Any volunteers? Infecting healthy people to test medicines *p. 882*

Hanging around is cheaper than hovering *pp. 895 & 978*

Preprints for the life sciences *p. 899*

Science

$15
20 MAY 2016
sciencemag.org

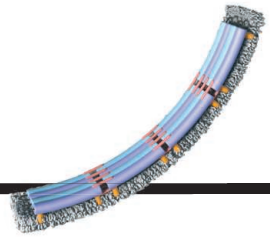
AAAS

SPECIAL ISSUE

URBAN PLANET

CONTENTS

20 MAY 2016 • VOLUME 352 • ISSUE 6288



890
Membrane-bending DNA nanostructures

SPECIAL SECTION

URBAN PLANET

INTRODUCTION

904 Cities are the future

NEWS

906 Rise of the city

908 Roots of the urban mind *By G. Miller*

912 A plague of rats *By W. Cornwall*

916 China rethinks cities *By D. Normile*

918 Vancouver's green dream *By K. R. Weiss*

REVIEWS

922 City-integrated renewable energy for urban sustainability *D. M. Kammen and D. A. Sunter*

928 Emerging solutions to the water challenges of an urbanizing world *T. A. Larsen* et al.

PERSPECTIVES

934 Transport solutions for cleaner air *F. J. Kelly and T. Zhu*

936 The ecological future of cities *M. J. McDonnell and I. MacGregor-Fors*

938 Living in cities, naturally *T. Hartig and P. H. Kahn Jr.*

940 Meta-principles for developing smart, sustainable, and healthy cities *A. Ramaswami* et al.

943 Hidden linkages between urbanization and food systems *K. C. Seto and N. Ramankutty*

946 Building functional cities *J. V. Henderson* et al.

ON THE COVER

Massive fog rolls into Dubai, United Arab Emirates. As host to more than half of the world's population, cities have footprints that reach well beyond their municipal boundaries. Efforts are now under way to make urban areas more sustainable and efficient. The challenges and opportunities of our urban planet are highlighted in a special section beginning on page 904 and online at http://scim.ag/1Tb0WdZ. *Photo: © Sebastian Opitz/Novarc Images*

SEE ALSO ▸ EDITORIAL P. 873 ▸ LETTERS P. 886 ▸ PODCAST ▸ VIDEO

NEWS

IN BRIEF

874 News at a glance

IN DEPTH

876 **TRACKING HOW HUMANS EVOLVE IN REAL TIME**
Analyses of thousands of sequenced genomes show changes in as little as a generation *By E. Pennisi*

877 **'CELL PAINTING' HIGHLIGHTS RESPONSES TO DRUGS AND TOXINS**
Machine learning combined with molecular paints allows rapid screening of changes in cell shape and function *By E. Pennisi*

878 **RESEARCHERS DECRY CONSENT PROPOSAL**
U.S. would require patient permission for use of anonymous tissue samples *By J. Kaiser*

879 **A MODULAR ROUTE TO NEW ANTIBIOTICS**
Divide-and-conquer chemistry yields hundreds of candidate medicines *By R. F. Service*

880 **DATA CHECK: 'EMPLOYMENT CRISIS' FOR NEW PH.D.S IS AN ILLUSION**
NSF surveys generate what seem to be conflicting data on the status of those entering the scientific job market *By J. Mervis*

881 **CAN BRAIN SCANS REVEAL CONCUSSION-LINKED DISEASE?**
By highlighting condition typically confirmed by autopsy, protein tracer could alter medical and legal landscape *By E. Underwood*

FEATURE

882 **THE TRUEST TEST**
Studies that intentionally infect people with pathogens have a checkered past, but they are seeing a resurgence *By J. Cohen*

INSIGHTS

LETTERS

886 **INGENUITY: NEXTGEN'S VISION**
▸ URBAN PLANET SECTION P. 904

PERSPECTIVES

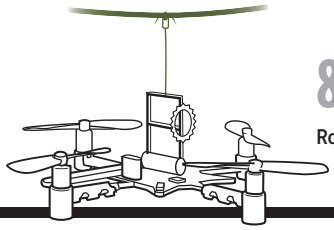
890 **CHANGING OF THE GUARD**
DNA nanostructures mimic membrane proteins *By S. Howorka*

891 **SUPERCONDUCTIVITY ON THE EDGE**
Hybrid correlated systems could reveal exotic quantum transport behavior *By N. Mason*
▸ REPORT P. 966

892 **METHANE—MAKE IT OR BREAK IT**
Biochemical data resolve the controversy over how methanogenic archaea produce methane *By T. J. Lawton and A. C. Rosenzweig*
▸ RESEARCH ARTICLE P. 953

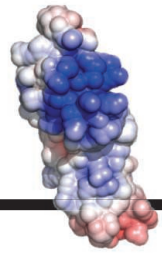
Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS



895 & 978

Robots that perch



986

Sensing cellular phosphate

894 FORMIN' FILAMENTS AT A FASTER CLIP
A microtubule-tracking protein amplifies actin assembly *By K. Rottner*
▸ REPORT P. 1004

895 LEARNING FROM NATURE HOW TO LAND AERIAL ROBOTS
Smaller robots can use mechanical intelligence to simplify the task of perching on a target *By M. Kovac*
▸ REPORT P. 978

897 INTERNATIONAL MIGRATION UNDER THE MICROSCOPE
Fragmented research and limited data must be addressed *By F. Willekens* et al.

899 PREPRINTS FOR THE LIFE SCIENCES
The time is right for biologists to post their research findings onto preprint servers *By J. M. Berg* et al.

BOOKS *ET AL.*

902 THE DANCING BEES
By T. Munz, reviewed by V. Callier

903 THE BIG PICTURE
By S. Carroll, reviewed by B. Loewer

RESEARCH

IN BRIEF

949 From *Science* and other journals

RESEARCH ARTICLES

952 CANCER IMMUNOTHERAPY
The tumor microenvironment underlies acquired resistance to CSF-1R inhibition in gliomas *D. F. Quail* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: dx.doi.org/10.1126/science.aad3018

953 ENZYME MECHANISMS
The radical mechanism of biological methane synthesis by methyl-coenzyme M reductase *T. Wongnate* et al.
▸ PERSPECTIVE P. 892

REPORTS

958 SUPERCONDUCTIVITY
Ubiquitous signatures of nematic quantum criticality in optimally doped Fe-based superconductors *H.-H. Kuo* et al.

962 ARTIFICIAL SPIN ICE
Rewritable artificial magnetic charge ice *Y.-L. Wang* et al.

966 SUPERCONDUCTIVITY
Supercurrent in the quantum Hall regime *F. Amet* et al.
▸ PERSPECTIVE P. 891

969 CATALYSIS
Quantifying the promotion of Cu catalysts by ZnO for methanol synthesis *S. Kuld* et al.

974 CATALYSIS
Self-assembly of noble metal monolayers on transition metal carbide nanoparticle catalysts *S. T. Hunt* et al.

978 ROBOTICS
Perching and takeoff of a robotic insect on overhangs using switchable electrostatic adhesion *M. A. Graule* et al.
▸ PERSPECTIVE P. 895; VIDEO

982 SYNAPSE FORMATION
Control of neuronal synapse specification by a highly dedicated alternative splicing program *L. Traunmüller* et al.

986 PLANT SCIENCE
Control of eukaryotic phosphate homeostasis by inositol polyphosphate sensor domains *R. Wild* et al.

990 SIGNAL TRANSDUCTION
Thresholds and ultrasensitivity from negative cooperativity *S. H. Ha and J. E. Ferrell Jr.*

993 DNA REPAIR
ppGpp couples transcription to DNA repair in *E. coli V. Kamarthapu* et al.

HIV-1 ANTIBODIES

997 HIV-1 therapy with monoclonal antibody 3BNC117 elicits host immune responses against HIV-1 *T. Schoofs* et al.

1001 Enhanced clearance of HIV-1–infected cells by broadly neutralizing antibodies against HIV-1 in vivo *C.-L. Lu* et al.

1004 DYNAMIC CYTOSKELETON
Accelerated actin filament polymerization from microtubule plus ends *J. L. Henty-Ridilla* et al.
▸ PERSPECTIVE P. 894

1009 GENE EVOLUTION
Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals *X. Lan and J. K. Pritchard*

DEPARTMENTS

873 EDITORIAL
Leave no city behind
By Michele Acuto and Susan Parnell
▸ URBAN PLANET SECTION P. 904

1022 WORKING LIFE
A problem by any other name
By Paul Rogge and Emily Regier

Science Staff 870
New Products 1014
Science Careers 1015

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the 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 © 2016 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS.

Domestic individual membership and subscription (51 issues): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1522; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $89. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. 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:** $15.00 current issue, $20.00 back issue prepaid includes surface postage; bulk rates on request. **Authorization to photocopy** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $35.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

Science
AAAS
1200 New York Avenue, NW, Washington, DC 20005
Bateman House, 82-88 Hills Road, Cambridge, UK CB2 1LQ

Editor-in-Chief Marcia McNutt
Executive Editor Monica M. Bradford **News Editor** Tim Appenzeller
Managing Editor, Research Journals Katrina L. Kelner
Deputy Editors Lisa D. Chong, Andrew M. Sugden(UK), Valda J. Vinson, Jake S. Yeston

Research and Insights

DEPUTY EDITOR, EMERITUS Barbara R. Jasny **SR. EDITORS** Caroline Ash(UK), Gilbert J. Chin, Julia Fahrenkamp-Uppenbrink(UK), Pamela J. Hines, Stella M. Hurtley(UK), Paula A. Kiberstis, Marc S. Lavine(Canada), Kristen L. Mueller, Ian S. Osborne(UK), Beverly A. Purnell, L. Bryan Ray, Guy Riddihough, H. Jesse Smith, Jelena Stajic, Peter Stern(UK), Phillip D. Szuromi, Sacha Vignieri, Brad Wible, Nicholas S. Wigginton, Laura M. Zahn **ASSOCIATE EDITORS** Brent Grocholski, Keith T. Smith **ASSOCIATE BOOK REVIEW EDITOR** Valerie B. Thompson **LETTERS EDITOR** Jennifer Sills **CHIEF CONTENT PRODUCTION EDITOR** Cara Tate **SR. CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmec **CONTENT PRODUCTION EDITORS** Jeffrey E. Cook, Chris Filiatreau, Cynthia Howe, Barbara P. Ordway, Catherine Wolner **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Lisa Johnson, Anita Wynn **PUBLICATIONS ASSISTANTS** Jeffrey Hearn, Dona Mathieu, Le-Toya Mayne Flood, Shannon McMahon, Scott Miller, Caitlyn Phillips, Jerry Richardson, Alice Whaley(UK), Brian White **EXECUTIVE ASSISTANT** Anna Bashkirova **ADMINISTRATIVE SUPPORT** Janet Clements(UK), Lizanne Newton(UK), Maryrose Madrid

News

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** Richard Stone **DEPUTY NEWS EDITORS** Robert Coontz, Elizabeth Culotta, David Grimm, David Malakoff, Leslie Roberts **CONTRIBUING EDITOR** Martin Enserink(Europe) **SR. CORRESPONDENTS** Daniel Clery(UK), Jeffrey Mervis, Elizabeth Pennisi **NEWS WRITERS** Adrian Cho, Jon Cohen, Jennifer Couzin-Frankel, Carolyn Gramling, Eric Hand, Jocelyn Kaiser, Catherine Matacic, Kelly Servick, Robert F. Service, Erik Stokstad(Cambridge, UK), Emily Underwood **INTERNS** Patrick Monahan **CONTRIBUTING CORRESPONDENTS** John Bohannon, Warren Cornwall, Ann Gibbons, Mara Hvistendahl, Sam Kean, Eli Kintisch, Kai Kupferschmidt(Berlin), Andrew Lawler, Christina Larson(Beijing), Mitch Leslie, Charles C. Mann, Eliot Marshall, Virginia Morell, Dennis Normile(Shanghai), Heather Pringle, Tania Rabesandratana(London), Gretchen Vogel(Berlin), Lizzie Wade(Mexico City) **CAREERS** Donisha Adams, Rachel Bernstein(Editor) **COPY EDITORS** Julia Cole, Dorie Chevlen, Jennifer Levin (Chief) **ADMINISTRATIVE SUPPORT** Jessica Adams

Executive Publisher Rush D. Holt
Interim Publisher Bill Moran **Chief Digital Media Officer** Rob Covey

BUSINESS OPERATIONS AND PORTFOLIO MANAGEMENT DIRECTOR Sarah Whalen **PRODUCT DEVELOPMENT DIRECTOR** Will Schweitzer **PRODUCT DEVELOPMENT ASSOCIATE** Hal Moore **BUSINESS SYSTEMS AND FINANCIAL ANALYSIS DIRECTOR** Randy Yi **MANAGER OF FULFILLMENT SYSTEMS** Neal Hawkins **SYSTEMS ANALYST** Nicole Mehmedovich **ASSISTANT DIRECTOR, BUSINESS OPERATIONS** Eric Knott **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **SENIOR BUSINESS ANALYST** Cory Lipman **BUSINESS ANALYSTS** David Garrison, Meron Kebede, Sandy Kim **FINANCIAL ANALYST** Drew Sher **DIRECTOR, COPYRIGHTS LICENSING SPECIAL PROJECTS** Emilie David **PERMISSIONS ASSOCIATE** Elizabeth Sandler **RIGHTS, CONTRACTS, AND LICENSING ASSOCIATE** Lili Kiser

MARKETING DIRECTOR Elise Swinehart **ASSOCIATE MARKETING DIRECTOR** Christina Schlecht **MARKETING MANAGER** Elizabeth Sattler **MARKETING ASSOCIATE** Steven Goodman **CREATIVE DIRECTOR** Scott Rodgerson **SENIOR ART ASSOCIATES** Seil Lee, Paula Fry **ART ASSOCIATE** Kim Huynh **PROGRAM DIRECTOR, AAAS MEMBER CENTRAL** Peggy Mihelich **FULFILLMENT SYSTEMS AND OPERATIONS** membership@aaas.org **MANAGER, MEMBER SERVICES** Pat Butler **SPECIALISTS** Terrance Morrison, Latasha Russell **MANAGER, DATA ENTRY** Mickie Napoleoni **DATA ENTRY SPECIALISTS** Brenden Aquilino, Fiona Giblin

PUBLISHER RELATIONS MANAGER Catherine Holland **PUBLISHER RELATIONS, EASTERN REGION** Keith Layson **PUBLISHER RELATIONS, WESTERN REGION** Ryan Rexroth **SALES RESEARCH COORDINATOR** Aiesha Marshall **MANAGER, SITE LICENSE OPERATIONS** Iquo Edim **SENIOR OPERATIONS ANALYST** Lana Guz **FULFILLMENT ANALYST** Judy Lillibridge **ASSOCIATE DIRECTOR, MARKETING** Christina Schlecht **MARKETING ASSOCIATE** Isa Sesay-Bah

WEB TECHNOLOGIES PORTFOLIO MANAGER Trista Smith **TECHINCAL MANAGER** Chris Coleman **PROJECT MANAGER** Nick Fletcher **DEVELOPERS** Ryan Jensen, Jimmy Marks, Brandon MOrrison

DIGITAL MEDIA DIRECTOR OF ANALYTICS Enrique Gonzales **DIGITAL REPORTING ANALYST** Eric Hossinger **SR. WEB PRODUCER** Sarah Crespi **WEB PRODUCER** Alison Crawford **VIDEO PRODUCER** Nguyen Nguyen **SOCIAL MEDIA PRODUCER** Brice Russ

