroxytoluene BHT] had no impact on the decarboxylative reactivity of **1** nor was cyclization of the pendant olefin in **34** detected.

Exchange of CO_2 via carbanion equivalents without competing quenching by other electrophiles (ketones, aldehydes, weak Brønsted acids) likely stems in part from the relatively high solubility of CO_2 in DMF and the slow kinetics of CO_2 evaporation into the reaction vessel headspace. For example, a 0.25 M solution of $^{13}CO_2$ in DMF retains a concentration of 0.2 M under a headspace of N_2 over 1 day [measured by ^{13}C nuclear magnetic resonance (NMR)]. For substrate **1**, the rate of CO_2 exchange at 70°C was ~10-fold faster than for reaction with benzaldehyde. Perhaps counterintuitively, the rate of protodecarboxylation by weak Brønsted acids is inversely related to acidity (Fig. 3A; in order of decreasing rate of protonolysis: piperidine, aniline, methanol, phenol; see fig. S7 for details). This observation could be attributed to the relative capacity of these species to act as nucleophiles to sequester the liberated CO_2. Trapping of CO_2 prevents back reaction of the carbanion to the carboxylate leading to an increase in the observed rate of protodecarboxylation.

At 70°C under N_2, **1** underwent slow net carboxylate-proton metathesis to generate a half-equivalent of the protodecarboxylated product **62** and a half-equivalent of the CO_2-trapped malonate **63** (Fig. 3B). Product **62** likely arises from deprotonation of a second equivalent of aryl acetate to generate a dienolate nucleophile. The dienolate intermediate can react with the CO_2 released by the initial decarboxylation event. This observation demonstrates the marked efficiency of CO_2 capture by carbon nucleophiles under suitable conditions. Alkyl arenes generated by protodecarboxylation do not convert back to the carboxylate under the conditions where carboxylate exchange is observed (Fig. 3C). Carbonic anhydride intermediates are likely generated under the reaction conditions on the basis of the observed increase in α-carboxyl H/D exchange rates with **1-H$_2$** and **1-D$_2$** under CO_2 (Fig. 3D; see fig. S9 for details). The generation of a dienolate from the more acidic potassium carbonic anhydride may explain these reactivity differences. Direct detection of anhydride intermediates was not achieved. The capacity of nonenolizable carboxylates, such as **23** and **24**, to undergo reversible decarboxylation indicates that dienolate or enol intermediates are not essential for carboxylate exchange.

Finally, with an understanding of the factors that contribute to transient substrate decarboxylation, conditions were identified that allowed for the direct decarboxylative trapping of alternative classes of electrophiles (Fig. 3E). Carbon-carbon bond forming reactions by the trapping of aldehydes (**64** to **69**), a trifluoromethyl ketone (**70**), and an α,β-unsaturated ester (**71**) occurred in reasonable yields under conditions similar to those for reversible decarboxylation. The rates of product formation in aldehyde trapping experiments correlate with substrate electrophilicity. In some cases, it was beneficial (but not essential) to add 18-C-6 to improve decarboxylative reactivity or aniline to sequester liberated CO_2 (see supplementary materials for complete details). H/D exchange followed by deuterodecarboxylation provides a simple approach to prepare CD_3-labeled toluenes and heterocycles from D_2O (**72** to **75**).

Efficient reversible decarboxylation-carboxylation masks the inherent reactivity of otherwise stable carboxylates. An appreciation of this phenomenon enables simple, direct protocols for isotopic exchange of carboxylic acids with $^{13}CO_2$ and methods for decarboxylative carbon-carbon bond forming reactions. The potential for reversible decarboxylation should be considered more generally when designing and executing decarboxylative functionalization processes.

REFERENCES AND NOTES

1. T. Li, L. Huo, C. Pulley, A. Liu, *Bioorg. Chem.* **43**, 2–14 (2012).
2. P. A. Frey, A. D. Hegeman, *Enzymatic Reaction Mechanisms* (Oxford Univ. Press, 2007).
3. M. H. O'Leary, in *The Enzymes*, D. S. Sigman, Ed. (Academic Press, 1992), vol. 20, pp. 235–269.
4. R. Kluger, *Acc. Chem. Res.* **48**, 2843–2849 (2015).
5. A. J. Colussi, V. T. Amorebieta, M. A. Grela, *J. Chem. Soc., Faraday Trans.* **88**, 2125–2127 (1992).
6. N. Rodríguez, L. J. Goossen, *Chem. Soc. Rev.* **40**, 5030–5048 (2011).
7. A. K. Vijh, B. E. Conway, *Chem. Rev.* **67**, 623–664 (1967).
8. Z. W. Zuo *et al.*, *Science* **345**, 437–440 (2014).
9. D. H. R. Barton *et al.*, *Pure Appl. Chem.* **60**, 1549–1554 (1988).
10. T. Qin *et al.*, *Science* **352**, 801–805 (2016).
11. M. B. Smith, *March's Advanced Organic Chemistry – Reactions, Mechanisms, and Structure* (Wiley, ed. 7, 2013).
12. J. Schwarz, B. Konig, *Green Chem.* **20**, 323–361 (2018).
13. R. B. Silverman, Ed., *Organic Chemistry of Enzyme-Catalyzed Reactions* (Academic Press, ed. 2, 2002), pp. 289–320.
14. S. S. Bailey *et al.*, *Nat. Chem.* **11**, 1049–1057 (2019).
15. T. Nunoura *et al.*, *Science* **359**, 559–563 (2018).
16. A. Mall *et al.*, *Science* **359**, 563–567 (2018).
17. X. Sheng *et al.*, *Biochemistry* **57**, 3167–3175 (2018).
18. Q. Liu, L. Wu, R. Jackstell, M. Beller, *Nat. Commun.* **6**, 5933 (2015).
19. A. Tortajada, F. Juliá-Hernández, M. Börjesson, T. Moragas, R. Martin, *Angew. Chem. Int. Ed.* **57**, 15948–15982 (2018).
20. D. J. Darensbourg, E. M. Longridge, M. W. Holtcamp, K. K. Klausmeyer, J. H. Reibenspies, *J. Am. Chem. Soc.* **115**, 8839–8840 (1993).
21. A. Häussermann, F. Rominger, B. F. Straub, *Chemistry* **18**, 14174–14185 (2012).
22. A. Szabolcs, J. Szammer, L. Noszkó, *Tetrahedron* **30**, 3647–3648 (1974).
23. R. Nakai, M. Sugii, H. Nakao, *J. Am. Chem. Soc.* **81**, 1003–1006 (1959).
24. P. J. Moon, S. Yin, R. J. Lundgren, *J. Am. Chem. Soc.* **138**, 13826–13829 (2016).
25. P. J. Moon, A. Fahandej-Sadi, W. Qian, R. J. Lundgren, *Angew. Chem. Int. Ed.* **57**, 4612–4616 (2018).
26. G. Destro *et al.*, *Angew. Chem. Int. Ed.* (2020).
27. D. H. Hunter, M. Hamity, V. Patel, R. A. Perry, *Can. J. Chem.* **56**, 104–113 (1978).
28. C. S. Elmore, R. A. Bragg, *Bioorg. Med. Chem. Lett.* **25**, 167–171 (2015).
29. K. Hinsinger, G. Pieters, *Angew. Chem. Int. Ed.* **58**, 9678–9680 (2019).
30. C. Kingston *et al.*, *J. Am. Chem. Soc.* **141**, 774–779 (2019).

31. A. Tortajada *et al.*, *ACS Catal.* **9**, 5897–5901 (2019).
32. R. Voges, J. R. Heyes, T. Moenius, *Preparation of Compounds Labeled with Tritium and Carbon-14* (Wiley, 2009).
33. D. U. Nielsen, K. T. Neumann, A. T. Lindhardt, T. Skrydstrup, *J. Labelled Comp. Radiopharm.* **61**, 949–987 (2018).
34. D. R. Gauthier Jr., N. R. Rivera, H. Yang, D. M. Schultz, C. S. Shultz, *J. Am. Chem. Soc.* **140**, 15596–15600 (2018).
35. G. Destro *et al.*, *J. Am. Chem. Soc.* **141**, 780–784 (2019).
36. I. Shiina, K. Ono, K. Nakata, *Catal. Sci. Technol.* **2**, 2200–2205 (2012).
37. P. J. Harrington, E. Lodewijk, *Org. Process Res. Dev.* **1**, 72–76 (1997).
38. T. Ooi, K. Maruoka, *Angew. Chem. Int. Ed.* **46**, 4222–4266 (2007).

ACKNOWLEDGMENTS

We thank A. Speed for analysis of this manuscript and experimental suggestions. We thank J. Takats for the gift of $^{13}CO_2$ to initiate this project and B. Reiz for assistance in mass spectrometry data analysis. **Funding:** Support was provided by NSERC Canada (RGPIN-2019-06050 and RGPAS-2019-00051), the Canadian Foundation for Innovation (IOF 32691), the University of Alberta, and the Killam Trusts. **Author contributions:** P.J.M. and R.J.L. conceived the study. Experiments and data analysis were conducted by D.K., P.J.M., E.K.J.L., and O.B. The manuscript was written by R.J.L. with consultation from all co-authors. **Competing interests:** P.J.M. and R.J.L. (University of Alberta) are pursuing a provisional patent on this work.

Data and materials availability: All experimental data are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/557/suppl/DC1
Materials and Methods
Figs. S1 to S9
HRMS and NMR Spectra
References (*39*–*49*)

25 February 2020; accepted 9 June 2020
Published online 18 June 2020
10.1126/science.abb4129

NEUROGENOMICS

Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants

Siwei Zhang[1]*, Hanwen Zhang[1]*, Yifan Zhou[2,3]*, Min Qiao[2,4]*, Siming Zhao[2]*, Alena Kozlova[1], Jianxin Shi[5], Alan R. Sanders[1,6], Gao Wang[2], Kaixuan Luo[2], Subhajit Sengupta[1], Siobhan West[1], Sheng Qian[2], Michael Streit[1], Dimitrios Avramopoulos[7], Chad A. Cowan[8], Mengjie Chen[9], Zhiping P. Pang[10], Pablo V. Gejman[1,6], Xin He[2,11]†, Jubao Duan[1,6]†

Most neuropsychiatric disease risk variants are in noncoding sequences and lack functional interpretation. Because regulatory sequences often reside in open chromatin, we reasoned that neuropsychiatric disease risk variants may affect chromatin accessibility during neurodevelopment. Using human induced pluripotent stem cell (iPSC)–derived neurons that model developing brains, we identified thousands of genetic variants exhibiting allele-specific open chromatin (ASoC). These neuronal ASoCs were partially driven by altered transcription factor binding, overrepresented in brain gene enhancers and expression quantitative trait loci, and frequently associated with distal genes through chromatin contacts. ASoCs were enriched for genetic variants associated with brain disorders, enabling identification of functional schizophrenia risk variants and their cis-target genes. This study highlights ASoC as a functional mechanism of noncoding neuropsychiatric risk variants, providing a powerful framework for identifying disease causal variants and genes.

The difficulty of assigning molecular function to noncoding disease risk variants has hindered the identification of causal variants and genes of complex disorders. Noncoding regulatory sequences are marked by open chromatin, and open chromatin regions (OCRs) are enriched for single-nucleotide polymorphisms (SNPs) associated with gene expression and complex disorders or traits (*1*, *2*). For neuropsychiatric disorders, disease-associated variants are also enriched in OCRs found not only in adult brains but also in developing brains and induced pluripotent stem cell (iPSC)–derived neuronal (iN) cells (*3*–*6*), suggesting the importance of neurodevelopmental chromatin regulation. However, as not all OCR variants are functional, it remains largely unknown whether neuropsychiatric disease risk variants affect chromatin accessibility during neurodevelopment.

[1]Center for Psychiatric Genetics, NorthShore University HealthSystem, Evanston, IL 60201, USA. [2]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA. [3]The Graduate Program in Biophysical Sciences, University of Chicago, Chicago, IL 60637, USA. [4]Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA. [5]Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA. [6]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL 60637, USA. [7]Department of Genetic Medicine and Psychiatry, Johns Hopkins University, Baltimore, MD 21205, USA. [8]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA. [9]Department of Medicine, University of Chicago, Chicago, IL 60637, USA. [10]Department of Neuroscience and Cell Biology, Child Health Institute of New Jersey, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ 08901, USA. [11]Grossman Institute for Neuroscience, Quantitative Biology and Human Behavior, University of Chicago, Chicago, IL 60637, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: xinhe@uchicago.edu (X.H.); jduan@uchicago.edu (J.D.)

To tie risk alleles to function, we identified allele-specific open chromatin (ASoC) variants displaying allelic imbalance of chromatin accessibility in assay for transposase-accessible chromatin using sequencing (ATAC-seq) at heterozygous SNP sites (fig. S1) (*7*, *8*). We focused on 20 individuals carrying heterozygous (i.e., informative for ASoC assay) genome-wide association study (GWAS) index SNPs at ~70 of 108 schizophrenia (SZ) loci (fig. S2 and table S2) (*7*, *9*, *10*).

We differentiated iPSCs of these 20 individuals into neural progenitor cells (NPCs), glutamatergic neurons (iN-Glut), γ-aminobutyric acid–releasing (GABAergic) neurons (iN-GA), and dopaminergic neurons (iN-DN) with high purity (75 to 90%) (Fig. 1A and fig. S3) (*7*). We performed ATAC-seq and RNA sequencing (RNA-seq) in each cell type for eight lines ("core-8") and ATAC-seq for 12 additional lines in NPCs and iN-Glut (Fig. 1A, figs. S4 and S5, and table S3). Pooling the reads of core-8 lines, we identified between 256,000 and 337,000 OCRs [false discovery rate (FDR) < 0.05] in each cell subtype (fig. S6, A to C; table S4; and data S1).

The OCRs of cell type–specific marker genes were confirmed (fig. S6, D and E), and principal component analysis (PCA) of ATAC-seq and RNA-seq samples showed cell type–specific clustering (Fig. 1B and fig. S7, A to D). OCRs of iPSC-derived neural cells were more similar to fetal brains (*11*) and cortical organoids (*12*) than to adult brains (*3*) (Fig. 1B and fig. S7, C and D). Neuronal OCRs overlapped with ~53% of the 117,935 adult brain OCR peaks from PsychENCODE (*3*); however, the 62,813 overlapping peaks only accounted for ~20% of our neuronal OCRs (table S4), which is consistent with the observation that the fetal cerebral cortex has more active enhancers than the adult brain (*12*).

We then tested ASoC variants in each iPSC-derived cell type and examined cell type specificity (fig. S8) (*7*). We used a sample pooling approach (*8*, *13*) after confirming the inter-individual concordance of the allelic imbalance directions of chromatin accessibility (fig. S9A). We identified between 920 and 2392 ASoC SNPs (FDR < 0.05) in each core-8 cell type, and 5611 and 3547 ASoC SNPs in iN-Glut-20 and NPC-20 cells (n = 20 lines each), respectively (fig. S9, B to E, and tables S5 to S12). These ASoC SNPs showed cell type–specific patterns (Fig. 1C and fig. S9B). Using Storey's π1 analysis (*7*), we found a low percentage of estimated pairwise ASoC SNP sharing between neurons and iPSCs (10 to ~20%) (Fig. 1D). Even within neuronal cell types, ASoCs differed substantially (30 to 70% sharing) (Fig. 1D), suggesting cell type specificity of ASoC.

Cell type–specific ASoC may be driven by either OCRs or different SNP effects (allelic ratios) across cell types. About half of the ASoC SNPs in neurons exhibited low or undetectable OCRs (read depth < 20 at SNP sites) in iPSCs (fig. S10). On the other hand, ~50% of neuron-specific ASoC SNPs showed comparable or more accessible OCR in iPSCs (Fig. 1E and fig. S11), suggesting that variations in OCR read depth and SNP effect size both contribute to cell type–specific ASoC. Thus, cell type–specific regulation likely occurs even in the absence of chromatin state changes.

To examine the cell type–specific regulatory potential of ASoC SNPs, we compared the genomic and epigenomic features of cell type–specific versus shared ASoC variants (*7*). ASoC SNPs were found to be enriched in both brain promoters and enhancers (Fig. 2A and fig. S12A) (*4*, *14*). However, compared with shared ASoC SNPs, cell type–specific SNPs showed stronger enrichment (approximately twofold) in enhancers (Fig. 2A and fig. S12, B to E). The enrichment analysis indicates that ASoC SNPs are associated with brain regulatory sequences and likely exhibit cell type–specific enhancer activity.

The effect of ASoC SNPs on enhancer and promoter activity was confirmed by reporter gene assay (*7*). In a massively parallel reporter assay (MPRA) SNP dataset from nonneural cells (*15*), 26 to 29% of neural ASoC SNPs altered reporter gene expression (approximately twofold enrichment compared with non-ASoC SNPs; P = 0.001, Fisher's exact test) (fig. S13, A to D, and tables S13 and S14). Using a reporter gene assay in NPCs, we found that five of eight SZ-associated ASoC SNPs affected reporter gene expression (fig. S13, E and F, and table S15). These results suggest that a proportion of ASoC SNPs affect gene expression.

We then jointly analyzed our data with recent studies of genetic variants associated with

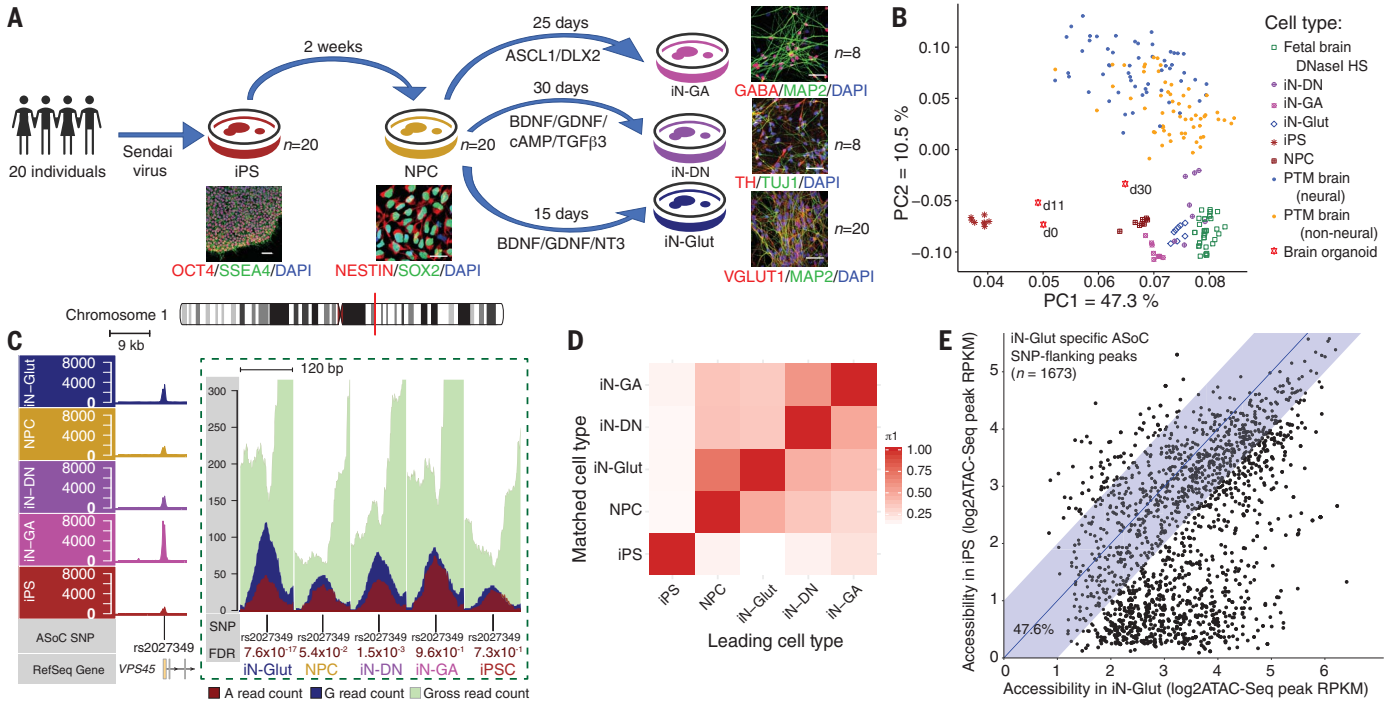


Fig. 1. Mapping ASoC variants. (**A**) Schematic of iPSC differentiation strategies. OCT4, octamer-binding transcription factor 4; SSEA4, stage-specific embryonic antigen 4; DAPI, 4′,6-diamidino-2-phenylindole; ASCL1, achaete-scute homolog 1; DLX2, distal-less homeobox 2; BDNF, brain-derived neurotrophic factor; GDNF, glial cell line–derived neurotrophic factor; cAMP, adenosine 3′,5′-monophosphate; TGFβ3, transforming growth factor–β3; NT3, neurotrophin-3; MAP2, microtubule-associated protein 2; TUJ1, class III beta-tubulin; TH, tyrosine hydroxylase; VGLUT1, vesicular glutamate transporter 1. (**B**) PCA of OCRs from iNs, fetal brains or day-30 organoids, and postmortem (PTM) adult brain. (**C**) Cell type–specific ASoC rs2027349 at the SZ-associated *VPS45* locus. The dashed box shows the ATAC-seq read pileups flanking rs2027349 and cell type–specific allelic ratios. (**D**) Pairwise π1 estimation of ASoC sharing across cell types (*7*). ASoC SNPs were ascertained in the leading cell type, and π1 was estimated in the matched one. (**E**) Chromatin accessibility in iN-Glut versus iPSC for ASoC SNP-flanking peaks. The shaded area shows comparable (less than twofold difference) accessibility between cell types. RPKM, reads per kilobase million.

brain gene expression [expression quantitative trait loci (eQTL)], histone acetylation (haQTL), and DNA methylation (meQTL) (*7*, *16*). Putative causal variants underlying these QTL are enriched (>30-fold) in iN-Glut-20 ASoC SNPs (Fig. 2B and table S12). Intersecting with brain eQTL (*17*) and brain or neuronal Hi-C chromatin contact (*18*) annotations (*7*), we found that out of 5611 iN-Glut-20 ASoC SNPs, 26% are also eQTL and 35% demonstrate chromatin contacts with distal promoters (Fig. 2C; fig. S14, A and B; and table S16). Together with promoter annotation, ~59% of ASoC SNPs were resolved to affect at least one cis-target gene (Fig. 2C and fig. S14C). Integrating ASoC SNPs with brain eQTL and Hi-C data, we delineated putative causal eQTL SNPs and their affected cis-target genes (e.g., Fig. 2D). These results provide supporting evidence that many ASoCs affect distal genes through long-range chromatin contacts.

Mapping the transcription factor (TF)–binding footprints from ATAC-seq (*7*), we found that 32% of the 5611 ASoC SNPs of iN-Glut-20 neurons were inside TF-binding sites (TFBSs), representing a 1.6-fold enrichment (compared with non-ASoC SNPs; Fisher's exact test, $P = 2.6 \times 10^{-58}$) (fig. S15A). Furthermore, most ASoC SNPs were within 200 base pairs (bp) of TF-binding footprints, and footprint analyses in other cell types gave similar results (fig. S15, B to G).

To test whether TF-motif disruptions by ASoC SNPs near footprints caused matched changes in chromatin accessibility, we performed a motif-break analysis (*7*). In iN-Glut-20 and NPC-20 cells, we identified 48 TFs with SNP motif-disruption scores correlated with allelic imbalances of ASoC (Fig. 2, E and F, and fig. S16). Comparing our ASoC SNPs with an allele-specific TF-binding SNP dataset (*19*), we found that 70 to 80% of overlapping SNPs showed matched direction of allelic effects on chromatin accessibility and TF binding (fig. S17 and table S17). These results suggest that genetic variants that alter TF binding result in ASoC.

We then performed TF-motif enrichment analysis for ASoC SNPs (±50 bp) and OCR (*7*) and identified cell type–specific patterns of enriched TF motifs (fig. S18 and tables S18 and S19). Restricting the ASoC analysis specifically to TF footprints gave similar patterns (Fig. 2G and fig. S19) and included *TCF4* and *SP4*, which are both SZ risk genes (*10*). Thus, ASoC SNPs may affect cell type–specific binding of TFs important for cell fate commitment and neurodevelopment.

Given the brain-relevant regulatory effects of ASoC SNPs, we reasoned that ASoCs could help infer noncoding risk variants for neuropsychiatric disorders. Of the 5611 iN-Glut-20 ASoC SNPs, 21 were SZ-associated SNPs (*10*) at 17 independent SZ loci (Fig. 3A and tables S20 and S21). Accounting for the uncertainty of GWAS risk variants due to linkage disequilibrium (LD) with TORUS (*7*), neuronal ASoC SNPs showed ~70-fold enrichment for SZ risk variants, compared with approximately three- to sixfold enrichments in neuronal OCRs (Fig. 3B). By computationally fine-mapping (*7*) SZ-associated loci (*10*), we estimated that ASoC SNPs, despite constituting only 0.08% of all SNPs, explain 5.5% of SZ risk variants, while non-ASoC SNPs in OCRs that account for 9.4% of all SNPs explain 22.5% of risk variants (fig. S20A).

Expanding this analysis to nine other brain disorders or traits (*7*), we observed enrichment of GWAS variants for bipolar disorder (BP), major depressive disorder (MDD), intelligence, educational attainment, and neuroticism in ASoC variants (iN-Glut and NPC) (Fig. 3C; fig. S20, B to M; and table S22). Enrichment of ASoC SNPs (only iN-Glut) was also seen for Alzheimer's disease but not for

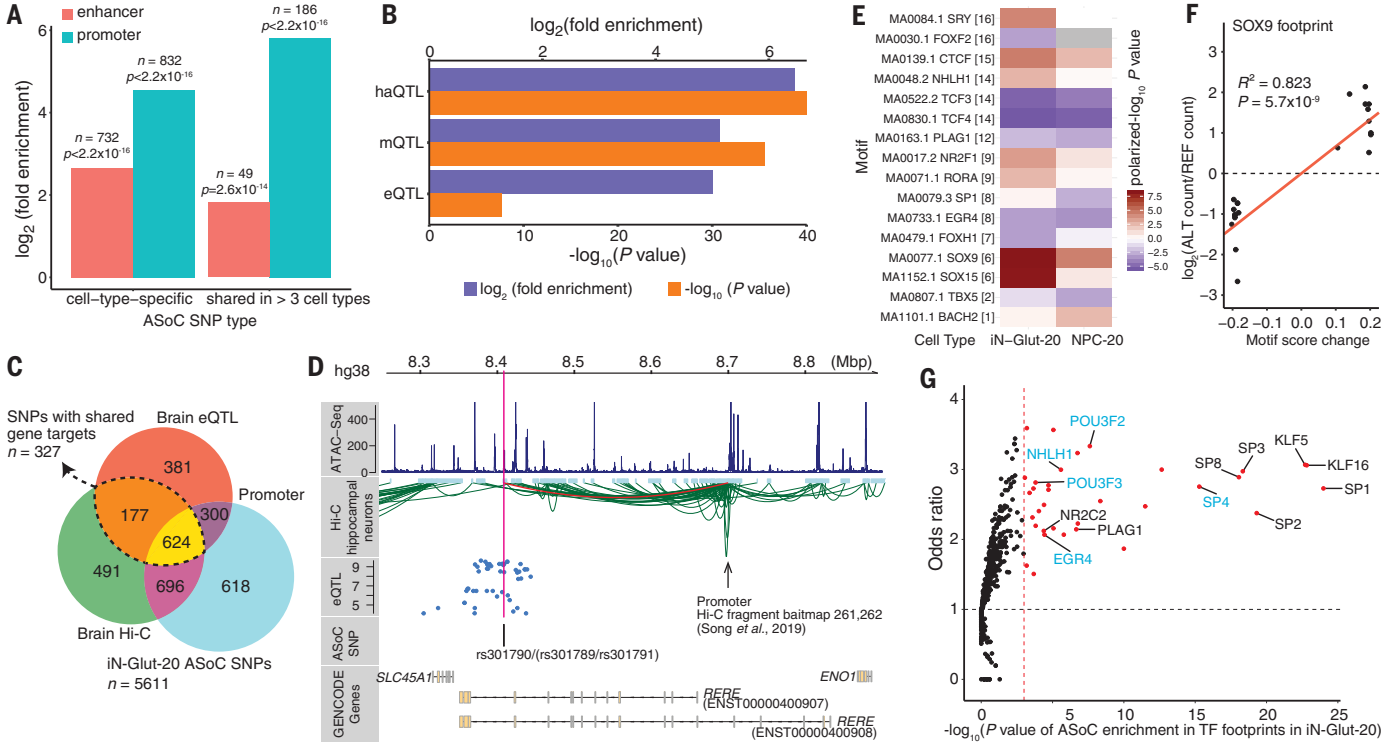


Fig. 2. Characteristics of ASoC SNPs. (**A**) Enrichment of ASoC SNPs (versus genomic background) in the chromatin state–based annotations of promoter and enhancer. *n* represents the number of SNPs. (**B**) Enrichment of ASoC SNPs (versus genomic background) for brain QTL, including histone acetylation (haQTL), DNA methylation (mQTL), and expression (eQTL). (**C**) Venn diagram showing iN-Glut-20 ASoC SNPs mapped to cis-target genes through brain eQTL (*17*), brain or neural Hi-C (*18*), and promoter annotations. (**D**) An ASoC SNP mapped to *RERE* through eQTL and Hi-C. ATAC-seq peak: iN-Glut. (**E**) Selected TFs with significant correlations (FDR < 0.05) of motif-disrupting scores and ASoC allelic imbalance. Numbers in brackets are motif cluster numbers. (**F**) *SOX9* as an example of a putative chromatin activator showing positive correlation in (E). R^2, coefficient of determination. (**G**) TFs with binding footprints enriched for neural ASoC SNPs (red dots; versus non-ASoC SNPs). Blue indicates neurodevelopmental TFs.

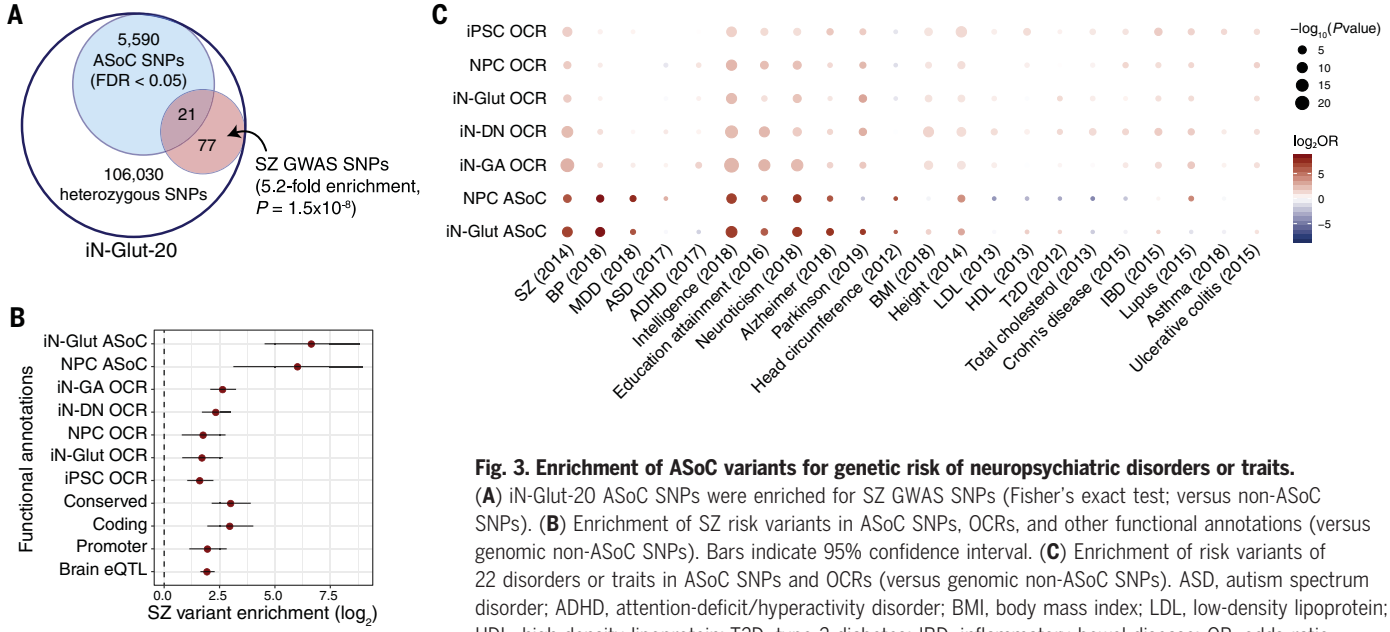


Fig. 3. Enrichment of ASoC variants for genetic risk of neuropsychiatric disorders or traits. (**A**) iN-Glut-20 ASoC SNPs were enriched for SZ GWAS SNPs (Fisher's exact test; versus non-ASoC SNPs). (**B**) Enrichment of SZ risk variants in ASoC SNPs, OCRs, and other functional annotations (versus genomic non-ASoC SNPs). Bars indicate 95% confidence interval. (**C**) Enrichment of risk variants of 22 disorders or traits in ASoC SNPs and OCRs (versus genomic non-ASoC SNPs). ASD, autism spectrum disorder; ADHD, attention-deficit/hyperactivity disorder; BMI, body mass index; LDL, low-density lipoprotein; HDL, high-density lipoprotein; T2D, type 2 diabetes; IBD, inflammatory bowel disease; OR, odds ratio.