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

DESIGN DIRECTOR Beth Rakouskas **DESIGN EDITOR** Marcy Atarod **SENIOR DESIGNERS** Garvin Grullón, Chrystal Smith **GRAPHICS MANAGING EDITOR** Alberto Cuadra **SENIOR SCIENTIFIC ILLUSTRATORS** Chris Bickel, Katharine Sutliff **SCIENTIFIC ILLUSTRATOR** Valerie Altounian **INTERACTIVE GRAPHICS EDITOR** Jia You **SENIOR ART ASSOCIATES** Holly Bishop, Nathalie Cary, Preston Huey **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **PHOTO EDITORS** Leslie Blizard, Christy Steele

DIRECTOR, GLOBAL COLLABORATION, CUSTOM PUBLICATIONS, ADVERTISING Bill Moran **EDITOR, CUSTOM PUBLISHING** Sean Sanders: 202-326-6430 **ADVERTISING MARKETING MANAGER** Justin Sawyers: 202-326-7061 science_advertising@aaas.org **ADVERTISING SUPPORT MANAGER** Karen Foote: 202-326-6740 **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **SR. PRODUCTION SPECIALIST/GRAPHIC DESIGNER** Amy Hardcastle **SR. TRAFFIC ASSOCIATE** Christine Hall **SALES COORDINATOR** Shirley Young **ASSOCIATE DIRECTOR, COLLABORATION, CUSTOM PUBLICATIONS/CHINA/TAIWAN/KOREA/SINGAPORE** Ruolei Wu: +86-186 0082 9345, rwu@aaas.org **COLLABORATION/CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu + 81532-81-5142 asandhu@aaas.org **EAST COAST/E. CANADA** Laurie Faraday: 508-747-9395, FAX 617-507-8189 **WEST COAST/W. CANADA** Lynne Stickrod: 415-931-9782, FAX 415-520-6940 **MIDWEST** Jeffrey Dembksi: 847-498-4520 x3005, Steven Loerch: 847-498-4520 x3006 **UK EUROPE/ASIA** Roger Goncalves: TEL/FAX +41 43 243 1358 **JAPAN** Katsuyoshi Fukamizu(Tokyo): +81-3-3219-5777 kfukamizu@aaas.org **CHINA/TAIWAN** Ruolei Wu: +86-186 0082 9345, rwu@aaas.org

WORLDWIDE ASSOCIATE DIRECTOR OF SCIENCE CAREERS Tracy Holmes: +44 (0) 1223 326525, FAX +44 (0) 1223 326532 tholmes@science-int.co.uk **CLASSIFIED** advertise@sciencecareers.org **U.S. SALES** Tina Burks: 202-326-6577, Nancy Toema: 202-326-6578 **EUROPE/ROW SALES** Sarah Lelarge **SALES ASSISTANT** Kelly Grace **JAPAN** Hiroyuki Mashiki(Kyoto): +81-75-823-1109 hmashiki@aaas.org **CHINA/TAIWAN** Ruolei Wu: +86-186 0082 9345 rwu@aaas.org **MARKETING MANAGER** Allison Pritchard **MARKETING ASSOCIATE** Aimee Aponte

AAAS BOARD OF DIRECTORS, CHAIR Geraldine L. Richmond **PRESIDENT** Barbara A. Schaal **PRESIDENT-ELECT** Susan Hockfield **TREASURER** David Evans Shaw **CHIEF EXECUTIVE OFFICER** Rush D. Holt **BOARD** Cynthia M. Beall, May R. Berenbaum, Carlos J. Bustamante, Stephen P.A. Fodor, Claire M. Fraser, Michael S. Gazzaniga, Laura H. Greene, Elizabeth Loftus, Mercedes Pascual

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

INSTITUTIONAL SITE LICENSES 202-326-6730 **REPRINTS:** Author Inquiries 800-635-7181 **COMMERCIAL INQUIRIES** 803-359-4578 **PERMISSIONS** 202-326-6765, permissions@aaas.org **AAAS Member Services** 202-326-6417 or http://membercentral.aaas.org/discounts

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 of conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

INFORMATION FOR AUTHORS See pages 624 and 625 of the 5 February 2016 issue or access www.sciencemag.org/authors/science-information-authors

SENIOR EDITORIAL BOARD
Gary King, *Harvard University,* **Susan M. Rosenberg,** *Baylor College of Medicine,*
Ali Shilatifard, *Northwestern University Feinberg School of Medicine*

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 Edinburgh*
Sonia Altizer, *U. of Georgia*
Sebastian Amigorena, *Institut Curie*
Kathryn Anderson, *Memorial Sloan-Kettering Cancer Center*
Meinrat O. Andreae, *Max-Planck Inst. Mainz*
Paola Arlotta, *Harvard U.*
Johan Auwerx, *EPFL*
David Awschalom, *U. of Chicago*
Clare Baker, *University of Cambridge*
Nenad Ban, *ETH Zurich*
Jordi Bascompte, *University of Zurich*
Franz Bauer, *Instituto de Astrofisica*
Ray H. Baughman, *U. of Texas, Dallas*
David Baum, *U. of Wisconsin*
Carlo Beenakker, *Leiden U.*
Kamran Behnia, *ESPCI-ParisTech*
Yasmine Belkaid, *NIAID, NIH*
Philip Benfey, *Duke U.*
May Berenbaum, *U. of Illinois*
Gabriele Bergers, *U. of California, San Francisco*
Bradley Bernstein, *Massachusettes General Hospital*
Peer Bork, *EMBL*
Bernard Bourdon, *Ecole Normale Supérieure de Lyon*
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, *U. Hamburg-Eppendorf*
Joseph A. Burns, *Cornell U.*
Carter Tribley Butts, *U. of California, Irvine*
Gyorgy Buzsaki, *New York U. School of Medicine*
Blanche Capel, *Duke U.*
Mats Carlsson, *U. of Oslo*
Ib Chorkendorff, *U. of Denmark*
David Clapham, *Children's Hospital Boston*
Joel Cohen, *Rockefeller U., Columbia U.*
James J. Collins, *MIT*
Robert Cook-Deegan, *Duke U.*
Lisa Coussens, *Oregon Health & Science U.*
Alan Cowman, *Walter & Eliza Hall Inst.*
Robert H. Crabtree, *Yale U.*
Roberta Croce, *Vrije Universiteit*
Janet Currie, *Princeton U.*
Jeff L. Dangl, *U. of North Carolina*
Tom Daniel, *U. of Washington*
Frans de Waal, *Emory U.*
Stanislas Dehaene, *Collège de France*
Robert Desimone, *MIT*
Claude Desplan, *New York U.*
Dennis Discher, *U. of Pennsylvania*
Gerald W. Dorn II, *Washington U. School of Medicine*
Jennifer A. Doudna, *U. of California, Berkeley*
Bruce Dunn, *U. of California, Los Angeles*
William Dunphy, *Caltech*
Christopher Dye, *WHO*
Todd Ehlers, *U. of Tuebingen*
David Ehrhardt, *Carnegie Inst. of Washington*
Tim Elston, *U. of North Carolina at Chapel Hill*
Gerhard Ertl, *Fritz-Haber-Institut, Berlin*
Barry Everitt, *U. of Cambridge*
Ernst Fehr, *U. of Zurich*
Anne C. Ferguson-Smith, *U. of Cambridge*
Michael Feuer, *The George Washington U.*
Toren Finkel, *NHLBI, NIH*
Kate Fitzgerald, *U. of Massachusetts*
Peter Fratzl, *Max-Planck Inst.*
Elaine Fuchs, *Rockefeller U.*
Daniel Geschwind, *UCLA*
Karl-Heinz Glassmeier, *TU Braunschweig*
Ramon Gonzalez, *Rice U.*
Julia R. Greer, *Caltech*
Elizabeth Grove, *U. of Chicago*
Nicolas Gruber, *ETH Zurich*
Kip Guy, *St. Jude's Children's Research Hospital*
Taekjip Ha, *U. of Illinois at Urbana-Champaign*
Wolf-Dietrich Hardt, *ETH Zurich*
Christian Haass, *Ludwig Maximilians U.*
Sharon Hammes-Schiffer, *U. of Illinois at Urbana–Champaign*
Michael Hasselmo, *Boston U.*
Martin Heimann, *Max-Planck Inst. Jena*
Yka Helariutta, *U. of Cambridge*
James A. Hendler, *Rensselaer Polytechnic Inst.*
Janet G. Hering, *Swiss Fed. Inst. of Aquatic Science & Technology*
Kai-Uwe Hinrichs, *U. of Bremen*
David Hodell, *U. of Cambridge*
Lora Hooper, *UT Southwestern Medical Ctr. at Dallas*
Tamas Horvath, *Yale University*
Raymond Huey, *U. of Washington*
Fred Hughson, *Princeton U.*
Auke Ijspeert, *EPFL Lausanne*
Stephen Jackson, *USGS and U. of Arizona*
Steven Jacobsen, *U. of California, Los Angeles*
Kai Johnsson, *EPFL Lausanne*
Peter Jonas, *Inst. of Science & Technology (IST) Austria*
Matt Kaeberlein, *U. of Washington*
William Kaelin Jr., *Dana-Farber Cancer Inst.*
Daniel Kahne, *Harvard U.*
Daniel Kammen, *U. of California, Berkeley*
Abby Kavner, *U. of California, Los Angeles*
Masashi Kawasaki, *U. of Tokyo*
V. Narry Kim, *Seoul National U.*
Joel Kingsolver, *U. of North Carolina at Chapel Hill*
Robert Kingston, *Harvard Medical School*
Etienne Koechlin, *Ecole Normale Suprieure*
Alexander Kolodkin, *Johns Hopkins U.*
Thomas Langer, *U. of Cologne*
Mitchell A. Lazar, *U. of Pennsylvania*
David Lazer, *Harvard U.*
Thomas Lecuit, *IBDM*
Virginia Lee, *U. of Pennsylvania*
Stanley Lemon, *U. of North Carolina at Chapel Hill*
Ottoline Leyser, *Cambridge U.*
Wendell Lim, *U.C. San Francisco*
Marcia C. Linn, *U. of California, Berkeley*
Jianguo Liu, *Michigan State U.*
Luis Liz-Marzan, *CIC biomaGUNE*
Jonathan Losos, *Harvard U.*
Ke Lu, *Chinese Acad. of Sciences*
Christian Lüscher, *U. of Geneva*
Laura Machesky, *CRUK Beatson Inst. for Cancer Research*
Anne Magurran, *U. of St. Andrews*
Oscar Marin, *CSIC & U. Miguel Hernández*
Charles Marshall, *U. of California, Berkeley*
C. Robertson McClung, *Dartmouth College*
Graham Medley, *U. of Warwick*
Tom Misteli, *NCI*
Yasushi Miyashita, *U. of Tokyo*
Mary Ann Moran, *U. of Georgia*
Richard Morris, *U. of Edinburgh*
Alison Motsinger-Reif, *NC State U.* **(S)**
Thomas Murray, *The Hastings Center*
Daniel Neumark, *U. of California, Berkeley*
Kitty Nijmeijer, *U. of Twente*
Helga Nowotny, *European Research Advisory Board*
Ben Olken, *MIT*
Joe Orenstein, *U. of California Berkeley & Lawrence Berkeley National Lab*
Harry Orr, *U. of Minnesota*
Pilar Ossorio, *U. of Wisconsin*
Andrew Oswald, *U. of Warwick*
Margaret Palmer, *U. of Maryland*
Steve Palumbi, *Stanford U.*
Jane Parker, *Max-Planck Inst. of Plant Breeding Research*
Giovanni Parmigiani, *Dana-Farber Cancer Inst.* **(S)**
John H. J. Petrini, *Memorial Sloan-Kettering Cancer Center*
Samuel Pfaff, *Salk Institute for Biological Studies*
Joshua Plotkin, *U. of Pennsylvania*
Albert Polman, *FOM Institute AMOLF*
Philippe Poulin, *CNRS*
Jonathan Pritchard, *Stanford U.*
David Randall, *Colorado State U.*
Felix Rey, *Institut Pasteur*
Trevor Robbins, *U. of Cambridge*
Jim Roberts, *Fred Hutchinson Cancer Research Ctr.*
Barbara A. Romanowicz, *U. of California,Berkeley*
Amy Rosenzweig, *Northwestern University*
Mike Ryan, *U. of Texas, Austin*
Mitinori Saitou, *Kyoto U.*
Shimon Sakaguchi, *Kyoto U.*
Miquel Salmeron, *Lawrence Berkeley National Lab*
Jürgen Sandkühler, *Medical U. of Vienna*
Alexander Schier, *Harvard U.*
Vladimir Shalaev, *Purdue U.*
Robert Siliciano, *Johns Hopkins School of Medicine*
Denis Simon, *Arizona State U.*
Uri Simonsohn, *U. of Pennsylvania*
Alison Smith, *John Innes Centre*
Richard Smith, *U. of North Carolina* **(S)**
John Speakman, *U. of Aberdeen*
Allan C. Spradling, *Carnegie Institution of Washington*
Jonathan Sprent, *Garvan Inst. of Medical Research*
Eric Steig, *U. of Washington*
Paula Stephan, *Georgia State U. and National Bureau of Economic Research*
Molly Stevens, *Imperial College London*
V. S. Subrahmanian, *U. of Maryland*
Ira Tabas, *Columbia U.*
Sarah Teichmann, *Cambridge U.*
John Thomas, *North Carolina State U.*
Shubha Tole, *Tata Institute of Fundamental Research*
Christopher Tyler-Smith, *The Wellcome Trust Sanger Inst.*
Herbert Virgin, *Washington U.*
Bert Vogelstein, *Johns Hopkins U.*
Cynthia Volkert, *U. of Göttingen*
David Wallach, *Weizmann Inst. of Science*
Ian Walmsley, *U. of Oxford*
Jane-Ling Wang, *U. of California, Davis* **(S)**
David A. Wardle, *Swedish U. of Agric. Sciences*
David Waxman, *Fudan U.*
Jonathan Weissman, *U. of California, San Francisco*
Chris Wikle, *U. of Missouri* **(S)**
Ian A. Wilson, *The Scripps Res. Inst.* **(S)**
Timothy D. Wilson, *U. of Virginia*
Rosemary Wyse, *Johns Hopkins U.*
Jan Zaanen, *Leiden U.*
Kenneth Zaret, *U. of Pennsylvania School of Medicine*
Jonathan Zehr, *U. of California, Santa Cruz*
Len Zon, *Children's Hospital Boston*
Maria Zuber, *MIT*

BOOK REVIEW BOARD
David Bloom, *Harvard U.,* **Samuel Bowring,** *MIT,* **Angela Creager,** *Princeton U.,*
Richard Shweder, *U. of Chicago,* **Ed Wasserman,** *DuPont*

EDITORIAL

Leave no city behind

Close to 4 billion people live in cities. As the driver of environmental challenges, accounting for nearly 70% of the world's carbon emissions, and as sites of critical social disparities, with 863 million dwellers now living in slums, urban settlements are at the heart of global change. This momentum is unlikely to disappear, as approximately 70 million more people will move to cities by the end of this year alone. The good news is that recent multilateral processes are now appreciating this key role of cities and are increasingly prioritizing urban concerns in policy-making. Yet, how can we ensure that these steps toward a global urban governance leave no city, town, or urban dweller behind?

The third United Nations (UN) Conference on Housing and Sustainable Urban Development (Habitat III) will set out a "new urban agenda" (NUA) this October in Quito, Ecuador. Twenty years after Habitat II, the NUA aims to inspire nations, cities, and towns to pursue sustainable urban development. Habitat III's focus on cities reinforces recent wider multilateral attention to cities through the Sendai Framework for Disaster Risk Reduction, the 2030 Agenda of the UN Sustainable Development Goals (SDGs), the Addis Ababa Action Agenda on financing sustainable development, and the COP21 Paris climate agreement [including a call for a city emphasis by the Intergovernmental Panel on Climate Change (IPCC)].

Despite these good efforts, and the potential for Habitat III and the NUA to coalesce action, the road ahead is complex. Consensus that cities are critical pathways of change does not indicate agreement on what the priorities should be, how they are assessed, and how policy and implementation might be refined. The envisioned assimilation of city, national, and international indicators to track progress across sectors and scales is far ahead of the science-policy capability on the ground. Until recently, there was limited discussion of any practical approach to achieving and assessing the transformations required. Yet the metrics selected to track progress, and the credibility of the organizations that implement them, will determine how this broader global urban agenda unfolds.

"...urban settlements are at the heart of global change."

Many organizations are already assessing sustainable urban development: multilateral bodies (e.g., UN-Habitat and the World Bank), city networks (e.g., United Cities and Local Governments and C40 Cities Climate Leadership Group), academia (through the International Council for Science), think tanks (e.g., Adelphi), and foundations (e.g., the Prince's Trust). Indeed, there is no shortage of urban expertise, but there is little clarity about how concerned and knowledgeable parties might engage. The NUA's zero draft points to evidence-based policy as critical to its success and proposes an International Multi-stakeholder Panel on Sustainable Urbanization, led by UN-Habitat. Some experts suggest creating an even broader monitoring body, akin to an "IPCC for cities." Others urge caution and raise concerns about scientific input. The SDGs, Paris agreement, and Sendai Framework all call for data that are increasingly spatial (e.g., geographic information system–based) rather than statistical (e.g., demographic) to achieve the granularity that is necessary to understand cities.

Data-gathering capacity is underdeveloped, weak, or dysfunctional in many parts of the world. Building credible local data systems requires strong governmental data institutions and university-city collaborations that, with the increasing influence of large private-sector interest and capacity, are rarely in place. Africa, Asia, and Latin America are especially data (infrastructure) poor. There is no consensus on who should set metrics, who might generate and monitor data, or what the architecture of the science-policy interface underpinning global urban governance should be. Implementing a global monitoring mechanism for cities acknowledges that there are transnational drivers of urban change and embraces the idea that the way all cities are run will determine our common future. If the Post-2030 Agenda logic of "leave no one behind" is to incorporate the logic of "leave no city behind," then fundamental attention to fair, accessible, and effective monitoring and mechanisms is imperative.

– Michele Acuto and Susan Parnell

Michele Acuto is director of the City Leadership Initiative and senior lecturer at STEaPP, University College London. Email: m.acuto@ucl.ac.uk; Twitter: @Pacioz

Susan Parnell is a professor in the Department of Environmental and Geographical Sciences and the African Centre for Cities at the University of Cape Town, South Africa. Email: susan.parnell@uct.ac.za

10.1126/science.aag1385

PHOTOS: (INSET) DANIEL J. GROSHONG/BLOOMBERG VIA GETTY IMAGES; (TOP RIGHT) UNIVERSITY COLLEGE LONDON, UK/CITY LEADERSHIP INITIATIVE; (BOTTOM RIGHT) UNIVERSITY OF CAPE TOWN

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

NEWS

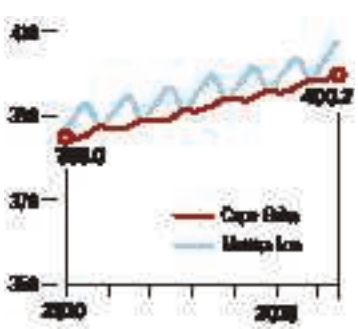
> “This is purely a thought experiment. It may not be as big news as it sounds.”
>
> **Harvard University biologist George Church** to STAT, on criticism that a 10 May meeting he co-organized on synthesizing a human genome was not open to public scrutiny.

IN BRIEF

Atmospheric CO_2 reaches a milestone

CO_2 measurements at Cape Grim on Tasmania passed 400 parts per million for the first time.

Earth has passed an unfortunate milestone. Last week, carbon dioxide (CO_2) levels at Cape Grim, an observatory on Tasmania in Australia, rose above 400 parts per million (ppm), according to scientists with the country's Commonwealth Scientific and Industrial Research Organisation. Cape Grim is one of three baseline observatories in the Global Atmosphere Watch program of the World Meteorological Organization, along with stations in Mauna Loa, Hawaii, and Point Barrow, Alaska. The measurement was expected: CO_2 measurements at Mauna Loa first surpassed 400 ppm in May 2013 and have topped that number every year since. But Northern Hemisphere observatories are subject to large seasonal cycles (see inset graph) because of the push-pull between CO_2 emissions from burning fossil fuels and the uptake by plants in spring and summer. In contrast, Cape Grim has a very small seasonal cycle—so if it's 400 Day at Cape Grim, it's 400 Day worldwide. Scientists estimate that 2°C of global warming will occur at 450 ppm.



http://scim.ag/Grim400

AROUND THE WORLD

Microbiome initiative launched

WASHINGTON, D.C. | President Barack Obama's administration last week rolled out a new plan to fund cross-disciplinary projects that would study microbes in all Earth's environments, including the human body. The National Microbiome Initiative would commit $121 million—from funding already appropriated and included in the president's 2017 budget request—to microbiome-focused grants at NASA, the Department of Energy, the National Science Foundation, the National Institutes of Health, and the U.S. Department of Agriculture. Private foundations, companies, and academic institutions have pledged another $400 million in research funding. The initiative aims to unite scientists from diverse fields to develop tools for understanding the function of individual microbes and mapping how they interact in communities. http://scim.ag/Microbelnit

Everest experiment cut short

MOUNT EVEREST, IN NEPAL | U.K. mountaineer Richard Parks has prematurely ended his team's expedition to the summit of Mount Everest, where he planned to take the highest ever blood sample and muscle biopsy (*Science*, 1 April, p. 14). Parks and his team had first climbed smaller peaks to acclimatize to high altitude, but a blood test taken last week at the mountain's Base Camp (5380 meters) showed that the acclimation was too much of a good thing. Parks's body was producing

Richard Parks being monitored in April.

CREDITS: (PHOTOS, LEFT TO RIGHT) CSIRO; RICHARD PARKS; (GRAPHIC) A. CUADRA/*SCIENCE*

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

extra red blood cells, allowing more oxygen to reach his brain—but the blood cell count was so high that it put him at increased risk of a stroke or heart attack, although he showed no outward signs of illness. This suggests that it's possible to "overacclimatize" to high altitudes—something that could inform both climbers and high-altitude clinicians, says the leader of the expedition's science team, physiologist Damian Bailey of the University of South Wales, Pontypridd, in the United Kingdom. http://scim.ag/ParksEverest

Report: U.K. research needs E.U.

LONDON | One month before the referendum that will determine whether the United Kingdom remains in the European Union, a study by a U.K.-based company has found that the country is "significantly more dependent on EU funding than other countries such as Germany." The 18 May report, by research software company Digital Science, states that "our success in gaining European funding is masking serious deficiencies" in government and business commitment to research funding. Between 2006 and 2015, U.K. organizations received £8 billion in competitive research grants ($11.6 billion) from the European Union, about a third of the amount that came from research councils within the United Kingdom, the study says. http://scim.ag/Brexitresearch

The head—including teeth, mandible, and some golden ornaments—of a horse found in the tomb.

Scythian stallions offer heads-up on horse domestication

Many biologists have thought that humans first domesticated equines 6000 years ago by breeding a few isolated individuals. "But that's not how it happened," says Ludovic Orlando, an evolutionary geneticist at the University of Copenhagen. Previous studies have shown that modern stallions share a similar Y chromosome, suggesting that breeding started with just a few males; other researchers have reported that centuries of inbreeding may have taken a toll, littering horse genomes with detrimental DNA (*Science*, 19 December 2014, p. 1439). But when Orlando and colleagues sequenced the genomes of 11 stallions found in a 2300-year-old Scythian prince's tomb (in what is now Kazakhstan), the new data suggested that these horses came from plentiful male stock, and that detrimental DNA had not yet begun to accumulate. In addition, wild horses had continued to interbreed with domestic ones, at least up until 2300 years ago, the team reported last week at the Biology of Genomes meeting in Cold Spring Harbor, New York.

NEWSMAKERS

Three Qs

Epidemiologist **Chen Chien-jen** will be inaugurated as Taiwan's vice president on 20 May. Educated at National Taiwan University in Taipei and Johns Hopkins University in Baltimore, Maryland, Chen, 64, earned scientific attention for work that led to new standards for arsenic exposure. His public profile soared when he stepped in to head the Department of Health and ended Taiwan's SARS crisis in 2003. He was minister of the National Science Council and a vice president of Academia Sinica, Taiwan's national research organization. Last fall he joined presidential candidate Tsai Ing-wen in representing the Democratic Progressive Party, known for defending the independence of the island, which China considers a breakaway province.

Q: As a scientist, what are you bringing to your new job?
A: [Plans for an] innovation-based economy and evidence-based policymaking. An innovation-based economy will be the first priority of the new Taiwan government.

Q: Are you planning any scientific initiatives?
A: We are going to initiate five research and innovation projects—green energy, biopharmaceuticals, smart machinery, information technologies, and defense—by strengthening R&D, recruiting international and domestic talent, governmental investment, international collaboration, and establishing research and innovation centers.

Q: Can scientific cooperation between Taiwan and mainland China help ease tensions?
A: Greater R&D cooperation may help maintain the status quo, which is the mainstream public will. Any bilateral collaboration [will] build a consistent, predictable, and sustainable relationship across the Taiwan Strait.

NIH hospital chief out

The National Institutes of Health (NIH) is replacing **John Gallin**, the director of the NIH Clinical Center, after an outside review found widespread problems with patient safety. With 240 in-patient beds, the Clinical Center is the largest hospital in the world devoted to research, NIH says. But the discovery last year of fungal contamination in two vials of drugs produced at the center led to an April report that found "substantial operations issues." The report blamed "fragmented governance" for some problems, referring to the fact that researchers working at the center report to leaders within their NIH institute, not the center's leadership. NIH Director Francis Collins plans to create a leadership structure similar to that of most hospitals: a chief executive, a chief operating officer, and a chief medical officer. Gallin, who has led the center since 1994, will stay on during the transition. http://scim.ag/Gallinout

PHOTO: SEBASTIEN LEPETZ/MUSEUM NATIONAL D'HISTOIRE NATURELLE, PARIS

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

PHOTO: GEORGE CLERK/ISTOCKPHOTO

HUMAN EVOLUTION

Tracking how humans evolve in real time

Analyses of thousands of sequenced genomes show changes in as little as a generation

By **Elizabeth Pennisi**, *in Cold Spring Harbor, New York*

Many people think evolution requires thousands or millions of years, but biologists know it can happen fast. Now, thanks to the genomic revolution, researchers can actually track the population-level genetic shifts that mark evolution in action—and they're doing this in humans. Two studies presented at the Biology of Genomes meeting here last week show how our genomes have changed over centuries or decades, charting how since Roman times the British have evolved to be taller and fairer, and how just in the last generation the effect of a gene that favors cigarette smoking has dwindled in some groups.

"Being able to look at selection in action is exciting," says Molly Przeworski, an evolutionary biologist at Columbia University. The studies show how the human genome quickly responds to new conditions in subtle but meaningful ways, she says. "It's a game-changer in terms of understanding evolution."

Evolutionary biologists have long concentrated on the role of new mutations in generating new traits. But once a new mutation has arisen, it must spread through a population. Every person carries two copies of each gene, but the copies can vary slightly within and between individuals. Mutations in one copy might increase height; those in another copy, or allele, might decrease it. If changing conditions favor, say, tallness, then tall people will have more offspring, and more copies of variants that code for tallness will circulate in the population.

With the help of giant genomic data sets, scientists can now track these evolutionary shifts in allele frequencies over short timescales. Jonathan Pritchard of Stanford University in Palo Alto, California, and his postdoc Yair Field did so by counting unique single-base changes, which are found in every genome. Such rare individual changes, or singletons, are likely recent, because they haven't had time to spread through the population. Because alleles carry neighboring DNA with them as they circulate, the number of singletons on nearby DNA can be used as a rough molecular clock, indicating how quickly that allele has changed in frequency.

Pritchard's team analyzed 3000 genomes collected as part of the UK10K sequencing project in the United Kingdom. For each allele of interest in each genome, Field calculated a "singleton density score" based on the density of nearby single, unique mutations. The more intense the selection on an allele, the faster it spreads, and the less time there is for singletons to accumulate near it. The approach can reveal selection over the past 100 generations, or about 2000 years.

Stanford graduate students Natalie Telis and Evan Boyle and postdoc Ziyue Gao found relatively few singletons near alleles that confer lactose tolerance—a trait that enables adults to digest milk—and that code for particular immune system receptors. Among the British, these alleles have evidently been highly selected and have spread rapidly. The team also found fewer singletons near alleles for blond hair and blue eyes, indicating that these traits, too, have rapidly spread over the past 2000 years, Field reported in his talk and on 7 May in the preprint server bioRxiv.org. One evolutionary driver may have been Britain's gloomy skies: Genes for fair hair also cause lighter skin color, which allows the body to make more vitamin D in conditions of scarce sunlight. Or sexual selection could have been at work, driven by a preference for blond mates.

Other researchers praise the new technique. "This approach seems to allow much more subtle and much more common signals of selection to be detected," says evolutionary geneticist Svante Pääbo of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany.

In a sign of the method's power, Pritchard's team also detected selection in traits controlled not by a single gene, but by tiny changes in hundreds of genes. Among them are height, head circumference in infants, and hip size in females—crucial for giving birth to those infants. By looking at the density of singletons flanking more than 4 million DNA differences, Pritchard's

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

With large genomic databases, researchers can detect evolving traits, such as blond hair in the British.

team discovered that selection for all three traits occurred across the genome in recent millennia.

Joseph Pickrell, an evolutionary geneticist at the New York Genome Center in New York City, has used a different strategy to put selection under an even keener microscope, detecting signs of evolution on the scale of a human lifetime. He and Przeworski took a close look at the genomes of 60,000 people of European ancestry who had been genotyped by Kaiser Permanente in Northern California, and 150,000 people from a massive U.K. sequencing effort called the UK Biobank. They wanted to know whether genetic variants change frequency across individuals of different ages, revealing selection at work within a generation or two. The biobank included relatively few old people, but it did have information about participants' parents, so the team also looked for connections between parental death and allele frequencies in their children.

In the parents' generation, for example, the researchers saw a correlation between early death in men and the presence in their children (and therefore presumably in the parents) of a nicotine receptor allele that makes it harder to quit smoking. Many of the men who died young had reached adulthood in the United Kingdom in the 1950s, a time when many British men had a pack-a-day habit. In contrast, the allele's frequency in women and in people from Northern California did not vary with age, presumably because fewer in these groups smoked heavily and the allele did not affect their survival. As smoking habits have changed, the pressure to weed out the allele has ceased, and its frequency is unchanged in younger men, Pickrell explains. "My guess is we are going to discover a lot of these gene-by-environment effects," Przeworski says.

Indeed, Pickrell's team detected other shifts. A set of gene variants associated with late-onset menstruation was more common in longer-lived women, suggesting it might help delay death. Pickrell also reported that the frequency of the ApoE4 allele, which is associated with Alzheimer's disease, drops in older people because carriers died early. "We can detect selection on the shortest timeframe possible, an individual's life span," he says.