Parkinson's disease (Fig. 3C). In contrast, 12 control disorders or traits showed little or no enrichment in neuronal ASoC SNPs or OCRs (Fig. 3C), although this may reflect limited statistical power. These results suggest that neuronal ASoCs are linked to neuropsychiatric disorders and highlight the value of using ASoCs in prioritizing studies of noncoding risk variants.

We selected the top 20 ASoC SNPs associated with SZ (tables S20 and S21) to assess the regulatory potential of their tagged ASoC sequences and cis-target genes. We used a

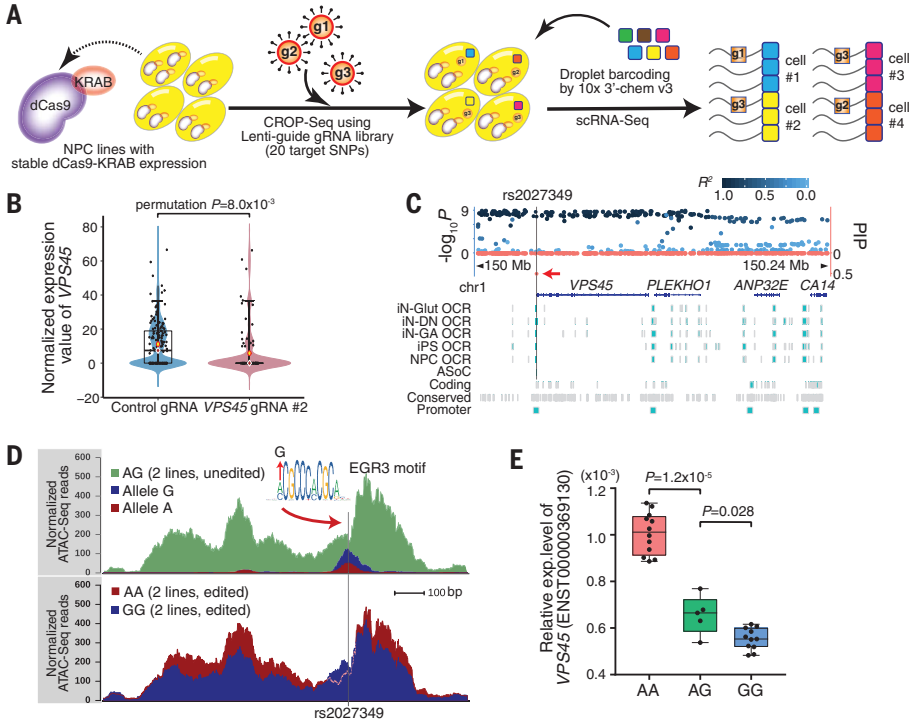


Fig. 4. Functional characterization of SZ-associated ASoC SNPs. (**A**) Schematic of the modified CROP-seq (*7*) to screen cis-targets of the 20 ASoC sites. (**B**) Reduced *VPS45* expression in single NPCs expressing gRNAs targeting rs2027349. (**C**) Fine-mapping the *VPS45* (rs2027349) locus. Left *y* axis: $-\log_{10} P$ values from SZ GWAS (*10*) (points above the *x* axis). Right *y* axis: PIPs (points below the *x* axis, red). Vertical bar: rs2027349. (**D**) CRISPR-editing rs2027349 altered local OCR accessibility in isogenic NPCs (AA versus GG; two lines per genotype). The G allele disrupts the EGR3 motif; note that AA lines showed a higher (more accessible) peak but with a lower AAC-seq signal at the EGR3-binding site likely because of A allele's stronger EGR3-binding, which prevents transposase from accessing the EGR3-binding site. (**E**) Editing rs2027349 altered *VPS45* expression in NPCs (major transcript ENST00000369130; by qPCR). *VPS45* expression was normalized to *GAPDH*. Two iPSC lines (AG at rs2027349) were edited to isogenic lines with AA or GG genotype, each with two or three clones (n = 2 or 3 independent cultures). Student's t test (Welch's correction) was used. The upper and lower box-whiskers represent the 95% and 5% quantile range.

modified CRISPR droplet sequencing (CROP-seq) approach (*20*) (Fig. 4A) in NPCs stably expressing dCas9/Krüppel associated box (KRAB) which represses possible cis-genes in specific ASoC sites targeted by guide RNAs (gRNAs) (fig. S21, A to C, and table S23) (*7*). We analyzed 4099 cells, 2522 of which were assigned to unique gRNAs (fig. S21, D to F). We identified putative cis-targets as differentially expressed genes (permutation $P < 0.05$) for a given gRNA and ASoC sequence (Fig. 4B, fig. S22, and table S24) and then validated by independent CRISPR interference and quantitative polymerase chain reaction (CRISPRi/qPCR) (fig. S23 and table S25) (*7*). In total, CROP-seq identified cis-targets of 10 ASoC sequences, four of which share targets with brain eQTL and brain or neural Hi-C contacts (*18*, *21*, *22*) (fig. S24 and table S26).

CROP-seq screening of cis-targets of an ASoC SNP-flanking sequence did not inform SNP allelic effect. To link SNP allelic function to a cis-target gene, we carried out precise SNP editing by CRISPR-Cas9 (*7*). We prioritized an ASoC SNP for editing on the basis of support from ASoC mapping, CROP-seq screening, and the posterior inclusion probability (PIP) (*7*), which is the statistical evidence of an SNP being a disease-causal variant on the basis of fine-mapping (*23*). Among the 20 targeted ASoC SNPs, nine had PIP > 0.10 (table S27), and six of those nine had at least one CROP-seq cis-gene (table S24). Of these, rs2027349 at the *VPS45* locus showed the strongest ASoC in NPC-20 (table S21), exhibited a repressive effect on *VPS45* in CROP-seq (permutation $P = 0.008$) (Fig. 4B), and was the only local SNP with a high PIP (0.45) (Fig. 4C). rs2027349 is within a TF-binding footprint, with the G allele predicted to disrupt the motif of *EGR3/4* (Fig. 4D), a TF involved in neurodevelopment and SZ (*24*). rs2027349 was thus prioritized for CRISPR-Cas9 editing.

We edited two iPSC lines heterozygous for rs2027349 from AG to AA and GG (fig. S25, A and B, and table S28) and characterized the SNP allelic effect on *VPS45* expression in NPCs. Allele A was associated with increased expression of *VPS45*, especially its major transcript isoform (Fig. 4E and fig. S25, C and D). The increased *VPS45* expression by allele A is consistent with the allelic effect of rs2027349 on *VPS45* in PsychENCODE brain RNA-seq data (fig. S25E). In addition, rs2027349-flanking OCR in CRISPR-edited AA lines showed higher chromatin accessibility than in GG (Fig. 4D). We did not observe similar effects on two other cis-genes (*ANP32E* and *SF3B4*) nominated by CROP-seq (fig. S22, A and B). Taken together, we identified rs2027349 as a likely SZ risk variant that affects the expression of *VPS45*, a gene potentially relevant to SZ given its involvement in vesicle-mediated protein trafficking and neurotransmitter release (*25*).

We also performed CRISPR-Cas9 editing of rs12895055 (C→T) (fig. S26), an SZ-associated SNP that did not have a CROP-seq cis-target but showed the strongest ASoC in iN-Glut (table S20) and a high probability of being a causal SZ variant (PIP = 0.33) (table S27). Because rs12895055 is in an intron of *BCL11B* (also known as CTIP2, a deep-layer neuronal marker) with undetectable expression in NPCs (table S24), we examined its allelic effect on *BCL11B* in CRISPR-edited iN-Glut cells (fig. S26 and table S29). The risk allele T was associated with reduced expression of *BCL11B* and is predicted to alter binding-motifs of multiple TFs (fig. S26). Notably, rs12895055 is in perfect LD with the BP-associated rs11624408 ($P = 5 \times 10^{-8}$) (*26*). *BCL11B* is thus a candidate risk gene for both SZ and BP, warranting follow-up.

We have found that genetic variation commonly affects chromatin accessibility and exhibits cell type–specific ASoC in an iPSC-based neurodevelopmental model. The enrichments of neuronal ASoC SNPs for brain enhancers, TFBSs, and brain eQTL suggest mechanistic links between chromatin accessibility and gene expression. Our study identified ASoC variants that have plausible causal associations with SZ and demonstrates that some neuropsychiatric disease risk variants alter chromatin accessibility.

REFERENCE AND NOTES

1. M. T. Maurano *et al.*, *Science* **337**, 1190–1195 (2012).
2. H. K. Finucane *et al.*, *Nat. Genet.* **47**, 1228–1235 (2015).
3. J. Bryois *et al.*, *Nat. Commun.* **9**, 3121 (2018).
4. L. de la Torre-Ubieta *et al.*, *Cell* **172**, 289–304.e18 (2018).
5. M. P. Forrest *et al.*, *Cell Stem Cell* **21**, 305–318.e8 (2017).
6. J. F. Fullard *et al.*, *Genome Res.* **28**, 1243–1252 (2018).
7. Materials and methods are available as supplementary materials.
8. J. Xu *et al.*, *Nat. Genet.* **49**, 377–386 (2017).
9. J. Shi *et al.*, *Nature* **460**, 753–757 (2009).
10. Schizophrenia Working Group of the Psychiatric Genomics Consortium, *Nature* **511**, 421–427 (2014).

11. B. E. Bernstein *et al*., *Nat. Biotechnol.* **28**, 1045–1048 (2010).
12. A. Amiri *et al*., *Science* **362**, eaat6720 (2018).
13. V. Onuchic *et al*., *Science* **361**, eaar3146 (2018).
14. R. E. Handsaker *et al*., *Nat. Genet.* **47**, 296–303 (2015).
15. R. Tewhey *et al*., *Cell* **165**, 1519–1529 (2016).
16. B. Ng *et al*., *Nat. Neurosci.* **20**, 1418–1426 (2017).
17. M. Fromer *et al*., *Nat. Neurosci.* **19**, 1442–1453 (2016).
18. M. Song *et al*., *Nat. Genet.* **51**, 1252–1262 (2019).
19. O. Wagih, D. Merico, A. Delong, B. J. Frey, Allele-specific transcription factor binding as a benchmark for assessing variant impact predictors. bioRxiv 253427 [Preprint]. 1 February 2018. https://doi.org/10.1101/253427.
20. P. Datlinger *et al*., *Nat. Methods* **14**, 297–301 (2017).
21. P. Rajarajan *et al*., *Science* **362**, eaat4311 (2018).
22. H. Won *et al*., *Nature* **538**, 523–527 (2016).
23. G. Wang, A. Sarkar, P. Carbonetto, M. Stephens, A simple new approach to variable selection in regression, with application to genetic fine-mapping. bioRxiv 501114 [Preprint]. 27 June 2020. https://doi.org/10.1101/501114.
24. K. K. Marballi, A. L. Gallitano, *Front. Behav. Neurosci.* **12**, 23 (2018).
25. I. Dulubova *et al*., *EMBO J.* **21**, 3620–3631 (2002).
26. E. A. Stahl *et al*., *Nat. Genet.* **51**, 793–803 (2019).
27. S. Zhang, Y. Zhou, M. Qiao, S. Zhao, endeneon/ASoC_functional_disease_variants: Allele-specific open chromatin inhuman iPSC neurons elucidates functional disease variants, Version 1.1, Zenodo (2020); http://doi.org/10.5281/zenodo.3728243.

ACKNOWLEDGMENTS

We thank W. J. Greenleaf for helping on ATAC-seq and Rutgers University Cell and DNA Repository (RUCDR; 2U24MH068457) for producing iPSC lines. **Funding:** Funding was provided by NIH grants R01AA023797 (to Z.P.P.); R01MH113215 (to D.A.); R01MH110531 (to X.H.); and R01MH106575, R01MH116281, and R01AG063175 (to J.D.). **Author contributions:** S.Zhang analyzed the ATAC-seq, bulk and single-cell RNA sequencing data, CROP-seq, and wrote the manuscript. S.Zhang and H.Z. performed the experiments, analyzed the data, and wrote the manuscript. Y.Z., M.Q., S.Zhao, G.W., K.L., and S.Q. performed computational analyses of tissue-specific ASoC, TF footprints, GWAS enrichment, fine-mapping, and CROP-seq. A.K. and M.S. carried out dopaminergic neuron differentiation. J.S. and S.S. helped with data quality control and statistical analyses. S.W. helped with CROP-seq gRNA design and library preparation. A.R.S. and P.V.G. helped with clinical phenotypes, data interpretation, and manuscript writing. C.A.C. and Z.P.P. guided the gene editing and neuron differentiation, respectively. D.A. helped with MPRA and reporter gene analyses. M.C. helped with the CROP-seq data analysis. X.H. supervised the data analyses of M.Q., S.Zhao, and Y.Z. and wrote the manuscript. J.D. conceived of the study, supervised the experiments and data analyses, and wrote the manuscript. **Competing interests:** The authors declare no conflicts of interest. **Data and materials availability:** All sequencing data are available at NCBI GEO under accession numbers GSE129017 and GSE132757. Codes have been deposited at Zenodo (*27*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/561/suppl/DC1
Materials and Methods
Figs. S1 to S26
Tables S1 to S29
References (*28*–*87*)
MDAR Reproducibility Checklist
Data S1

View/request a protocol for this paper from *Bio-protocol*.

17 June 2019; resubmitted 16 September 2019
Accepted 1 June 2020
10.1126/science.aay3983

BIOTECHNOLOGY

DNA capture by a CRISPR-Cas9–guided adenine base editor

Audrone Lapinaite[1]*†, Gavin J. Knott[1,2]*, Cody M. Palumbo[3], Enrique Lin-Shiao[1], Michelle F. Richter[4,5,6], Kevin T. Zhao[4,5,6], Peter A. Beal[3], David R. Liu[4,5,6], Jennifer A. Doudna[1,7,8,9,10,11]‡

CRISPR-Cas–guided base editors convert A•T to G•C, or C•G to T•A, in cellular DNA for precision genome editing. To understand the molecular basis for DNA adenosine deamination by adenine base editors (ABEs), we determined a 3.2-angstrom resolution cryo–electron microscopy structure of ABE8e in a substrate-bound state in which the deaminase domain engages DNA exposed within the CRISPR-Cas9 R-loop complex. Kinetic and structural data suggest that ABE8e catalyzes DNA deamination up to ~1100-fold faster than earlier ABEs because of mutations that stabilize DNA substrates in a constrained, transfer RNA–like conformation. Furthermore, ABE8e's accelerated DNA deamination suggests a previously unobserved transient DNA melting that may occur during double-stranded DNA surveillance by CRISPR-Cas9. These results explain ABE8e-mediated base-editing outcomes and inform the future design of base editors.

The site-specific conversion of A•T to G•C base pairs and C•G to T•A base pairs in genomic DNA could correct ~60% of all known pathogenic single nucleotide polymorphisms (SNPs) in humans (*1*). Such conversions can be achieved with CRISPR-Cas9 base editors, RNA-guided Cas proteins fused to a single-stranded DNA (ssDNA) deaminase (*2*, *3*). A•T to G•C transitions in DNA require the deamination of adenosine to inosine, which is recognized by cellular machinery as guanosine (Fig. 1, A and B). With no enzyme capable of deaminating adenine in DNA, *Escherichia coli* tRNA adenosine deaminase (TadA) was fused to Cas9 and evolved into ABE7.10, which catalyzes the targeted deamination of deoxyadenosine (*3*–*5*). ABE7.10 encodes two copies of TadA, an N-terminal wild-type (WT) TadA linked to an evolved TadA (TadA-7.10), which is C-terminally linked to a "nickase" [i.e., cuts one strand of the double-stranded DNA (dsDNA)] version of *Streptococcus pyogenes* Cas9 (nSpCas9) (Fig. 1C). ABE7.10 has since been widely applied as a tool to deaminate genomic DNA in many cell types and organisms (*6*–*12*). The ABE7.10 variant was found to catalyze off-target RNA editing in cells, an activity that was reduced by mutations to the TadA-7.10 domain (*13*) or by removing the N-terminal WT TadA domain, generating a truncated version miniABEmax (Fig. 1C) (*14*). ABE7.10 was further evolved to generate ABE8e, which encodes a single TadA domain (TadA-8e) broadly compatible with eight tested Cas effectors (Fig. 1C) (*15*). In vitro, ABE8e deaminates DNA at a rate 590- and 1170-fold higher than ABE7.10 and miniABEmax, respectively (*15*) (Fig. 1D and fig. S1). ABE7.10 and ABE8e contain 14 and 22 substitutions, respectively, within their evolved TadA domains relative to WT TadA (*3*, *15*) (fig. S2A). How these mutations enabled evolved TadA to catalyze deamination of DNA, and the molecular basis for the different catalytic rates of ABE7.10 and ABE8e, remain unknown.

To advance our understanding of DNA substrate recognition during SpCas9-guided and TadA-catalyzed DNA adenosine deamination, we used single-particle cryo–electron microscopy (cryo-EM) to determine the structure of a DNA-bound ABE8e complex. The ABE8e complex was trapped in a catalytic conformation using a DNA nontarget strand (NTS) containing a transition state analog for adenosine deamination reactions, 8-azanebularine (8Az) (Fig. 1E) (*16*). At an overall resolution of 3.2 Å, we resolved the ABE8e complex containing Cas9 bound to a single-guide RNA (sgRNA) hybridized to a 20-nucleotide DNA target-strand (TS) within a dsDNA substrate (Fig. 1F, figs. S3 and S4, and table S1). This structure provides the most complete molecular model for a Cas9 R-loop complex, showing Cas9 engaged with the GG dinucleotide protospacer-adjacent motif (PAM), whereas both PAM-distal and PAM-proximal DNA duplexes are arranged on either side of the complex. The PAM-distal duplex is positioned against SpCas9-Y1016, whereas the sgRNA-TS heteroduplex is arranged against SpCas9-Y1013, generating a noncoaxially stacked duplex junction separated by SpCas9 (fig. S5 and fig. S6A). The ABE8e construct encodes a single TadA-8e domain (Fig. 1A). Cryo-EM density supported unambiguous modeling of secondary structure elements of a TadA-8e domain engaged with the 5′ end of the NTS emerging from SpCas9 (Fig. 1F and fig. S6B). We observed poorly resolved density below the aforementioned TadA-8e domain containing secondary structural features, consistent with a second copy of TadA-8e (Fig. 1F and fig. S6B). Further analysis revealed that the second TadA-8e domain was contributed in *trans* from another ABE8e molecule (fig. S7A), which is unsurprising given that the dimerization interface was unchanged between TadA-WT and TadA-8e (fig. S2). Although the linker region was unresolved, measured distances between Cas9 and each TadA-8e domain suggest that the *cis*-TadA-8e domain is engaged with the NTS (Fig. 1F and see the supplementary text, section 1).

The cryo-EM structure of the ABE8e complex revealed that the TadA-8e domain engages with the exposed single-stranded region of the PAM-distal NTS (Fig. 2A). Consistent with this, in vitro DNA adenosine deamination assays demonstrated that the ABE8e complex rapidly deaminates adenines present in accessible *trans*-ssDNA but not *trans*-dsDNA (Fig. 2B). Examining the TadA-8e dimer contacts revealed that one TadA-8e domain contacts the NTS but has no interface with SpCas9 (fig. S8A). By contrast, the second TadA-8e domain does not contact the NTS but forms a nonspecific interface with the SpCas9 RuvC domain (fig. S8B). These observations suggested that the presentation of single-stranded NTS by SpCas9 drives accelerated, TadA-8e–catalyzed *cis*-DNA deamination. To explore this idea, we compared the kinetics of adenosine deamination by TadA-8e of a guide-complementary dsDNA (NTS presented as *cis*-ssDNA) and noncomplementary *trans*-ssDNA (Fig. 2C and figs. S1, E and F, and S9A). The rate of *cis*-ssDNA deamination was ~3.7-fold faster relative to *trans*-ssDNA deamination, confirming that the presentation of the NTS by SpCas9 accelerates the kinetics of TadA-8e. Given that ABE8e was shown to edit more positions within an R-loop than previous iterations of ABEs (*15*), we hypothesized that the TadA-8e domain of ABE8e is a multiple-turnover enzyme. To test this hypothesis, we performed in vitro DNA deamination assays under multiple-turnover conditions (Fig. 2D and fig. S9B). The extracted turnover number for ssDNA in *trans* was 4.5 ± 0.1, in contrast to 0.97 ± 0.03 for *cis*-ssDNA. Although the observed single-turnover deamination kinetics on *cis*-DNA reflect the single-turnover kinetics of SpCas9 forming an R-loop (*17*), the rapid multiple-turnover kinetics are consistent with the higher processivity of ABE8e observed in cells relative to previous ABEs (*15*).

The rapid multiple-turnover *trans*-ssDNA deamination kinetics of ABE8e suggested that any adenosine in ssDNA, including those

[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. [2]Monash Biomedicine Discovery Institute, Department of Biochemistry & Molecular Biology, Monash University, Victoria 3800, Australia. [3]Department of Chemistry, University of California, Davis, CA 95616, USA. [4]Merkin Institute of Transformative Technologies in Healthcare, Broad Institute of Harvard and MIT, Cambridge, MA 02141, USA. [5]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. [6]Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138, USA. [7]Innovative Genomics Institute, University of California, Berkeley, CA 94720, USA. [8]MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [9]Department of Chemistry, University of California, Berkeley, CA 94720, USA. [10]Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA. [11]Gladstone Institute of Data Science and Biotechnology, Gladstone Institutes, San Francisco, CA 94158, USA.
*These authors contributed equally.
†Present address: School of Molecular Sciences, Arizona State University, Tempe, AZ 85281, USA.
‡Corresponding author. Email: doudna@berkeley.edu

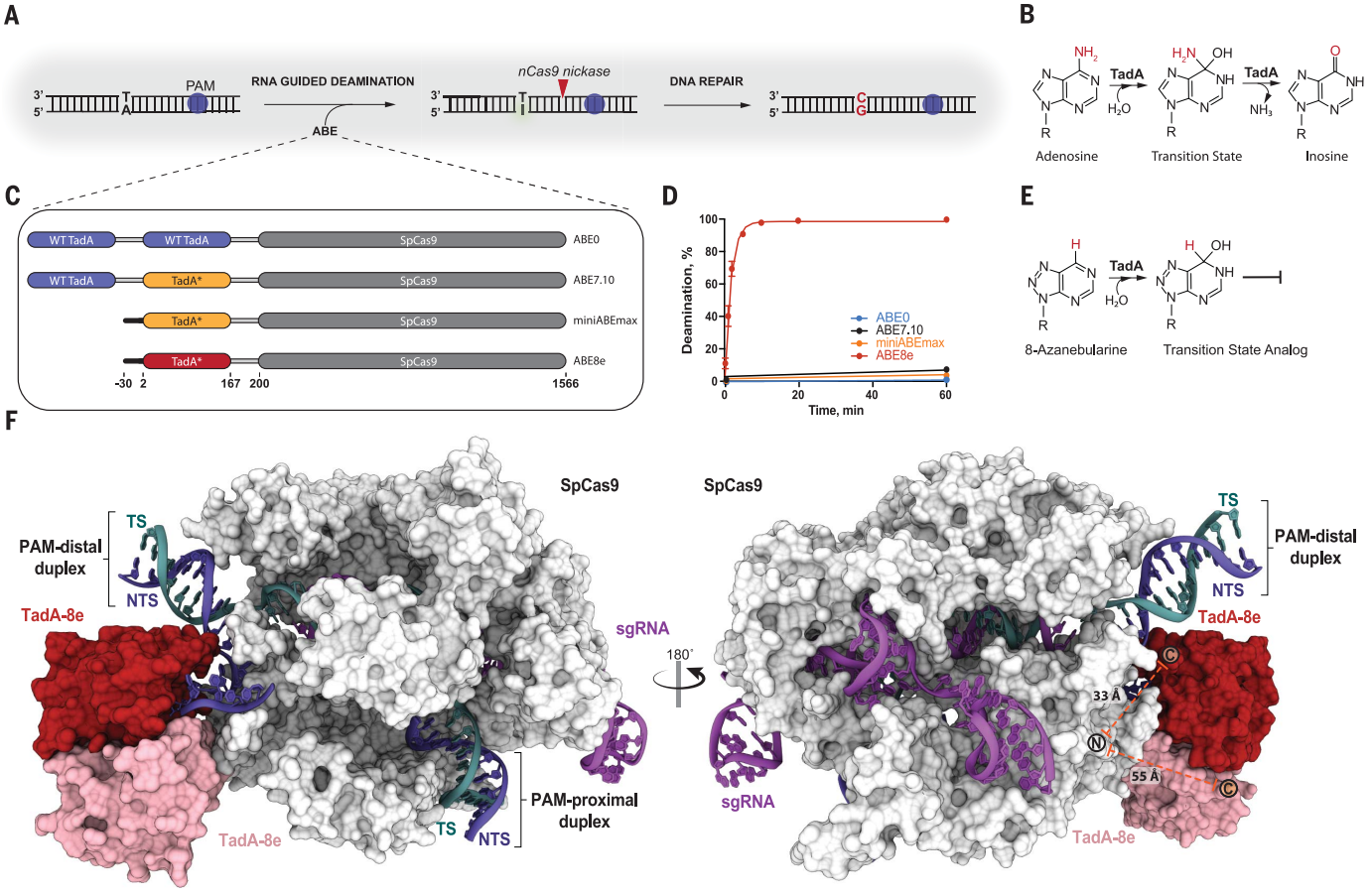


Fig. 1. Cryo-EM structure of ABE8e in a substrate-bound state. (**A**) Schematic representation of RNA-guided DNA adenosine deamination in cells. (**B**) Mechanism of *E. coli* TadA-catalyzed tRNA adenosine deamination. (**C**) Domain architecture of four generations of ABEs encoding WT TadA, evolved TadA7.10 (TadA*, yellow), or evolved TadA8e (TadA*, red) N-terminally linked to SpCas9. (**D**) Single-turnover dsDNA deamination kinetics of targeting ABE RNPs measured using dsDNA containing a single adenine. The fraction of deaminated dsDNA is plotted as a function of time and fitted to a single-phase exponential equation. The extracted apparent deamination rates of ABE7.10 (black), miniABEmax (orange), and ABE8e (red) are 0.0010 ± 3 × 10^{-4} min^{-1}, 0.0005 ± 1 × 10^{-4} min^{-1}, and 0.585 ± 0.034 min^{-1}, respectively. Data are represented as the mean ± SD from three independent experiments. The deamination data of ABE8e and ABE7.10 are reproduced and were originally published in (15). (**E**) Mechanism of inhibition of adenosine deamination by 8Az that mimics the adenosine deamination reaction intermediate. (**F**) The 3.2 Å resolution cryo-EM structure of the SpCas9-ABE8e complex. Subunits are colored as follows: SpCas9, gray; sgRNA, purple; TS, teal; NTS, blue; and TadA-8e dimer, red and pink. The direct distances between the Cas9 N-terminus and either TadA-8e C-termini are shown as dashed orange lines.

transiently exposed to solvent by Cas9, might be deaminated by TadA-8e. To explore this idea, we compared the in vitro editing characteristics of ABE7.10, miniABEmax, and ABE8e using dsDNA in which the NTS contained multiple adenines (Fig. 2E). Whereas ABE7.10 and miniABEmax predominantly deaminated a single PAM-distal adenine, ABE8e deaminated multiple adenines, including those present within double-stranded regions in the R-loop (Fig. 2E and fig. S10, A and B). Moreover, adenines were deaminated sequentially starting at the 3′ end of the NTS (Fig. 2E and fig. S10, A and B), reminiscent of the biased 3′-to-5′ diffusion of SpCas9 during search of a PAM along dsDNA (*18*). Thus, we hypothesized that the directional DNA adenosine deamination might be related to the SpCas9 ribonucleoprotein (RNP) target search mechanism.

Binding to an sgRNA remodels SpCas9 to allow target interrogation through a biased directional search along dsDNA (Fig. 3A) (*17*, *18*). Stable interaction with a PAM initiates probing for complementarity and stable R-loop formation on matched substrates to activate Cas9 for DNA cleavage (Fig. 3B) (*17*, *19*). With this model in mind, we tested whether directional adenosine deamination by ABE8e might be related to the Cas9 target search mechanism. Deamination assays were conducted using either radiolabeled ssDNA or dsDNA substrates and ABE8e in three different states: apo-ABE8e, ABE8e complexed with a targeting sgRNA, or ABE8e complexed with a nontargeting sgRNA (Fig. 3C). Apo-ABE8e was able to deaminate only ssDNA, whereas ABE8e RNP was able to modify adenines in ssDNA and dsDNA, in contrast to the absence of *trans*-dsDNA editing by ABE8e engaged in an R-loop complex (Fig. 2B). Moreover, the dsDNA deamination is sgRNA sequence independent because a nontargeting ABE8e RNP also deaminates dsDNA adenines (Fig. 2B), suggesting that stable R-loop formation is not required. To test whether Cas9's interaction with the PAM affects the observed dsDNA editing by ABE8e, we performed an in vitro DNA deamination assay with radiolabeled dsDNA devoid of consensus PAMs (Fig. 3D). To our surprise, ABE8e RNPs containing both the targeting and nontargeting sgRNA deaminated adenine in dsDNA lacking PAMs in the NTS (Fig. 3A). We speculate that during the PAM search, SpCas9 could transiently melt DNA, providing a ssDNA substrate for rapid deamination by the highly active TadA-8e domain. We note that under conditions tested in mammalian cells, ABE8e activity remains strongly guide RNA dependent with no detectable deamination of adenines flanking target protospacers (*15*), potentially because of differences in protein or DNA concentration or accessibility between in vitro and cellular conditions (see the

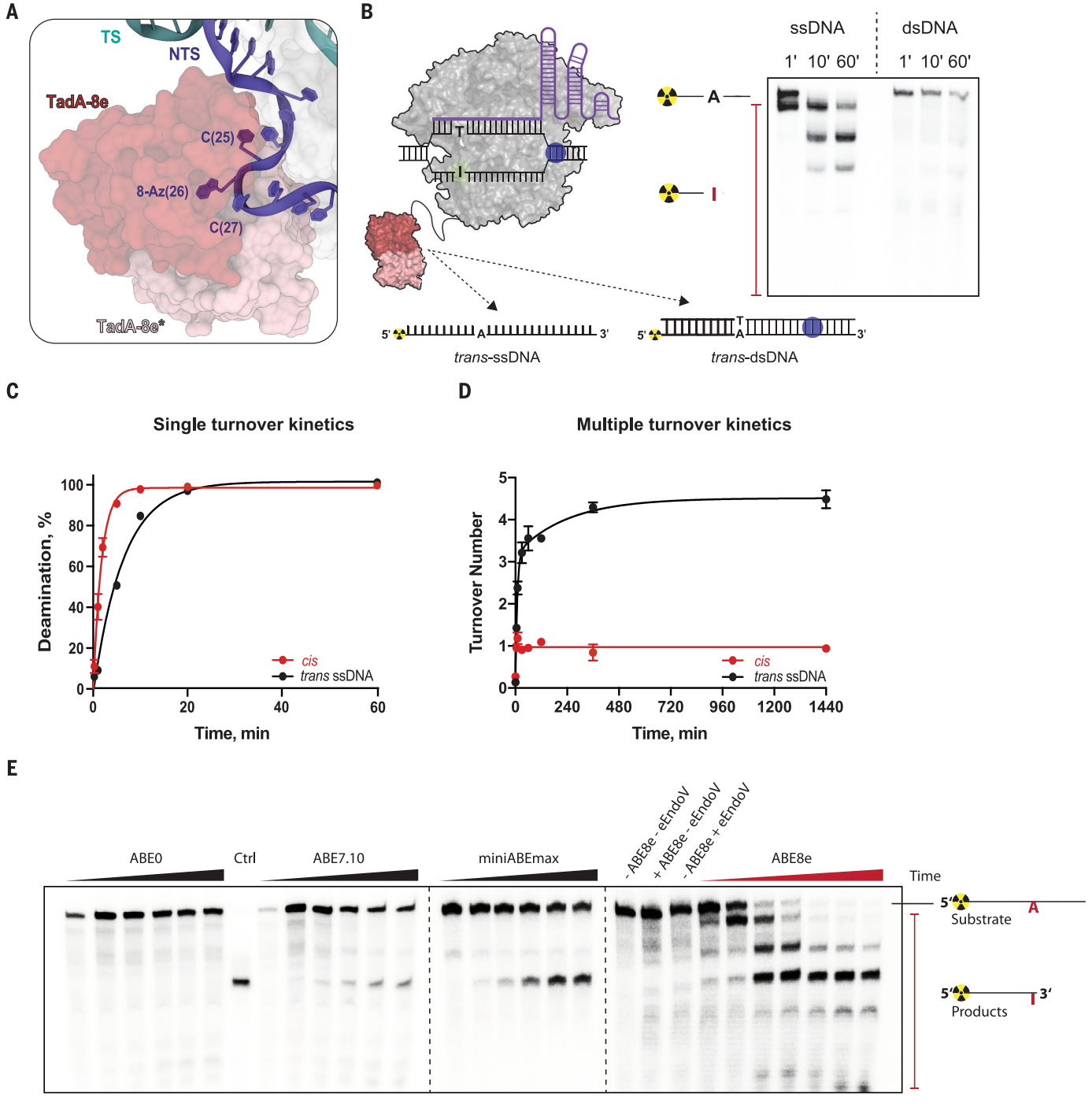


Fig. 2. ABE8e is a multiple-turnover enzyme with a wide editing window. (**A**) Single-stranded topology of the NTS (diagram, blue) engaged with the TadA8e domain (surface, red). (**B**) Schematic representation of ABE8e RNP R-loop complex acting on radiolabeled *trans*-ssDNA or *trans*-dsDNA substrate containing a single adenine. (Right) A representative gel showing ABE8e RNP in R-loop state single-turnover deamination of radiolabeled ssDNA and dsDNA over time (min) in *trans*. (**C**) Single-turnover kinetics of ABE8e RNP measured using either dsDNA (deamination in *cis*) or ssDNA (deamination in *trans*) containing a single adenine. The fraction of deaminated DNA plotted as a function of time and fitted to a single exponential equation. Data are represented as the mean ± SD from three independent experiments. The apparent rate of DNA editing in *trans* is ~3.7-fold lower than the apparent rate of DNA editing in *cis* (k_{app} = 0.16 ± 0.01 min^{-1} versus k_{app} = 0.59 ± 0.03 min^{-1}). (**D**) DNA deamination assay in multiple-turnover conditions using ABE8e RNP and either ssDNA (*trans*) or dsDNA (*cis*) substrate. Turnover number (a ratio of the deaminated DNA concentration and total concentration of ABE8e) was plotted as a function of time. The extracted turnover number for ssDNA deamination in *trans* is 4.5 ± 0.1, whereas for DNA in *cis*, it is 0.97 ± 0.03. (**E**) Gel representing single-turnover kinetics of ABE0, ABE7.10, miniABEmax, and ABE8e measured using dsDNA (*cis*) containing multiple adenines. Assays were performed in three independent replicates, and time points for the ABE0, ABE7.10, and miniABEmax assays were taken at 0, 1, 3, 8, 24, and 32 hours; for ABE8e, they were taken at 0, 1, 5, 10, 20, 60, and 180 min. Concentrations of ABE RNPs were 1 μM for single-turnover assays and 25 nM for multiple-turnover assays. Concentrations of DNA were 2.5 nM for single-turnover assays and 250 nM for multiple-turnover assays.

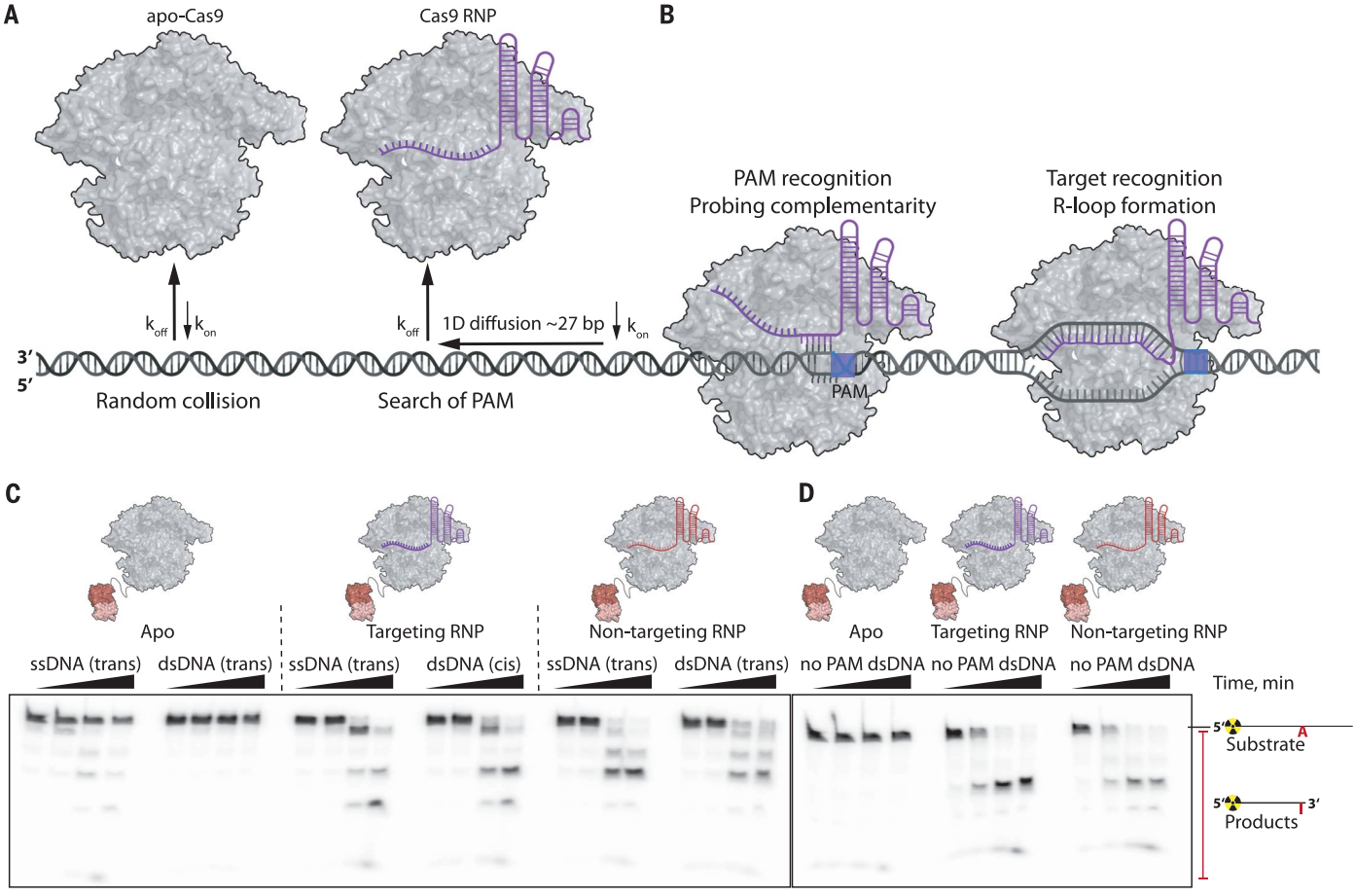


Fig. 3. Fast ABE8e kinetics reports on a new step in the SpCas9 target search pathway. Shown is a model for the SpCas9 target search mechanism. (**A**) The apo-Cas9–dsDNA interaction is a random collision; Cas9 bound to gRNA gets primed for a target search, which initiates through random interaction with the DNA, followed by unidirectional (NTS-3′ to 5′) Cas9 RNP diffusion along dsDNA (for ~27 bp) in search of PAM and rapid dissociation from DNA unless PAM is encountered. (**B**) Once Cas9 RNP comes across PAM, its residence time on DNA increases and it unwinds the first few nucleotides adjacent to PAM to probe for guide RNA complementarity. Cas9 RNP at correct target sites initiates an R-loop formation through sequential unwinding (*9*, *10*). (**C**) Gel representing single-turnover kinetics of apo-ABE8e acting on either ssDNA (*trans*) or dsDNA (*trans*), of ABE8e RNP programmed with targeting sgRNA deaminating ssDNA (*trans*) or dsDNA (*cis*), and of ABE8e RNP programmed with nontargeting sgRNA deaminating ssDNA (*trans*) or dsDNA (*trans*). Time points were taken at 0, 1, 10, and 60 min. (**D**) Gel representing single-turnover kinetics of apo-ABE8e, ABE8e-targeting RNP, and ABE8e-nontargeting RNP acting on dsDNA that lacks a PAM sequence. Time points were taken at 0, 1, 10, and 60 min. Concentrations of ABE RNPs were 1 μM and concentrations of DNA were 1 nM.

supplementary text, section 2). Future biophysical and single-cell whole-genome sequencing studies are necessary to explore the proposed mechanism and to identify potential unanticipated ABE8e activity.

Examining the overall topology of the TadA-8e homodimer relative to WT TadA revealed a substantial difference in the conformation of the C-terminal α-helix, α5 (Fig. 4A and fig. S2B). The α5-helix of TadA-8e undergoes a sharp 180° turn at P152 (Fig. 4A), a mutation introduced early during ABE evolution in addition to five other substitutions along the α5-helix of TadA-7.10 (fig. S2). To investigate the importance of the P152 substitution and the cumulative changes within the α5-helix of TadA-8e, we purified the single P152R variant and a stacked variant REIK (P152R, V155E, F156I, N157K), both of which displayed reduced rates of *cis*-DNA deamination relative to WT (Fig. 4B and fig. S11). These structural and biochemical observations demonstrate that cumulative substitutions within the α5-helix likely enable effective access to the NTS presented by SpCas9 within an R-loop.

Directed evolution of TadA-7.10 yielded the TadA-8e enzyme capable of rapid adenosine deamination in ssDNA. The ABE8e structure shows 8Az buried deep within the active-site pocket of TadA-8e, where it is oriented by the zinc-containing active site, consistent with previous reports for *E. coli* TadA and ADAR2 (*20*, *21*) (Fig. 4C). The nucleotides flanking the editing site, C(25) and C(27), are splayed against TadA-8e, generating a sharp U-turn geometry in the NTS, a conformation not unlike that observed in the native tRNA anticodon (Fig. 4C). Throughout ABE8e evolution, two residues were particularly conserved, R111 and Y149, both of which are oriented at the entrance to the active-site cleft and near the distorted NTS backbone flanking the editing site (Fig. 4C). To explore the role of R111 and Y149 in TadA-8e–catalyzed DNA deamination, we reverted these amino acids back to ABE7.10 (R111T, Y149F). Under single-turnover conditions, we observed that Y149F had a 7-fold decrease in the apparent rate of *cis*-DNA deamination, whereas the rate of the R111T mutant resembled that of ABE7.10 (Fig. 4B and fig. S11). Collectively, our structural and biochemical data support a role for R111 and Y149 in ssDNA substrate capture through stabilization of the U-turn conformation and induced fit into the deaminase active center.

WT TadA catalyzes site-specific deamination of tRNA by recognizing both the tRNA structure and the anticodon YAC motif (Fig. 4C) (*20*, *21*). A106V and L84F, a pair of substitutions introduced into ABE7.10 and preserved in ABE8e, provide steric bulk within the *E. coli*

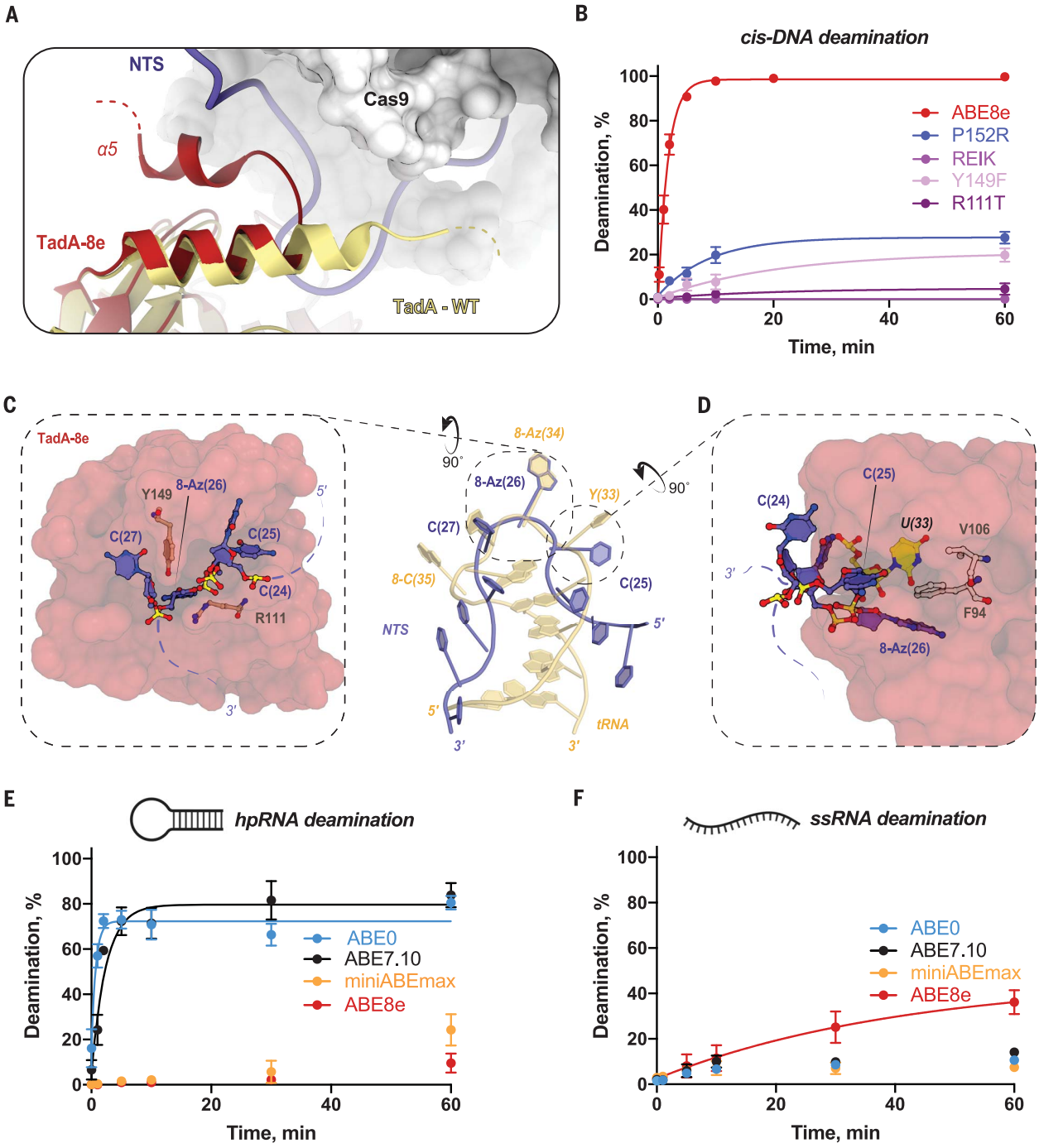


Fig. 4. Mechanism of ABE8e substrate specificity. (**A**) Superposition of TadA-8e (red) and TadA-WT (yellow) showing the altered disposition of the α5-helix. (**B**) Single-turnover kinetics of ABE8e and four ABE8e variants using dsDNA (containing a single adenosine in the NTS). The apparent deamination rates are: ABE8e 0.59 ± 0.04 min^{-1}; ABE8e-P152R 0.11 ± 0.02 min^{-1}; ABE8e-Y149F 0.05 ± 0.02 min^{-1}; ABE8e-R111T 0.05 ± 0.04 min^{-1} and not determined for ABE8e-REIK. Concentrations of ABE RNPs were 1 μM and concentrations of DNA were 2.5 nM. (**C**) TadA-8e active site showing the NTS (sticks, blue) entering the TadA-8e substrate-binding pocket (surface, red) with evolved residue R111 and Y149 shown as sticks (brown). (**D**) TadA-8e active-site pocket (surface, red) with evolved residues F84 and V106 shown as sticks (white). The *E. coli* tRNA U(33) is shown in yellow. (**E**) Single-turnover RNA deamination kinetics of all four ABEs programmed with targeting sgRNA (table S2) using hpRNA as a substrate that contains a single adenine. The fraction of deaminated hpRNA is plotted as a function of time and fitted to a single exponential equation. The extracted apparent deamination rates of ABE0 (blue) and ABE7.10 (black) are 1.38 ± 0.28 min^{-1} and 0.41 ± 0.08 min^{-1}, respectively. The kinetics of miniABEmax and ABE8e were much slower and nonexponential. Data are represented as the mean ± SD from three independent experiments. Concentrations of ABE RNPs were 1 μM and concentrations of RNA were 1 nM. (**F**) Single-turnover kinetics of all four ABEs programmed with targeting sgRNA (table S2) using ssRNA as a substrate that contains a single adenine. The fraction of deaminated ssRNA is plotted as a function of time and fitted to a single exponential equation. The extracted apparent deamination rate of ABE8e (red) is 0.02 ± 0.01 min^{-1}. The kinetics of ABE0, ABE7.10, and miniABEmax were much slower and nonexponential. Data are represented as the mean ± SD from three independent experiments. Concentrations of ABE RNPs were 1 μM and concentrations of RNA were 1 nM.

tRNA U(33)-binding pocket (Fig. 4D). Coupled with the loss of the α5-helix that inserts into the tRNA major groove (*20*), the steric clash at *E. coli* tRNA U(33) could prevent tRNA deamination by ABE8e. Consistent with this structural observation, the kinetics of in vitro hairpin RNA (hpRNA) adenosine deamination revealed that miniABEmax (TadA-7.10) had a substantially reduced rate of hpRNA deamination relative to ABE0 (WT TadA – WT TadA) and ABE7.10 (WT TadA – TadA-7.10) (Fig. 4E and fig. S12, A and D). The kinetics of ssRNA deamination demonstrated that ABE7.10 and miniABEmax are as slow and inefficient as WT TadA (Fig. 4F and fig. S12, A and E), consistent with data showing that miniABEmax has reduced RNA off-target editing in cells (*14*). By contrast, the additional eight substitutions in ABE8e (relative to miniABEmax) do not affect hpRNA editing but induce a two-fold faster ssRNA deamination rate (Fig. 4, E and F, and fig. S12, D and E). These data suggest that the active site of TadA-8e has evolved to capture single-stranded nucleic acids of any sequence and to exclude more rigid tRNA substrates.

Our results demonstrate how the laboratory-evolved base editor ABE8e achieves rapid deamination on DNA. Because of substitutions near the active center, TadA-8e effectively remodels and induces fit of structurally flexible ssDNA into a narrow substrate-binding pocket. The structure of ABE8e also demonstrated that the TadA-8e domain (and likely that of Tad-7.10) does not specifically interact with SpCas9, suggesting that the broader Cas-effector compatibility of TadA-8e (*15*) stems from the rapid deamination kinetics. Furthermore, ABE8e's accelerated DNA deamination suggested a previously unobserved transient melting of DNA that may occur during dsDNA surveillance by SpCas9. Site-specific base editors are possible by constraining the deaminase domain within the Cas9 R-loop structure to enable transmission of Cas9 conformational changes upon interaction with on-target DNA.

REFERENCES AND NOTES

1. M. J. Landrum *et al.*, *Nucleic Acids Res.* **44**, D862–D868 (2016).
2. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, *Nature* **533**, 420–424 (2016).
3. N. M. Gaudelli *et al.*, *Nature* **551**, 464–471 (2017).
4. J. Wolf, A. P. Gerber, W. Keller, *EMBO J.* **21**, 3841–3851 (2002).
5. J. Kim *et al.*, *Biochemistry* **45**, 6407–6416 (2006).
6. S.-M. Ryu *et al.*, *Nat. Biotechnol.* **36**, 536–539 (2018).
7. W. Qin *et al.*, *BMC Biol.* **16**, 139 (2018).
8. P. Liang *et al.*, *Protein Cell* **9**, 808–813 (2018).
9. Y. Ma *et al.*, *Cell Discov.* **4**, 39 (2018).
10. L. Yang *et al.*, *Protein Cell* **10**, 700 (2019).
11. C.-Q. Song *et al.*, *Nat. Biomed. Eng.* **4**, 125–130 (2020).
12. J. M. Levy *et al.*, *Nat. Biomed. Eng.* **4**, 97–110 (2020).
13. H. A. Rees, C. Wilson, J. L. Doman, D. R. Liu, *Sci. Adv.* **5**, eaax5717 (2019).
14. J. Grünewald *et al.*, *Nat. Biotechnol.* **37**, 1041–1048 (2019).
15. M. F. Richter *et al.*, *Nat. Biotechnol.* 10.1038/s41587-020-0453-z (2020).
16. B. L. Haudenschild *et al.*, *J. Am. Chem. Soc.* **126**, 11213–11219 (2004).
17. S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, J. A. Doudna, *Nature* **507**, 62–67 (2014).
18. M. Yang, L. Zhang, R. Sun, W. Zhong, Y. Yang, C. Chen, *Streptococcus pyogenes* Cas9 displays biased one-dimensional diffusion on dsDNA to search for a target. bioRxiv 754564 [Preprint]. 4 September 2019. https://doi.org/10.1101/754564.
19. S. H. Sternberg, B. LaFrance, M. Kaplan, J. A. Doudna, *Nature* **527**, 110–113 (2015).
20. H. C. Losey, A. J. Ruthenburg, G. L. Verdine, *Nat. Struct. Mol. Biol.* **13**, 153–159 (2006).
21. M. M. Matthews *et al.*, *Nat. Struct. Mol. Biol.* **23**, 426–433 (2016).

ACKNOWLEDGMENTS

We thank N. Orlova, J. Cofsky, J.-J. Liu, K. Soczek, S. Eitzinger, C. Huang, J. Grunewald, K. J. Joung, and members of the Doudna laboratory for comments and discussions; A. T. Iavarone from QB3/Chemistry Mass Spectrometry Facility [National Institutes of Health (NIH) grant no. 1S10OD020062-01] at UC Berkeley for native mass spectrometry of ABE8e; D. B. Toso for expert electron microscopy assistance; and P. Tobias for computational support. EM data were collected at the Bay Area Cryo-EM (BACEM) facility located at UC Berkeley. **Funding:** Research reported in this publication was supported by the Centers for Excellence in Genomic Science of the National Institutes of Health under award no. RM1HG009490 and the Somatic Cell Genome Editing Program of the Office of the Director, National Institutes of Health (OD), under award no. U01AI142817-02. J.A.D. receives funding from the William M. Keck Foundation, the National Multiple Sclerosis Society, the Paul Allen Frontiers Group, and the Howard Hughes Medical Institute. This material is based upon work supported by the National Science Foundation under award no. 1817593. A.L. was supported by the Centers for Excellence in Genomic Science of the National Institutes of Health under award no. RM1HG009490. G.J.K. was supported by an NHMRC Investigator Grant (EL1, 1175568) and previously by an American Australian Association Fellowship. E.L.-S. was supported by the Paul G. Allen Frontiers Group. P.A.B. was supported by NIH grant no. R01GM061115. C.M.P. was supported by NIH Predoctoral Training Grant no. T32GM113770. D.R.L. was supported by NIH grant nos. U01 AI142756, RM1 HG009490, R01 EB022376, and R35 GM118062; the St. Jude Collaborative Research Consortium; the Bill and Melinda Gates Foundation; and the Howard Hughes Medical Institute. M.F.R. was supported by an HHMI Hanna Gray Fellowship. K.T.Z. was supported by Harvard Chemical Biology Training Grant no. T32 GM095450. **Author contributions:** A.L. and G.J.K. conceived of and designed experiments with input from C.M.P., M.F.R., K.T.Z., P.A.B., D.R.L., and J.A.D. A.L. performed protein expression, purification, and biochemical experiments. G.J.K. performed cryo-EM data collection, processing, and model building with assistance from A.L. C.M.P. synthesized 8Az-modified substrates. E.L.-S. cloned ABE8e constructs containing various amino acid substitutions. M.F.R. and K.T.Z. provided the evolved ABE8e construct. A.L., G.J.K., and J.A.D. wrote the manuscript and all authors edited the manuscript. **Competing interests:** The Regents of the University of California have patents issued and pending for CRISPR technologies on which J.A.D. and G.J.K. are inventors. J.A.D. is a cofounder of Caribou Biosciences, Editas Medicine, Scribe Therapeutics, and Mammoth Biosciences. J.A.D. is a scientific advisory board member of Caribou Biosciences, Intellia Therapeutics, eFFECTOR Therapeutics, Scribe Therapeutics, Mammoth Biosciences, Synthego, Algen Biotechnologies, Felix Biosciences, and Inari. J.A.D. is a Director at Johnson & Johnson and has research projects sponsored by Biogen, Pfizer, AppleTree Partners, and Roche. P.A.B. is a consultant with stock options in Beam Therapeutics. D.R.L. is a consultant and cofounder of Editas Medicine, Pairwise Plants, Beam Therapeutics, and Prime Medicine, companies that use genome editing. The authors have filed patent applications on evolved ABEs. **Data availability:** The cryo-EM map of the ABE8e-sgRNA-dsDNA complex at 3.2-Å resolution and model coordinates have been deposited to the EMDB and PDB databases with accession codes EMD-21308 and 6VPC. All other data generated or analyzed during this study are included in the manuscript or supporting files. Plasmids are available from Addgene.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/566/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Tables S1 and S2
References (*22*–*42*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

3 February 2020; accepted 27 May 2020
10.1126/science.abb1390

NANOMATERIALS

Rational synthesis of atomically precise graphene nanoribbons directly on metal oxide surfaces

Marek Kolmer[1], Ann-Kristin Steiner[2], Irena Izydorczyk[3], Wonhee Ko[1,4], Mads Engelund[5], Marek Szymonski[3], An-Ping Li[1,4]*, Konstantin Amsharov[2,6]*

Atomically precise graphene nanoribbons (GNRs) attract great interest because of their highly tunable electronic, optical, and transport properties. However, on-surface synthesis of GNRs is typically based on metal surface–assisted chemical reactions, where metallic substrates strongly screen their designer electronic properties and limit further applications. Here, we present an on-surface synthesis approach to forming atomically precise GNRs directly on semiconducting metal oxide surfaces. The thermally triggered multistep transformations preprogrammed in our precursors' design rely on highly selective and sequential activations of carbon-bromine (C-Br) and carbon-fluorine (C-F) bonds and cyclodehydrogenation. The formation of planar armchair GNRs terminated by well-defined zigzag ends is confirmed by scanning tunneling microscopy and spectroscopy, which also reveal weak interaction between GNRs and the rutile titanium dioxide substrate.

Atomically precise graphene nanoribbons (GNRs) (*1*, *2*) can host nontrivial electronic structures (*3*–*5*), magnetic edge states (*6*–*8*), and coupled spin centers (*9*–*11*). These properties make GNRs promising candidates as functional building blocks for applications in emerging quantum information science (QIS) (*12*). The rational synthesis of GNRs has been achieved by designing organic molecules to undergo surface-catalyzed reactions under ultrahigh vacuum (UHV) conditions after deposition on a crystalline substrate (*13*, *14*). The commonly used strategies are based on intramolecular carbon-carbon (C-C) bond formations leading to annulation through thermally triggered cyclodehydrogenation reactions (*15*), which are preceded by intermolecular C-C coupling (polycondensation of precursor units; see Fig. 1A).

A two-step approach reported by Cai *et al.* (*16*) works reliably on noble metal surfaces. However, the high density of low-energy electronic states inherent to a metal substrate does not allow for sufficient electronic decoupling of the synthesized GNRs, which necessitates a transfer onto semiconducting or insulating substrates for detailed characterization or prototypical device fabrication. So far, these transfer processes have been performed locally by scanning tunneling microscopy (STM) tip–induced manipulation of single molecules (*6*, *17*) or globally either by in situ intercalation of a decoupling layer (*18*) or ex situ by using wet-chemistry processing (*19*–*21*). For spintronic and QIS applications, which rely on atomically defined nanostructures and their subtle interactions, establishing reliable on-surface chemistry methods for nonmetallic substrates is a desirable solution.

In this regard, the versatile chemical activity of transition-metal oxides that finds use in heterogeneous catalysis is attractive. In particular, aryl halide precursors undergo surface-assisted polycondensation on rutile TiO_2 surfaces through thermal activation of carbon-bromine (C-Br) or carbon-iodine (C-I) bonds (Fig. 1A) (*22*–*24*). However, for intramolecular annulation, a lack of catalytic activity of the metal oxide surface demands alternatives for the cyclodehydrogenation reaction (*22*). Thus, direct synthesis of planar GNRs by mimicking the two-step triggering strategy used for metal substrates could not be accomplished on nonmetallic substrates. Previously, we demonstrated that intramolecular aryl-aryl coupling in polycyclic aromatic hydrocarbon precursors can be achieved by the activation of carbon-fluorine (C-F) bonds on various metal oxides (*25*–*29*), including the TiO_2 surface (*30*). The high efficiency of this approach enabled on-surface synthesis of two confined nanographenes through thermally initiated sequential cyclodehydrofluorination reactions (Fig. 1B).

However, the formation of extended sp^2 carbon systems, like GNRs, by combining the reported polycondensation and cyclization processes remains an open challenge. Here, we report a methodology for rational precursor design and direct synthesis of atomically precise GNRs on the rutile TiO_2(011)-(2 × 1) surface. The multistep on-surface synthesis programmed in our precursors consisted of sequential polycondensation, cyclodehydrofluorination, and cyclodehydrogenation reactions. These thermally triggered processes led to formation of finite-length seven-armchair graphene nanoribbons (7a-GNRs) terminated by well-defined zigzag ends (*31*). Such finite 7a-GNRs are open-shell character molecules with a single electron localized at each end (*7*, *32*). Our scanning tunneling microscopy and spectroscopy (STM/S) analysis of this model system revealed weak screening of the intrinsically magnetic end states by substrate electrons, which is a prerequisite for spintronic applications.

To realize the on-surface synthesis strategy, we used 10,10''-dibromo-1',4'-difluoro-9,9':10',9''-teranthracene (DBDFTA, **1**; see details of organic synthesis and characterization in supplementary materials) presented in Fig. 1C. The rational design of (**1**) was imbued with our conception of controlled consecutive reactions.