Signs of selection on short timescales will always be prey to statistical fluctuations. But together the two projects "point to the power of large studies to understand what factors determine survival and reproduction in humans in present-day societies," Pääbo says. ■

IMAGES: BROAD INSTITUTE (3)

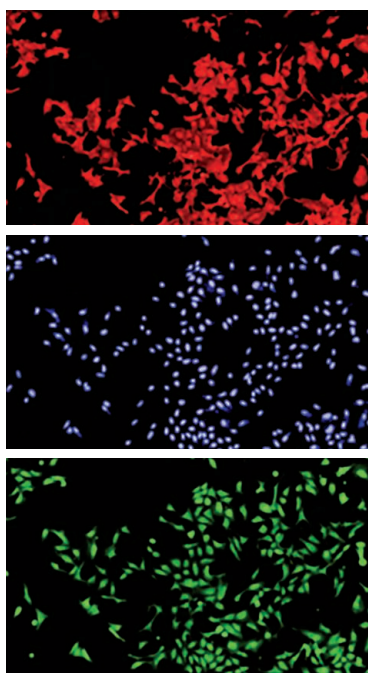
IMAGING

'Cell painting' highlights responses to drugs and toxins

Machine learning combined with molecular paints allows rapid screening of changes in cell shape and function

By **Elizabeth Pennisi**

A slight lift of an eyebrow, a tightening of the lips, a red tint rising in the cheeks—a careful watcher can gauge the emotional state of a person from such signs. Scientists are becoming similarly adept at discerning the state of a cell, with the help of molecular "paints" that highlight more than a thousand cellular features, together with computerized image processing. By assessing multiple clues at once, the technique can pick up subtle changes in cell function in response to drugs, toxins, and other factors. Last week, cell biologists, software engineers, image analysts, and machine learning experts met in Cambridge, Massachusetts, for a "hackathon" to refine this approach for profiling cells, known in its latest incarnation as cell painting.



Separate dyes used for cell painting stain components including the cell membrane, Golgi body, and actin filaments (top); DNA (middle); and mitochondria (bottom).

Their goal is to streamline and standardize the process, which they argue could cut the costs of drug development and speed discoveries in basic biology. It's "an exciting method to investigate a host of questions relating to cell biology," says Paul Rees, a physicist at Swansea University in the United Kingdom. "It has the possibility of being used routinely in every biology lab."

More than a decade ago, Steven Altschuler, a computational biologist now at the University of California, San Francisco, and colleagues showed that cells could be swiftly classified through automated analysis of more than 100 visual "features," from DNA density to membrane texture. The idea was seductive, promising a faster, cheaper way to analyze cells than methods based on gene expression or specific proteins.

Yet Altschuler was ahead of his time: The microscopes, cameras, and computational tools needed to analyze and handle the images were still immature. Since then, several teams have improved those tools. Automated image-based techniques are now commonplace in university labs and in industry, where, for example, they allow for rapid assessment of potential drugs against particular cell targets. But the amount of information that such techniques can glean from a cell is limited.

In 2013, Anne Carpenter, a cell biologist at the Broad Institute of the Massachusetts Institute of Technology and Harvard University in Cambridge, and her colleagues took the next step. They "painted" cells with six distinct molecular labels and enlisted new image analysis software to detect and track about 1400 morphological features, including the concentrations and locations of certain biomolecules. The approach allowed them to assess subtle changes in cell shape and biochemistry.

The researchers image cells before and after exposing them to a compound—a potential drug, perhaps—or altering gene activity, for instance by knocking out an unknown gene or interfering with a genetic regulator. Cameras record enough features to generate a signature for that perturbation. Machine learning algorithms can then suggest what is happening in the cell biologically by comparing that signature to ones produced by molecules and genes whose functions are known.

"It's one of the only ways you can see in a completely unbiased way how a chemical affects a biological system," explains

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

natural product chemist Roger Linington of Simon Fraser University, Burnaby, in Canada. "You can use those data to make hypotheses about what the compounds are doing."

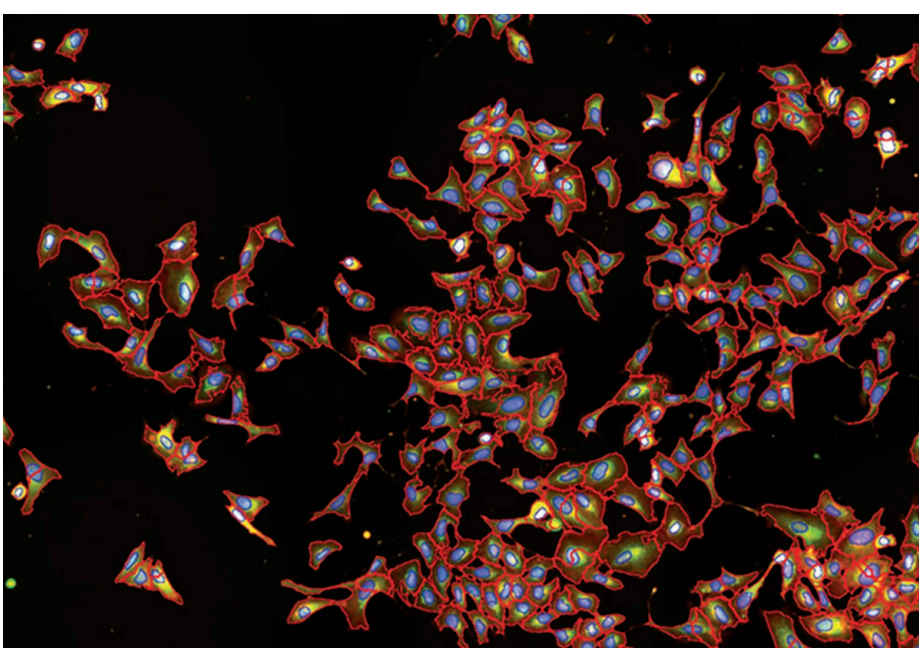
To help others embrace the approach, Carpenter's team last month posted an online preprint that essentially offers a do-it-yourself user manual for cell painting. (The paper is slated for publication in *Nature Protocols*.) "It can be implemented in anyone's laboratory," says Carpenter, who is on the scientific advisory board of a new drug-discovery company aiming to commercialize the method. A dozen academic researchers and several pharma companies have already started to do their own cell painting. Because it delivers fast, reproducible results, "you can start to think about it as an industrial process," says Richard Conroy, a program officer at the National Institute of Biomedical Imaging and Bioengineering in Bethesda, Maryland.

Drug developer Eugene Butcher, a pathologist at Stanford University in Palo Alto, California, contends that cell painting will be a better and cheaper way of evaluating drugs than the more reductionist approach of finding compounds that interact with a molecular target, and then evaluating how they alter overall cell biology. Cell biologists, too, are intrigued. At Maastricht University in the Netherlands, Jan de Boer constructs microscopic structures to force cells into myriad shapes that may exist in a body. Carpenter's approach is helping him efficiently quantify the effects of each shape on a cell's biology. "It brings to morphological screening what DNA microarrays brought to gene expression profiling," De Boer says.

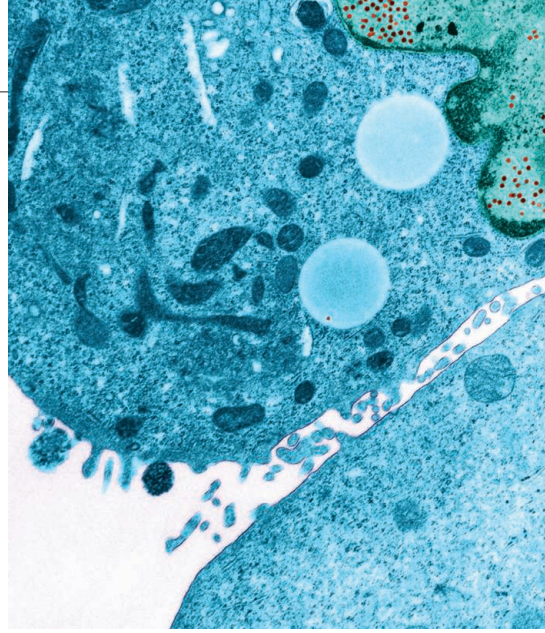
Linington predicts that researchers will begin combining cell painting with other methods. To streamline the discovery and isolation of useful compounds from plants, animals, and microbes, he'd like to marry the cell-profiling technique with mass spectrometry–based analysis of the compounds. The mass spectrometer would characterize the molecules in each extract tested, and computers would match them to the cellular changes revealed by cell painting, pairing chemistry and phenotype earlier in the discovery stage. Cell painting "offers a lot as a companion to other profiling tools," Linington says.

The technique still has a ways to go. For now, the microscopy methods only allow cells to be imaged in a single layer; Carpenter and others would like to look at cells in their natural, 3D configurations. Also, to get the most biology out of the method, researchers will have to build large databases of information about cells' reactions to known molecules and other influences, so they have baselines for comparison.

Perhaps most important, Carpenter and others have yet to show the biological importance of many of the morphological changes they're tracking—the "roughening" of a membrane when a cell is exposed to certain compounds, for example. But the cell painters expect to fill out the picture. "All of a sudden you have many more things you can measure," says Ellen Berg, a cell biologist at BioSeek in Fremont, California, a drug discovery company that she co-founded with Butcher. "I think it has a lot of power." ■



Dyes that highlight specific cell components make it possible to image a total of 1400 cell features at once.



HUMAN SUBJECTS

Researchers decry consent proposal

U.S. would require patient permission for use of anonymous tissue samples

By **Jocelyn Kaiser**

For her research on a rare childhood brain cancer, neuropathologist Jennifer Baccon needs a scarce and precious resource: tumor samples from patients. But her main source of these anonymous samples—a tissue archive at Penn State Health in Pennsylvania—could dwindle. New federal ethics rules might soon require that researchers use tissue samples only if they come from patients who have given written consent. The rules "will profoundly impact the collection of rare tumors" and "limit my ability to help patients," predicts Baccon, who works at Pennsylvania State University's College of Medicine in Hershey.

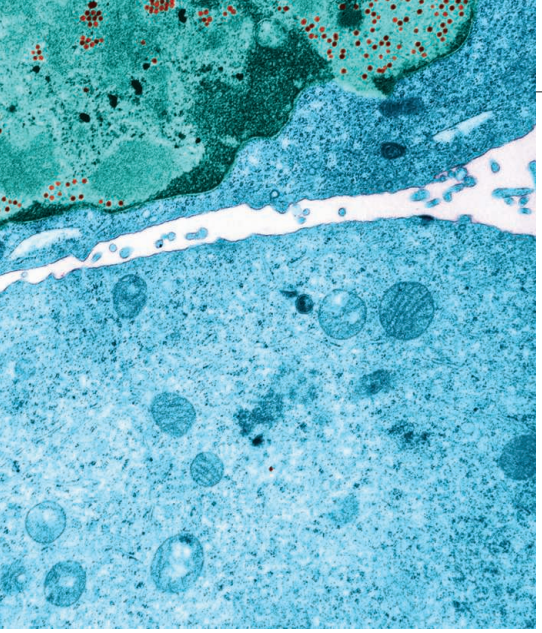
She is not alone in her dismay. The changes, proposed last fall by the U.S. Department of Health and Human Services (HHS), affect the so-called Common Rule, which protects human research subjects. A key provision would bestow on deidentified biological specimens from patients the same ethical status as a living person.

That change is "revolutionary," said one government bioethics committee. Some patient advocacy groups and bioethicists

IMAGES: (LEFT TO RIGHT) BROAD INSTITUTE; DAVID M. PHILLIPS/SCIENCE SOURCE

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS



HeLa cells, a research workhorse, were derived from Henrietta Lacks's cancer without her permission.

welcome it, saying it is necessary to protect privacy and patient rights. "The convenience of researchers is not the only interest that should count," says Hank Greely, a bioethicist at Stanford University in Palo Alto, California, who generally agrees with the proposal. "People have interests in how their DNA, blood, and tissues are used."

But research institutions warned last week that the biospecimen changes would create an expensive logistical nightmare—and asked HHS to reconsider. The proposed rule would be "damaging to science, medicine and human health" without improving patient safety, said a 9 May joint statement from three groups that speak for U.S. universities—the Council on Governmental Relations (COGR), the Association of Public and Land-grant Universities, and the Association of American Universities (AAU), all based in Washington, D.C. The statement came after a COGR-sponsored analysis of nearly 2200 public comments on the plan found "overwhelming" opposition from researchers, says AAU's Lizbet Boroughs.

The debate dates back to 2011, when HHS and other agencies floated several revisions to the Common Rule, including the new biospecimen requirement. One impetus was a 2010 bestseller, *The Immortal Life of Henrietta Lacks* by Rebecca Skloot, which traced how a woman's cervical cancer cells became a widely used research tool without her consent or the knowledge of her family. Another was a 2013 paper in *Science* questioning whether specimens—which are typically stripped of identifying information such as the patient's name and address—can be kept anonymous: Geneticists showed that if the DNA in a sample has been fully sequenced, it's potentially possible to identify the donor using public information such as genealogy databases (*Science*, 18 January 2013, p. 262).

To close such ethical gaps, HHS proposed the new rule this past September. It would apply to millions of tumor, blood, urine, and other samples gathered during routine clinical care that become part of archives used by researchers. The rule would not be retroactive—researchers would be free to use existing biospecimen collections—and would not go into effect for 3 years. And it would exempt samples used to develop a diagnostic test when researchers already know the donor's disease status, reducing privacy risks.

Under the rule, doctors would ask patients for written consent for the broad use of their deidentified samples in future research. Some patients may decline, so clinics would need to track the consent data.

That is a costly requirement that will put "extraordinary stresses" on the research community, said the Association of American Medical Colleges (AAMC), a Washington, D.C.–based group, in written comments filed earlier this year. AAMC and other groups argue that only academic research centers have the resources to obtain patients' consent and create the consent databases, which will add billions of dollars to research costs. (HHS's own estimate is $12 billion over 10 years.) The groups warn that many community hospitals and clinics will simply stop providing samples, which will skew collections toward whiter, wealthier, more urban populations.

Critics of the plan also say that there's no evidence that the current system harms patients. They argue that it's the medical data attached to specimens that should be protected, and that HHS could achieve that goal by stiffening penalties for breaches.

If HHS leaders decide to go ahead, some groups—including HHS's own human research protections advisory board—have urged the agency to simply have health care providers tell patients that their tissue might be used for research and give them the option of excluding their samples. This would reduce the burden on institutions because they would have to track only samples from patients who opted out, some proponents say. HHS has also asked for feedback on the idea of applying the rules only to samples already accompanied by whole genome sequencing data—which, at the moment, would exempt most samples.

HHS is expected to issue a final rule by the end of the year, but it is now "looking at [the comments] carefully," says Jerry Menikoff, director of the HHS Office for Human Research Protections. Research institutions, meanwhile, are hoping "common sense will prevail," AAU's Boroughs says. ■

ORGANIC CHEMISTRY

A modular route to new antibiotics

Divide-and-conquer chemistry yields hundreds of candidate medicines

By **Robert F. Service**

This week, researchers unveiled a way of cranking out hundreds of new members of one of the most popular classes of antibiotics. Macrolides—drugs that include erythromycin and azithromycin—were first developed in the 1950s and have proven vital in combatting a wide variety of infections. But microbes have become increasingly resistant to them, and coming up with new variants has proven difficult and expensive.

The new technique could help replenish medicine's supply of effective antibiotics, says Phil Baran, an organic chemist at the Scripps Research Institute in San Diego, California, who was not involved in the work. "This is a great example of beautiful chemistry that will have a tangible societal benefit," Baran says.

Chemically, macrolides are giant rings of 14 to 16 carbon atoms dangling sugar appendages. Bacteria synthesize them as weapons against their neighbors. Medicinal chemists try to make the natural versions safer and more effective by tweaking their bonds one at a time. But most of the time multiple bonds react, resulting in junk.