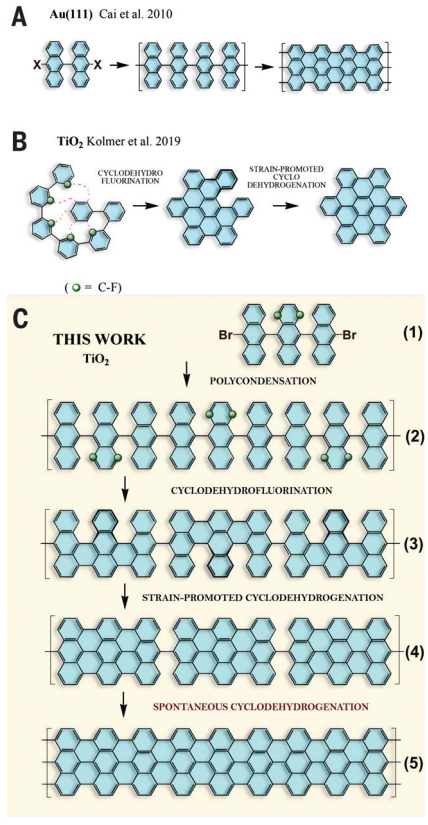


Fig. 1. On-surface synthesis strategies for nanographene formation. (**A**) Scheme of the most common two-step strategy of graphene nanoribbons formation introduced by Cai *et al.* on noble metal substrates (*16*). X represents halogen atom (typically Br or I). (**B** and **C**) Multistep metal oxide surface-assisted strategies for formation of compact nanographenes [(B), from (*30*)] and extended graphene nanoribbons [(C), this work]. C-F bonds are shown as green circles for clarity.

[1]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [2]Department of Organic Chemistry, Friedrich Alexander University Erlangen-Nuremberg, Nikolaus-Fiebiger Strasse 10, 91058 Erlangen, Germany. [3]Centre for Nanometer-Scale Science and Advanced Materials, NANOSAM, Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Łojasiewicza 11, 30-348 Kraków, Poland. [4]Department of Physics and Astronomy, The University of Tennessee, Knoxville, TN 37996, USA. [5]Espeem S.A.R.L. (espeem.com), 9 rue de Haut-Fournaux, L-4365 Esch-sur-Alzette, Luxembourg. [6]Institute of Chemistry, Organic Chemistry, Martin-Luther-University Halle-Wittenberg, Kurt-Mothes-Strasse 2, 06120 Halle, Germany.
*Corresponding author. Email: apli@ornl.gov (A.-P.L.); konstantin.amsharov@fau.de (K.A.)

As shown previously (*22*), terminal bromine-atom functionalization ensured the polycondensation step, which was facilitated by nonplanar geometry of the teranthracene chassis. Such precursor conformation weakened interactions with the substrate to allow sufficient precursor mobility and minimized steric hindrance during intermolecular C-C bond formation. As the intermolecular coupling reactions on TiO_2 can be optimized (*24*) or achieved by other approaches (*33*), our main focus was on the effective transformation of resulting oligoanthracenes into planar GNRs.

Following our previous reports, we initiated this process by metal oxide surface–induced C-F bonds activation (*30*). During rational design of precursors, the substrate-activated C-F bond must be paired with an electron-rich π-system to undergo effective cyclodehydrofluorination (*34*). For larger π-extended systems, the cyclization reaction proceeded more effectively (*35*). In the case of compact nanographenes, this effect could be used to implement self-accelerating, consecutive reactions with each subsequent cyclization step proceeding faster because of the increased size of the corresponding π-system involved (Fig. 1B). However, the presence of electron-withdrawing F atoms in a given aryl group reduced reactivity. These mechanistic effects put some limitations on the design of precursors suitable for rational synthesis of extended sp^2 carbon systems. Additionally, steric hindrance in access to surface atoms within an extended molecular system may cause difficulties in activation of C-F bonds. Our rational design of precursor aimed to address all of the above considerations.

The cyclization of oligomers (**2**) was initiated through implementation of two F atoms placed in the same central anthracene unit of the precursor (Fig. 1C). Each C-F bond had a separate electron-rich anthracene system required for HF elimination, and they occupied para positions within the same central anthracene block that provided the required configuration for effective cyclization. Moreover, because of the free rotation of single bonds between anthracene units, activation of all C-F bonds by surface Ti atoms was not limited by geometric constraints. Successful cyclodehydroflourination reactions were then followed by a series of programmed cyclodehydrogenations. In this case, we used structural-strain engineering within corresponding GNR intermediates, which, combined with lack of TiO_2 catalytic activity for H_2 elimination, provided high selectivity in the formation of intramolecular C-C bonds.

As presented in Fig. 1C, intermediates (**3**) possessed highly strained double helicene-like structures in single oligomer units, which facilitated closure of these two fjord regions (*30*). These transformations finally led to formation

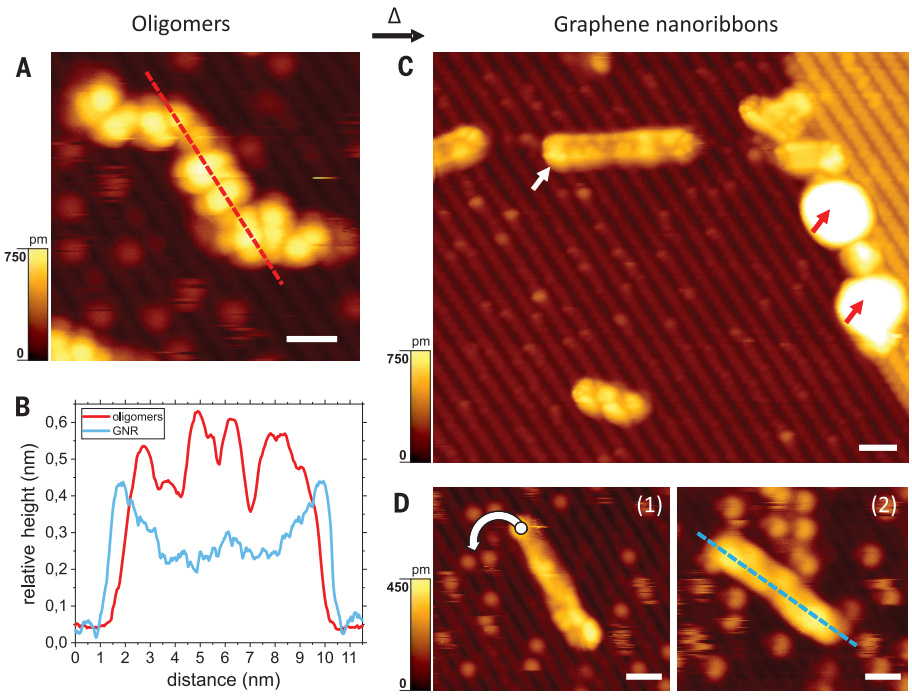


Fig. 2. Transformation of oligoanthracenes into graphene nanoribbons. (**A**) STM image (+3 V; 2 pA) of oligoanthracenes obtained after DBDFTA deposition on the rutile (011)-(2 × 1) kept at 470 K. (**B**) Cross sections along dashed lines in (A) (red) and (D) (blue) obtained in comparable STM imaging conditions. (**C**) STM image (+2.5 V; 10 pA) of seven-armchair graphene nanoribbon (7a-GNR) obtained after DBDFTA deposition on the rutile (011)-(2 × 1) m at 470 K followed by annealing at 670 K. White arrow marks about 9-nm-long GNR adsorbed on a domain boundary. Observed defects are related to polycondensation reaction by-products (red arrows). (**D**) Two sequential STM images (1; +2 V, 5 pA) and (2; +3 V; 2 pA) of the same 7a-GNR on a flat terrace. GNR is rotated by an STM bias voltage pulse of −3 V applied in the location marked by the white circle in (1). The initial adsorption site observed in (2) is covered by surface hydroxyl groups. Scale bars in all STM images represent 2 nm.

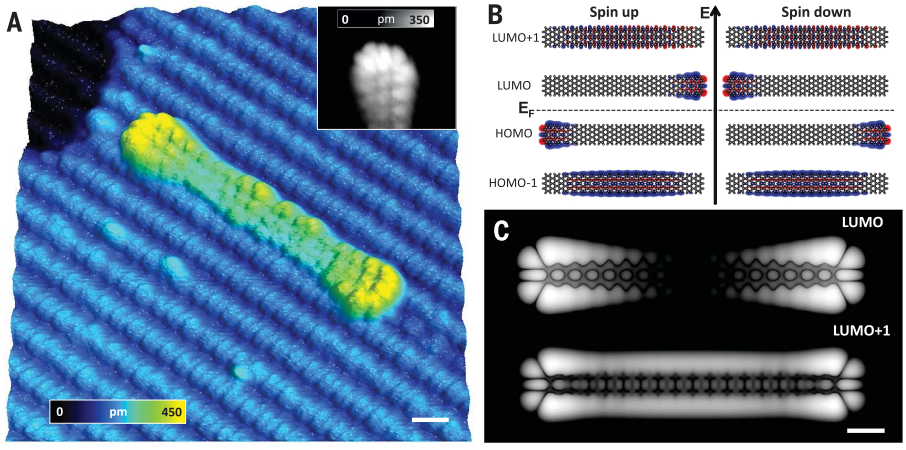


Fig. 3. Spin-polarized end states in finite 7a-GNRs. (**A**) Three-dimensional visualization of a high-resolution STM image (+0.9 V; 2 pA) presenting finite 7a-GNR synthesized from six DBDFTA precursors with clearly symmetric edge electronic states (inset). GNR is adsorbed on a domain boundary between two reconstructed rutile (011)-(2 × 1) terraces (blue color). Black region in the corner is the lower-laying terrace. (**B**) Real-space visualization of spin-polarized electronic wave functions for finite 7a-GNR synthesized from six DBDFTA precursors and terminated by perfect ZZ edges. Two occupied and unoccupied electronic states in single-electron molecular orbitals correspond to HOMO+1, HOMO, LUMO, and LUMO+1, respectively. Symmetric HOMO and LUMO states represent in this case spin-polarized zigzag edge states of 7a-GNR. (**C**) STM image simulation obtained for a bias voltage corresponding to the first and second excited unoccupied state (LUMO and LUMO+1). Scale bars in (A) and (C) represent 1 nm.

of an oligomer composed of teranthene units (Fig. 1C, intermediate **4**). As a result of the high biradical character of teranthene fragments (y = 42%) (*36*), we expected that further cyclodehydrogenation could violate the Woodward-Hoffmann rule given the low activation barrier (*37*–*39*). Thus, the resulting intermediate (**4**) should form target 7a-GNRs through a formally symmetry-forbidden thermal electrocyclization process.

On-surface synthesis procedures and low-temperature STM characterization were performed in situ under UHV conditions (for details, see supplementary materials) to evaluate the reaction products. The rutile TiO_2 monocrystals exposing (011) face were prepared by several sputtering and annealing cycles that led to a high-quality (2 × 1) reconstructed surface, as confirmed by low-energy electron diffraction (LEED) and STM imaging (see fig. S1, a and b). To induce the polymerization step, we deposited submonolayer coverages of DBDFTA molecules using a standard Knudsen cell on rutile (011) substrate kept at an elevated temperature of ~470 K (*22*). Subsequent STM characterization revealed short oligoanthracene chains present on the reconstructed terraces as depicted in Fig. 2A and fig. S1d. Nonuniform STM contrast and apparent STM profile height of about 0.5 nm (Fig. 2B, red line) show flexibility between single anthracene units that gives rise to nonplanar geometry of the molecules.

By following our previously reported conditions for triggering C-F bond activation processes (*30*), we annealed the substrate at ~670 K. Corresponding large-scale STM images obtained after this step show straight and elongated features located mainly on chemically active sites such as terrace step edges or domain boundaries (Fig. 2C and fig. S2), which suggests successful on-surface synthesis of 7a-GNRs. To further verify this claim, we focused on single molecules on flat terraces. As shown in Fig. 2D, the rigid sp^2 carbon-based structure can be confirmed in this case by STM tip–induced manipulation (*40*). Application of a negative-bias voltage pulse resulted in rotation of the whole molecule without distorting its straight geometry. Moreover, this rotated molecule became flatter, as plotted in Fig. 2B (blue line), with ~50% reduction of the apparent height in comparison to oligoanthracene intermediate (**2**). The height contrast was consequently only dominated by both ends, proving the completed planarization of GNR.

The manipulation event also exposed the native adsorption site of the GNR molecule, which was initially immobilized by surface hydroxyl groups. The presence of hydroxyl groups beneath newly formed GNRs (other examples in fig. S3) was consistent with the previously reported manifestation of HF-elimination reaction on the rutile TiO_2(011)

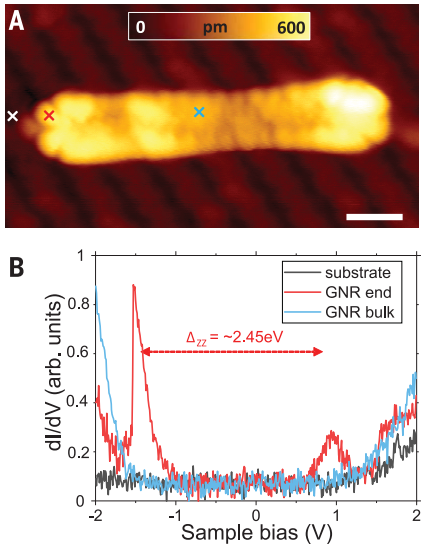


Fig. 4. Electronic structure characterization. (**A**) High-resolution STM image (+1.5 V, 10 pA) of 7a-GNR synthesized on the rutile (011). The molecule is adsorbed on a domain boundary between two reconstructed rutile (011)-(2 × 1) terraces (see also Fig. 2B). Scale bar represents 1 nm. (**B**) Scanning tunneling spectroscopy *dI/dV* data obtained in locations marked by three crosses in (A). Two resonances marked by red arrow at about −1.45 and +1.0 eV correspond to occupied and unoccupied zigzag end state, respectively.

(*30*). Our postulated GNR synthesis pathway shown in Fig. 1C was additionally confirmed in observation of the corresponding intermediates (**4**), which have not completed the last cyclodehydrogenation steps (figs. S3 and S4).

Detailed analysis of end states observed in most of the on-surface–synthesized molecules could provide more insight into the exact morphology and electronic properties of finite 7a-GNRs on the rutile TiO_2 substrate. A low-bias STM image (+0.9 V, Fig. 3A) showed a finite 7a-GNR composed of six precursor units adsorbed on a domain boundary between two (2 × 1) reconstructed terraces. Under these imaging conditions corresponding to unoccupied states around the rutile TiO_2 conduction band edge (CBE), both GNR ends exhibited characteristic three lobe features (see Fig. 3A, inset). This observation agreed with the expected symmetry of the lowest unoccupied molecular orbital (LUMO) for finite 7a-GNR terminated by two hydrogen-passivated zigzag ends, as captured by our spin-dependent density functional theory calculations (Fig. 3B) and corresponding unoccupied-states STM image simulations (Fig. 3C). Stability of zigzag ends was also consistent with previous reports on 7a-GNRs synthesized on metal substrates (*7*, *17*, *41*).

Because the local electronic density of states related to LUMO resonance decayed in the central part of the GNR (see Fig. 3, B and C), the experimentally obtained STM image contrast was in this case determined by nonresonant tunneling through the molecule to the underlying substrate. Thus, it presented additional corrugation coming from the domain boundary structure. The good agreement between our STM results and gas-phase STM image simulations from Fig. 3C confirmed not only successful realization of the atomically precise GNR synthesis but also the weak coupling between obtained GNR and the rutile TiO_2 substrate.

Moreover, the calculated electronic structure presented in Fig. 3B revealed two spatially and energetically localized zigzag end states consisting of spin-polarized wave functions. Thus, the electronic ground state can be seen as two single-electron occupied molecular orbitals reflecting the open-shell character of the finite-length 7a-GNR molecules. The antiferromagnetically coupled end states offers a model system for studying nonlocal spin entanglement, as proposed previously (*32*). However, with increasing length of GNRs, the corresponding open-shell singlet to triplet excitation energy exponentially decayed to zero, and our examples of 7a-GNRs at 77 K became paramagnetic.

Characterization of such magnetic zigzag end states requires sufficient electronic decoupling from the substrate, which has been difficult to realize on reactive noble metals (*42*). Wang *et al.* (*7*) used insulating sodium chloride layers on Au(111) to decouple finite 7a-GNRs. Their systematic STS characterization showed that spectroscopic resonances of occupied and unoccupied end-localized states were split by an energy gap of Δ_{zz} = 1.9 eV. Because of intramolecular electron-electron correlation effects, self-consistent many-body perturbation theory calculations predict for free-standing finite 7a-GNRs Δ_{zz} of about 2.8 eV (*7*). The ~0.9 eV difference in experimentally detected and calculated fundamental gaps was in that case explained by additional screening coming from electrons of the NaCl/Au(111) substrate.

Figure 4 shows our STS differential conductance *dI/dV* data obtained on finite 7a-GNRs adsorbed on the rutile TiO_2 (see also fig. S5). Two zigzag end resonances (red curve) were localized within the energy gap of bulk armchair GNR states (blue curve), in excellent agreement with previous reports (*7*). The measured quasiparticle gap Δ_{zz} = 2.45 ± 0.10 eV was near the values expected for the free-standing system, indicating minimal interaction with the substrate electrons. A strong intensity asymmetry between registered resonances came from the difference in their relative energy positions with respect to the rutile TiO_2 electronic structure. An unoccupied edge state at +0.9 V was located near the CBE

of the substrate. By contrast, an occupied edge state at −1.45 V fell into a wide TiO_2 bandgap that caused a sharp, Lorentzian-like shape of the corresponding *dI/dV* resonance.

Our results show that the magnetic ground state of the synthetized model open-shell GNRs was electronically decoupled from the substrate, which opens future possibilities for probing their intrinsic transport properties (*43*, *44*). Moreover, recent developments of scanning probe microscopes (SPMs) compatible with the microwave transmission have demonstrated detection of single-electron spin resonances (*45*, *46*). Application of these techniques to decoupled spin centers in graphitic molecular systems would enable the complete processing of quantum information including qubit initialization, logic gate operation, and readout (*10*, *47*). Finally, atom-manipulation protocols and multiprobe SPMs could further advance these efforts toward a more complex quantum circuitry (*40*).

REFERENCES AND NOTES

1. L. Talirz, P. Ruffieux, R. Fasel, *Adv. Mater.* **28**, 6222–6231 (2016).
2. K. Müllen, X. Feng, *From Polyphenylenes to Nanographenes and Graphene Nanoribbons*, Advances in Polymer Science (Springer International, 2017).
3. T. Cao, F. Zhao, S. G. Louie, *Phys. Rev. Lett.* **119**, 076401 (2017).
4. O. Gröning *et al*., *Nature* **560**, 209–213 (2018).
5. D. J. Rizzo *et al*., *Nature* **560**, 204–208 (2018).
6. P. Ruffieux *et al*., *Nature* **531**, 489–492 (2016).
7. S. Wang *et al*., *Nat. Commun.* **7**, 11507 (2016).
8. S. Mishra *et al*., *J. Am. Chem. Soc.* **142**, 1147–1152 (2020).
9. J. Li *et al*., *Nat. Commun.* **10**, 200 (2019).
10. F. Lombardi *et al*., *Science* **366**, 1107–1110 (2019).
11. S. Mishra *et al*., *Nat. Nanotechnol.* **15**, 22–28 (2020).
12. S. von Kugelgen, D. E. Freedman, *Science* **366**, 1070–1071 (2019).
13. S. Clair, D. G. de Oteyza, *Chem. Rev.* **119**, 4717–4776 (2019).
14. X. Zhou, G. Yu, *Adv. Mater.* **32**, e1905957, 1905957 (2019).
15. M. Treier *et al*., *Nat. Chem.* **3**, 61–67 (2011).
16. J. Cai *et al*., *Nature* **466**, 470–473 (2010).
17. P. H. Jacobse, M. J. J. Mangnus, S. J. M. Zevenhuizen, I. Swart, *ACS Nano* **12**, 7048–7056 (2018).
18. O. Deniz *et al*., *Nano Lett.* **17**, 2197–2203 (2017).
19. J. P. Llinas *et al*., *Nat. Commun.* **8**, 633 (2017).
20. G. Borin Barin *et al*., *ACS Appl. Nano Mater.* **2**, 2184–2192 (2019).
21. A. Radocea *et al*., *Nano Lett.* **17**, 170–178 (2017).
22. M. Kolmer *et al*., *Angew. Chem. Int. Ed.* **52**, 10300–10303 (2013).
23. M. Kolmer *et al*., *Chem. Commun.* **51**, 11276–11279 (2015).
24. G. Vasseur *et al*., *J. Am. Chem. Soc.* **138**, 5685–5692 (2016).
25. K. Y. Amsharov, M. A. Kabdulov, M. Jansen, *Angew. Chem. Int. Ed.* **51**, 4594–4597 (2012).
26. K. Y. Amsharov, P. Merz, *J. Org. Chem.* **77**, 5445–5448 (2012).
27. K. Amsharov, *Phys. Status Solidi, B Basic Res.* **253**, 2473–2477 (2016).
28. A. K. Steiner, K. Y. Amsharov, *Angew. Chem. Int. Ed.* **56**, 14732–14736 (2017).
29. D. Sharapa, A. K. Steiner, K. Amsharov, *Phys. Status Solidi, B Basic Res.* **255**, 1800189 (2018).
30. M. Kolmer *et al*., *Science* **363**, 57–60 (2019).
31. L. Talirz *et al*., *J. Am. Chem. Soc.* **135**, 2060–2063 (2013).
32. M. Golor, C. Koop, T. C. Lang, S. Wessel, M. J. Schmidt, *Phys. Rev. Lett.* **111**, 085504 (2013).
33. L. Benz, J. Haubrich, R. G. Quiller, S. C. Jensen, C. M. Friend, *J. Am. Chem. Soc.* **131**, 15026–15031 (2009).
34. V. Akhmetov, M. Feofanov, S. Troyanov, K. Amsharov, *Chemistry* **25**, 7607–7612 (2019).
35. V. Akhmetov, M. Feofanov, O. Papaianina, S. Troyanov, K. Amsharov, *Chemistry* **25**, 11609–11613 (2019).
36. A. Konishi *et al*., *J. Am. Chem. Soc.* **132**, 11021–11023 (2010).
37. T. Šolomek, P. Ravat, Z. Mou, M. Kertesz, M. Juríček, *J. Org. Chem.* **83**, 4769–4774 (2018).
38. T. Šolomek, P. Ravat, M. Juríček, *Trends Chem.* **1**, 705–706 (2019).
39. P. Ravat, T. Šolomek, D. Häussinger, O. Blacque, M. Juríček, *J. Am. Chem. Soc.* **140**, 10839–10847 (2018).
40. W. Ko, C. X. Ma, G. D. Nguyen, M. Kolmer, A. P. Li, *Adv. Funct. Mater.* **29**, 1903770 (2019).
41. M. Koch, F. Ample, C. Joachim, L. Grill, *Nat. Nanotechnol.* **7**, 713–717 (2012).
42. Y. Li, W. Zhang, M. Morgenstern, R. Mazzarello, *Phys. Rev. Lett.* **110**, 216804 (2013).
43. M. Kolmer *et al*., *Nat. Commun.* **10**, 1573 (2019).
44. A. L. Miettinen *et al*., *Phys. Rev. B* **100**, 045425 (2019).
45. F. D. Natterer *et al*., *Nature* **543**, 226–228 (2017).
46. P. Willke, K. Yang, Y. Bae, A. J. Heinrich, C. P. Lutz, *Nat. Phys.* **15**, 1005–1010 (2019).
47. K. Yang *et al*., *Science* **366**, 509–512 (2019).

ACKNOWLEDGMENTS

Funding: This work was supported in part by the Center for Nanophase Materials Sciences (CNMS), which is a U.S. Department of Energy Office of Science User Facility (M.K., W.K., A.-P.L.). Research was partially funded by grants ONR N00014-16-1-3213 and ONR N00014-20-1-2302 from the University of Tennessee (W.K., A.-P.L.). Preliminary research was supported by the Polish Ministry of Science and Higher Education, contract no. 0341/IP3/2016/74 (I.I.). A.K.S. and K.A. thank the Deutsche Forschungsgemeinschaft for financial support (Projektnummer 182849149–SFB 953 A6 and AM407). **Author contributions:** M.K., A.-P.L., and K.A. conceived the project. K.A. and A.K.S. carried out the precursor synthesis and analysis. M.K. conducted the on-surface synthesis and characterization experiments with support from W.K, and A.-P.L., M.K., I.I., and M.S. performed the preliminary experiments. M.E. conducted the computations. M.K., A.-P.L., and K.A. wrote the manuscript with feedback from all other authors; **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/369/6503/571/suppl/DC1
Materials and Methods
Figs. S1 to S13
References (*48*–*52*)

25 March 2020; accepted 3 June 2020
Published online 25 June 2020
10.1126/science.abb8880

AIR POLLUTION

Disparities in $PM_{2.5}$ air pollution in the United States

Jonathan Colmer[1]*, Ian Hardman[2], Jay Shimshack[3], John Voorheis[4]

Air pollution at any given time is unequally distributed across locations. Average concentrations of fine particulate matter smaller than 2.5 micrometers in diameter ($PM_{2.5}$) have fallen over time. However, we do not know how the spatial distribution of $PM_{2.5}$ has evolved. Here, we provide early evidence. We combine 36 years of $PM_{2.5}$ concentrations measured over ~8.6 million grid cells with geographic, economic, and demographic data from ~65,000 U.S. census tracts. We show that differences in $PM_{2.5}$ between more and less polluted areas declined substantially between 1981 and 2016. However, the most polluted census tracts in 1981 remained the most polluted in 2016. The least polluted census tracts in 1981 remained the least polluted in 2016. The most exposed subpopulations in 1981 remained the most exposed in 2016. Overall, absolute disparities have fallen, but relative disparities persist.

Fine particulate matter concentrations in the United States have declined by ~70% since 1981 (*1*, *2*). These improvements in air quality are associated with greater life expectancy, reduced infant mortality, higher property values, increased productivity, higher earnings, and other benefits (*3*–*11*). The existing literature documenting changes in air pollution over time focuses on concentrations averaged across locations. However, little evidence documents how the distribution of pollution concentrations has changed over time.

Air pollution in any given time period is unequally distributed over space (*12*). Particulate matter and other criteria air pollutants vary geographically because of differences in population density, emissions sources, economic activity, climate, and geophysical conditions (*13*). Pollution is associated with race, poverty, and demographic factors owing to sociopolitical forces, residential choice, and other influences (*12*, *14*, *15*). However, related scholarship does not systematically characterize how this spatial variation has evolved over time. We have limited information on whether policy or other factors have reduced pollution disparities across locations or whether such disparities persist through time.

Here, we provide evidence on how the spatial distribution of fine particulate matter smaller than 2.5 μm in diameter ($PM_{2.5}$) has changed over recent decades. To do this, we compile 36 years of annual $PM_{2.5}$ data from chemical transport modeling, satellite remote sensing, and ground-based measurements over ~8.6 million U.S. grid cells that measure 0.01° by 0.01° (0.9 km by 1.1 km) (*2*) (supplementary text, section S1). We focus on $PM_{2.5}$ because of its high morbidity and mortality risk, because of its disproportionate impact on the monetized benefits of federal regulation, and because newly available data provide fine-scale estimates spanning nearly four decades (*2*, *8*, *10*, *16*, *17*). We combine these $PM_{2.5}$ estimates with time-varying administrative, geographic, policy, and sociodemographic data for each of the ~65,000 census tracts in the contiguous United States (supplementary text, sections S2 and S3).

We first characterize how the distribution of $PM_{2.5}$ concentrations across U.S. census tracts changed between 1981 and 2016. Figure 1 documents a 69% decline in mean $PM_{2.5}$ between 1981 and 2016, consistent with (*1*, *2*). Figure 1 also illustrates that the gaps in $PM_{2.5}$ concentrations between more and less polluted census tracts narrowed considerably. The $PM_{2.5}$ concentration gap between the 90th and 10th percentiles was 15.66 μg/m^3 in 1981 and 4.16 μg/m^3 in 2016. $PM_{2.5}$ concentration gaps between census tracts with high and low fractions of disadvantaged subpopulations have generally, but not universally, narrowed over time as well (fig. S1).

The reductions in absolute disparities documented in Fig. 1 provide limited information on relative disparities across locations. We do not know whether locations that were historically the most polluted are still the most polluted. It is not clear a priori whether one might expect persistence or reversals in relative disparities (supplementary text, section S4). To explore the issue, we assign each census tract to a percentile of the distribution and examine the correlation between $PM_{2.5}$ percentile rank in 1981 and the average corresponding $PM_{2.5}$ percentile rank in 2016. Rank-rank comparisons are common for analyzing distributional changes over time and have statistical advantages (*18*). On average, we estimate a rank-rank correlation coefficient of 0.80 ($p < 0.01$), indicating a very strong positive association between historical relative pollution and more recent relative pollution (table S1). The modal change in percentile rank between 1981 and 2016 was 0, the mean change was −0.00006, and the median change was −1 percentile point. On average, the least polluted census tracts in 2016 were the least polluted census tracts in 1981. On average, the most polluted census tracts in 2016 were the most polluted census tracts in 1981.

Figure 2 plots percentiles of census tracts according to $PM_{2.5}$ in 2016 (vertical axis) against $PM_{2.5}$ percentiles in 1981 (horizontal axis). First, we observe widespread relative rank preservation across the entire distribution. All percentile pairs lie near the 45° line, such that, on average, a census tract's $PM_{2.5}$ percentile rank in 2016 was similar to its $PM_{2.5}$ rank in 1981. Second, we observe modest median reversion. On average, a census tract with below-median $PM_{2.5}$ in 1981 experienced greater relative pollution in 2016, and a census tract with above-median $PM_{2.5}$ in 1981 experienced lower relative pollution in 2016. Third, the tails of the distribution are particularly stable. Values on the y

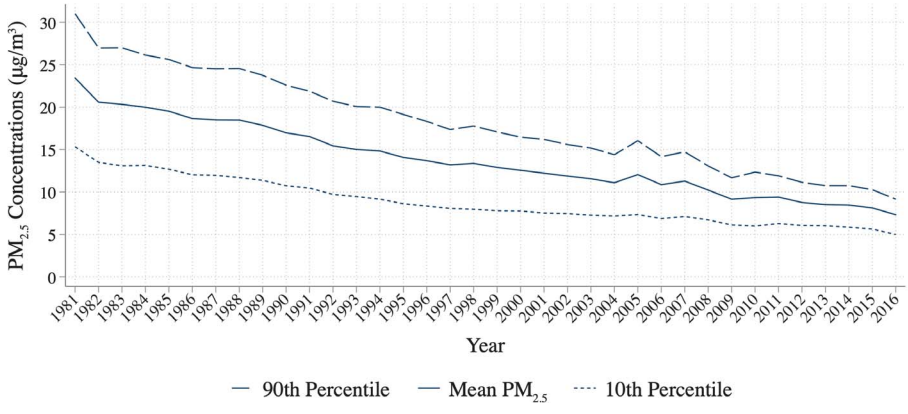


Fig. 1. Change in $PM_{2.5}$ concentrations between 1981 and 2016 in μg/m^3. We observe that mean $PM_{2.5}$ concentrations have fallen substantially over time, from 23.43 μg/m^3 in 1981 to 7.32 μg/m^3 in 2016, a reduction of 16.11 μg/m^3. In addition, we observe that the gap between the 90th and 10th percentiles has also shrunk. In 1981, the difference between the 90th and 10th percentiles was 15.66 μg/m^3. In 2016, the difference between the 90th and 10th percentiles was 4.16 μg/m^3. Gaps between the 90th and 10th percentiles are arbitrarily chosen but illustrative of broader patterns.

[1]Department of Economics, University of Virginia, Charlottesville, VA, USA. [2]Program on Energy and Sustainable Development, Stanford University, Stanford, CA, USA. [3]Frank Batten School of Leadership and Public Policy, University of Virginia, Charlottesville, VA, USA. [4]U.S. Census Bureau, Washington, DC, USA.
*Corresponding author. Email: jonathan.colmer@virginia.edu

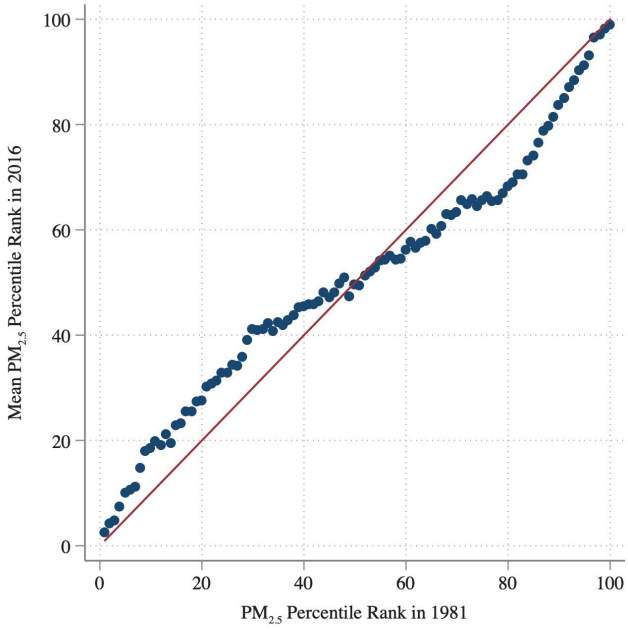


Fig. 2. Relationship between $PM_{2.5}$ rank in 1981 and $PM_{2.5}$ rank in 2016. The red line is the 45° line. If all points lay on the red line, this would indicate that there was no change in rank between 1981 and 2016; for example, a census tract at the 20th percentile of the $PM_{2.5}$ distribution in 1981 would also be at the 20th percentile of the $PM_{2.5}$ distribution in 2016. Deviations from the red line represent the average change in rank between 1981 and 2016. Points above the red line represent an increase in relative rank and a decrease in relative air quality. Points below the red line represent a reduction in relative rank and an improvement in relative air quality. The figure indicates that there is very high persistence in rank over time. The highest-rank locations in 1981 are the highest-rank locations in 2016. The lowest-rank locations in 1981 are the lowest-rank locations in 2016. Intermediate ranks experience some change; however, the average change is small and follows a pattern consistent with reversion to the median. Below-median *x*-axis values (1981 ranks) correspond to *y*-axis values (2016 ranks) above the 45° line, and above-median *x*-axis values (1981 ranks) correspond to *y*-axis values (2016 ranks) below the 45° line.

axis (2016 ranks) for particularly low and high 1981 ranks are systematically near the 45° line. Of the most polluted 5% of census tracts in 1981, 93.9% were among the most polluted 10% of census tracts in 2016. Of the least polluted 5% of census tracts in 1981, 89.7% were among the least polluted 10% of census tracts in 2016. These summary statistics are arbitrarily chosen but illustrative of broader patterns (fig. S2).

A possible explanation for the persistence in relative disparities across locations is that reductions in $PM_{2.5}$ were proportional. Figure 3A illustrates that $PM_{2.5}$ concentrations fell more in census tracts that were more polluted in 1981. However, larger reductions came from much higher baselines. Figure 3B indeed documents that observed $PM_{2.5}$ concentration reductions were approximately proportional across U.S. census tracts. The most polluted locations did not experience disproportional reductions in $PM_{2.5}$.

Our findings suggest that relative disparities persist for most U.S. census tracts. Nevertheless, it remains possible that aggregation obscures regional or local differences. Figure 4 summarizes the geographic distribution of changes in $PM_{2.5}$ percentile rank over time. The dominant message remains that relatively more polluted areas in 1981 were still more polluted in 2016 and relatively less polluted areas in 1981 were still less polluted in 2016. However, some local variation in rank change exists. Ohio, West Virginia, eastern Kentucky, and the Northeast Corridor became relatively less polluted. California's Central and Imperial Valleys, southwestern Arizona, southern Texas, and western Arkansas and eastern Oklahoma became relatively more polluted. Mechanisms for differential trends in these specific locations may include import competition, the decline in manufacturing, changes in the structure of energy production (e.g., the hydraulic fracking boom, a decline in coal production), and differential environmental regulation (*19*–*26*).

Our results are robust to population weighting, which rebalances the analysis to apply to the average American rather than the average U.S. census tract (table S1). From a socioeconomic and demographic perspective, the most exposed subpopulations also remain similar through time (fig. S3 and table S2). Census tracts' baseline 1981 black population shares, educational attainment, unemployment rates, and poverty rates are unrelated statistically to changes in rank between 1981 and 2016. More populated, whiter, higher income, and less Hispanic areas at baseline in 1981 are associated with reductions in rank over time. A 1-SD increase in a census tract's baseline 1981 population, white share, and income is associated with an average decrease of 1.77, 2.36, and 0.93 percentile rank points between 1981 and 2016. A 1-SD increase in a census tract's baseline 1981 Hispanic share is associated with an average increase of 3.92 percentile rank points between 1981 and 2016. Areas becoming more populated, poorer, more Hispanic, less white, and less educated between 1981 and 2016 are associated with increases in rank over time. A 1-SD increase in a census tract's growth in population, poverty rate, unemployment rate, and Hispanic share is associated with an average increase of 3.27, 1.98, 2.01, and 3.68 percentile rank points between 1981 and 2016. A 1-SD increase in a census tract's growth in income, educational attainment, and white share is associated with an average decrease of 1.60, 2.56, and 2.52 percentile rank points between 1981 and 2016. We caution that these are descriptive correlations, documenting differences across subpopulations.

We find somewhat larger associations between $PM_{2.5}$ percentile rank and market or regulatory factors (fig. S3 and table S2). A 1-SD decrease in a census tract's manufacturing employment share over time is associated with an average increase of 4.68 percentile rank points between 1981 and 2016. Census tracts in nonattainment with $PM_{2.5}$ ambient air-quality standards before 2016 are associated with an average decrease of 7.09 percentile rank points between 1981 and 2016.

We explore sensitivity. Our findings are robust to using alternative measures of relative disparities (table S3). Our results are not driven by differences in population density or urban-rural distinctions (fig. S4 and table S4). Our results are robust to comparisons within states; rank-rank correlations within each state range from 0.72 to 0.99 and average 0.91 (fig. S5 and table S5). Our results are robust to choosing comparison years other than 1981. Our approach generates conservative estimates (fig. S6). Our results are not sensitive to the choice of single years as points of comparison. Analyses using multiyear averages of $PM_{2.5}$ generate similar findings (fig. S7 and table S6). Correlations between changes in $PM_{2.5}$ percentile rank and socioeconomic and demographic factors are robust to using the rank of, and rank changes in, tract-level characteristics (fig. S8 and table S7). Our findings are robust to using pollution monitor data (figs. S9 and S10). Although $PM_{2.5}$ monitor data are only available between 2000 and 2016 for a comparatively small number of (mostly urban) locations, we estimate a rank-rank correlation coefficient of 0.79, which is statistically indistinguishable from our primary correlation coefficient. Last, our results are not sensitive to using different moments of the $PM_{2.5}$ distribution, including the median and each decile (10th through 90th percentile) (figs. S11 to S13 and table S8). Rank dependence is weaker when using annual maximum $PM_{2.5}$ concentrations, but we still estimate a correlation coefficient of 0.5 to 0.7.

We note caveats. First, we study the distribution of $PM_{2.5}$ across the contiguous United States. We do not inform distributional changes for other pollutants or other locations. Second, we study changes in the distribution of $PM_{2.5}$

across locations; we do not observe individual-level pollution exposure. Our results are robust to population weighting. Third, our analysis is descriptive. We do not directly explore hypotheses related to demographic change (i.e., "moving to the nuisance" or "residential sorting") (*15*, *27*–*30*). In our aggregate analysis, changes in relative pollution are not strongly correlated with socioeconomic and demographic factors. We do not provide causal mechanisms for our empirical regularities.

In sum, we explore how the spatial distribution of $PM_{2.5}$ has changed in recent decades. We first document that absolute disparities are shrinking. Differences in $PM_{2.5}$ concentrations between more and less polluted census tracts declined between 1981 and 2016. All else equal, absolute differences between areas in particulate-induced morbidity, mortality, and earnings losses should be expected to decline (*3*–*5*, *7*, *8*, *10*, *11*). By contrast, we show that relative disparities in $PM_{2.5}$ concentrations across space are notably persistent. Areas that were more polluted in 1981 are still more polluted in 2016. Areas that were less polluted in 1981 are still less polluted in 2016. Fairness, equity, and justice are often inherently comparative, and

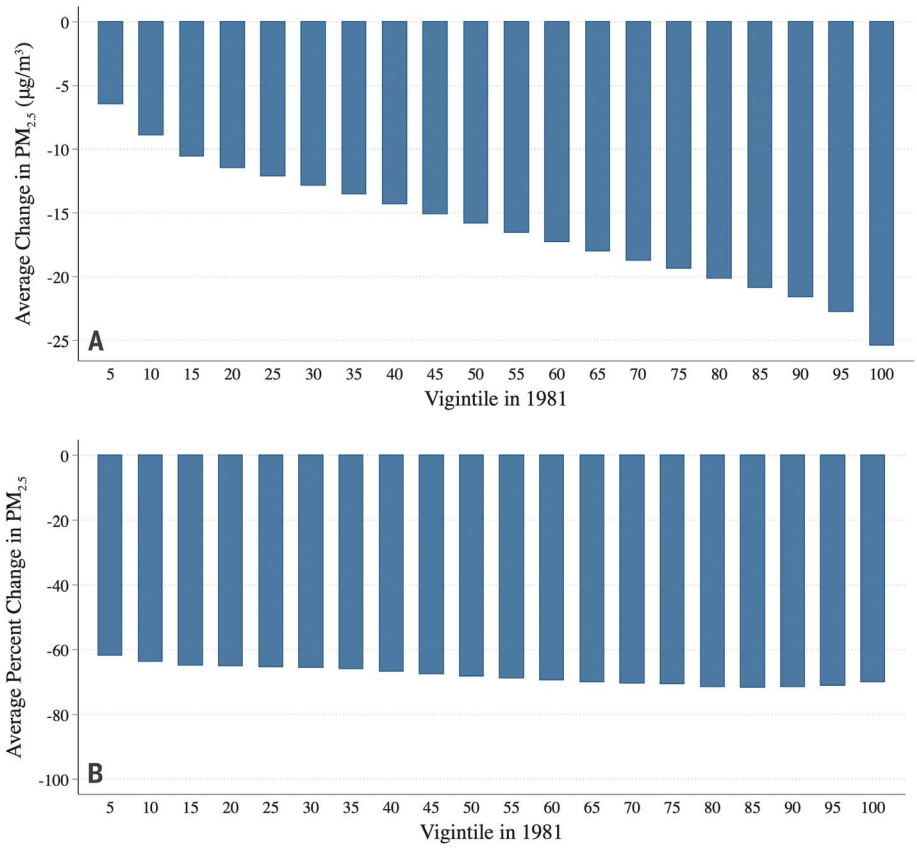


Fig. 3. Changes in $PM_{2.5}$ between 1981 and 2016 for each vigintile bin in 1981. (**A**) Reduction in $PM_{2.5}$ concentrations ($\mu g/m^3$) between 1981 and 2016 for each vigintile bin in 1981. We observe that the largest reductions in $PM_{2.5}$ concentrations were experienced in census tracts that were the most polluted in 1981. (**B**) Percent reduction in $PM_{2.5}$ concentrations between 1981 and 2016 for each vigintile bin in 1981. We observe that reductions in $PM_{2.5}$ were largely proportional throughout the distribution. The most polluted census tracts experienced approximately the same relative reductions in $PM_{2.5}$ as the least polluted census tracts.

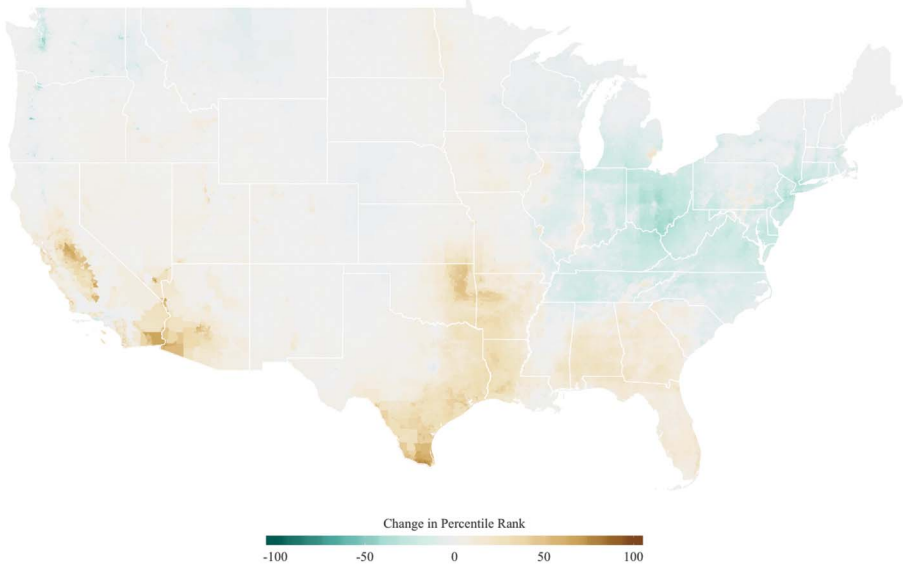


Fig. 4. Map of the change in $PM_{2.5}$ rank between 1981 and 2016 for each census tract in the contiguous United States. Green represents tracts where the rank has declined over time. Brown represents tracts where the rank has increased over time. Darker shades represent greater absolute changes. Most census tracts have experienced small changes in rank. The modal change is 0, the mean change is −0.00006, and the median change is −1. Ohio, West Virginia, eastern Kentucky, and the Northeast Corridor became relatively less polluted. California's Central and Imperial Valleys, southwestern Arizona, southern Texas, and western Arkansas and eastern Oklahoma became relatively more polluted.

these types of relative disparities are important for normative aspects of social welfare (*31*). In this sense, we build on the growing literature documenting relative disparities in income, health, and other contexts (*32*–*37*). We also illustrate that changes in relative $PM_{2.5}$ concentrations are not strongly correlated with socioeconomic and demographic factors. If anything, relative disparities are growing for vulnerable subpopulations. From a socioeconomic and demographic perspective, the most exposed subpopulations to fine particulate matter remain constant over time.

We do not attempt a formal welfare analysis. Nevertheless, environmental justice and reducing disparities have been stated objectives of U.S. environmental policy for decades (*12*, *14*, *15*). U.S. Environmental Protection Agency and state guidelines aim for all people to enjoy the same degree of protection from environmental hazards and stipulate that no groups should bear a disproportionate share of pollution. Although absolute $PM_{2.5}$ disparities have fallen substantially over the past four decades, relative disparities persist.

REFERENCES AND NOTES

1. U.S. Environmental Protection Agency (EPA), "Our nation's air: Status and trends through 2018" (EPA, 2019).
2. J. Meng *et al*., *Environ. Sci. Technol.* **53**, 5071–5079 (2019).
3. D. W. Dockery *et al*., *N. Engl. J. Med.* **329**, 1753–1759 (1993).
4. C. A. Pope III *et al*., *JAMA* **287**, 1132–1141 (2002).
5. K. Chay, M. Greenstone, *Q. J. Econ.* **118**, 1121–1167 (2003).
6. K. Chay, M. Greenstone, *J. Polit. Econ.* **113**, 376–424 (2005).
7. J. Currie, M. Neidell, *Q. J. Econ.* **120**, 1003–1030 (2005).
8. C. A. Pope III, D. W. Dockery, *J. Air Waste Manag. Assoc.* **56**, 709–742 (2006).
9. J. G. Zivin, M. Neidell, *Am. Econ. Rev.* **102**, 3652–3673 (2012).
10. A. J. Cohen *et al*., *Lancet* **389**, 1907–1918 (2017).
11. A. Isen, M. Rossin-Slater, R. Walker, *J. Polit. Econ.* **125**, 848–902 (2017).
12. P. Mohai, D. Pellow, J. T. Roberts, *Annu. Rev. Environ. Resour.* **34**, 405–430 (2009).
13. J. P. Pinto, A. S. Lefohn, D. S. Shadwick, *J. Air Waste Manag. Assoc.* **54**, 440–449 (2004).
14. J. Agyeman, D. Schlosberg, L. Craven, C. Matthews, *Annu. Rev. Environ. Resour.* **41**, 321–340 (2016).
15. S. Banzhaf, L. Ma, C. Timmins, *J. Econ. Perspect.* **33**, 185–208 (2019).
16. F. Dominici, M. Greenstone, C. R. Sunstein, *Science* **344**, 257–259 (2014).
17. R. Burnett *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9592–9597 (2018).
18. R. Chetty, N. Hendren, P. Kline, E. Saez, *Q. J. Econ.* **129**, 1553–1623 (2014).
19. B. Bolin, S. Grineski, T. Collins, *Hum. Ecol. Rev.* **12**, 156–168 (2005).
20. T. W. Collins, S. E. Grineski, M. D. L. R. Aguilar, *Appl. Geogr.* **29**, 448–461 (2009).
21. D. D. Parrish, H. B. Singh, L. Molina, S. Madronich, *Atmos. Environ.* **45**, 7015–7025 (2011).
22. D. Autor, D. Dorn, G. Hanson, *Am. Econ. Rev.* **103**, 2121–2168 (2013).
23. R. Jackson *et al*., *Annu. Rev. Environ. Resour.* **39**, 327–362 (2014).
24. J. Pierce, P. Schott, *Am. Econ. Rev.* **106**, 1632–1662 (2016).
25. J. S. Shapiro, R. Walker, *Am. Econ. Rev.* **108**, 3814–3854 (2018).
26. A. W. Bartik, J. Currie, M. Greenstone, C. R. Knittel, *Am. Econ. J. Appl. Econ.* **11**, 105–155 (2019).
27. V. Been, *Yale Law J.* **103**, 1383–1422 (1994).
28. V. Been, F. Gupta, *Ecol. Law Q.* **24**, 1–56 (1997).
29. H. S. Banzhaf, R. P. Walsh, *Am. Econ. Rev.* **98**, 843–863 (2008).
30. S. Gamper-Rabindran, C. Timmins, *Am. Econ. Rev.* **101**, 620–624 (2011).
31. L. Temkin, *Econ. Philos.* **19**, 61–87 (2003).
32. A. Atkinson, *J. Econ. Theory* **2**, 244–263 (1970).
33. Y. Amiel, F. Cowell, *Thinking About Inequality* (Cambridge Univ. Press, 1999).
34. P. Braveman, *Annu. Rev. Public Health* **27**, 167–194 (2006).
35. S. Harper *et al*., *Milbank Q.* **88**, 4–29 (2010).
36. N. Eyal, S. Hurst, O. Norheim, D. Wikler, *Inequalities in Health: Concepts, Measures, and Ethics* (Oxford Univ. Press, 2013).
37. T. Piketty, E. Saez, G. Zucman, *Q. J. Econ.* **133**, 553–609 (2018).
38. J. Colmer, Replication data for: PM2.5 disparities in the United States, version 1, UVA Dataverse (2020); https://doi.org/10.18130/V3/QLBNQJ.

ACKNOWLEDGMENTS

We thank J. Meng for assistance in accessing the historical $PM_{2.5}$ data. We thank anonymous reviewers for helpful comments. Any opinions and conclusions expressed herein are those of the authors and do not necessarily reflect the views of the U.S. Census Bureau. **Funding:** This study was not externally funded. **Author contributions:** All authors contributed equally to this work. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data and code used in our analysis are available at (*38*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/575/suppl/DC1
Supplementary Text
Figs. S1 to S13
Tables S1 to S8
References (*39*–*55*)

21 October 2019; accepted 9 June 2020
10.1126/science.aaz9353

PALEOBOTANY

Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora

Wen-Na Ding[1,2], Richard H. Ree[3]*, Robert A. Spicer[1,4,5], Yao-Wu Xing[1,4]*

Understanding how alpine biotas formed in response to historical environmental change may improve our ability to predict and mitigate the threats to alpine species posed by global warming. In the world's richest temperate alpine flora, that of the Tibet-Himalaya-Hengduan region, phylogenetic reconstructions of biome and geographic range evolution show that extant lineages emerged by the early Oligocene and diversified first in the Hengduan Mountains. By the early to middle Miocene, accelerated diversification and colonization of adjacent regions were likely driven jointly by mountain building and intensification of the Asian monsoon. The alpine flora of the Hengduan Mountains has continuously existed far longer than any other alpine flora on Earth and illustrates how modern biotas have been shaped by past geological and climatic events.

Terrestrial biodiversity is disproportionately concentrated in mountains (*1*, *2*), but our understanding about when and how these exceptionally species-rich biotas were formed, with their characteristic altitudinal zonation, remains limited (*3*). Of particular interest is the alpine biome, above the natural tree line, which harbors distinctive communities that are adapted to the biome's characteristically stressful environmental conditions (*4*) while being especially sensitive to global warming, underscoring the urgency of understanding the natural history of these biotas (*3*, *4*). For plants, the world's most species-rich temperate alpine biota occurs in the Qinghai-Tibet Plateau (QTP), Himalaya, and Hengduan Mountains (*5*) (Fig. 1), hereafter collectively referred to as the Tibet-Himalaya-Hengduan (THH) region. In particular, the alpine flora of the Hengduan Mountains is extraordinarily diverse, harboring an estimated 3030 species of alpine seed plants (*5*), making it comparable in area and richness to the Páramos in the tropical Andes (*6*). Threatened by climate change and anthropogenic disturbances, the Hengduan Mountains region is considered one of the most vulnerable of all global biodiversity hotspots (*7*).

Understanding the temporal and geographic origins of alpine biotas requires consideration of how geophysical processes such as orogeny and climate change could have influenced the biological processes of species accumulation, diversification, and persistence (*1*, *2*). In this respect, the geophysical history of the THH region is complex (*8*). For instance, one might ask when the mountains in this area became high enough for tree line formation. In the generally accepted time line, the southern part of the QTP exceeded 4.5 km in elevation by the Eocene (*9*), but the high central plateau was not fully formed until the Neogene (*10*), and the Himalaya did not project above the plateau until the middle Miocene (*11*). The Hengduan Mountains were long thought to have formed more recently, in the late Miocene to the Pliocene (*12*), and there is evidence of rapid uplift during this period (table S4). However, the latest fossil and tectonic evidence suggests that part of the Hengduan Mountains attained near-modern elevations by the early Oligocene (*13*), with continued deformation through the late Oligocene to the early Miocene (fig. S1 and table S4). Of course, high elevation alone is insufficient to establish an alpine biota; sufficiently cool temperatures are also needed. At the latitude of the THH region, alpine habitats at high elevations likely became possible from the late Eocene to the early Oligocene, when paleoclimates shifted through the late Cenozoic from a warm greenhouse world to one with colder and increasingly seasonal conditions (*14*).

In this study, we connected the dynamic tectonic and climatological history of the THH region to the biological processes that have driven the development of its alpine biota. We focus specifically on whether phylogenetic estimates of alpine ancestry are temporally and spatially consistent with geological evidence of alpine habitat availability, as well as whether major tectonic events in the QTP, Himalaya, and Hengduan Mountains left discernible imprints in the tempo and mode of alpine biotic assembly.

We analyzed time-calibrated phylogenies of 18 groups of flowering plants by using a joint model of biome occupation, geographic range evolution, and lineage diversification that allowed us to infer how alpine biotas of the Hengduan Mountains, Himalaya, and QTP were assembled through time (*15*). Specifically, we estimated rates of species accumulation through time in alpine and non-alpine biomes for each area from model-based simulations of historical sequences of speciation, extinction, colonization, and local recruitment (evolutionary shifts in biome tolerance) that are consistent with the observed data (*15*).

Our historical reconstructions indicate that an alpine flora had emerged in the THH region by the early Oligocene (fig. S2), mainly by ancestral lineages of Delphineae (fig. S11) and Saxifragaceae (fig. S24). This is much earlier than estimated origins of other extant alpine floras, such as those of the Andean Páramos [~3.5 million years ago (Ma)] (*6*, *16*), Mediterranean Alps of Europe (<1.75 Ma) (*17*), Southern Alps of New Zealand (<5 Ma) (*18*), and Afro-alpine zone of East Africa (8 to 7 Ma) (*19*). In some locations (e.g., Africa), the mountains themselves do not predate the Oligocene (*20*). By contrast, in the Rocky Mountains, an early alpine flora likely existed in the Eocene but did not persist, owing to climate-driven turnover (*21*). The modern alpine flora of the Rocky Mountains was assembled over the past 10 million years (*16*, *22*).

Overall rates of in situ alpine speciation began to increase from the late Oligocene to the early Miocene, with a pulse in the Hengduan Mountains at 27 to 24 Ma and in the Himalaya at 19 to 17 Ma, peaking in the middle Miocene (Fig. 2C). For both regions, but especially the Hengduan Mountains, in situ diversification was the dominant mode of alpine assembly from the middle Miocene to the present, exceeding rates of colonization and local recruitment. By contrast, alpine assembly within the QTP was dominated by colonization; the rate of in situ diversification was relatively low through the Miocene and increased moderately over only the past 5 Ma (Fig. 2C). All of the regions experienced a relatively low rate of local recruitment compared with rates of in situ speciation and colonization (Fig. 2C).

Our results, derived from analyses of time-scaled molecular phylogenies and not in situ fossil evidence of alpine ancestry, are nevertheless temporally consistent with the latest geological evidence that active orogeny associated with widespread crustal shortening and thickening established highlands from eastern Tibet to the Hengduan Mountains by the end of the Eocene (table S4). Although warm climates in the Eocene would have limited the extent of alpine habitat, even at such high elevations, further uplift and climate cooling by the late Oligocene to the early Miocene would have contributed to alpine expansion (*10*, *23*) and stimulated the diversification of alpine plants during this period (Fig. 2). Continuing high rates of in situ alpine diversification after the mid-Miocene climatic optimum (17 to 14 Ma) are temporally congruent with intensification of the Asian summer monsoon (*24*), which

[1]CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan 666303, China. [2]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China. [3]Negaunee Integrative Research Center, Field Museum, Chicago, IL 60605, USA. [4]Center of Plant Ecology, Core Botanical Gardens, Chinese Academy of Sciences, Mengla, Yunnan 666303, China. [5]School of Environment, Earth, and Ecosystem Sciences, The Open University, Milton Keynes MK7 6AA, UK.
*Corresponding author. Email: ywxing@xtbg.org.cn (Y.-W.X.); rree@fieldmuseum.org (R.H.R.).

Fig. 1. Topographic map of the Tibet-Himalaya-Hengduan (THH) region. Inset images show representatives of the 18 clades of flowering plants that we analyzed (clockwise from top left: *Rheum*, *Leontopodium*, *Saussurea*, *Salix*, Caryophyllaceae, *Anaphalis*, *Gentiana*, Saxifragaceae, Delphineae, *Pedicularis*, *Allium*, *Rhododendron*, Primuloideae, *Rhodiola*, *Draba*, *Chesneya*, *Meconopsis*, and *Pleurispermum*). Pie charts show the proportions of sampled alpine species in a clade that occur in each area of the THH region. Red circles denote sites from which data were contributed to estimate the paleoelevations and geological events of the THH region (1, Kailas; 2, Qiabulin; 3, Linzhou; 4, Nima; 5, Lunpola; 6, Markam; 7, Jianchuan; 8, Everest; 9, Namling; 10, Zanda; 11, Gonggar; and 12, Yangtze River). The Himalaya is delimited by the Main Boundary Thrust (MBT) and the Yarlung-Tsangpo suture (YTS). The Qinghai-Tibet Plateau proper is bounded by the YTS to the south, the Karakomrum to the west, the Altyn Tagh to the northwest, the Qilian Shan to the northeast, and the Hengduan Mountains to the southeast.

increased moisture availability for sustaining vegetation and associated biota (*1*), as well as erosion through river incision, leading to greater topographic relief (*25*). This would have promoted the differentiation of microhabitats associated with elevational gradients and slope aspect, increasing the potential for local adaptation, ecological divergence, and lineage diversification. In particular, the pulse in per-capita speciation rate [to 0.21 species per million years (spp/Myr)] at ~7 Ma is coincident with high monsoon intensity (Fig. 2A) and a period of heightened tectonic activity in the Hengduan Mountains in the late Miocene (Fig. 2B) (*26*). Our results thus strongly suggest that alpine diversification in the Hengduan Mountains was jointly triggered by orogeny and monsoon intensification in cooler post-Eocene climates, which together facilitate ecological and evolutionary opportunities that have continued to the present (*27*, *28*). Similarly, in the Himalaya, accelerating in situ alpine diversification from ~19 Ma through 17 Ma coincided with intense Himalayan orogeny (*11*) and peaked first (0.12 spp/Myr) at 17 to 14 Ma, when elevations were nearing present-day levels (*11*). Subsequent peaks (0.16 spp/Myr) at 13 Ma and 10 to 5 Ma occurred during periods of continued orogeny (*29*) and summer monsoon intensification (Fig. 2).

In the QTP, by contrast, the dominance of colonization in alpine biotic assembly relative to in situ speciation since 18 Ma (Fig. 2C) reflects orogenic and climatological differences from the Hengduan Mountains and the Himalaya. Whereas the latter regions experienced increasing topographical dissection and summer precipitation, the QTP has seen gradual aridification since the retreat of the Paratethys Sea (*30*), extensive planation since the Oligocene-Miocene transition (OMT) (*23*), and a strong orographic Himalaya-induced rain shadow since the middle Miocene (*11*). Paleontological studies show a substantial turnover of both flora and fauna, in which a warm and humid lowland environment was replaced by one that was high, cold, and dry (*23*). Our results thus corroborate geological evidence showing that pronounced environmental change since the OMT reshaped the biota of the QTP, with extinction of previous residents and colonization by cold-adapted species from neighboring regions. This finding echoes other Neogene instances of major floristic turnover in response to extrinsic climatic changes—e.g., in the Rocky Mountains (*21*). The modern biota of the QTP is thus newly developed since ~18 Ma, driven by uplift-induced aridification and subsequent Quaternary climate change (Fig. 2).