To solve that problem, Harvard University organic chemist Andrew Myers and colleagues adapted a divide-and-conquer strategy they had earlier applied to tetracycline antibiotics (*Science*, 15 April 2005, p. 395). They took three basic macrolide ring structures, broke each one down into eight molecular "modules," and mapped out or invented reactions that would put the pieces back together. That way they could tinker with the modules individually, then reassemble them. By repeating this strategy over and over, the researchers reported this week in *Nature*, they forged more than 300 entirely new macrolides—many of which have already shown potent antibiotic activity in lab tests. Myers has set up a company, Macrolide Pharmaceuticals, to commercialize the new work. ■

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

DATA CHECK

BEHIND THE NUMBERS

'Employment crisis' for new Ph.D.s is an illusion

By **Jeffrey Mervis**

Last month the National Science Foundation (NSF) reported that U.S. universities awarded a record number of Ph.D. degrees in 2014: 54,070, with 75% conferred in science and engineering. Those new graduates face gloomy job prospects, media reports said—"The Ever-Tightening Job Market for Ph.D.s" was a typical headline—citing data that only 61% had lined up a "definite employment commitment." And things have gotten worse in recent years, many stories added, noting the 2014 employment number was lower than the 69% reported for the class of 2009.

Don't believe everything you read. Dig deeper into the available data and you'll discover that almost all of those new Ph.D.s are gainfully employed, thank you very much. But it helps to know what to look for, and where.

The 61% figure comes from the Survey of Earned Doctorates (SED), an annual census of soon-to-be graduates. But NSF also conducts a biennial sampling of all Ph.D.-level scientists and engineers in the United States. It's called the Survey of Doctorate Recipients (SDR), and the 2013 report pegged the unemployment rate for the entire sector at a minuscule 2.1%. (By comparison, the national jobless rate that year stood at about 7.5%.)

The jobless rate for young scientists is no higher. A special analysis of the 2010 SDR data found that only 2.1% of Ph.D. scientists and engineers were unemployed 2 years after earning their degrees. And that number drops to 1.9% for those 3 to 5 years beyond their degree.

The nonexistent jobs crisis is a reminder of the dangers of taking government data at face value and using them for unintended purposes. The SED—the survey that prompted the press coverage—was never designed to measure the employment status of new graduates, says Mark Fiegener, a project officer at NSF's headquarters in Arlington, Virginia. (NSF actually contracts out both surveys to the National Opinion Research Center [NORC], an independent research organization at the University of Chicago in Illinois. The SED reaches every graduating Ph.D. student, and has a response rate of 93%. The SDR is a sampling of 120,000 people with science and engineering Ph.D.s.)

One problem with trying to use the SED to gauge the job status of graduates is that students may fill out the survey months before they actually receive a degree and, thus, may not yet be focused on the next stage of their career. Another problem is that the students' answers may not be an accurate depiction of their job prospects.

Take someone who defended her dissertation and filled out the SED in December, Fiegener says. That would be before the winter-spring academic hiring season. So she would be more likely to report "no commitment"—even if she lands a job by the time she graduates in May. In contrast, someone who completes the survey in March may have just returned from a job fair with a firm offer. So his answer would be yes.

Fiegener says a true employment survey would avoid that problem by using a reference date, such as, "Were you working for pay on 1 April?" But the SED does not, because postgraduation plans are only part of a broader suite of questions about the cost of education, field of study, and other aspects of the graduate experience.

Another complicating factor is how the SED defines unemployment. The government standard requires meeting two criteria: someone who is looking for a job and doesn't have one. But for the SED, it's enough to not have a job. That leaves out people who aren't looking for work—for whatever reason—a category that applies to 4.5% of the supposed "jobless" graduates.

An even bigger factor is those who are actively negotiating with an employer but haven't closed the deal. Some 8.8% of the respondents find themselves in that situation. Removing both categories from the SED tally would lower the 2014 Ph.D. jobless rate to 25.3%, far below the 38.6% that spawned the dire media warnings about a lost generation of Ph.D.s.

A closer look at the SED also casts doubt on the media's assertion that conditions are worsening. The 61% that the 2014 SED report counted as employed is actually a composite of both those with a firm offer and those returning to a job they held before entering graduate school. And although the total dropped from 69% in 2009, the percentage heading to a new job after their Ph.D. has stayed the same for the past decade, at about 50%. It's also worth noting that the overall total held steady for 15 years, a period spanning both the 2001 dot-com bust and the 2008 financial collapse, before dipping in 2014. That suggests the recent drop may be an anomaly.

The SDR—NSF's broader survey, which paints a sunnier picture of Ph.D. employment—follows a sample of Ph.D.s for their entire careers. Although it reveals some volatility over the last 2 decades in the jobless rate for early-career scientists, the rate remains very low. For those within 2 years of earning their Ph.D., it stood as low as 1.3% in 1999 and 2010, and reached a recent high of 3.4% in 2012, found an analysis by NORC for the Federation of American Societies for Experimental Biology in Bethesda, Maryland. Still, even that number means more than 96 out of 100 new Ph.D.s were employed.

Demographers may be aware of all these caveats. But they can easily trip up a reporter trying to dig into the latest data on employment trends. The ambiguity is worth keeping in mind the next time you read an article warning about the dismal job market for scientists. ■

How numbers can mislead

Tracking what happens to students after they graduate and enter the workforce isn't easy. These two ways of keeping score give results that appear to conflict.

38.6

percent

of Ph.D.s who graduated in 2014 didn't have a firm job commitment, according to one National Science Foundation (NSF) survey that seems to paint a bleak picture of their employment status.

2.1

percent

The actual jobless rate for scientists and engineers within 2 years of getting their Ph.D., from a second NSF survey designed to track workforce trends.

DATA: SED, SDR

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

NEUROSCIENCE

Can brain scans reveal concussion-linked disease?

By highlighting condition typically confirmed by autopsy, protein tracer could alter medical and legal landscape

By **Emily Underwood**

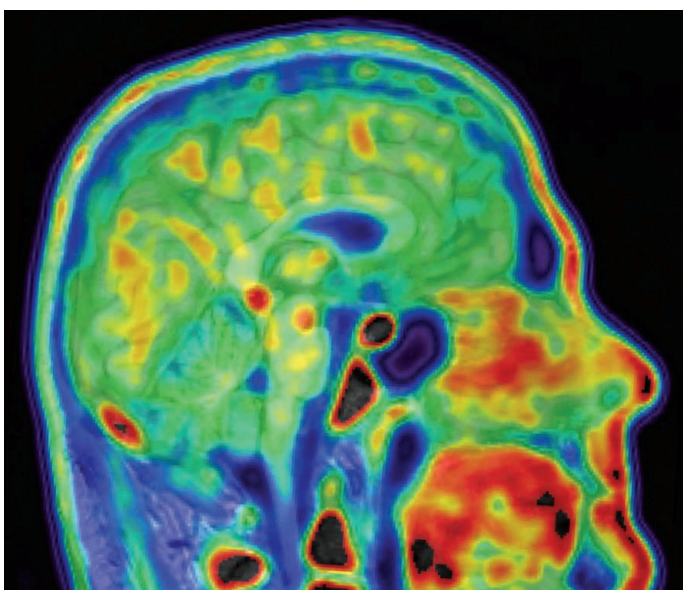
The former National Football League (NFL) player sustained at least 22 concussions over the course of his 11-year career, including four so violent they knocked him out cold. His brain shows no scar tissue or atrophy in an MRI scan, and he has the mental acumen to run his own business. But neurologist Samuel Gandy of Mount Sinai Hospital in New York City says that when lighted up by an experimental imaging agent, the 39-year-old's brain shows ominous signs of chronic traumatic encephalopathy (CTE), a brain disease that is thought to be caused by repetitive blows to the head and has been linked to the deaths of NFL players such as linebacker Junior Seau.

Last week, at the Sixth Annual Traumatic Brain Injury Conference in Arlington, Virginia, Gandy presented the former player's positron emission tomography (PET) scan as the "most dramatic" evidence yet of CTE in a living person. "I've never seen anything like it," he said of the scan, which used a PET tracer called T807 to reveal deposits of a sticky, helical protein called tau in the player's brain.

The announcement could represent a milestone for tau imaging, a promising but controversial strategy for diagnosing neurodegenerative diseases such as Alzheimer's in living patients. If the science pans out, it could also transform the medical and legal status of CTE, which at present can only be officially diagnosed after death, when a pathologist looks for tau in brain tissue. The stakes are high: The existence of a definitive CTE diagnosis could affect who gets compensated in NFL's billion-dollar concussion settlement.

For now there's no way to confirm Gandy's diagnosis, and some critics suggest it could be another example of wishful thinking in the tau imaging field. Last year the U.S. Food and Drug Administration forced researchers at the University of California (UC), Los Angeles, to dial back claims that their commercial tau PET scan, TauMark, could provide a definitive CTE diagnosis (*Science*, 24 April 2015, p. 378).

Until recently, the only way to measure tau levels in a living person was to take a sample of their cerebrospinal fluid, which doesn't reveal brain deposition patterns. In the past several years, however, researchers have developed several imaging agents, or ligands, that bind to tau in the brain and light up in a PET scan. The new ligands have quickly become powerful tools for distinguishing between a range of tau-linked diseases, including Alzheimer's, frontotemporal dementia, and Huntington disease.



The yellow-red splotches in the brain of this former football player may herald chronic traumatic encephalopathy, a concussion-linked degenerative disease.

CTE is an appealing target for tau imaging because it has a distinctive pattern of tau deposition. In 2015 an expert panel commissioned by the National Institute of Neurological Disorders and Stroke (NINDS) laid out diagnostic criteria for the disease based on samples of postmortem tissue. From the tau patterns, the experts were able to distinguish CTE brain tissue from tissue affected by other neurodegenerative diseases.

The 39-year-old athlete's scan is the first to faithfully reflect the new NINDS criteria for CTE, Gandy says. Unlike the patchy, uneven tau deposits seen in previous studies using the T807 tracer, it reveals distinct bands of tau "beautifully" outlining the brain's folds, he says. Bolstering the case for CTE are the former NFL player's behavioral symptoms, which include bouts of rage so intense that he has to live apart from his wife when they occur, Gandy says.

Other researchers using T807 are starting to see similar tau patterns in athletes at high risk for CTE. That suggests Gandy is on the right track and that T807 will be a useful diagnostic test for CTE, says Robert Stern, a neurologist at Boston University. But skeptics worry about the tendency of T807 and other ligands to bind to substances other than tau. "Whilst we might be keen to have a magic scan" to diagnose CTE, "putting patients into PET scans for tau might be misleading," says Willie Stewart, a neurologist at the University of Glasgow in the United Kingdom. Although T807 appears to bind to tau and eschew other proteins in the lab, it may not be so choosy in a living brain, he says. Stewart thinks that careful clinical workups are likely to be more useful than scans in diagnosing CTE.

The most convincing evidence that T807 does bind selectively to tau comes from studies of Alzheimer's, says neurologist Gil Rabinovici of UC San Francisco. In postmortem studies of Alzheimer's brain tissue, the tracer seems to have a special affinity for the helical filaments making up the tau tangles that are a hallmark of the disease. T807 does sometimes bind to other targets, such as a dark pigment called neuromelanin, but Rabinovici thinks "tau PET could be a real game-changer" in CTE and other neurodegenerative diseases.

Confirming that PET imaging of tau can diagnose CTE will require tracking many people over time to determine who actually develops symptoms, as well as checking for CTE-like pathology in their brains after they die, Stern says. Gandy is currently recruiting professional hockey players for such a study and is tracking the 39-year-old football player as he ages. "I'm keen to rescan him," he says. Ultimately, however, only time—and an autopsy—will tell whether the athlete has the devastating brain disease. ■

IMAGE: CRC COMMENTARY/DARA DICKSTEIN

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

FEATURES

NIH researchers infect volunteers with the flu virus in an ongoing effort to improve vaccines.

PHOTO: AP PHOTO/CHARLES DHARAPAK

Downloaded from http://science.sciencemag.org/ on May 19, 2016

THE TRUEST TEST

Studies that intentionally infect people with pathogens have a checkered past, but they are seeing a resurgence

By **Jon Cohen**

Vibrio cholerae, a comma-shaped bacterium that contaminates water and food, can kill fast. Acids in the stomach can wipe out billions of the bacteria, but if a person swallows as few as 1000 with food, some may survive the swim to the small intestine. There, the invaders will release enzymes to penetrate a thick mucous layer that lines the epithelium. Once through, the bacteria will attach to epithelial cells and begin dividing, establishing microcolonies that secrete toxins. Then the death clock begins ticking.

Irritated by the main cholera toxin, the intestine will gush fluid, and a person will develop cramping, vomiting, and, most notoriously, diarrhea that rapidly becomes a staggering volume of "rice-water" stool: a watery liquid filled with mucous flakes and epithelial cells. In severe cases, people lose a liter of rice-water stool per hour and, without rehydration to replace lost body water and electrolytes, can die within half a day.

In 1976, at the behest of a U.S. government panel, Myron "Mike" Levine of the University of Maryland School of Medicine in Baltimore began intentionally giving humans *V. cholerae*. He is still doing so today.

Forty years ago Levine was one of a tiny cadre of researchers doing so-called human challenge studies—intentionally infecting people with *V. cholerae* and other pathogens to test drugs and vaccines. But in the past few decades, this practice, which has a long and checkered past, "has become much

Published by AAAS

more mainstream," Levine says. Stricter safety procedures and new ways to weaken pathogens to reduce their risks are leading investigators in industry, universities, and government to take a new look at human challenge trials, which offer a powerful tool for studying diseases and potential therapies. There's even a commercial company, hVIVO in London, that specializes in human challenges. Today, people are being deliberately infected with malaria, influenza, shigella, dengue, norovirus, tuberculosis, rhinovirus, *Escherichia coli*, typhoid, giardia, and campylobacter.

The risks are obvious: Otherwise healthy people can suffer harm and, if the disease is contagious, potentially sicken others. But if done right, the benefits are compelling, a growing number of researchers say. The standard pharmaceutical development path for products that target pathogens moves slowly from studying safety, dosing, and biological responses in hundreds of people to an expensive efficacy trial with thousands of participants at high risk of becoming naturally infected. Human challenge studies, which only involve a few dozen volunteers, speed the process of deciding whether to scrap or pursue a promising lead, saving time and money. And tests that intentionally infect people can quickly and efficiently flag potential side effects, advocates say. "You certainly can't do a $100 million study for every candidate vaccine that appears safe and immunogenic," says Mark Mulligan, a molecular virologist who heads the vaccine center at Emory University in Atlanta and does human challenges with norovirus and tuberculosis.

Insights from human challenges reach far beyond drugs and vaccine development. Christine Moe, another Emory University researcher, has shown that norovirus more readily transmits via vomit than diarrhea, and that this "Ferrari of viruses," famous for the speed at which it races through vacationers on cruise ships, is impervious to alcohol-based hand sanitizers and to power-washing the oysters that carry it. She notes that "sometimes human challenge studies are the only way to answer critical questions."

Ripley Ballou led malaria challenge trials to test vaccines at the Walter Reed Army Institute of Research starting in 1985. Volunteers—including Ballou—willingly exposed their arms to malarial mosquitoes.

PHOTO: WALTER REED ARMY INSTITUTE OF RESEARCH

HUMAN CHALLENGES date back to the 18th century and the first vaccine, when English physician Edward Jenner attempted to persuade the world that infecting a person with harmless cowpox could prevent disease from its dreaded cousin, smallpox. Jenner scraped "matter" taken from a cowpox sore on a dairymaid's hand into the skin of 8-year-old James Phipps, the son of his occasional gardener, and then repeatedly tried to infect him with smallpox. "Poor Phipps," as Jenner later referred to the boy, never came down with smallpox. Jenner reported that some 6000 other people were vaccinated and then the "far greater part of them" were challenged with smallpox. Two centuries later, the vaccine Jenner pioneered eradicated the virus from the human population.