The Hengduan Mountains have been the primary source of alpine lineages colonizing

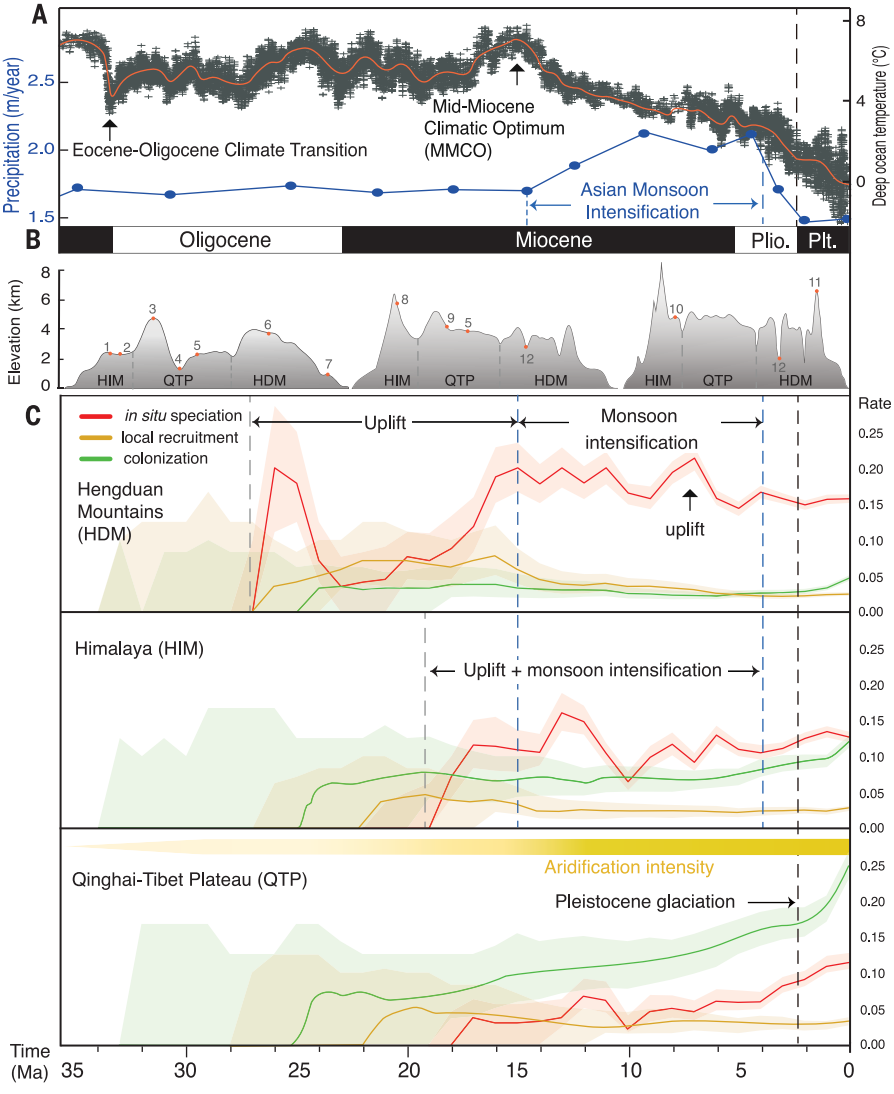


Fig. 2. Rates of biotic assembly in relation to climate and geological history in the Hengduan Mountains, Himalaya, and Qinghai-Tibet Plateau (QTP). (**A**) Evolution of global climate is represented by deep-sea oxygen isotope records (*33*) and estimated deep ocean temperature (*34*). Monsoon conditions are indicated by the modeled mean annual precipitation for each geologic stage, denoted by blue lines at idealized CO_2 levels (blue circles) [modified from (*24*)]. Plio., Pliocene; Plt., Pleistocene. (**B**) Schematic representation of the topography of the Himalaya (HIM), QTP, and Hengduan Mountains (HDM) in three phases, from the late Eocene to the present [modified from (*10*)]. Selected sites of reconstructed paleoelevation, based on the latest evidence (table S4), are represented by red circles with numbers, as in Fig. 1. (**C**) Rolling estimates of per-capita rate through time in the Hengduan Mountains, Himalaya, and QTP. Solid lines indicate median values; shaded regions indicate 25 to 75% quantile intervals estimated from 1000 replicate simulations of joint biome-range evolution and diversification. Rate estimates from the period before the earliest time at which the upper quantile value remains above zero continuously to the present are not shown (see fig. S4). The yellow color gradient in the bottom panel represents the intensity of aridification in QTP since the Paleogene.

the Himalaya and QTP (Fig. 3 and fig. S3), suggesting that across the THH region, extinction driven by Quaternary glacial cycles was lowest in the southeast. Most dispersal events (60%) (table S3) from the alpine biome of the Hengduan Mountains were to alpine biomes in the Himalaya and QTP (Fig. 3). By contrast, assembly of the Hengduan alpine flora occurred mostly by local recruitment from lower elevations (*31*) and, to a lesser extent, by colonization events from adjacent alpine zones (112 from Himalaya and QTP versus 51 from other alpine regions) (Fig. 3 and table S3). The Hengduan Mountains are similar in this respect to the Páramos, where local recruitment from lower-elevation Andean grasslands and Amazonia has been prevalent (*32*). As in the model proposed for the New Zealand alpine flora (*18*), local recruitment may have been accelerated by adaptive differentiation across contrasting habitats (forest versus alpine) over short geographic distances, facilitated by habitat fragmentation and elevational gradients caused by active orogeny and climate change–induced landscape modifications.

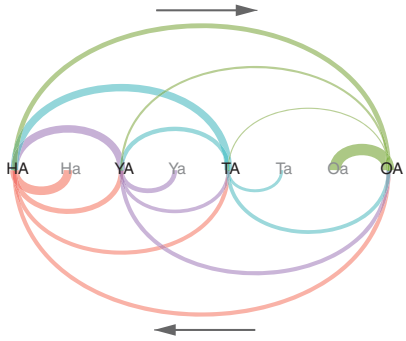


Fig. 3. Assembly of alpine biotas in the THH, estimated from local recruitment and colonization events among regions. Line color indicates the target biota of assembly events; line weight indicates the relative numbers of assembly events. Arrows indicate the direction of events, from source to target. H, Hengduan Mountains; Y, Himalaya; T, Qinghai-Tibet Plateau; O, Other regions; A, Alpine; a, non-alpine.

The rich alpine flora of the THH region has been shaped by a long and complex history of colonization, local recruitment, and in situ diversification driven by mountain building and climate change. The pivotal role of the Hengduan Mountains—a cradle of rapid origination and a notable example of alpine species persistence since the Oligocene—emphasizes the importance of conservation in this temperate biodiversity hotspot.

REFERENCES AND NOTES

1. A. Antonelli *et al.*, *Nat. Geosci.* **11**, 718–725 (2018).
2. C. Rahbek *et al.*, *Science* **365**, 1114–1119 (2019).
3. C. Rahbek *et al.*, *Science* **365**, 1108–1113 (2019).
4. C. Körner, in *Alpine Plant Life: Functional Plant Ecology of High Mountain Ecosystems*, C. Körner, Ed. (Springer, ed. 2, 2003), pp. 9–20.
5. X.-H. Li, X.-X. Zhu, Y. Niu, H. Sun, *J. Syst. Evol.* **52**, 280–288 (2014).
6. S. Madriñán, A. J. Cortés, J. E. Richardson, *Front. Genet.* **4**, 192 (2013).
7. D. E. Boufford, *Arnoldia* **72**, 24–35 (2014).
8. V. Mosbrugger, A. Favre, A. N. Muellner-Riehl, M. Päckert, A. Mulch, in *Mountains, Climate, and Biodiversity*, C. Hoorn, A. Perrigo, A. Antonelli, Eds. (Wiley Blackwell, 2018), pp. 429–448.
9. L. Ding *et al.*, *Earth Planet. Sci. Lett.* **392**, 250–264 (2014).
10. T. Su *et al.*, *Sci. Adv.* **5**, eaav2189 (2019).
11. L. Ding *et al.*, *Geology* **45**, 215–218 (2017).
12. A. Favre *et al.*, *Biol. Rev. Camb. Philos. Soc.* **90**, 236–253 (2015).

13. T. Su *et al*., *Natl. Sci. Rev.* **6**, 495–504 (2018).
14. J. S. Eldrett, D. R. Greenwood, I. C. Harding, M. Huber, *Nature* **459**, 969–973 (2009).
15. Materials and methods are available as supplementary materials.
16. C. E. Hughes, G. W. Atchison, *New Phytol.* **207**, 275–282 (2015).
17. P. Vargas, *Taxon* **52**, 463–476 (2003).
18. P. B. Heenan, M. S. McGlone, *N. Z. J. Ecol.* **37**, 105–113 (2013).
19. H. P. Linder, *Front. Ecol. Evol.* **2**, 38 (2014).
20. B. Gehrke, H. P. Linder, *Alp. Bot.* **124**, 165–177 (2014).
21. J. A. Wolfe, *Ann. Mo. Bot. Gard.* **74**, 785–803 (1987).
22. A. N. Muellner-Riehl *et al*., *J. Biogeogr.* **46**, 2826–2838 (2019).
23. T. Deng, F. Wu, S. Wang, T. Su, Z. Zhou, *Chin. Sci. Bull.* **64**, 2894–2906 (2019).
24. A. Farnsworth *et al*., *Sci. Adv.* **5**, eaax1697 (2019).
25. J. Nie *et al*., *Nat. Geosci.* **11**, 944–948 (2018).
26. Y. Xing, R. H. Ree, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E3444–E3451 (2017).
27. D. Luo *et al*., *J. Biogeogr.* **43**, 31–43 (2016).
28. X.-Y. Ye *et al*., *J. Biogeogr.* **46**, 2678–2689 (2019).
29. P. Kapp, P. G. DeCelles, *Am. J. Sci.* **319**, 159–254 (2019).
30. G. Ramstein, F. Fluteau, J. Besse, S. Joussaume, *Nature* **386**, 788–795 (1997).
31. X.-L. Jiang *et al*., *J. Biogeogr.* **46**, 2203–2214 (2019).
32. A. Antonelli *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6034–6039 (2018).
33. J. C. Zachos, G. R. Dickens, R. E. Zeebe, *Nature* **451**, 279–283 (2008).
34. J. Hansen, M. Sato, G. Russell, P. Kharecha, *Philos. Trans. R. Soc. A* **371**, 20120294 (2013).
35. W.-N. Ding, R. H. Ree, R. A. Spicer, Y.-W. Xing, Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora, v15, Dryad (2020); https://doi.org/10.5061/dryad.k6djh9w3s.

ACKNOWLEDGMENTS

We thank W. Freyman and M. Landis for helping with RevBayes and J. Liu for geological consultation and helping with fig. S1. We also thank the Germplasm Bank of Wild Species in Southwest China for providing barcoding sequences MT326201 to MT326205 and Herbarium (PE), Institute of Botany, CAS for the DNA material (MT332204). **Funding:** This work was supported by the National Natural Science Foundation of China (U1802242), the Strategic Priority Research Program of Chinese Academy of Sciences (XDB31000000), and the National Key R&D Program of China (2017YFC0505200). R.A.S. was supported by NERC-NSFC joint project NE/P013805/1 and a Xishuangbanna Tropical Botanical Garden Fellowship for Visiting Scientists. W.-N.D. was supported by the China Scholarship Council (201804910698). R.H.R. was supported by the Grainger Bioinformatics Center at the Field Museum. **Author contributions:** Y.-W.X., W.-N.D., and R.H.R. conceived the idea. W.-N.D. and R.H.R. analyzed the data and contributed to modeling. W.-N.D., R.H.R., and Y.-W.X. interpreted the results and wrote the paper, with substantial contributions from R.A.S. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data generated, analyzed, and presented in this study are available at Dryad (*35*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/578/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S24
Tables S1 to S4
References (*36*–*151*)

1 March 2020; accepted 8 June 2020
10.1126/science.abb4484

CORONAVIRUS

Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California

Xianding Deng[1,2]*, Wei Gu[1,2]*, Scot Federman[1,2]*, Louis du Plessis[3]*, Oliver G. Pybus[3,4], Nuno R. Faria[3,5], Candace Wang[1,2], Guixia Yu[1,2], Brian Bushnell[6], Chao-Yang Pan[7], Hugo Guevara[7], Alicia Sotomayor-Gonzalez[1,2], Kelsey Zorn[8], Allan Gopez[1], Venice Servellita[1], Elaine Hsu[1], Steve Miller[1], Trevor Bedford[9,10], Alexander L. Greninger[9,11], Pavitra Roychoudhury[9,11], Lea M. Starita[10,12], Michael Famulare[13], Helen Y. Chu[10,14], Jay Shendure[10,11,15], Keith R. Jerome[9,11], Catie Anderson[16], Karthik Gangavarapu[16], Mark Zeller[16], Emily Spencer[16], Kristian G. Andersen[16], Duncan MacCannell[17], Clinton R. Paden[17], Yan Li[17], Jing Zhang[17], Suxiang Tong[17], Gregory Armstrong[17], Scott Morrow[18], Matthew Willis[19], Bela T. Matyas[20], Sundari Mase[21], Olivia Kasirye[22], Maggie Park[23], Godfred Masinde[24], Curtis Chan[24], Alexander T. Yu[7], Shua J. Chai[7,17], Elsa Villarino[25], Brandon Bonin[25], Debra A. Wadford[7], Charles Y. Chiu[1,2,26]†

The coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread globally, with >365,000 cases in California as of 17 July 2020. We investigated the genomic epidemiology of SARS-CoV-2 in Northern California from late January to mid-March 2020, using samples from 36 patients spanning nine counties and the Grand Princess cruise ship. Phylogenetic analyses revealed the cryptic introduction of at least seven different SARS-CoV-2 lineages into California, including epidemic WA1 strains associated with Washington state, with lack of a predominant lineage and limited transmission among communities. Lineages associated with outbreak clusters in two counties were defined by a single base substitution in the viral genome. These findings support contact tracing, social distancing, and travel restrictions to contain the spread of SARS-CoV-2 in California and other states.

The novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes coronavirus disease 2019 (COVID-19), is a pandemic that has infected more than 13.8 million people around the world and caused more than 591,000 deaths as of 17 July 2020 (*1*), including >3.5 million cases in the United States and >365,000 in California. An exponential growth in the number of cases has overburdened clinical care facilities and threatens to overwhelm the medical workforce. The reported case numbers also underestimate the true number of infections because of the presence of asymptomatic or mild cases who do not get tested (*2*–*4*). As a result, California, along with many other states and countries, has issued a "shelter-in-place" policy for all residents, effective 20 March 2020 and ongoing at the time of this report. These unprecedented measures have disrupted daily life for ~40 million inhabitants of the state and have incurred profound economic losses (*5*).

Until late February 2020, the majority of infections identified in the United States were related to travelers returning from high-risk countries, repatriated citizens under quarantine, or close contacts of infected patients. Community spread, in which the source of the infection is unknown, has since been documented in multiple states. In particular, Washington state reported a series of COVID-19 cases from 21 January to 18 March, following the identification of the earliest case reported in the United States, WA1, on 19 January; this suggests that a persistent WA1 lineage transmission chain was present in the community during that time period (*6*, *7*).

Genomic epidemiology of emerging viruses has proven to be a useful tool for outbreak investigation and for tracking virus evolution and spread (*7*–*9*). During the Ebola virus disease epidemic of 2013–2016 in West Africa, genomic analyses established that the outbreak had a single zoonotic origin (*9*), that two major viral lineages were circulating (*10*), and that sexual transmission played a role in maintaining some transmission chains (*11*). Viral genome sequencing also uncovered the route that Zika virus traveled from northern Brazil to other regions (*12*), including Central America and Mexico (*13*) and the Caribbean and United States (*14*). However, real-time genomic epidemiology data for COVID-19 to inform public health interventions in California have been limited to date.

We recently developed a method called MSSPE (metagenomic sequencing with spiked primer enrichment) to rapidly enrich and assemble viral genomes directly from clinical samples (*15*). Here, we used this method and/or tiling multiplex polymerase chain reaction (PCR) to recover viral genomes from COVID-19 patients in Northern California and to perform phylogenetic analyses, with the goal of better understanding the genetic diversity of SARS-CoV-2 in the United States and the nature of transmission of virus lineages in the community.

We screened a total of 62 respiratory swab samples from 54 COVID-19 patients available from hospitals and clinics at the University of California, San Francisco (UCSF), the California Department of Public Health (CDPH), and eight county public health departments in Northern California (table S1). Presumptive positive cases were confirmed to be SARS-CoV-2–infected by a U.S. Centers for Disease Control and Prevention (CDC) assay approved by a Food and Drug Administration (FDA) Emergency Use Authorization (EUA) on 4 February 2020 (*16*). SARS-CoV-2 genomes (>65% coverage) were recovered from 36 patients (Fig. 1A and table S2). The 36 infected patients for whom viral genomes were obtained were collected from 29 January to 20 March 2020 and spanned nine counties in Northern California (Fig. 1B and table S2). The patient samples included (i) 11 samples collected from the Grand Princess cruise ship during its two voyages from San Francisco to Mexico and Hawaii in February and March 2020, (ii) three samples from a Solano County cluster that included the first reported case of community transmission in the United States with subsequent spread to two health care workers, (iii) seven samples from a local outbreak cluster in Santa Clara County associated with workspace transmission, (iv) three samples from patients who contracted the infection from a sick contact, (v) five samples related to domestic or international travel, and (vi) seven samples from additional cases of community transmission.

We performed MSSPE (*15*) and/or tiling multiplex PCR (*17*) on each sample to enrich for the

[1]Department of Laboratory Medicine, University of California, San Francisco, CA, USA. [2]UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, CA, USA. [3]Department of Zoology, University of Oxford, Oxford, UK. [4]Department of Pathobiology and Population Sciences, Royal Veterinary College, London, UK. [5]MRC Centre for Global Infectious Disease Analysis, J-IDEA, Imperial College London, London, UK. [6]Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [7]California Department of Public Health, Richmond, CA, USA. [8]Department of Biochemistry and Biophysics, University of California, San Francisco, CA, USA. [9]Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [10]Brotman Baty Institute for Precision Medicine, Seattle, WA, USA. [11]Department of Laboratory Medicine, University of Washington, Seattle, WA, USA. [12]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [13]Institute for Disease Modeling, Bellevue, WA, USA. [14]Department of Medicine, University of Washington, Seattle, WA, USA. [15]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. [16]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA, USA. [17]U.S. Centers for Disease Control and Prevention, Atlanta, GA, USA. [18]San Mateo County Department of Public Health, San Mateo, CA, USA. [19]Marin County Division of Public Health, San Rafael, CA, USA. [20]Solano County Department of Public Health, Fairfield, CA, USA. [21]Sonoma County Department of Public Health, Santa Rosa, CA, USA. [22]Sacramento County Division of Public Health, Sacramento, CA, USA. [23]San Joaquin County Department of Public Health, Stockton, CA, USA. [24]San Francisco County Department of Public Health, San Francisco, CA, USA. [25]County of Santa Clara, Public Health Department, Santa Clara, CA, USA. [26]Department of Medicine, Division of Infectious Diseases, University of California, San Francisco, CA, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: charles.chiu@ucsf.edu

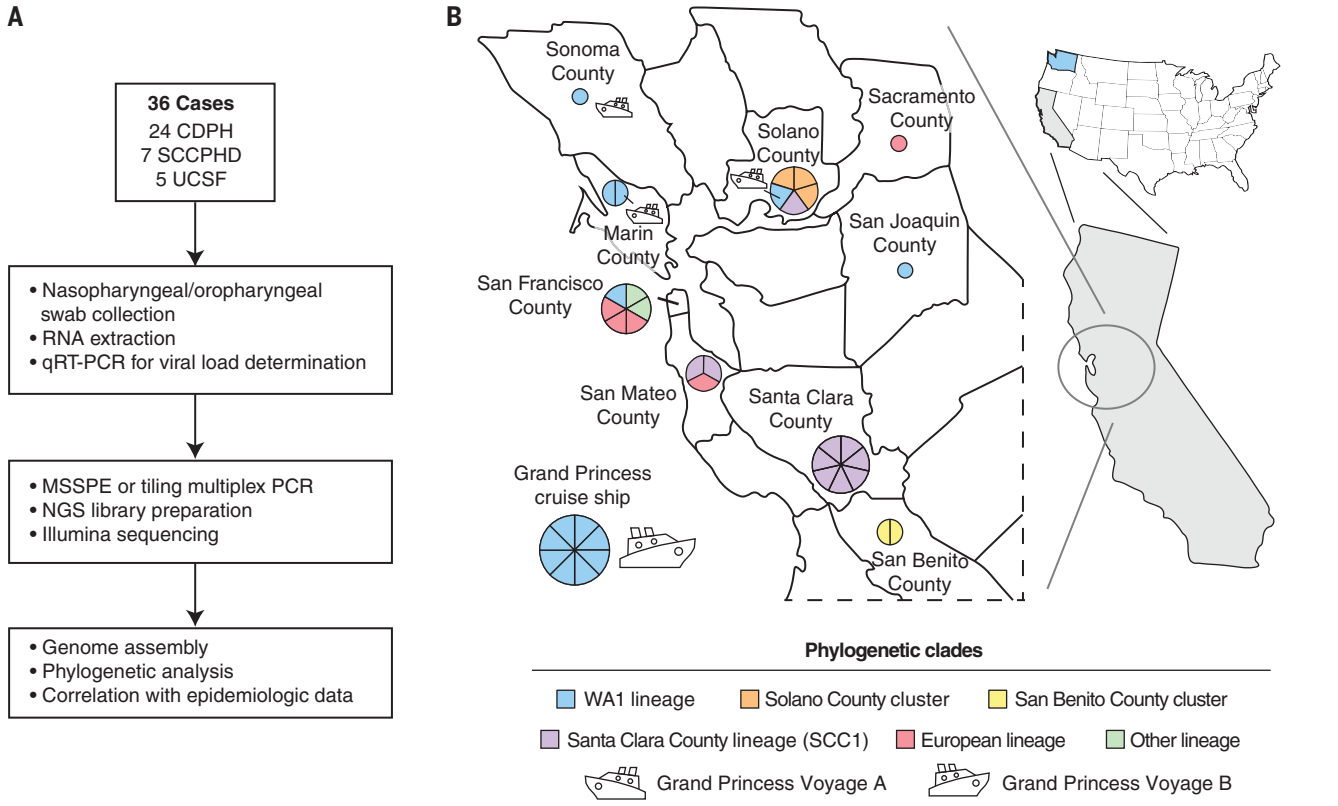


Fig. 1. Genomic survey of SARS-CoV-2 genomes in Northern California. (**A**) Analysis workflow. (**B**) Map of the Northern California survey region divided by county. The pie charts for each county are subdivided according to the number of patients whose viral genome was sequenced; colors correspond to viral lineages as determined by phylogenetic analysis. Passengers (n = 3) who were on the Grand Princess cruise ship during voyage A to Mexico and disembarked to return to their home communities are denoted by a ship icon facing left; passengers (n = 8) aboard the Grand Princess during voyage B to Hawaii are denoted by a ship icon facing right. Abbreviations: SCCPHD, Santa Clara County Public Health Department; CDPH, California Department of Public Health; UCSF, University of California, San Francisco; MSSPE, metagenomic sequencing with spiked primer enrichment; NGS, next-generation sequencing.

SARS-CoV-2 RNA genome, followed by metagenomic next-generation sequencing (mNGS) of pooled and indexed samples on Illumina NextSeq, HiSeq, or MiSeq instruments (*18*, *19*). The PCR cycle thresholds ranged from 15.3 to 33.4, corresponding to virus loads of 1.1×10^4 to 2.7×10^8 copies/ml (fig. S1 and table S2). An average of 31 million [interquartile ratio (IQR), 23 million to 57 million] and 2.2 ± 0.2 million reads were generated per sample for use in MSSPE and tiling multiplex PCR, respectively, and virus genomes were assembled by mapping to reference genome NC_045512 (Wuhan-Hu-1). The assembly yielded 34 SARS-CoV-2 genomes with genome coverage exceeding 65%, and these were included in the study. An additional two genomes sequenced from samples of a returning traveler from Wuhan, China and a household contact collected on 29 January by the CDC (CA3 and CA4) were also included in the analysis. The median coverage achieved across all samples was 97.7% (IQR, 90.4% to 99.7%).

Phylogenetic analysis revealed that the 36 SARS-CoV-2 genomes from California generated in this study were dispersed across the evolutionary tree of SARS-CoV-2 that was constructed from 789 worldwide genomes deposited into GISAID as of 20 March 2020 (Fig. 2A). The 36 genomes included 14 in the Washington state (WA1) lineage, 10 in a lineage associated with the Santa Clara County outbreak cluster (henceforth the SCC1 lineage), three from a Solano County cluster of three individuals, five related to lineages circulating in Europe and New York, and four related to early lineages from Wuhan or other regions of China (including two patients from San Benito County with identical genomes) (Figs. 1, 2A, and 3 and table S2).

A large outbreak was associated with travel on the Grand Princess cruise ship (with at least 78 confirmed positive cases out of 469 tested) as of 26 March (*20*). The Grand Princess undertook two consecutive voyages from San Francisco (voyage A to Mexico, 11 to 21 February; voyage B to Hawaii, 22 February to 4 March), with much of the same crew and a shared subset of passengers. Samples from 11 infected patients were sequenced, three of whom had been on voyage A and became sick after returning to their home county, and eight from crew members and passengers aboard the cruise ship on voyage B. Note that all 11 available sequenced genomes from the Grand Princess were part of the WA1 lineage (Fig. 2, A and B, and Fig. 3). In addition to sharing three single-nucleotide variants (SNVs) that define WA1 (C8782T, C18060T, and T28144C), the sequences from cruise ship passengers and crew also shared two additional SNVs, C17747T and A17858G, common to nearly all WA1 sequences sampled from Washington and California but not the basal WA1 case (Figs. 2B and 3).

The WA1 case was reported on 19 January (*6*) and thus substantially predated the voyages of the Grand Princess cruise ship (*7*, *20*). In addition, six of eight passengers on voyage B (UC7 to UC11 and UC30) each carried at least two new mutations (G16975T and C23185T) not observed in UC1, UC19, and UC20, who were all on voyage A (Fig. 3). This suggested that the virus from UC19 could be basally positioned relative to the cruise ship strains from voyage B, and that COVID-19 infections associated with voyage A may have been passed on to passengers and crew on voyage B. However, because of sequencing artifacts from areas of low coverage, the initial WA1 subtree extracted from the global maximum likelihood phylogenetic tree did not place UC19 basal to

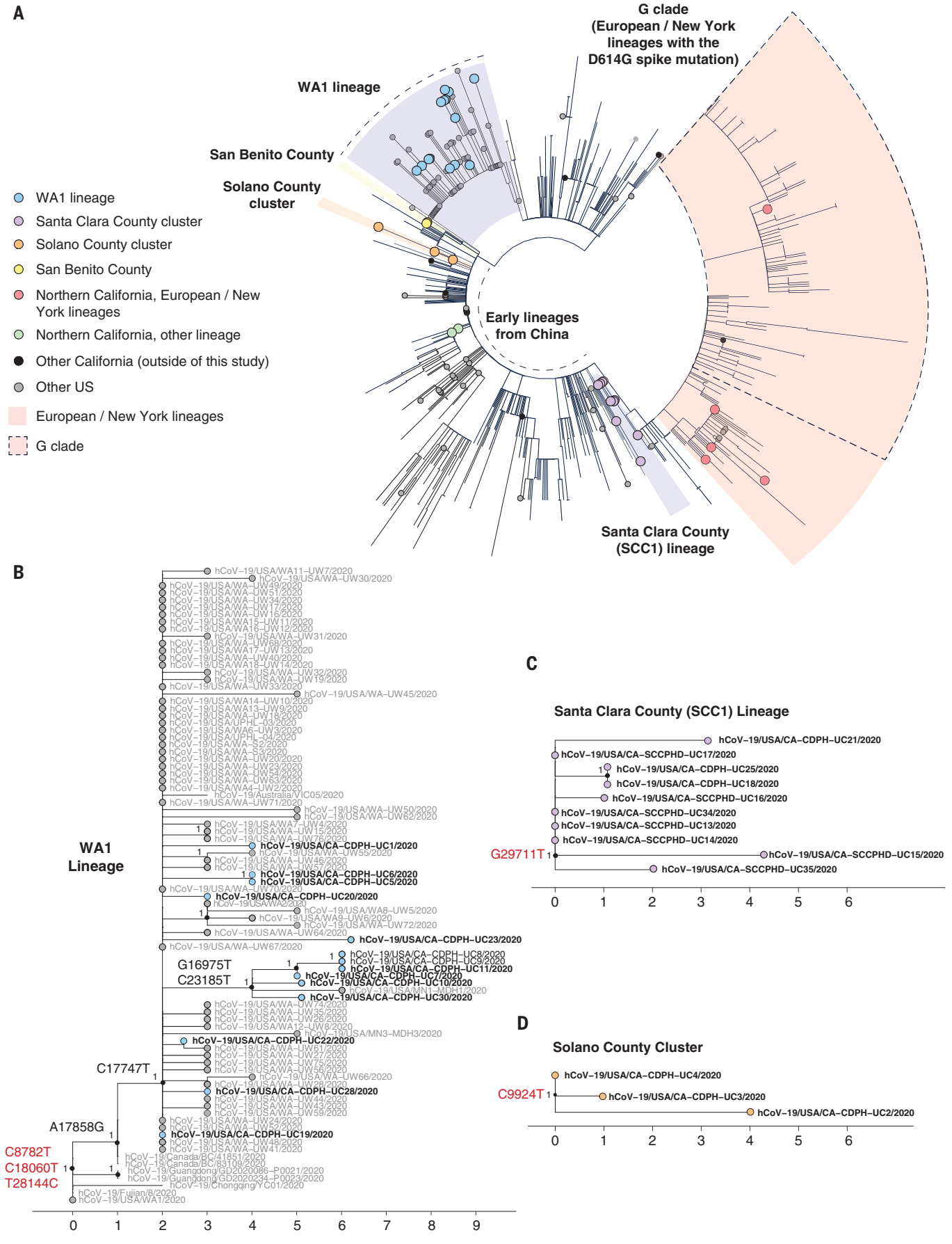

A
G clade
(European / New York
lineages with the
D614G spike mutation)
WA1 lineage
San Benito County
Solano County
cluster
Early lineages
from China
Santa Clara County
(SCC1) lineage
WA1 lineage
Santa Clara County cluster
Solano County cluster
San Benito County
Northern California, European / New York lineages
Northern California, other lineage
Other California (outside of this study)
Other US
European / New York lineages
G clade
B
WA1
Lineage
G16975T
C23185T
C17747T
A17858G
C8782T
C18060T
T28144C
hCoV-19/USA/CA-CDPH-UC1/2020
hCoV-19/USA/CA-CDPH-UC6/2020
hCoV-19/USA/CA-CDPH-UC5/2020
hCoV-19/USA/CA-CDPH-UC20/2020
hCoV-19/USA/CA-CDPH-UC23/2020
hCoV-19/USA/CA-CDPH-UC8/2020
hCoV-19/USA/CA-CDPH-UC9/2020
hCoV-19/USA/CA-CDPH-UC11/2020
hCoV-19/USA/CA-CDPH-UC7/2020
hCoV-19/USA/CA-CDPH-UC10/2020
hCoV-19/USA/CA-CDPH-UC30/2020
hCoV-19/USA/CA-CDPH-UC22/2020
hCoV-19/USA/CA-CDPH-UC28/2020
hCoV-19/USA/CA-CDPH-UC19/2020
0 1 2 3 4 5 6 7 8 9
C
Santa Clara County (SCC1) Lineage
G29711T
hCoV-19/USA/CA-CDPH-UC21/2020
hCoV-19/USA/CA-SCCPHD-UC17/2020
hCoV-19/USA/CA-CDPH-UC25/2020
hCoV-19/USA/CA-CDPH-UC18/2020
hCoV-19/USA/CA-SCCPHD-UC16/2020
hCoV-19/USA/CA-SCCPHD-UC34/2020
hCoV-19/USA/CA-SCCPHD-UC13/2020
hCoV-19/USA/CA-SCCPHD-UC14/2020
hCoV-19/USA/CA-SCCPHD-UC15/2020
hCoV-19/USA/CA-SCCPHD-UC35/2020
0 1 2 3 4 5 6
D
Solano County Cluster
C9924T
hCoV-19/USA/CA-CDPH-UC4/2020
hCoV-19/USA/CA-CDPH-UC3/2020
hCoV-19/USA/CA-CDPH-UC2/2020
0 1 2 3 4 5 6

Fig. 2. Phylogeny of SARS-CoV-2 lineages in California. (**A**) Phylogenetic tree of 753 SARS-CoV-2 genomes from GISAID (until 20 March 2020) along with the 36 genomes in this survey (see supplementary materials for tree file). Genomes from Northern California sequenced in the current study are denoted by colored circles; other genomes sequenced from California and from other U.S. states are denoted by black and gray circles, respectively. The name of each lineage or outbreak cluster is shown next to the arc line. (**B**) Phylogenetic subtree corresponding to the WA1 lineage. This subtree was reconstructed from 88 SARS-CoV-2 genomes after removal of ambiguous nucleotide sites that had generated low-coverage artifacts (see text). The WA1 virus from Washington state (first case in the United States) is at the root of the subtree along with a virus sequenced in China. The UC19 virus (from a Grand Princess voyage A passenger) is basal to the viruses sequenced from crew members and passengers on voyage B. (**C**) Zoomed view of the SCC1 lineage associated with the Santa Clara County outbreak cluster. (**D**) Zoomed view of the Solano County cluster. The *x* axis shows the number of base substitutions relative to the root of the phylogenetic tree. In (B) to (D), the key single-nucleotide variants (SNVs) defining a lineage or cluster are shown in red or black text. Bootstrap values (converted from the approximate likelihood ratio test, or aLRT score) are displayed at each node, with a value of 1 indicating 100% support.

sequences from voyage B passengers (fig. S2). To establish a more accurate tree topology, we therefore reconstructed a new phylogenetic subtree of the WA1 lineage after excluding all ambiguous sites. In this new subtree (Fig. 2B), UC19 is basal to all other California genomes within the WA1 lineage. In addition, among the sequences from patients on voyage B, UC5 and UC6 group together, whereas UC7 to UC11 and UC30 group together with a sequence sampled in Minnesota.

The chronology and phylogeny of the cruise ship outbreak, and the predominance of the WA1 lineage in Washington state (*7*), together suggest that the virus on the Grand Princess likely came from Washington, although the cases may also have originated from a different region in which the WA1 strain is circulating. In addition to passengers and crew members aboard the Grand Princess, virus genomes sampled from three cases of community transmission in different counties of the San Francisco Bay Area (UC22, UC23, and UC28) were also of the WA1 lineage. UC22 was the son of an infected Grand Princess passenger (UC20) on voyage A and most likely contracted the virus from household contact. The UC23 and UC28 cases may also reflect transmission from disembarking Grand Princess passengers on voyage A, or preexisting circulation of the WA1 strain in the community.

Three patients examined in this study (CA3, CA4, and UC12) had COVID-19 infections associated with international travel or exposure to international travelers. CA3 corresponds to a resident of San Benito County who became sick shortly after returning from Wuhan, China in late January. The sequence of his SARS-CoV-2 genome is identical to that of CA4, a household contact who was also infected with the virus. Their viral genomes were found to be closely related to early lineages from China (Fig. 2A and data S1). UC12 had a prolonged exposure to a known positive traveler from Switzerland while attending a conference. The genome from UC12 fell within a lineage containing many sequences from European residents or travelers from Europe (Fig. 2A). Interestingly, four additional genomes (UC24, UC26, UC27, and UC36) were also grouped within the European lineage. UC27 and UC36 were both diagnosed shortly after returning to California from New York, consistent with reports that the New York outbreak that began in March 2020 originated with travelers coming from Europe (*21*, *22*). UC26 also reported domestic travel from Los Angeles, whereas UC24 had no known travel history.

In Santa Clara County, we sequenced seven genomes from individuals who were part of a local outbreak of COVID-19 at a large workplace facility with multiple employers, large areas of shared space, and heavy pedestrian traffic. The genomes all shared the G29711T SNV that defines the SCC1 lineage (Figs. 2C and 3). Four employees (UC13, UC14, UC15, and UC34) had dates of symptom onset within 2 weeks of each other, although they did not know each other. UC16 and UC17 were family members of UC13 and lived in the same residence, while UC35 transported UC14 to the hospital via emergency medical services. Notably, the genomes from a Solano County resident (UC21) and a San Mateo County couple (UC18 and UC25) were also placed in the SCC1 lineage, suggesting possible spread to different counties. Further epidemiological investigation found that UC21 had visited a merchant in Santa Clara County, during which he likely became infected.

In Solano County, a small cluster of three cases included the first reported instance of community transmission in the United States on 26 February (UC4) (Figs. 2D and 3). The two other cases (UC2 and UC3) were health care workers who were taking care of patient UC4 and likely contracted the disease in the hospital, consistent with transmission of the disease from patients to health care providers (*23*). The genomic epidemiology of the COVID-19 cases associated with community spread studied here do not show any predominant SARS-CoV-2 lineage circulating in Northern California. In California, multiple recent and unrelated introductions of SARS-CoV-2 into the state via different routes appear to have given rise to the diversity of virus lineages reported in this study, with no single predominant lineage observed. We note that this does not exclude the possibility of cryptic transmission of multiple lineages in California simultaneously, as the current level of sampling is not dense enough to confidently estimate the dates of the seeding events, nor the subsequent periods of cryptic transmission before a lineage was identified.

There is growing evidence that WA1 is now an established lineage of SARS-CoV-2 in the United States. Here, we found WA1-lineage viruses from Grand Princess cruise ship passengers as well as from residents of several Northern California counties. In addition, WA1-lineage viruses have been identified in COVID-19 cases from many states including Minnesota, Connecticut, Utah, Virginia, and New York (*24*, *25*). The early date and basal phylogenetic position of the WA1 virus make it likely that the direction of dissemination was from Washington state to California and other states. Notably, SARS-CoV-2 sequences from Connecticut (*25*) and British Columbia, Canada (Fig. 2B) are positioned close to the root of the subtree containing the WA1 sequences, raising the possibility that the virus may not have been first introduced into the United States via Washington state.

SARS-CoV-2, like other coronaviruses, contains a nonstructural gene with proofreading activity (*26*). Consequently, the virus evolves more slowly than many other human RNA viruses, on the order of one to two DNA base substitutions per month across its ~29-kb genome (*27*). Thus, only one to three SNVs in general are needed to define a distinct lineage. The WA1 lineage consists of three key SNVs (C8782T, C18060T, and T28144C), whereas the SCC1 lineage associated with the Santa Clara County cluster and the Solano County cluster are each defined by only one SNV (G29711T and C9924T), respectively (Figs. 2 and 3).

Our epidemiological and genomic survey of SARS-CoV-2 has several limitations. First, this initial analysis represents a relatively sparse sampling of cases. Undersampling of virus genomes is due in part to the high proportion of cases (~80%) with asymptomatic or mild disease who do not get tested (*2*–*4*). Second, the majority of samples analyzed were obtained from public health laboratories and thus may not be representative of the general population. Finally, phylogenetic grouping of viruses from different locations, such as Washington state and California in the same WA1 lineage, does not prove the directionality of spread. Despite this, our study shows that robust insights into COVID-19 transmission are achievable if virus genomic diversity is combined and jointly interpreted with detailed epidemiological

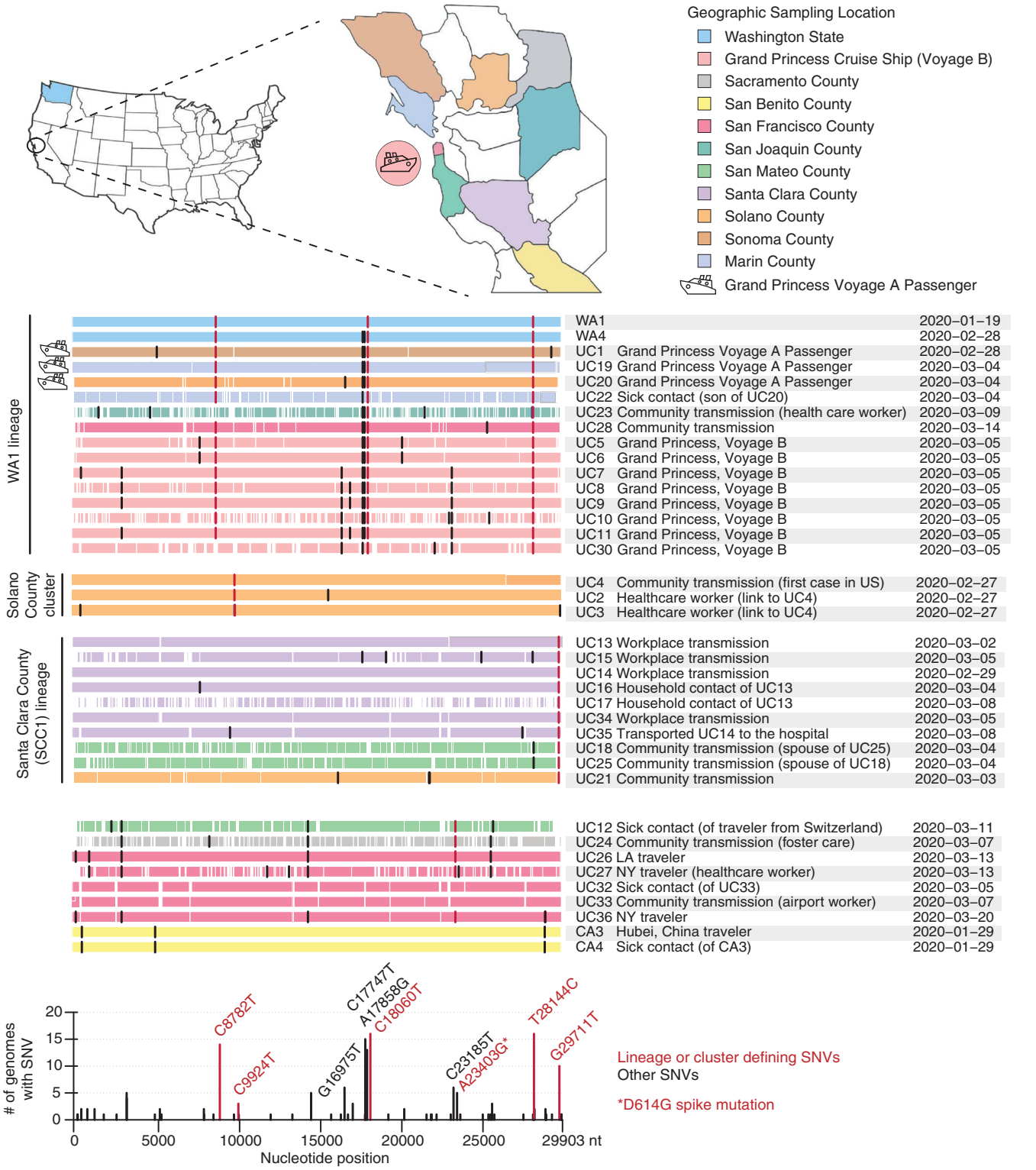


Fig. 3. Multiple sequence alignment of all SARS-CoV-2 genomes reported across nine Northern California counties and the Grand Princess cruise ship. SNVs with respect to the reference genome (NC_045512) are shown as vertical red and black lines for lineage-defining SNVs and other SNVs, respectively. Cases that are part of the WA1 lineage include the first case of COVID-19 infection (WA1) in the United States, eight passengers and crew members aboard the Grand Princess cruise ship during its second trip (voyage B), and three individuals surveyed from three Northern California counties as passengers on the ship's first trip (voyage A). The three SNVs C8782T, C18060T, and T28144C define the WA1 lineage, and the two SNVs C17747T and A17858G are common to Grand Princess passengers and crew. Viruses from voyage B passengers and crew share SNVs G16975T and C23185T that are lacking in viruses from voyage A passengers. Single SNV variants C9924T and G29711T define the lineages from Solano County and Santa Clara County, respectively. European lineages share SNV A23403G. The putative epidemiological link and sample collection date are shown beside each sequence alignment.

case data. In particular, we found that a returning traveler from New York was infected with a lineage circulating widely in Europe, thus suggesting an association between the New York outbreak and intercontinental travel to and from Europe before this was widely recognized (*21*, *22*).

Public health containment measures such as isolation and contact tracing, as performed in the Solano County and Santa Clara County outbreak clusters, become more difficult to maintain once a lineage becomes established in the community. Our data suggest trends in this direction, such as the association between the WA1 lineage and community-acquired COVID-19 cases in several counties of Northern California, and the detection of a virus from the SCC1 lineage in residents of neighboring San Mateo and Solano counties. Social distancing interventions, such as the "shelter-in-place" directive that was issued by the governor of California on 20 March 2020, may have assisted in stemming spread from community to community. Interstate dissemination of SARS-CoV-2 lineages has also been demonstrated coast-to-coast between Washington state and Connecticut (*25*), and from domestic and international travel into the Bay Area in the current study. Suspension of non-essential travel may help to prevent importation of new cases in California and other states.

REFERENCES AND NOTES

1. E. Dong, H. Du, L. Gardner, *Lancet Infect. Dis.* **20**, 533–534 (2020).
2. J. F. Chan *et al.*, *Lancet* **395**, 514–523 (2020).
3. R. Li *et al.*, *Science* **368**, 489–493 (2020).
4. R. Lu *et al.*, *Lancet* **395**, 565–574 (2020).
5. A. Otani, P. Santilli, *Wall Street Journal*, 29 February 2020.
6. T. Bedford *et al.*, medRxiv 20051417 [preprint]. 16 April 2020.
7. M. L. Holshue *et al.*, *N. Engl. J. Med.* **382**, 929–936 (2020).
8. C. Fraser *et al.*, *Science* **324**, 1557–1561 (2009).
9. J. L. Gardy, N. J. Loman, *Nat. Rev. Genet.* **19**, 9–20 (2018).
10. R. A. Urbanowicz *et al.*, *Cell* **167**, 1079–1087.e5 (2016).
11. S. E. Mate *et al.*, *N. Engl. J. Med.* **373**, 2448–2454 (2015).
12. N. R. Faria *et al.*, *Nature* **546**, 406–410 (2017).
13. J. Thézé *et al.*, *Cell Host Microbe* **23**, 855–864.e7 (2018).
14. N. D. Grubaugh *et al.*, *Nature* **546**, 401–405 (2017).
15. X. Deng *et al.*, *Nat. Microbiol.* **5**, 443–454 (2020).
16. Centers for Disease Control and Prevention, CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel, Revision 03 (30 March 2020); fda.gov/media/134922/download.
17. J. Quick *et al.*, *Nat. Protoc.* **12**, 1261–1276 (2017).
18. C. Y. Chiu, S. A. Miller, *Nat. Rev. Genet.* **20**, 341–355 (2019).
19. W. Gu, S. Miller, C. Y. Chiu, *Annu. Rev. Pathol.* **14**, 319–338 (2019).
20. A. S. Gonzalez-Reiche *et al.*, *Science* **368**, eabc1917 (2020).
21. M. T. Maurano *et al.*, medRxiv 20064931 [preprint]. 23 April 2020.
22. A. Brufsky, *J. Med. Virol.* jmv.25902 (2020).
23. L. F. Moriarty *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **69**, 347–352 (2020).
24. M. Klompas, *Ann. Intern. Med.* **172**, 619–620 (2020).
25. J. Hadfield *et al.*, *Bioinformatics* **34**, 4121–4123 (2018).
26. J. R. Fauver *et al.*, *Cell* **181**, 990–996.e5 (2020).
27. M. Bouvet *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 9372–9377 (2012).
28. V. Hill, A. Rambaut, Virological.org (6 March 2020); https://virological.org/t/phylodynamic-analysis-of-sars-cov-2-update-2020-03-06/420.
29. W. Gu *et al.*, *Clin. Infect. Dis.* ciaa599 (2020).
30. L. du Plessis *et al.*, Zenodo DOI:10.5281/zenodo.3922369 (2020).

ACKNOWLEDGMENTS

We acknowledge the help and advice of J. T. Lee at the U.S. CDC. We thank all of the authors who have contributed genome data on GISAID. Author credits for specific GISAID contributions can be found at www.gisaid.org/. Clinical samples from UCSF were processed according to protocols approved by the UCSF Institutional Review Board (protocol numbers 10-01116 and 11-05519). Samples collected by the CDPH were de-identified and deemed not research or exempt by the Committee for the Protection of Human Subjects (project number 2020-30) issued under the California Health and Human Services Agency's Federal Wide Assurance #00000681 with the Office of Human Research Protections. A nonresearch determination for this project was also granted by Sonoma County as SARS-CoV-2 genome sequencing was designated an epidemic disease control activity, with collected data directly related to disease control. **Funding:** Supported by NIH grants R33-AI129455 (C.Y.C.) from the National Institute of Allergy and Infectious Diseases and K08-CA230156 (W.G.) from the National Cancer Institute, the California Initiative to Advance Precision Medicine (C.Y.C.), the Charles and Helen Schwab Foundation (C.Y.C.), the Burroughs-Wellcome CAMS Award (W.G.), the Oxford Martin School and the European Research Council under the Seventh Framework Program of the European Commission (Pathogen Phylodynamics grant 614725) (O.G.P. and L.d.P.), Wellcome Trust and Royal Society (204311/Z/16/Z), and Medical Research Council–São Paulo Research Foundation (MR/S0195/1, FAPESP 18/14389-0). **Author contributions:** C.Y.C. conceived, designed, and supervised the study; A.G., A.S.-G., C.W., C.-Y.P., G.Y., H.G., V.S., and X.D. performed experiments; C.Y.C., S.F., and B.B.. assembled and curated the viral genomes; C.Y.C., W.G., and X.D. analyzed data; C.Y.C., E.H., K.Z., S.M., and W.G. collected patient samples at UCSF; D.M. and G.A. analyzed genomic and epidemiologic data; T.B., A.L.G., P.R., L.M.S., M.F., H.Y.C., J.S., and K.R.J. collected, assembled, and provided viral genome data from Washington and contributed to the phylogenetic analysis; C.A., K.G., M.Z., E.S., and K.G.A. provided viral genome data from Southern California; C.R.P., J.Z., S.T., and Y.L. sequenced and analyzed viral genomes at the CDC; L.d.P., N.F., and O.G.P. performed phylogenetic analysis of genomes; A.T.Y., B.T.M., B.B., C.C., D.A.W., E.V., G.M., M.P., M.W., O.K., S.J.C., and S.M. collected samples, extracted the viral RNA, and/or provided epidemiology data from counties in California; C.Y.C., W.G., and X.D. wrote the manuscript; and C.Y.C., X.D., S.F., C.W., D.A.W., L.d.P., O.G.P., and W.G. edited the manuscript. All authors read the manuscript and agree to its contents. **Competing interests:** C.Y.C. is the director of the UCSF-Abbott Viral Diagnostics and Discovery Center and receives research support funding from Abbott Laboratories. C.Y.C. and X.D. are inventors on a patent application on the MSSPE method titled "Spiked Primer Design for Targeted Enrichment of Metagenomic Libraries" (U.S. Application No. 62/667,344, filed 05/04/2018 by University of California, San Francisco). H.Y.C. is a consultant for Merck and GlaxoSmithKline and receives research funding from Sanofi-Pasteur, Ellume, and Cepheid, unrelated to this work. All other authors have no conflicts to declare. The opinions expressed by the authors contributing to this journal do not necessarily reflect the opinions of the Centers for Disease Control and Prevention or the institutions with which the authors are affiliated. **Data and materials availability:** Assembled SARS-CoV-2 genomes in this study were uploaded to GISAID (*28*, *29*) as FASTA files (accession numbers in table S2) and can be visualized on a continually updated phylogenetic tree on NextStrain (*24*). Viral genomes were submitted to the National Center for Biotechnology Information (NCBI) GenBank database (accession numbers MT419827 to MT419860). Raw sequence data were submitted to the NCBI Sequence Read Archive (SRA) database (BioProject accession number PRJNA 629889 and umbrella BioProject accession number PRJNA171119). Locations of SNVs aligned to the reference sequence (NC_045512), was done by custom scripts (*30*). This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/582/suppl/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 to S5
Data S1
References (*31*–*39*)

View/request a protocol for this paper from *Bio-protocol*.

27 March 2020; accepted 3 June 2020
Published online 8 June 2020
10.1126/science.abb9263

SOLAR PHYSICS

A physics-based method that can predict imminent large solar flares

Kanya Kusano[1]*, Tomoya Iju[2], Yumi Bamba[1,3], Satoshi Inoue[1]

Solar flares are highly energetic events in the Sun's corona that affect Earth's space weather. The mechanism that drives the onset of solar flares is unknown, hampering efforts to forecast them, which mostly rely on empirical methods. We present the κ-scheme, a physics-based model to predict large solar flares through a critical condition of magnetohydrodynamic instability, triggered by magnetic reconnection. Analysis of the largest (X-class) flares from 2008 to 2019 (during solar cycle 24) shows that the κ-scheme predicts most imminent large solar flares, with a small number of exceptions for confined flares. We conclude that magnetic twist flux density, close to a magnetic polarity inversion line on the solar surface, determines when and where solar flares may occur and how large they can be.

Solar flares are the sudden emission of electromagnetic radiation, and they eject plasma and energetic particles from the solar corona (the Sun's outer atmosphere). Solar flares are thought to be driven by the rapid release of magnetic energy stored in active regions (ARs) around visible sunspots. The energy and plasma released in solar flares (and associated phenomena) drive space weather disturbances experienced at Earth (*1*). Many physical models of solar flares have been proposed (*2*), but it remains unclear which parameters of the magnetic field determine the occurrence and properties of each flare. This makes it difficult to accurately predict the occurrence of solar flares (*3*), and most forecast methods rely on empirical methods (*4*).

The morphological structure of magnetic fields in eruptive solar flares are theoretically described by the Carmichael-Sturrock-Hirayama-Kopp-Pneuman (CSHKP) model (*5*), in which the reconnection of magnetic field lines (magnetic reconnection) in the solar corona produces a bright arcade over the polarity inversion line (PIL), the location where the vertical magnetic field on the solar surface changes sign. A plausible mechanism for the energy release in solar flares is the positive feedback between magnetohydrodynamic (MHD) instabilities and magnetic reconnection, which reinforce each other and result in runaway energy liberation from the magnetic field (*6*, *7*). Therefore, an MHD instability might be necessary for the onset of solar flares. Previous work has studied MHD instabilities as the driver of solar flares, suggesting the involvement of electric current–driven instabilities—torus mode and kink mode—in the onset of solar flares [e.g., (*8*–*10*)]. However, the relationship between the torus and kink instabilities and the occurrence of solar flares remains ambiguous, so predictions of solar flares based on the theory of MHD instabilities has not yet been demonstrated (*11*, *12*).

Other work has proposed that a different instability—the double-arc instability (DAI)—could be responsible for the initial ignition of a flare (*13*). The DAI is driven by the Lorenz force acting on electric current–carrying magnetic loops in the solar corona, which is similar to the torus instability (*14*). The DAI explicitly accounts for the triggering effect of magnetic reconnection, unlike the torus instability model. We consider the following scenario: A small-scale reconnection, which we refer to as the trigger-reconnection, occurs between electric current–carrying magnetic loops above the PIL, causing the formation of a double arc (DA) magnetic loop (Fig. 1, A and B). The DA provides the electric current–carrying magnetic flux, which contains part of the free energy that is released by flares and eruptive events. The DA loop with an electric current may correspond to previously observed structures known as sigmoids (*15*, *16*). If the DAI grows, the concave part of the DA begins to rise, causing further reconnection and a positive feedback that releases the free energy in the current-carrying overlying magnetic flux (Fig. 1, C and D). The topology of this scenario is the same as that of the tether-cutting reconnection (*17*), which has been used to explain the role of reconnection in reducing the inward magnetic tension force (*7*). If the DAI is responsible for the triggering of flares, the trigger-reconnection on (or near) the PIL could determine the occurrence of flares and their properties. Simulations support this interpretation (*18*).

If this scenario is correct, determining the conditions necessary for the trigger-reconnection

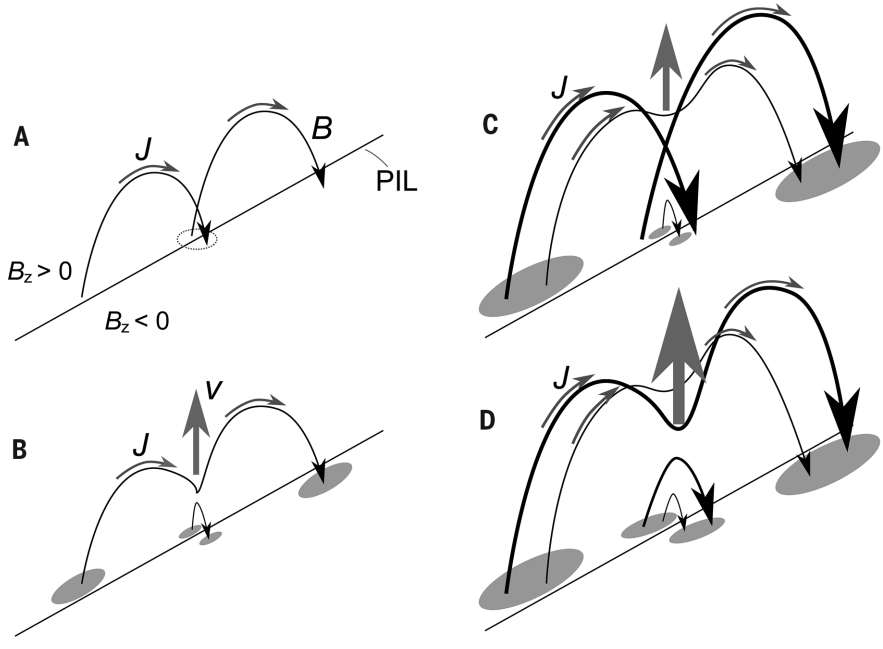


Fig. 1. Schematic illustration of our scenario for the magnetic energy release in a solar flare. (**A** to **D**) The proposed sequence of events in the initial stage of a solar flare, in which the connectivity of magnetic field lines changes through magnetic reconnection while magnetic flux on the solar surface is conserved. The curved black lines represent the magnetic field (*B*), and the curved gray arrows are the electric current (*J*). The vertical arrows show the ascending motion (*V*) caused by the magnetohydrodynamic instability. The straight line represents the magnetic PIL—the boundary between positive and negative values of the vertical magnetic field B_z on the solar surface. Ellipses illustrate the bright flare ribbons on the foot points of reconnected magnetic field lines.

[1]Institute for Space-Earth Environmental Research, Nagoya University, Nagoya 464-8601, Japan. [2]National Astronomical Observatory of Japan, Mitaka 181-8588, Japan. [3]Institute for Advanced Research, Nagoya University, Nagoya 464-8601, Japan.
*Corresponding author. Email: kusano@nagoya-u.jp

could allow us to infer how close the magnetic field is to the critical state of the instability. If the magnetic field is in a less stable configuration, such that a small-scale reconnection event could trigger a flare, then this would indicate that a solar flare is imminent. We propose a flare prediction model, which we call the κ-scheme, based on this concept.

The magnetic field becomes unstable with respect to the DAI when a parameter κ exceeds a threshold κ_0, which is ~0.1 (*13*). This condition is both necessary and sufficient for the DAI (*13*). Based on this criterion, the κ-scheme derives the critical length scale, r_c, for the region of the trigger-reconnection to become unstable to the DAI at any point on the PIL (*19*). We assume that the trigger-reconnection occurs within a circle centered on a point on the PIL (Fig. 1A) using a three-dimensional magnetic field model, based on the nonlinear force-free field (NLFFF) approximation, extrapolated from the observed vector magnetic field data for the solar surface (*20*). The κ-scheme can also be used to estimate the magnetic free energy, E_r, stored in the magnetic field overlying the DA magnetic loop (see eq. S4). This energy corresponds to the minimum energy available for release by the DAI and magnetic reconnection when the DAI grows. We also evaluate the free magnetic energy, E_{Hi}, contained in the i-th largest high free-energy regions (HiFERs), which are subdomains in an AR where the nonpotential magnetic field, B_{np}, is stronger than a threshold B_0 = 1000 G, and the free energy of a whole AR, E_{AR} (see eq. S5).

We investigate the utility of the κ-scheme for predicting solar flares by applying it to data from the Helioseismic and Magnetic Imager (HMI) on the Solar Dynamics Observatory (SDO) spacecraft (*21*). We adopt measurements of the vector magnetic field at the solar surface from the Space-weather HMI Active Region Patch (SHARP) dataset (*22*). First, we apply the κ-scheme to National Oceanic and Atmospheric Administration (NOAA) AR 12673, an AR that produced four X-class flares in 2017: X2.2 at 08:57 universal time (UT) and X9.3 at 11:53 UT on 6 September, X1.3 at 14:20 UT on 7 September (Fig. 2A), and X8.2 on the solar limb at 15:35 UT on 10 September (*23*). X-class is the largest classification of solar flare, defined as having peak energy flux in soft x-rays (wavelength 1 to 8 Å) larger than 10^{-4} W m^{-2}. The number in the classification indicates multiples of that threshold, e.g., the X9.3 flare reached a peak soft x-ray flux of 9.3×10^{-4} W m^{-2}. This X9.3 flare was the largest solar flare observed between 2006 and 2019.

Figure 2, C and D, and movie S1 show the evolution of E_r and r_c for points on the PILs in HiFERs of this AR from 1 day before the flares on 6 September to half a day later. The estimated free energy of this AR (E_{AR}) reached 10^{33} erg by 12:00 UT on 5 September, 21 hours before the X2.2 flare, then continued to increase. The minimum critical size for trigger-reconnection, r_c, for a DAI that can release energy ($E_r > 10^{32}$ erg) fell from about 1.5 Mm to <1 Mm over the 24 hours before the onset of the X2.2 flare at 08:57 UT (Fig. 2B). This indicates that the stability decreased shortly before the onset of the flare. The point of minimum r_c (for $E_r > 10^{32}$ erg) occurred at 08:48 UT on 6 September (labeled point a in Fig. 2C), reaching r_c ~ 0.7 Mm and $E_r \sim 2 \times 10^{32}$ erg and indicating that an X-class flare could occur at this point. Point a was located at the north end of the PIL in the largest HiFER of this AR (Fig. 3, A and C). This corresponds with the central portion of the first brightening ribbons of the X2.2 flare (Fig. 3B), i.e., the initial reconnection began around point a, as predicted. After the flare, the value of r_c at point a increased, and no points with $r_c < 1$ Mm and $E_r > 10^{31}$ erg remained by 10:00 UT (Fig. 2D). This implies that stability returned after the flare, even though free energy (E_{Hi} and E_{AR}) of more than 2×10^{33} erg remained.

Stability began to decrease again more than an hour later. The lowest value of r_c for $E_r > 10^{32}$ erg occurred at point b in Fig. 2D, reaching an r_c of 0.9 Mm while E_r was 4×10^{32} erg at 11:48 UT, 5 min before the X9.3

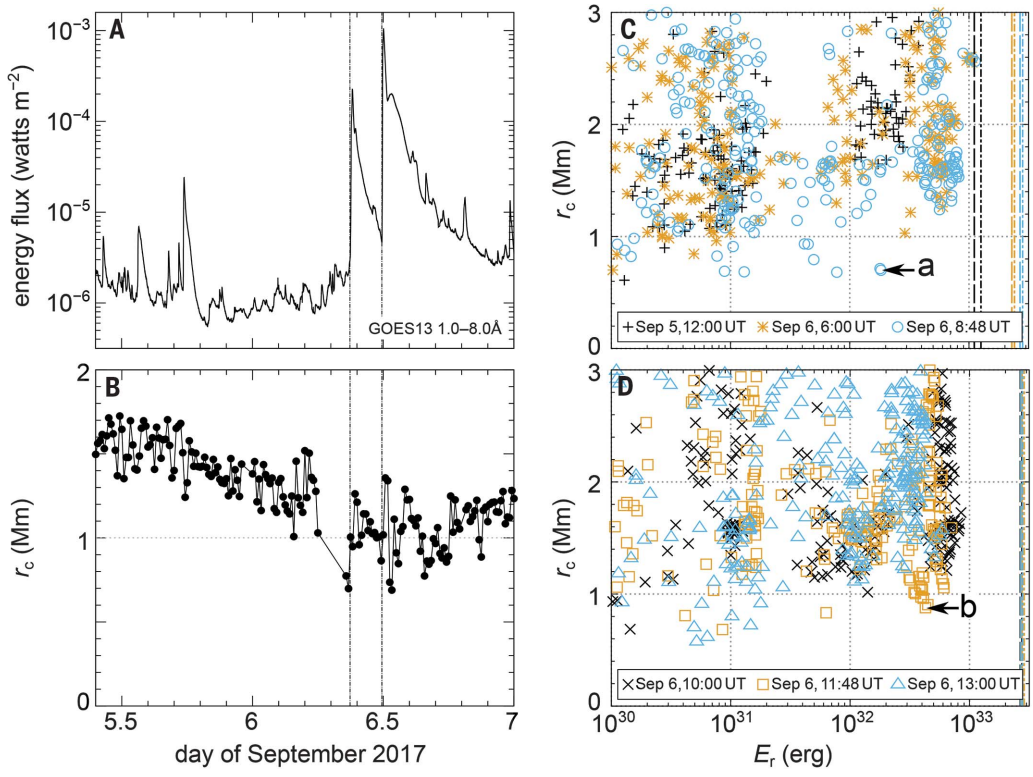


Fig. 2. Soft x-ray light curve and stability analysis results using the κ-scheme for AR 12673. (**A**) Light curve of the soft x-ray (1.0 to 8.0 Å) flux observed by the Geostationary Operational Environmental Satellite 13 (GOES-13) (*31*) from 09:36 UT on 5 September 2017 to 00:00 UT on 7 September. (**B**) The time evolution of the minimum critical radius of the trigger-reconnection, r_c, for $E_r > 10^{32}$ erg. The two vertical dot-dashed lines indicate the onset time of the two flares larger than class X2. (**C** and **D**) Scatter plots for E_r and r_c for points on the PIL in HiFERs. The black plus symbols, orange asterisks, and blue circles in (C) show the results at 12:00 UT on 5 September, 06:00 UT on 6 September, and 08:48 UT on 6 September 2017, respectively. The black ×'s, orange squares, and blue triangles in (D) show the results at 10:00 UT, 11:48 UT, and 13:00 UT on 6 September 2017, respectively. The vertical dashed and dot-dashed lines indicate the total free energies of the largest HiFER and the AR (E_{Hi} and E_{AR}), respectively, with the color of each line corresponding to the symbol colors. We identify points a and b as the locations predicted to produce large flares, because they have the lowest values of r_c for $E_r > 10^{32}$ erg at each moment. Movie S1 shows the time evolution of these values.