Intentionally infecting a human—let alone a child—with a disfiguring and even deadly disease would never pass ethical muster today. But as recently as the early 20th century, intentional infection was seen as cutting-edge: Austrian psychiatrist Julius Wagner-Jauregg won the 1927 Nobel Prize in Physiology or Medicine for injecting blood from people with malaria into patients with neurosyphilis, which putatively cured them of insanity and paralysis. As journalist Lawrence Altman documented in his book *Who Goes First?*, many investigators have challenged themselves with pathogens to prove the worth of their own experimental medicines or theories. Some died.

In the 1940s, the University of Chicago in Illinois and the U.S. Army collaborated on challenge experiments that tested malaria drugs in 400 Illinois prisoners. Nazi doctors, who horrified the world with their own medical experiments, including malaria tests that killed several hundred people, cited the U.S. studies in their defense when they were put on trial in Nuremberg, Germany, in 1947. This led to the Nuremberg Code, which spells out what are now the standard research principles of informed consent, voluntary participation, and the freedom to quit a study.

Yet U.S. experiments on prisoners continued, leading to investigative journalist Jessica Mitford's 1973 exposé in *The Atlantic Monthly*, "Experiments Behind Bars." Levine, who was just starting his career, was then challenging humans with shigella and typhoid at the Maryland House of Correction in Jessup—experiments he insists were conducted ethically. "The studies done at Jessup were 2 decades ahead of their time in terms of the methods of informed consent," he says. But in 1976, the U.S. Na-

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

Vaccine pioneer Edward Jenner put cowpox fluid into James Phipps's arm in 1796, then challenged him with smallpox.

tional Commission for the Protection of Human Subjects of Biomedical and Behavioral Research—the country's first bioethics policy effort—issued a report that effectively brought human challenge experiments in prisons to a halt.

Outside of prisons, however, research continued. In 1974 the U.S. National Institute of Allergy and Infectious Diseases (NIAID) in Bethesda, Maryland, awarded the University of Maryland a half-million dollars to create a new vaccine testing center, headed by Levine, that would recruit volunteers from colleges and church groups. The center began with influenza challenges, which were conducted in refurbished rooms at the University of Maryland Hospital that had bunk beds for 22 people and an isolated air system. The researchers had little trouble recruiting volunteers, who received the same fee as jurors ($20 a day), which the researchers deemed fair but not coercive. Volunteers had to take a written test to prove that they understood the risks.

Two years later, at the request of NIAID's Cholera Panel, Levine's group added challenges with *V. cholerae* to test cholera vaccines. "One of the really big questions was, 'Would anyone be willing to participate?'" Levine recalls. "It's one thing doing flu, which most people experience every other year or so, and it's another thing to take this exotic tropical infection and set that up."

Maryland required that the hospital fly a yellow flag to warn of a cholera quarantine area. Again, finding volunteers presented few obstacles. "These were the same young people who would go down the hairiest parts of rivers on rafts," Levine says.

The cholera studies led to the scuttling of a leading vaccine candidate, a finer understanding of effective immune responses, and, ultimately, compelling evidence that a different cholera vaccine worked. In June, the U.S. Food and Drug Administration (FDA) will consider licensing a cholera vaccine for travelers based largely on Levine's work. This is the most influential role the human challenge model has ever played in the FDA approval process.

OVER THE NEXT DECADE, Levine's group expanded to challenges with *Salmonella typhi*, *E. coli*, and rotavirus. The only other substantial human challenge operation was the Common Cold Unit established by the Medical Research Council in Salisbury, U.K. Then in 1985, a team led by Ripley Ballou began human challenges with malaria at the Walter Reed Army Institute of Research (WRAIR) in Silver Spring, Maryland. That program pioneered advances that have lowered the risks of human malaria challenges and increased the benefits, opening the way to the trials flourishing today in several places.

In a 1970s University of Maryland cholera study, this man needed 26 liters of intravenous electrolytes to replace lost fluids.

Ballou, who now heads vaccine R&D at GlaxoSmithKline (GSK), and his team bred mosquitoes in an insectary, and then fed them on human blood infected with the malaria parasite *Plasmodium falciparum*. He and five other Army colleagues each took a candidate malaria vaccine and then let five infected mosquitoes—which their group had determined was the number needed to reliably transmit the parasite—lunch on their arms. "I got a full-blown case of malaria and was never so sick in my life," Ballou says, even though he was promptly treated. "It made a huge impression on me and I was committed to finding a way to stop this disease."

In WRAIR's first vaccine trial, all the participants were "my friends in the laboratory or from down the hall," says Ballou, and they went home after being infected. Now, WRAIR recruits civilians—special policies govern participation of people serving in the military—who for up to 10 days stay in a hotel together, where they receive regular checkups. Technology has also made the experiment safer than when Ballou infected himself: The polymerase chain reaction test can detect minute amounts of parasite DNA and identify an infection 2 days earlier than traditional microscopy, and if volunteers receive immediate treatment, they rarely suffer any symptoms.

The Army's malaria challenge studies have yielded impressive dividends. "We trashed a whole bunch of vaccines," Ballou says. They also contributed to the development of GSK's RTS,S, the only malaria vaccine that has so far demonstrated efficacy, albeit modest, in a large-scale field trial.

TRIALS LAUNCHED more recently face greater regulatory scrutiny than Levine's and Ballou's did. Since the mid-1990s, FDA has deemed that organisms used in challenge studies are experimental medicines, and the agency has required researchers to submit Investigational New Drug applications before

IMAGES: (TOP TO BOTTOM) UNIVERSITY OF MICHIGAN HEALTH SYSTEM. GIFT OF PFIZER INC. UMHS.23; COURTESY OF MYRON M. LEVINE

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

conducting trials. Institutional reviews have intensified, too.

Human challenge studies with influenza provide a glimpse of the new landscape. In the 1980s and 1990s, for example, Frederick Hayden of the University of Virginia School of Medicine in Charlottesville conducted challenge studies with influenza that helped speed the development of Tamiflu and Relenza, drugs that have become the mainstays of treatment. But the work ground to a halt in 2000 after a volunteer in one of Hayden's studies experienced what FDA calls an "adverse event." A 21-year-old man testing a flu drug developed heart abnormalities after being challenged with the virus. "I still don't know what caused that episode," Hayden says. "There were a lot of sleepless nights." No long-term harm occurred, but the incident has led to a thorough review of cardiac events in other influenza challenge studies.

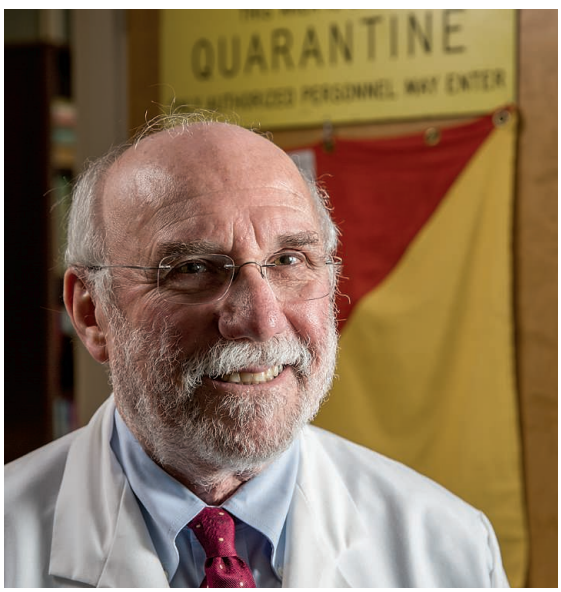
So there was considerable concern when NIAID's Matthew Memoli proposed new human challenge studies with influenza in 2011, which ultimately aimed to test novel treatments and vaccines. Some of his colleagues were so wary that the ethics department at the National Institutes of Health (NIH), NIAID's parent, was asked to conduct a formal review of the protocol. "We went through a lot of steps," Memoli says. The ethicists were particularly concerned about the proposed "high levels of payment"—up to $4000—but deemed this was not an "undue influence" because no one had an obligation to accept the offer.

Volunteers were "meticulously" screened, Memoli says: They had to be under 45 and undergo a battery of tests, including electrocardiograms. Memoli and colleagues also worked with FDA to grow a strain of the virus that met the agency's good manufacturing practices, and they precisely calculated the minimum dose needed to cause disease in most volunteers.

In ongoing studies, the researchers spray influenza virus into the noses of volunteers via a mist, created by an atomizer that only produces particles larger than 10 microns. These relatively fat particles can cause infections in the upper respiratory pathway but do not reach the lungs, where influenza virus can cause life-threatening pneumonia. To avoid infecting others, participants remain in hospital isolation rooms for 9 days. "I've challenged nearly 200 people and have had no serious complications," Memoli says. "The worst thing that happened is a guy slipped in a shower."

Memoli stresses that intentionally infecting people is an odd pursuit for a doctor. "We're purposefully making people sick," Memoli says. "It's a different idea than what you originally go to medical school for. But over the course of the next few years I think we're going to get information that's going to be tremendously helpful." In a study published online in *mBio* on 19 April, Memoli and his co-workers reported that their challenge studies indicated that a widely ignored antibody response to influenza vaccines might be a better predictor of effectiveness than the antibody routinely analyzed today.

Five years ago, the small community that studies dengue began discussing challenge trials, which made some people nervous, says Anna Durbin of Johns Hopkins Bloomberg School of Public Health in Baltimore. The mosquito-borne infection can trigger a high fever, serious joint pain, and intense rashes; in rare cases, it can lead to hemorrhaging and death. No drugs specifically target dengue virus. "We heard, 'You can't treat dengue so you can't do a human challenge model,'" Durbin says. Human challenges with dengue date back a century, but the last intentional infections of volunteers took place at WRAIR in 2001, and a few of the participants developed dengue fever.



"The first question I ask is, 'Would I want my kids ... to participate?'"
Myron "Mike" Levine, University of Maryland School of Medicine

In 2011, WRAIR and NIH sponsored a workshop to discuss "reintroducing" the human challenge model for dengue. Several attendees, including Durbin, argued that the trials could be done safely and would speed development of a badly needed vaccine for this disease.

With the blessing of FDA, Durbin in June 2013 began challenging volunteers who had received a dengue vaccine made by NIAID. Instead of using wild-type dengue virus, Durbin and her Hopkins team infected people with a naturally weak isolate of the virus that had been further attenuated in the lab. "I don't think you need to make people sick" to see whether they develop an infection, Durbin says.

As she and her group reported online on 16 March in *Science Translational Medicine*, none of the 21 people who received the vaccine became infected after the challenge, but all 20 controls had the virus in their blood, and 16 developed a rash. Based in part on these results, the Butantan Institute in São Paulo, Brazil, this year launched an efficacy trial of the vaccine that plans to enroll 17,000 people.

Regulations of human challenge studies differ from place to place. In the United Kingdom, for example, challenge agents are not considered drugs, and experiments with them thus don't require regulatory approval. A group at the University of Oxford led by pediatrician Andrew Pollard has conducted a challenge study of experimental vaccines against typhoid and paratyphoid. Although both diseases are contagious, the researchers allow volunteers to go home instead of staying in isolation. "They're potentially shedding organisms that are going into a flush toilet," says Levine, who collaborates with the Oxford group. "That's something that's not amenable to being carried out in the USA."

The United Kingdom is "much more permissive," agrees Pollard, but he says the trials go through extensive ethical reviews, and the risk of transmission is "near zero" if people have good hygiene.

The human challenge model has its limits, Levine stresses, noting that his group declined to participate in an experiment done elsewhere that put *Neisseria gonorrhoeae* in a penile catheter to study gonorrhea transmission. "The first question I ask is, 'Would I want my kids, siblings, or spouse to participate?'" Levine says. "If the answer is 'no,' we don't do it." And he worries that even though researchers today address risks more carefully than ever before, someone could push too far and undo the gains the field has made. "This should not be a Wild West show," he says. "Some newcomers may not be totally aware of the burden the pioneers went through. It has taken a lot of time to get buy-in from everyone imaginable." ■

PHOTO: PAUL FETTERS

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

INSIGHTS

LETTERS

Edited by Jennifer Sills

Downloaded from http://ScienceMag.org on May 19, 2016

INGENUITY: NEXTGEN'S VISION

Full speed ahead to the City on the Hill

We asked young scientists from a variety of fields this question: **According to the United Nations (http://esa.un.org/unpd/wup/), more than two-thirds of the human population will live in cities by 2050. How can scientists in your field help society prepare for an increasingly urbanized world?** Excerpts of our panel's proposed contributions follow, categorized by the challenges they address. For responses that span categories, we have printed the author's name with the first excerpt and initials thereafter. Read the full responses, and many more, at scim.ag/UPInGen.

POLLUTION AND WASTE MANAGEMENT

To efficiently overcome pollution, we could use engineered microorganisms with altered biochemical or metabolic pathways to degrade environmentally toxic compounds. Enhanced phytoremediation of toxic and heavy metals from the environment, through the assistance of engineered microorganisms, could facilitate the treatment of polluted soil and water on a large scale.

ONLINE

Follow *Science*'s InGenuity on Twitter with the hashtag #InGenSci

Bipin Singh

Center for Computational Natural Sciences and Bioinformatics (CCNSB), International Institute of Information Technology (IIIT), Hyderabad, Hyderabad, Telangana, 500032, India. Email: bipin.singh@research.iiit.ac.in

When a chemical has risk potential, chemists can formulate similar chemicals without negative health and environmental implications. Disposal of a chemical must involve a regenerative cycle rather than a destination, lest that destination become dangerously close in proximity to our living spaces.

Rose Joy Crocker

Adelaide, South Australia, 5022, Australia. Email: rose.crocker@me.com

Scientists have proposed the use of molecular sponges to purify land and sea vehicle emissions, to filter the emissions of large-scale industry, and to passively generate water using the humidity and temperature differences between day and night.

Timothy Easun

School of Chemistry, Cardiff University, Cardiff, CF10 3AT, UK. Email: easuntl@cardiff.ac.uk

The inside of factory pipelines can be coated with chemical materials to react with the waste moving through them and produce environmentally friendly materials. Water treatment with suitable chemical materials can allow water to be

PHOTO: SEAN PAVONE/ISTOCK PHOTO

Published by AAAS

reused. Plastic is a petrochemical nondecomposable material, but it can be reused through recycling and the help of chemical additives. Even food can be turned into biofuel with the right catalyst. Chemical filters can make emitted gasses safe or even recyclable.

Basant A. Ali

Department of Chemistry, Faculty of Science, Alexandria University, Alexandria, El-Montaza, 21611, Egypt.
Email: basant_walieldeen@alex-sci.edu.eg

Biologists can collaborate with app designers and hardware developers to create inexpensive cellphone (app)-based biosensors for easy water-quality analysis, analogous to the new generation of glucose-monitoring instruments. The water-quality analysis apps could be integrated with social media, alerting others if water tests positive for contaminants and flagging areas with polluted water supply on the map.

Gunjan Guha

Cellular Dyshomeostais Lab (CDHL), School of Chemical and Bio Technology, SASTRA University, Thanjavur, Tamil Nadu, 613401, India.
Email: gunjan.doc@gmail.com

RENEWABLE ENERGY

The development of novel chemical materials to be used in dye-sensitized solar cells or perovskite solar cells will improve efficiency and produce more energy.

B.A.A.

Waste products like rinds, peels, or bones will be saved for microbial fuel cells that biochemical engineers hope to place under the sink.

Michael A. Tarselli

NIBR Informatics, Novartis Institutes for BioMedical Research, Cambridge, MA 02139, USA. Email: mike.tarselli@novartis.com

Biochemists the world over are working to turn hydrogen-making microbes into large-scale hydrogen production systems.

M. Clayton Speed

Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, CO 80523, USA. Email: cspeed@rams.colostate.edu

Fusion promises to be a source of clean energy, but only with the hard work of our plasma and nuclear physicists. The creation of higher-capacity, safer batteries would go a long way toward solving our energy storage needs.

Congzhou Mike Sha

Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA. Email: consha@sas.upenn.edu

HEALTH CARE AND DISEASE PREVENTION

As the global population moves into cities and social networks grow denser, the spread of disease will likely pick up its pace. But so can the spread of health! Epidemiologic tools afford us a bird's eye view of the evolving patterns of disease distribution, which can then be harnessed to inform smart public health campaigns and urban planning initiatives.

Stella Aslibekyan

Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL 35205, USA.
Email: saslibek@uab.edu

Data scientists can use mobile health data from fitness devices and smartphones to model the optimal amount of walking required to minimize cardiovascular risks, which can help inform urban planning decisions.

Edward Lau

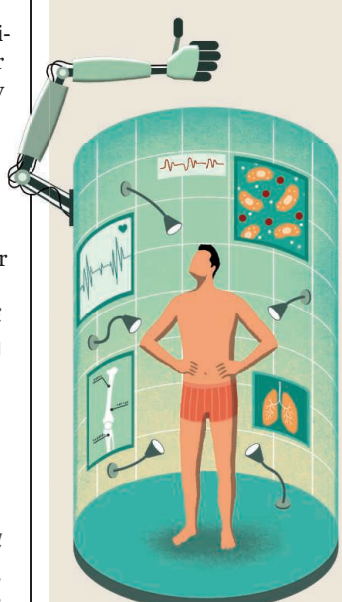
Department of Physiology, University of California at Los Angeles, Los Angeles, CA 90095, USA.
Email: edward.lau@me.com

With two-thirds of the world population living in big cities, carcinogens such as smoke, asbestos, radiation, and those derived from infection might increase. It would be crucial to provide early cancer diagnosis, mandatory tests, and public health efforts, along with personalized treatments. However, none of these policies would be useful unless Third World countries were included and protected from the high costs of tailor-made cancer medicine and regular clinical tests.

Ada G. Blidner

Laboratorio de Inmunopatología, Instituto de Biología y Medicina Experimental, CONICET, Buenos Aires City, Buenos Aires, C1428ADN, Argentina.
Email: adablidner@gmail.com

Biomedical informaticians can use mobile technology and the Internet of Things to track dietary patterns, build exercise break reminders, record blood pressure and blood sugar levels routinely, and monitor psychological stress. Harnessing big data analytics, we can identify lifestyle and biomarker patterns associated with the development of illnesses and establish an early-warning system for disease prevention. With secure data transmission and storage methods, we can link patients' lifestyle information with



FLASH FORWARD

Health in style

Yawning, I step into my in-home physician assistant (PA) station. I am just old enough to remember when urbanites, such as myself, needed in-person checkups for their medical peace of mind. How could a scant, single, annual visit to a physician possibly diagnose everything? Slowly, the in-home PA's sleek, brushed, chrome arm, embedded with diagnostic electronics, arcs around my body. Unlike antiquated medical equipment of the past, my in-home PA2050LE is subtle, dare I say even tasteful, and blends with my bathroom's decor. In the early 21st century, there were attempts to make wearable devices for more personalized wellness. However, the true breakthrough in personalized wellness came when scientists integrated engineered cells with the monitoring electronics. Now, my PA2050LE can interface with the cells on my skin and with the cells in my body, communicating with populations of benign engineered bacteria living among my gut's microbiome. "Alert, Alert" the PA2050LE blares. "Deficient Vitamin B Detected." Clearly, the engineered cells has identified an issue and relayed a phenotypic change to the PA2050LE. In the past, nutrition abnormalities could only be approximated by diet assessment. Now with automated daily checkups, any of my biochemical markers can be monitored, recorded to my cloud-based medical records, and... "Up-regulate Vitamin B production" I respond to the PA2050LE. Within hours my engineered microbiome will synthesize the vitamin I need. I leave the bathroom, thanking my in-home PA again for saving me, and millions of others, the time and money for a trip to the doctor's office.

Keith Cameron Heyde

Biomedical Engineering and Mechanics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24060, USA. Email: kch2118@vt.edu

ILLUSTRATION: DAVIDE BONAZZI/@SALZMANART

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

MODEL SYSTEMS

Lessons from the colony

The study of ants reveals important lessons for humans living in an increasingly urban ecosystem. There are many architectural and functional adaptations of ant nests that human city planners would be remiss to ignore. Ant colonies use thermal gradients to circulate air without expending energy, whereas our buildings often spend about 50% of total energy on heating and cooling. When escaping a high-density space, ant groups undergo a liquid-to-glass phase transition, facilitating a quick group exit. Human crowds and traffic jams, in contrast, undergo a liquid-to-solid phase transition at high densities, leading to trample damage and inefficiency. Ants maintain impeccable hygiene, preventing the spread of disease. We do not need to lick each other to stay clean, but we must reconsider public health surveillance strategies in light of ongoing human epidemics. The deepest lesson of all relates to the most crucial problem facing humanity today—overpopulation—addressed by the communal brood-rearing strategies of ants.

Daniel Ari Friedman

Department of Biology, Stanford University, Stanford, CA 94305, USA. Email: dfri@stanford.edu

their medical records, which will provide physicians with valuable longitudinal data, facilitate patient follow-up, and ensure the quality of home care.

Kun-Hsing Yu

Biomedical Informatics Training Program, Stanford University, Stanford, CA 94305, USA. Email: khyu@stanford.edu

Neuroscientists can help society combat the negative influence of urban living on inhabitants' brain biology by using research to determine how economic, environmental, and social stressors translate into cognitive and mental disorders, including anxiety and depression. Findings from the research can be used as a foundation to work with local community organizations and urban planning and governmental entities to promote and implement public health strategies that can reduce the risk for brain disorders in urban areas.

Joyonna Carrie Gamble-George

Vanderbilt Brain Institute, Vanderbilt University School of Medicine, Nashville, TN 37210, USA. Email: joyonna.c.gamble-george@vanderbilt.edu

With increasing digital interconnectivity in urban areas, electronic monitoring could reduce the frequency of outpatient visits a patient would require for chronic disease management, while allowing a physician to optimize drug therapy more effectively for patients.

Bryce W. Q. Tan

Department of Physiology, National University of Singapore, Singapore, 117597, Singapore. Email: brycetan03@hotmail.com

Analytical chemists will lead development of smartphone-based sensors for volatile organics in your breath, to diagnose disease before a hospital visit. Drinking water, long subject to fluoridation, may begin to incorporate trace amounts of statins, aspirin, or vitamins to improve general health.

M.A.T.

COMMUNITY BUILDING

From parks to vegetated strips or street trees, nature-based solutions help to build a sense of community and substantially improve our physical and mental health. Urban ecosystem services researchers evaluate urban nature's effect on communities and design more resilient and efficient urban environments in a way that is fair to all.

Perrine Hamel

Natural Capital Project–Woods Institute for the Environment, Stanford University, Stanford, CA 94305, USA. Email: perrine.hamel@stanford.edu

Social big data include demographic data, online social media data, and individ-uals' daily interaction data with external environment (such as household, transportation, and public service data). From the social big data analytics, the behavioral patterns of new citizens and the operation status of the urbanized world can be revealed.

Kaile Zhou

School of Management, Hefei University of Technology, Hefei, Anhui, 230009, China. Email: kailezhou@gmail.com

The results of cooperation analyses can help policy-makers to guide society to a more harmonious coexistence.

Ádám Kun

Plant Systematics, Ecology, and Theoretical Biology, Eötvös University, Budapest, 117, Hungary. Email: kunadam@elte.hu

CRIME PREVENTION AND SAFETY

To prepare for an increasingly urbanized and technically integrated world, digital forensic cyber-crime investigation and information security researchers are developing new techniques to quickly investigate rapidly emerging, large, distributed, complex systems and data in order to detect and reconstruct criminal activities. Especially important, researchers must be able to prove the reliability and integrity

ILLUSTRATION: DAVIDE BONAZZI/@SALZMANART

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

of evidence derived from these complex cyber-physical systems to meet the standards of admissibility in court.

Joshua I. James

Legal Informatics and Forensic Science Institute, Hallym University, Chuncheon, Gangwon, 24252, Korea. Email: joshua.i.james@hallym.ac.kr

A rise in crimes, drugs, and addictions could be tackled with the use of analytical tools available in miniaturized forms to the law-keepers to record the entire human metabolome snapshot at the site of crime, to test the victim's cause of demise and suffering, and possibly to deliver justice to convicts on-site in a timely fashion.

Biswapriya Biswavas Misra

Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX 78227, USA. Email: bbmisraccb@gmail.com

With increasing criminal rates in cities, security systems will become even more difficult to crack as they incorporate genomic information. By 2050, we might need to provide part of our genome sequence to enter workplaces, banks, and shops.

Klaudija Daugelaite

Lithuanian University of Health Sciences, Kaunas, LT-50393, Lithuania. Email: klaudija.daug@gmail.com

If more than two-thirds of the human population lives in cities by 2050, there will be an increasing number of high-rise buildings, which may cause more urban fire-safety problems. By collecting fire hazards data with high-precision and real-time monitors powered by big data and cloud computing technology, we can obtain maps of fire hazards and conduct comprehensive risk analyses. We can also focus on fire technology innovation for high-rise buildings, such as fire-fighting robots and heat-resisting escape compartments.

Jian Zhang

School of Safety and Environmental Engineering, Hunan Institute of Technology, Hengyang, Hunan, 421002, China. Email: zhangjian3954@126.com

FOOD SECURITY

In "Climate Smart Cities," parks and green areas must not only satisfy aesthetic and recreational demands, but also provide space to empower citizens to produce their own food.

Nicolas Eduardo Bambach

Centro de Cambio Global Pontificia, Universidad Catolica de Chile, Santiago, RM, Chile. Email: nbambach@ucdavis.edu

Molecular biologists can help by working together on a food source that is easily transportable, nutritious, palatable, and above all, sustainable. Such a source could draw from any number of organisms, such as algae, seaweed, rotifera, and other photosynthetic/bottom-dwelling organisms. This would enable the food source not to depend on the availability of feed (as is the case with farmed fish and livestock).

Michael Patrick Christopher Schwoerer

The Vagelos Scholars Program in the Molecular Life Sciences, University of Pennsylvania, Philadelphia, PA 19128, USA. Email: mschwo@sas.upenn.edu

Research on plant-microbe interactions has the potential to improve productivity of crops that are grown indoors in space-efficient and environmentally friendly "vertical" farms. Microbes can make nutrients available, increase growth rates, and stimulate disease resistance in plants. Indoor farming has grown rapidly in recent years, partly because of new LED lighting technology. Hydroponic and aeroponic growth methods offer substantial water savings.

J. Steen Hoyer

Computational and Systems Biology Program, Washington University in St. Louis, Donald Danforth Plant Science Center, St. Louis, MO 63132, USA. Email: j.s.hoyer@wustl.edu

Genetically engineered foods will be cheap, contain high nutritional value, and require no preparation time.

K.D.

INNOVATION

It is time for scientists to combine technologies on artificial intelligence (AI) and modern wireless communications to ease the traffic problem. AI makes cars smart; self-driving cars can make decisions automatically according to traffic conditions. Meanwhile, vehicle-to-vehicle communications and wireless systems make it possible for cars to talk with each other, regardless of the distance between them. With these technologies, massive amounts of information can be exchanged among cars, and a sophisticated scheduling and decision-making scheme can enable individual cars to make wise decisions and avoid traffic jams.

Lei Jiao

Department of Information and Communication Technology, University of Agder, Grimstad, 4885, Norway. Email: lei.jiao@uia.no

Research in carbon nanotube and graphene transistors could be the key to continuing the past 50 years of exponential growth in computational power. As our cities become larger, the distribution of resources will become too difficult a problem for humans to solve without the aid of robust, automated control systems, designed to be mathematically correct and physically optimal.

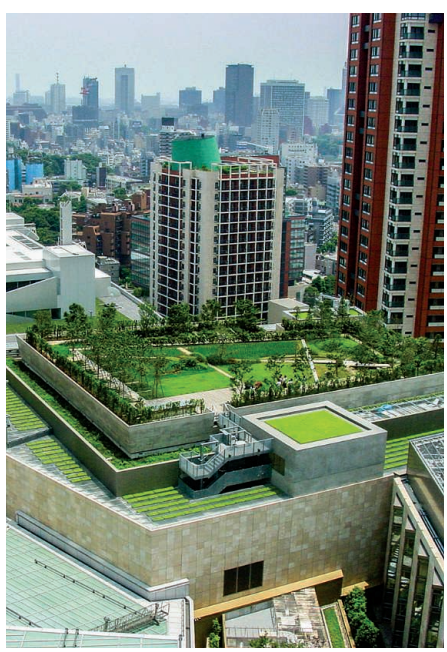
C.M.S.

With the help of synthetic biology, the city of the future can potentially be engineered less like a conglomerate of machines and more like a natural system. Cells can be engineered to sense and respond to environmental signals. Communication networks through sensors can be embedded in the city infrastructure. Protocell technology, a new material that possesses some of the properties of living systems, can be manipulated to grow architecture. We might see self-repairing architecture.

Gerd Moe-Behrens

Leukippos Institute, Berlin, Berlin, 10777, Germany. Email: gerdmoebehrens@icloud.com

Landscape architects and ecological engineering scientists can design networks of self-sustaining and functioning natural ecosystems within the urban matrix. Urban parks, green roofs, and vegetation corridors will be more common and better connected to ensure the continuity of the natural ecosystem in the urban environment. The paradigm needs to shift from planting trees and building parks within the city to building the city around the natural environment.



Stéphane Boyer

Department of Environmental and Animal Sciences, Unitec Institute of Technology, Auckland, 612, New Zealand. Email: sboyer@unitec.ac.nz

10.1126/science.aag1520

PHOTO: LAWRENCEKARN/ISTOCK PHOTO

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

PERSPECTIVES

NANOTECHNOLOGY

Changing of the guard

DNA nanostructures mimic membrane proteins

By **Stefan Howorka**

Membrane proteins control access of ions and molecules to a cell's interior, shuttle cargo and information across the cell boundary, and determine the cell's shape. Engineering the function of these proteins is key to the development of vaccines, biofuels, biosensor elements, and research tools. However, the range of accessible architectures is limited, because protein engineering usually involves making relatively modest structural changes to existing protein structures; folding extensively altered polypeptides into defined structures is very challenging. Recent studies have shown that some membrane-protein functions can be mimicked with DNA nanostructures, which are easier to manipulate than their natural templates.

The programmable base-pairing of DNA has previously been exploited to create nanoscale architectures with angstrom-scale precision (*1*). These nonmembrane DNA origami nanostructures self-assemble from mixtures of component strands to form interlinked duplexes that act as the main structural domains. Scientists have used this approach to create molecular motors and antibodies.

Department of Chemistry, Institute of Structural Molecular Biology, University College London, London WC1 H0AJ, UK. Email: s.howorka@ucl.ac.uk

However, creating biomimetic membrane structures poses a further challenge: how to insert the hydrophilic, negatively charged DNA into hydrophobic lipid bilayers. This challenge can be overcome by attaching hydrophobic membrane anchors, such as cholesterol, to the DNA nanostructures, as pioneered for DNA duplexes (*2*, *3*). The resulting biomimetic architectures float on bilayers without puncturing them (see the figure). These structures are analogous to natural membrane proteins that are components of the cytoskeleton or of signaling cascades.