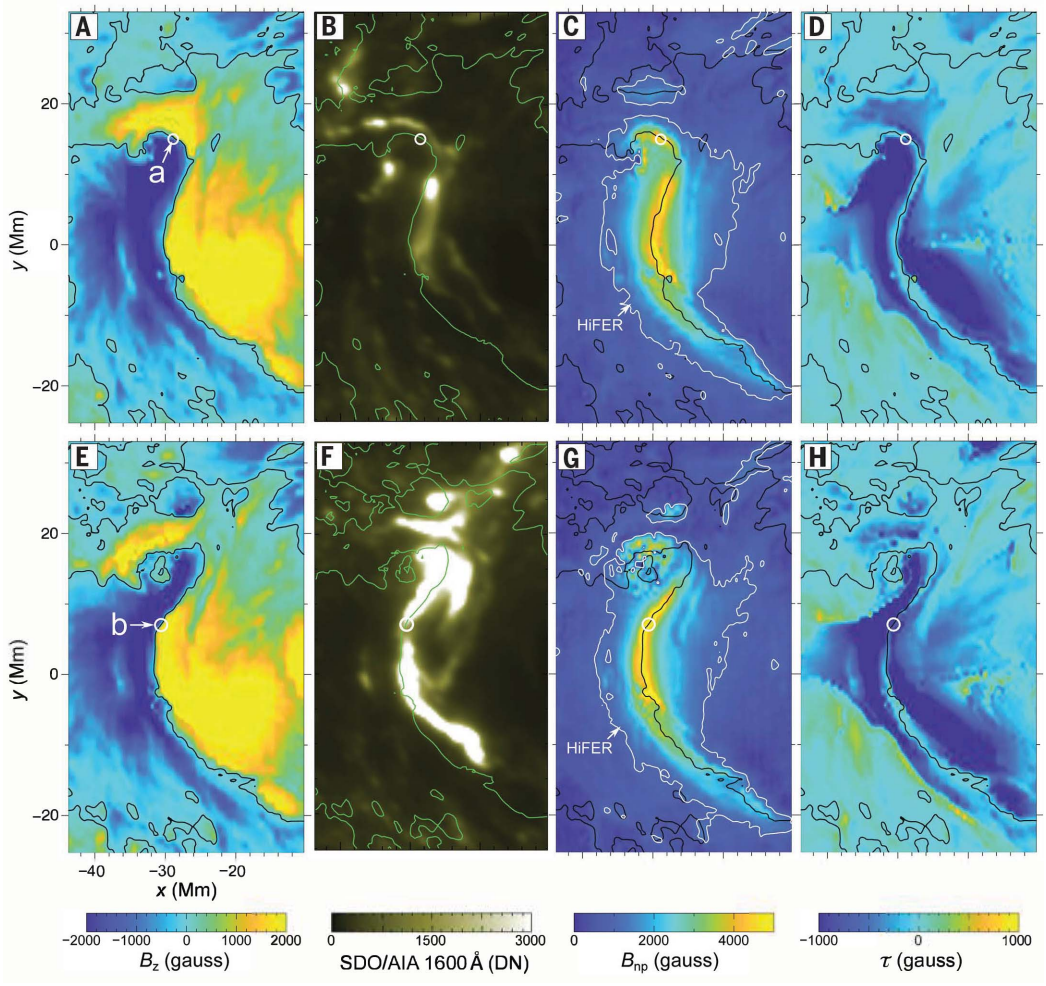


Fig. 3. Maps of the observed parameters for AR 12673 on 6 September 2017. (**A**, **C**, and **D**) The magnetic parameters B_z, B_{np}, and τ, respectively, at 08:48 UT before the X2.2 flare. (**B**) The initial flare ribbon of the X2.2 flare observed by SDO's Atmospheric Imaging Assembly (AIA) (*32*) at 1600 Å at 08:59:26 UT. The SDO-AIA 1600-Å signal is enhanced in the lower atmosphere of the Sun by the precipitation of nonthermal particles accelerated by magnetic reconnection, forming flare ribbons on the foot point of reconnected field lines (see Fig. 1). (**E**, **G**, and **H**) B_z, B_{np}, and τ, respectively, at 11:48 UT before the X9.3 flare. (**F**) The initial flare ribbon of the X9.3 flare observed by SDO-AIA 1600 Å at 11:55:02 UT. Points a and b (white circles) are the locations identified in Fig. 2, where a large flare is predicted to occur. The radius of each circle equals the critical radius (r_c) of the flare-triggering reconnection. Black contours in (A), (C) to (E), (G), and (H) and green contours in (B) and (F) are the PILs. White contours in (C) and (G) are the boundaries of the HiFER where $B_{np} = 10^3$ G. The coordinates *x* and *y* are measured from the center of the SHARP data.

flare. The initial two ribbons of the X9.3 flare expanded from point b, which was located south of point a (Fig. 3, E and F). After the flare, no points remained with $r_c < 1$ Mm and $E_r > 2 \times 10^{31}$ erg (Fig. 2C).

These results are consistent with the triggered instability model and suggest that the differing flare sizes may be due to the positional difference in the least stable points. Point a was located at the north edge of the HiFER (Fig. 3C), so the X2.2 flare could only release the free energy in the north part of the HiFER. By contrast, point b was located in the main part of the HiFER (Fig. 3G), allowing the release of free energy over a wider area during the X9.3 flare. The positional difference between points a and b was due to the distribution of the magnetic twist flux density τ (see equation S3) near the PIL (Fig. 3, D and H). The high τ region moved close to the PIL at points a and b, causing the reduction in r_c at these points. This suggests that the small-scale structure of τ near the PIL can determine the size of flares as well as their location.

We evaluated the sensitivity of r_c and E_r to uncertainties in the observations using a Monte Carlo analysis. We performed the same calculation as in Fig. 2, C and D, using 16 synthetic observations, constructed by adding noise to the real observation at each pixel (*14*). We find that the variance of r_c and E_r at points a and b is small enough that the prediction is insensitive to the observational uncertainties (fig. S1). This is because r_c and E_r are calculated from the integrals of magnetic quantities (equations S1 and S4).

The results for AR 12673 suggest that large flares are produced at the PIL if it satisfies the conditions $r_c < 1$ Mm and $E_r > 10^{32}$ erg. We extended this analysis to additional ARs in two groups. Group 1 consists of 198 ARs with the largest sunspot areas recorded from 2006 to 2019 during solar cycle 24 (table S1). Group 1 included only data for each AR observed at 00:00 UT for the day when the AR was located nearest the central meridian of the Sun. None of the ARs of group 1 produced solar flares larger than class X2 within 20 hours after the observation. Group 2 comprises the seven ARs that produced all flares of class X2 or stronger (nine events; table S2) within solar latitudes ±50° during solar cycle 24. We excluded flares occurring near and behind the solar limb because of the difficulty in observing the magnetic field. We analyzed the observations of the ARs of group 2 for 1.5 to 5.5 days, spanning the time of the large flares, at a cadence of 1 hour or 12 min according to the available observations (table S2).

Figure 4, A and B, shows r_c and E_r for all points on the PILs in HiFERs of groups 1 and 2. None of the ARs in group 1 satisfied the conditions $r_c < 1$ Mm and $E_r > 2 \times 10^{31}$ erg, and most of the points had a higher r_c than expected for self-similar scaling, i.e., $E_r \propto l^3$ and $r_c \propto l$ for a length scale l. By contrast, six of the seven ARs in group 2 (with the only exception being AR 12192) satisfied the condition $r_c < 1$ Mm and $E_r > 4 \times 10^{31}$ erg before their flares. Figure S2 shows the time evolution of the minimum r_c for $E_r > 10^{31}$ erg, and figs. S3 and S4 show the corresponding points where the flare ribbons began to brighten. For those six ARs, the minimum value of r_c for $E_r > 10^{31}$ erg decreased to less than 1 Mm before the onset of the large solar flares, and the minimum r_c position was located in the region of the flare ribbons, confirming the results for AR 12673 in Figs. 2 and 3. We conclude that if any point on a PIL satisfies the conditions $r_c < 1$ Mm and $E_r > 4 \times 10^{31}$ erg,

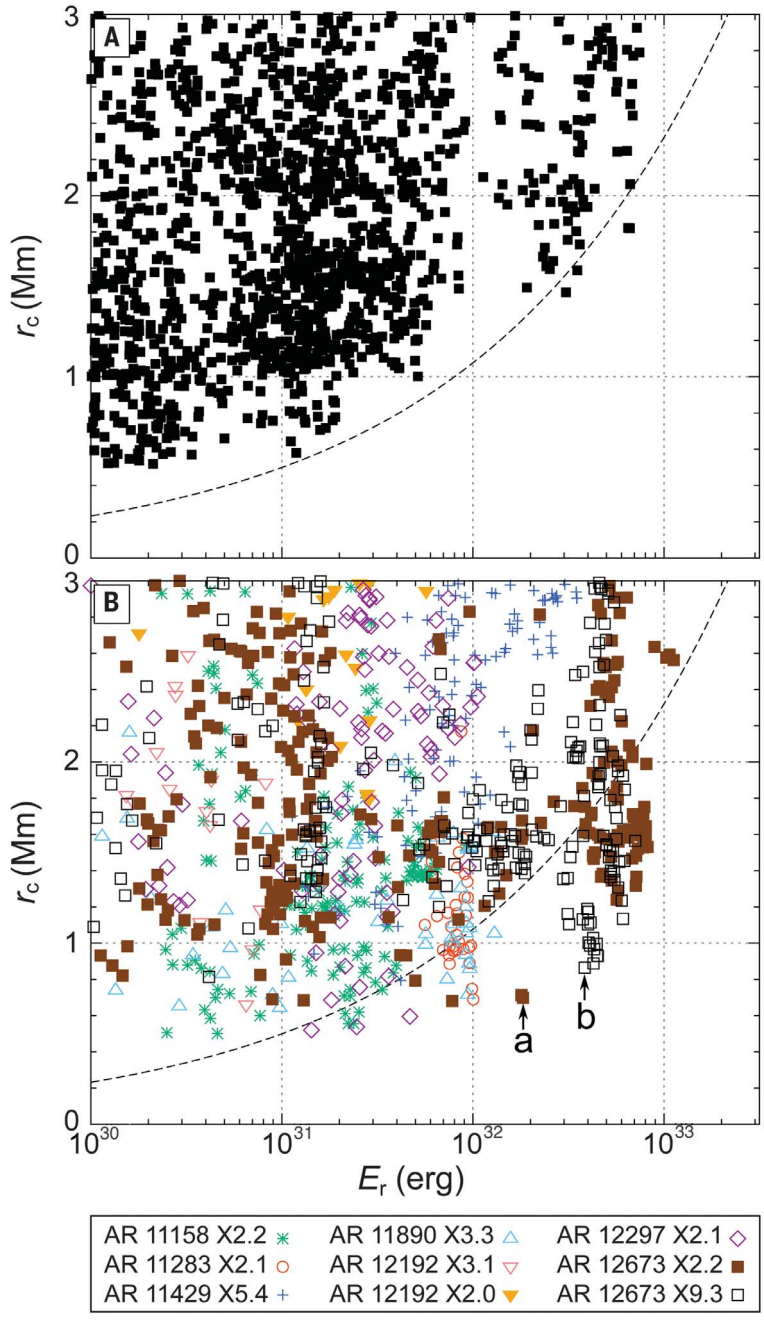


Fig. 4. Stability analysis results for the ARs in solar cycle 24. (**A** and **B**) r_c and E_r for all points on the PILs in HiFERs for groups 1 and 2 of the ARs which, respectively, did not and did produce flares larger than class X2 within 20 hours after the observation. Table S1 lists the AR numbers and times of observation for group 1. For group 2, AR numbers are indicated by the symbols shown in the legend. The observation times for group 2 are listed in table S2. Points a and b are the locations indicated in Figs. 2 and 3. The dashed curve represents the self-similar scaling relationship $r_c \propto E_r^{1/3}$.

an X-class flare usually occurs. These conditions were satisfied for periods ranging from several hours (AR 11283) to 1 day (AR 12297) before the onset of the X-class flares (Fig. 2 and fig. S2).

The decrease in r_c is a precursor of solar flares, suggesting a transition from a stable branch to an unstable branch (fig. S5). However, the κ-scheme cannot determine the lead time between when the condition is first satisfied and the onset of the flare. It is likely that small-scale magnetic structures favorable for magnetic reconnection (*18*, *24*), flux emergence and cancellation (*25*, *26*), and the collision of nonconjugated magnetic polarities (*27*) may determine the initiation of the trigger-reconnection.

We evaluated the false-positive rate for predicting flares using the κ-scheme (*14*). We found three false-positive predictions out of 198 large sunspot groups (group 1). For comparison, the number of true positives was seven correctly predicted flares in six ARs, out of a total of nine flares larger than class X2 from seven ARs.

The only AR in group 2 for which the κ-scheme incorrectly predicted no flare was AR 12192, in which the minimum r_c did not decrease, so no portion of the PIL satisfied the conditions above before the onset of the X3.1 flare (Fig. 4B and fig. S2). The reason could be related to the size of this AR, which is much larger than the others. It was also the only AR in group 2 that did not produce a coronal mass ejection coinciding with either of its two X-class flares (*28*). The magnetic reconnection of these flares may have initiated at a higher altitude above the solar surface. Two ribbons of the flares in this AR (fig. S4B) began with a longer interval between magnetic poles than the other flares in group 2 (*29*). The κ-scheme assumes that the first reconnection occurs at a location near the solar surface, so it cannot detect the onset conditions of flares triggered at a higher position. The magnetic field intensity is weaker further from the surface, so flares starting from a higher region must be generally weaker than those triggered at the surface. We speculate that AR 12192 may have produced X-class flares even though they started far from the solar surface because of the large size of this AR. The height of the initial reconnection point could also be related to whether the flare produces a coronal mass ejection.

Previous methods for predicting large solar flares relied on empirical relations, in which the predictions of the previous day tend to continue even if flare activity changes (*30*). By contrast, the κ-scheme predicts large solar flares through a physics-based approach, predicting the occurrence, location, and possible size of solar flares from observed magnetic data, regardless of previous flare activity.

REFERENCES AND NOTES

1. C. J. Schrijver, *Space Weather* **13**, 524–528 (2015).
2. K. Shibata, T. Magara, *Living Rev. Sol. Phys.* **8**, 6 (2011).
3. G. Barnes *et al.*, *Astrophys. J.* **829**, 89 (2016).
4. K. D. Leka *et al.*, *Astrophys. J.* **243**, 36 (2019).
5. Z. Svestka, E. W. Cliver, in *Eruptive Solar Flares*, Z. Svestka, B. V. Jackson, M. E. Machado, Eds. (Lecture Notes in Physics Series, vol. 399, Springer, 1992), pp. 1–11.
6. J. Zhang, K. P. Dere, *Astrophys. J.* **649**, 1100–1109 (2006).
7. B. T. Welsch, *Sol. Phys.* **293**, 113 (2018).
8. A. W. Hood, E. R. Priest, *Sol. Phys.* **64**, 303–321 (1979).
9. T. Török, B. Kliem, V. S. Titov, *Astron. Astrophys. Suppl. Ser.* **413**, L27–L30 (2004).
10. B. Kliem, T. Török, *Phys. Rev. Lett.* **96**, 255002 (2006).
11. J. Jing *et al.*, *Astrophys. J.* **864**, 138 (2018).
12. A. Duan *et al.*, *Astrophys. J.* **884**, 73 (2019).
13. N. Ishiguro, K. Kusano, *Astrophys. J.* **843**, 101 (2017).

14. Supplementary text is available as supplementary material.
15. D. M. Rust, A. Kumar, *Astrophys. J.* **464**, L199–L202 (1996).
16. R. C. Canfield, H. S. Hudson, D. E. McKenzie, *Geophys. Res. Lett.* **26**, 627–630 (1999).
17. R. L. Moore, A. C. Sterling, H. S. Hudson, H. R. Lemen, *Astrophys. J.* **552**, 833–848 (2001).
18. K. Kusano *et al*., *Astrophys. J.* **760**, 31 (2012).
19. Materials and methods are available as supplementary materials.
20. S. Inoue *et al*., *Astrophys. J.* **780**, 101 (2014).
21. P. H. Scherrer *et al*., *Sol. Phys.* **275**, 207–227 (2012).
22. M. G. Bobra *et al*., *Sol. Phys.* **289**, 3549–3578 (2014).
23. S. Yang, J. Zhang, X. Zhu, Q. Song, *Astrophys. J.* **849**, L21 (2017).
24. K. Kusano, T. Maeshiro, T. Yokoyama, T. Sakurai, *Astrophys. J.* **610**, 537–549 (2004).
25. P. F. Chen, K. Shibata, *Astrophys. J.* **545**, 524–531 (2000).
26. H. Wang, C. Liu, *Res. Astron. Astrophys.* **15**, 145–174 (2015).
27. G. Chintzoglou, J. Zhang, M. C. M. Cheung, M. Kazachenko, *Astrophys. J.* **871**, 67 (2019).
28. X. Sun *et al*., *Astrophys. J.* **804**, L28 (2015).
29. J. K. Thalmann, Y. Su, M. Temmer, A. M. Veronig, *Astrophys. J.* **801**, L23 (2015).
30. S.-H. Park *et al*., *Astrophys. J.* **890**, 124 (2020).
31. J. Machol, GOES x-ray sensor (XRS) measurements (2016); https://www.ngdc.noaa.gov/stp/satellite/goes/doc/GOES_XRS_readme.pdf
32. J. R. Lemen *et al*., *Sol. Phys.* **275**, 17–40 (2012).

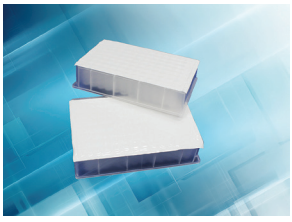
ACKNOWLEDGMENTS

We are grateful for productive discussions with K. D. Leka, S.-H. Park, A. C.-L. Chian, and many collaborators. We thank the anonymous reviewers for their careful reading of our manuscript and their insightful comments. Part of this work was carried out by using the computational resources of the Center for Integrated Data Science, Institute for Space-Earth Environmental Research (ISEE), Nagoya University; the FX100 and CX400 supercomputer systems at the Information Technology Center, Nagoya University; and the Earth Simulator at JAMSTEC. HMI and AIA are instruments on the SDO, a mission for NASA's Living with a Star Program. **Funding:** This work was supported by MEXT/JSPS KAKENHI JP15H05814, MEXT as "Exploratory Challenge on Post-K computer," and JAMSTEC. **Author contributions:** K.K. led the study and wrote the paper; T.I. contributed the statistical analysis; Y.B. contributed the analysis of SDO-AIA data; and S.I. developed the software for the NLFFF extrapolation. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data and software required to reproduce the results of this paper are available in the NAGOYA repository at http://hdl.handle.net/2237/00032309. The SDO-HMI and SDO-AIA data were taken from the Joint Science Operations Center http://jsoc.stanford.edu/ajax/lookdata.html; we used the data series hmi.sharp_cea_720s, hmi.M_45s, and aia.lev1_euv_12s at the observing times listed in tables S1 and S2.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/587/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
Tables S1 and S2
References (*33*–*42*)
Movie S1

24 August 2019; accepted 24 June 2020
10.1126/science.aaz2511



Microplate Heat-Sealing Films

Porvair Sciences expanded range of heat-sealing films for microplates and tubes now includes clear, strong (Dura), heavy-duty and gas-permeable products with several variations that add flexibility and compatibility to suit all requirements for sample storage and collection. When heat is applied to these products, a tight seal is formed on polypropylene, polyethylene, polystyrene, polycarbonate, and cyclin olefin copolymer (COC) microplates and tubes. Optically clear films from Porvair are available in a variety of formats, including peelable, pierceable, and strong nonpeelable. These seals are ideal for a wide range of applications, such as imaging, fluorescence, colorimetric assays, and PCR/quantitative PCR. All clear seals are also compatible with devices that use a heat-pressurized lid. The Dura range of durable, foil-based heat-sealing films are ideal for high-temperature applications and room-temperature storage. They come in a variety of formats, including peelable, solvent resistant (including DMSO) and pierceable. All Dura heat-sealing films are autoclavable.

Porvair Sciences
For info: +44-(0)-1978-661144
www.microplates.com/heat-sealing-films

SARS-CoV-2 Synthetic Nanobodies

Absolute Antibody Ltd., along with the University of Zurich, now offers synthetic nanobodies against the receptor binding domain of SARS-CoV-2, the coronavirus that causes COVID-19. Researchers are exploring the potential of nanobodies as inhalable COVID-19 drugs, which would be easier to administer and could reach patients' lungs faster than other treatments. Absolute Antibody recombinantly produces the SARS-CoV-2 synthetic nanobodies for ensured batch-to-batch reproducibility, high purity, and low endotoxin levels. In addition, we use antibody engineering to fuse the nanobodies to fragment crystallizable (Fc) domains in different species, isotypes, and subtypes for use as serological controls.

Absolute Antibody
For info: +44-(0)-1642-688810
absoluteantibody.com

Freshly Isolated Chondrocytes

CELLvo Human Articular Chondrocytes from StemBioSys are low-passage (P0 and P1) cells freshly isolated from young, healthy donors and available for regenerative medicine and basic research. When cultured with CELLvo Matrix, the cells home, attach, and proliferate more rapidly than chondrocytes cultured under standard conditions. They also exhibit similar morphology and phenotype to in vivo chondrocytes, such as a high ratio of type 2 collagen to type 1 collagen. Because the cells and the growth conditions are never exposed to nonhuman proteins, they offer a streamlined pathway to clinical studies. The CELLvo Human Chondrocyte cell culture system delivers improved performance and exceptional research potential to basic and translational studies of osteoarthritis, joint function, chondrocyte differentiation, tissue engineering, and chondrocyte implantation.

StemBioSys
For info: 210-877-9323
www.stembiosys.com

Automated Flow Chemistry Synthesis

The FlowSyn Automated Loop Filling (FlowSyn Auto-LF) system is a powerful, highly efficient module for running multiple experiments with multiple reagent inputs under different sets of chemical conditions. This product is ideal for robotic synthesis groups looking to generate combinatorial libraries utilizing the power of flow chemistry. With a choice of two or four channels, this intelligent, simultaneous loop-filling and fraction-collection system saves valuable time by loading the reagent loops for the next reaction while the current run is in progress. Fully integrated wash protocols minimize the risk of cross-contamination. The unit incorporates an automated liquid handler that can accommodate up to 96 samples.

Uniqsis
For info: +44-(0)-845-864-7747
www.uniqsis.com

Microplate Reader

The BioTek LogPhase 600 Microbiology Reader is the only four-plate microplate reader built to perform microbial growth curve analysis. Multiple plate measurements increase throughput and reduce the need for costly single-plate absorbance readers that take up valuable bench space. Additionally, highly consistent environmental conditions support high-quality absorbance results without growth curve artifacts, even in bacterial and yeast assays with extended incubation periods. The compact LogPhase 600 features a robust, stringently tested shaking mechanism to ensure that cells remain in suspension and that the hardware will not wear out, even over the course of long-term kinetic assays. Incubation is controlled by several sensors for even heating and includes Condensation Control, which prevents light scatter and artifacts due to condensation. The LogPhase 600 App controls the reader, capturing data and providing powerful analysis with an easy-to-use interface. This reader is ideal for yeast and bacterial growth assays, antimicrobial resistance studies, algal and biofuel research, and food and beverage testing.

BioTek Instruments
For info: 888-451-5171
www.biotek.com

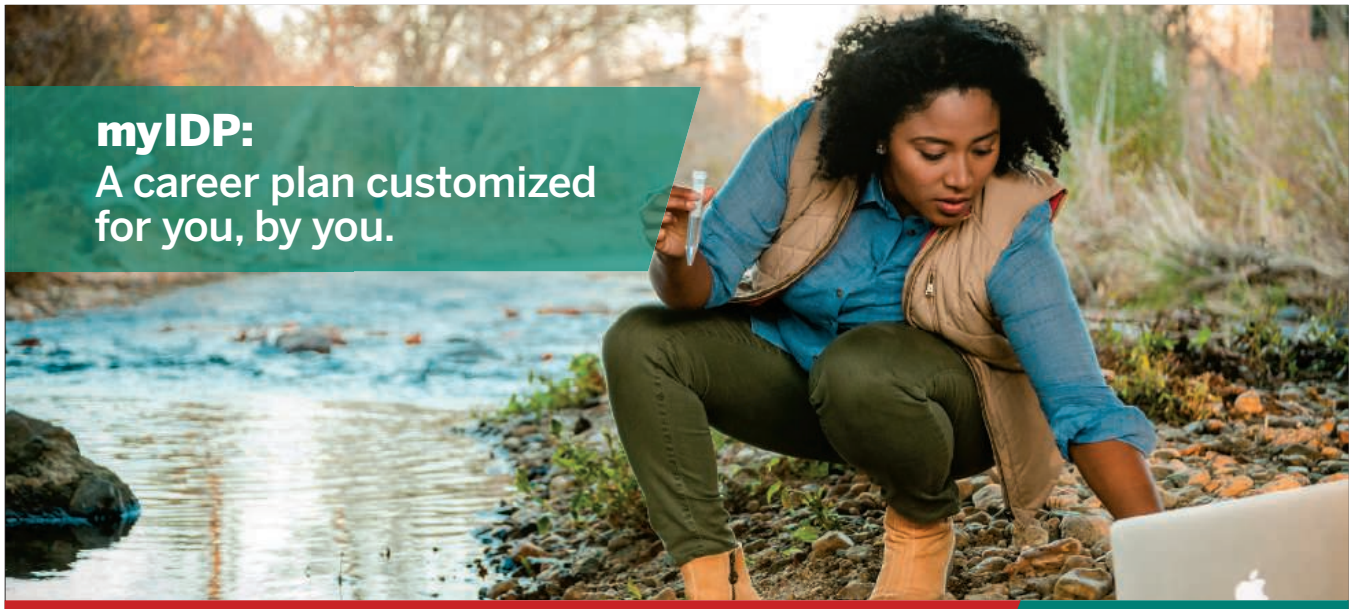
T-Cell Immune Response to SARS-CoV-2

The T-SPOT Discovery SARS-CoV-2 kit from Oxford Immunotec has been developed to aid in the investigation of the T-cell response to SARS-CoV-2 infection. The kit uses a modified enzyme-linked immunospot (ELISPOT) assay, where samples prepared from peripheral blood are stimulated in vitro by peptide pools from SARS-CoV-2. This restimulation allows SARS-CoV-2–specific T cells to be enumerated. The kit builds on our experience with tuberculosis diagnosis and the assessment of immune response to cytomegalovirus in transplant patients to apply our proprietary T-SPOT technology to the fight against COVID-19. While serology is able to detect antibodies to SARS-CoV-2 in the blood of some individuals after infection, little is currently known about how this confers immunity to COVID-19. The T-SPOT technology goes further than simple serology by interrogating the immune system's T-cell response, and will help researchers measure the strength of that response to SARS-CoV-2. The strength of this response may be linked to protection from reinfection.

Oxford Immunotec
For info: 877-208-7768
go.oxfordimmunotec.com/t-spot_discovery_sars_cov-2

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.



For your career in science, there's only one Science

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.
- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track.
- Articles and resources to guide you through the process.
- Options to save materials online and print them for further review and discussion.
- Ability to select which portion of your IDP you wish to share with advisors, mentors, or others.
- A certificate of completion for users that finish myIDP.

myIDP

Visit the website and start planning today!
myIDP.sciencecareers.org

Science Careers AAAS In partnership with:

FASEB Federation of American Societies for Experimental Biology
UMASS University of Massachusetts Medical School
UCSF University of California San Francisco
UIC
BURROUGHS WELLCOME FUND

By **Wen-Jing Lin**

My womb is my business

"Do you plan to have children?" The question came out of nowhere. I was sitting in a professor's office, interviewing for a postdoc position, when the topic suddenly shifted from my science to my womb. I paused, unsure how to respond. I didn't feel my personal life was any of the professor's business. But I was interested in the position, and I worried that I might not get it if I refused to answer. So I told him details about my personal life that I never imagined sharing with a stranger.

That interview was in Japan, where his question—I've since learned—isn't unusual. But I'm not from Japan. I grew up in Taiwan, and it was unusual for me.

I wished I told him I thought his line of questioning was inappropriate. I worried he'd ask the same question to other interviewees in the future and make them feel just as uncomfortable. It was also demoralizing to be judged based on the odds that I would have children. "Shouldn't my academic accomplishments be the only thing that matters?" I fumed.

At first, I told myself it was a one-off incident, an unlucky break with that particular professor. Then I got called to interview for a postdoc position at another institution in Japan. I was excited because it was at a top university and I thought it would be good for my career. But midway through the interview, a warning bell started to ring in my mind when the faculty member asked, "Do you live with your family?" Flustered, I responded, "Yes, I live with my family." He pressed further: "Do you have kids?" Once again, I didn't dare refuse to answer, so I found myself opening up about my personal life.

I received an offer for one of those positions, but I turned it down because I didn't want to work in an environment where I risked being judged differently for my parental status or for being a woman. I am still trying to secure a second postdoc, and I fear I will face the same question again.

At the time of those incidents, I didn't feel I could do anything about them, beyond complaining to friends and colleagues. But when I told my current postdoc adviser about my second interview, she was furious. She told me I could seek help from the university's harassment counselor and report the case.

With the help of the counselor, I sent a letter to the dean and department chair who oversee the second faculty member, describing what happened. I told them that I wanted all the employees in the department to go through bias training on a regular basis to learn about inappropriate questions and behavior. I had no plans to work there, but I wanted to ensure that other employees and interviewees had a discrimination-free work environment.

"The entire ordeal has left me feeling helpless."

They sent a reply 20 days later. The dean thanked me for raising the issue, and the department chair told me he would consider arranging training, but he didn't commit to any specific steps. The chair said the professor who made the comment was sorry his questions offended me, but that they weren't asked because of my gender—he asks every interviewee the same questions.

Unfortunately, the law is on his side: If he asks the questions of men and women, then they aren't against the law in Japan. That's frustrating for me because I don't think they are appropriate questions to ask anyone, regardless of their gender.

The entire ordeal has left me feeling helpless. I lost out on a potential postdoc opportunity. And when I complained about what I perceived to be unfair treatment, I didn't receive any assurances that the situation would change for the next scientist who interviews for a position.

In many countries, interviewers are taught not to ask interviewees about their home life. Where those kinds of questions are more culturally accepted, as in Japan, I'd argue that universities have a responsibility to push back against norms and ensure that job candidates are treated equally during the interview process. If they fail to do that, they'll risk driving away good candidates.

Above all, I want decision-makers to build a discrimination-free environment for all scientists. I hope that, one day, no interviewee will have to hear the question, "Do you plan to have children?" ■

Wen-Jing Lin is a postdoctoral researcher at the University of Tokyo.

ILLUSTRATION: ROBERT NEUBECKER

Published by AAAS