Self-assembly with DNA is highly modular, allowing membrane-floating nanostructures of different shapes to be built, including bricks (*4*), stars (*5*), and flat rectangles (*5*, *6*). At diameters of up to 80 nm, the structures are larger than amino acid–based engineered membrane proteins and can thus control the shape of lipid vesicles (*4*, *5*). The nanorectangles can be tuned to oligomerize into larger superstructures and undergo a nanoscale shape change (*4*–*6*). Furthermore, DNA origami can form ringlike scaffolds as large as 200 nm that enclose vesicles like a ring around Saturn (*7*, *8*). Similarly, spherical DNA scaffolds can function as endoskeletons inside vesicles (*9*).

DNA structures can also form membrane-puncturing pores (see the figure). Like natural channels, they control transport of water-soluble cargo across biological bilayers. DNA pores have a structural core of several interlinked DNA duplexes that encloses a hollow lumen. Membrane insertion is made possible by modifying the outside of the nanostructure with cholesterol (*10*) or by chemically removing negative charges on the DNA backbone (*11*). The pore diameter can be tuned by altering the number of DNA duplexes (*12*). My group has shown that a pore with a six-duplex core can be turned into a synthetic ligand-gated channel by adding a DNA lid at one end; the channel entrance can be reopened with the appropriate ligand. This channel is highly selective for the transport of small, charged organic molecules (*13*). We have also shown that the pores can be opened and closed via voltage changes, similar to biological ion channels (*14*).

These nanostructures are remarkable; DNA would not naturally interact in such a defined manner with bilayers. But there are also important applications. In research, membrane-anchored DNA plates are used as tools to reorganize the fluidity or local morphology of the lipid bilayer (*4*–*9*) and thereby influence biological behavior. In the study of cells and their activation, DNA plates may be used as molecular pegboards to present cell-activating proteins in stoichiometrically defined low numbers and at precise geometry and nanoscale distance to each other. Furthermore, vesicle-enclosing DNA rings with optional membrane proteins can help to induce contact with other vesicles or planar membranes, providing an important tool for studying membrane fusion (*7*, *8*).

In biotechnology, DNA scaffolds can control the size and stabilize bilayer vesicles that serve as bioimaging agents or deliver encapsulated drugs (*7*, *9*). DNA pores could also serve as components in portable, label-free biosensors. Currently, few protein pores are used for electrical DNA sequencing. But to extend electrical biosensing to other diagnosti-

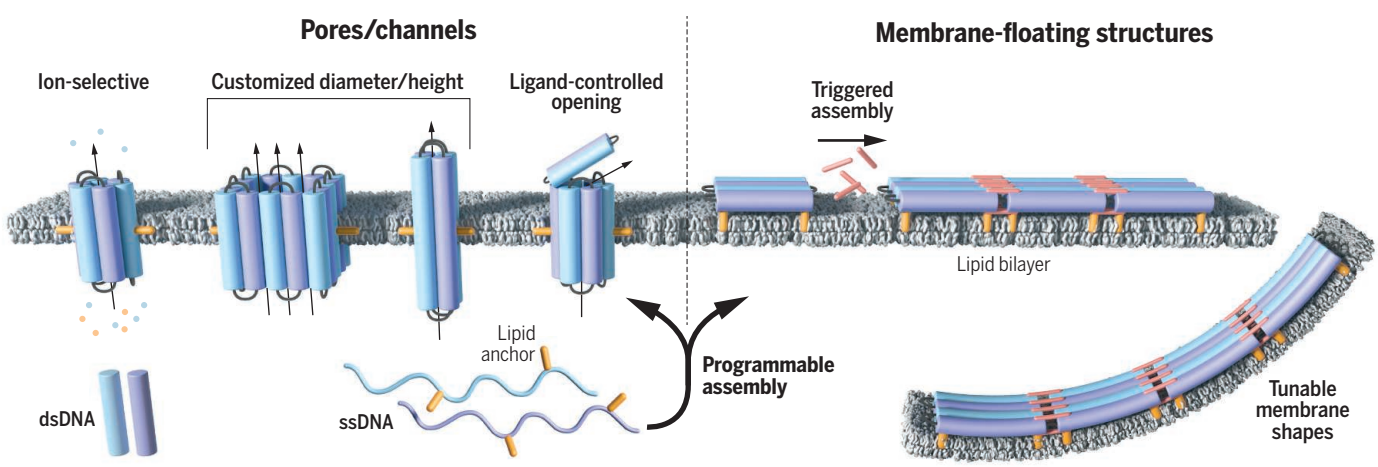


ILLUSTRATION: C. BICKEL/*SCIENCE*

DNA-based gatekeepers. DNA nanostructures have been rationally designed to form pores and channels or to adhere to the membrane bilayer, thereby mimicking membrane proteins in living systems. The DNA nanostructures are of interest for a range of research and biotechnology applications. Only selected DNA structures are shown; they are drawn schematically to maintain visual clarity. dsDNA and ssDNA, double- and single-stranded DNA.

Published by AAAS

Downloaded from http://science.sciencemag.org/ on May 19, 2016

cally important analytes, rational DNA design could construct more suitable pores. Ligand-gated DNA channels could further be used in vesicles for controlled drug release (*13*).

Membrane DNA nanostructures expand the toolkit of synthetic biology. In classical synthetic biology, engineered DNA or protein parts are used in their traditional biological roles to create artificial viruses, biocatalytic microcompartments, or cells with designed genetic circuits. The DNA rafts and pores break with this convention by using DNA outside its usual functional spectrum. They thus complement other engineered DNA units that mimic the biological function of soluble proteins in catalysis or signal processing within lipid membrane compartments (*15*). Currently, these systems are rudimentary but can help us to understand natural ones by providing a simplified and modular model. Variants composed solely of DNA might also shed light on the origin of life.

Future research in this young field will explore DNA designs of different size, geometry, and lipid anchoring. This will help to answer questions concerning how nanostructures interact with and deform membranes. Of additional interest is how the structurally flexible and porous DNA nanostructures compare to more rigid proteins, and how to combine DNA with other protein or polymeric components for hybrid or biomimetic nanostructures. To realize the potential of DNA nanomaterials in biotechnology, it will be important to produce them more cheaply (*1*). Finally, to create self-replicating synthetic cells, membrane-anchored DNA nanostructures must be made solely from biological components. With further progress in these research areas, DNA-based gatekeepers at biological membranes are well positioned to exploit powerful engineering with DNA nanotechnology. ■

REFERENCES AND NOTES

1. Y. J. Chen *et al.*, *Nat. Nanotechnol.* **10**, 748 (2015).
2. C. Yoshina-Ishii, S. G. Boxer, *J. Am. Chem. Soc.* **125**, 3696 (2003).
3. F. Baldelli Bombelli *et al.*, *Soft Matter* **5**, 1639 (2009).
4. A. Czogalla *et al.*, *Angew. Chem. Int. Ed.* **54**, 6501 (2015).
5. S. Kocabey *et al.*, *ACS Nano* **9**, 3530 (2015).
6. A. Johnson-Buck *et al.*, *ACS Nano* **8**, 5641 (2014).
7. Y. Yang *et al.*, *Nat. Chem.* **8**, 476 (2016).
8. W. Xu *et al.*, *J. Am. Chem. Soc.* **138**, 4439 (2016).
9. S. D. Perrault, W. M. Shih, *ACS Nano* **8**, 5132 (2014).
10. M. Langecker *et al.*, *Science* **338**, 932 (2012).
11. J. Burns, E. Stulz, S. Howorka, *Nano Lett.* **13**, 2351 (2013).
12. K. Gopfrich *et al.*, *Nano Lett.* **15**, 3134 (2015).
13. J. R. Burns *et al.*, *Nat. Nanotechnol.* **11**, 152 (2016).
14. A. Seifert *et al.*, *ACS Nano* **9**, 1117 (2015).
15. M. Weitz *et al.*, *Nat. Chem.* **6**, 295 (2014).

ACKNOWLEDGMENTS

Supported by UK EPSRC grant EP/N009282/1, BBSRC grants BB/M012700/1, BB/M025373/1, and BB/N017331/1, UCL Chemistry, Oxford Nanopore Technologies, and the National Physical Laboratory. I thank R. Caygill, A. Sella, and Y. Xing for critically reading the manuscript. Illustration concept by J. R. Burns and A. Hodel.

10.1126/science.aaf5154

PHYSICS

Superconductivity on the edge

Hybrid correlated systems could reveal exotic quantum transport behavior

By **Nadya Mason**

Standard superconductors consist of a condensate of paired electrons, called Cooper pairs. The transport behavior of these pairs at junctions can produce exotic effects that are of fundamental and practical interest. When two superconductors are in contact via a normal metal, the pairs must convert to single-particle states to traverse between superconductors. This occurs by the Andreev process, whereby a low-energy electron in the normal metal injects a Cooper pair into the superconductor and generates a hole that reflects back into the metal; coherent, opposite-momentum electron-hole pairs then carry the supercurrent across the metallic junction (*1*) (see the figure, panel A). In the case of superconductors connected to a quantum Hall state, there are only one-way paths along the junction edges. Here, a new type of Andreev process is predicted to occur, whereby electron and hole states on opposite sample edges carry the supercurrent. This prediction was made more than 20 years ago (*2*), but clear observation of the effect was frustrated by the difficulty of creating coexisting superconducting and quantum Hall states. On page 966 of this issue, Amet *et al.* (*3*) report on the interplay between these two states, finding evidence for the unconventional Andreev process. Their results confirm new physics that appears when two correlated states are connected, and opens the door to a range of novel excitations and exotic devices.

Superconductivity and quantized Hall systems are both macroscopic quantum phenomena that allow electrical currents to flow without resistance. The quantum Hall effect (QHE) occurs in two-dimensional systems subjected to perpendicular magnetic fields, where electrons undergo cyclotron motion (*4*). The QHE appears for fully filled cyclotron orbitals, or when the density of electrons per unit area is a value ν times the density of magnetic flux, where ν is an integer in the integer QHE and a small rational factor for the fractional QHE. In the QHE regime, the current is fully carried by dissipationless edge states that travel in only one direction along the edge (a chiral edge current).

Studies of the integer and fractional QHE have yielded many surprises over the years, and there has been great interest in understanding the emergent physics from coupling the QHE to other correlated states. But this coupling has proven experimentally difficult: The best examples of QHE occurred in semiconductors such as Si and GaAs, which have large Schottky barriers to metals, and in heterostructures where the interface is difficult to access. More recently, however, a new, surface-accessible QHE system was discovered—graphene (*5*). As a monolayer sheet of hexagonally bonded carbon atoms, graphene exhibits both integer and fractional QHE. It can be coupled to superconducting leads, demonstrating both Josephson effects and supercurrents. In addition, recent progress has been made in creating ultrahigh-quality, low-noise electron systems in graphene by encasing it in hexagonal boron nitride (HBN) (*6*). Amet *et al.* use these fabrication advances to create high-quality graphene samples encased in HBN and contacted by a superconductor along the edges. Sensitive transport measurements reveal not just superconducting and QHE regimes in tandem, but an interspersing of the effects.

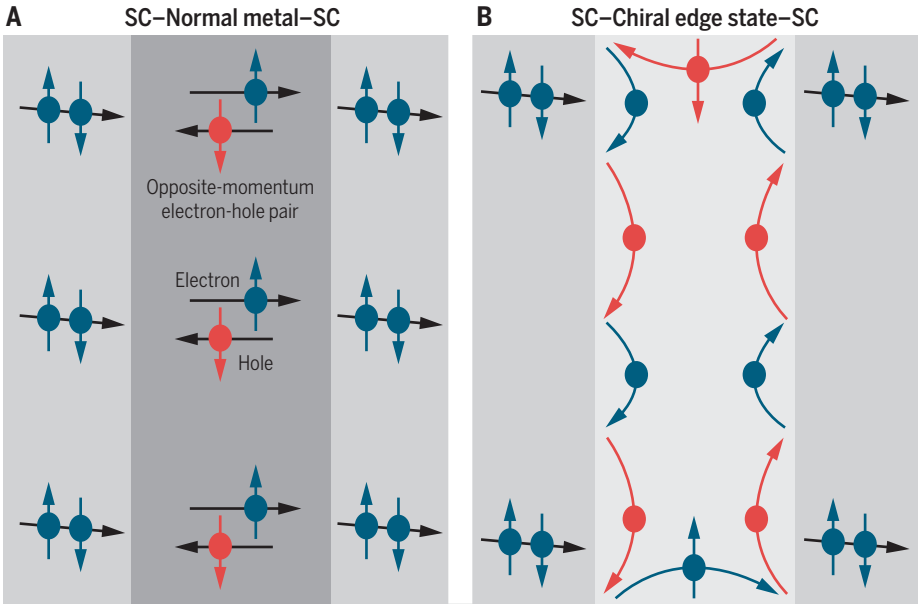
> ***"...measurements reveal not just superconducting and QHE regimes in tandem, but an interspersing of the effects."***

Amet *et al.* measure resistance as a function of gate voltage and magnetic field, B, and observe pockets of superconductivity deep in the QHE regime, at fields as high as 2 T. The effects coexist: When the applied current is reduced below the superconducting critical current I_c, regions of suppressed differential resistance, which show clear superconducting gap structure in current bias measurements, emerge on top of the QHE plateau. (It should be noted

Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801-3080, USA. Email: nadya@illinois.edu

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS



Andreev propagation. (**A**) Quasiclassical schematic of a typical superconductor (SC)–normal metal (N)–superconductor (SC) junction, where electrons (blue) are retroreflected as holes (red) at the SC-N interface, and the supercurrent is carried uniformly across the junction by the resultant opposite-momentum electron-hole pairs. (**B**) Quasiclassical schematic of a SC–chiral edge state–SC junction, where electrons Andreev reflected as holes, and vice versa, propagate in the same direction via partial cyclotron orbits, and the supercurrent is carried across the junction by electron-hole pairs at opposite edges.

that the measurements are not performed in the QH spin-singlet regime, so the QH bulk should be fully gapped to standard superconducting transport.) The coexistence is also evident in the structure of I_c versus B, which exhibits interference analogous to Fraunhofer diffraction through a single slit. The periodicity of I_c versus B transitions from irregular at moderate fields (when cyclotron orbits r_c are larger than the sample size L, so nonuniform current paths can cross the sample) to regular at larger fields (when $r_c < L$, so the bulk is gapped and the current has fixed paths at the edges). Further work is needed to understand the Fraunhofer periodicity in the QHE regime, as well as the effect of disorder on the appearance and behavior of I_c.

How the supercurrent is transmitted in the quantum Hall regime can be explained by a coupling of electron and hole edge states: An electron completing a partial orbit along the superconducting interface is Andreev reflected as a hole; the hole orbits in the same direction and then Andreev reflects as an electron, and so on (see the figure, panel B). The coherent intermixing of these states allows a chiral supercurrent to be passed: Along one edge, the electron component of this mixed state crosses the junction, while along the opposite edge, the hole component returns (*7*). This novel type of Andreev process involves electrons and holes binding along a single superconducting interface. It also represents a new type of edge state, involving spatially separated electron-hole pairs.

The mixed electron-hole modes at the superconducting interface have not previously been studied, and could enable exotic behavior such as a spin-triplet supercurrent (*8*) or Majorana modes (*9*). There is particular excitement about superconductivity within the fractional quantum Hall regime, where excitations of the mixed modes could be parafermions (*10*) or Fibonacci anyons (*11*). Such non-Abelian excitations (behaving as neither bosons nor fermions) have generated tremendous interest for their potential utility in fault-tolerant topological quantum computation. The results of Amet *et al.* demonstrate an exciting new hybrid correlated system that could enable such theoretical proposals to become reality. ■

REFERENCES

1. A. F. Andreev, *Zh. Eksp. Teor. Fiz.* **46**, 1823 (1964).
2. M. Ma, A. Y. Zyuzin, *Europhys. Lett.* **21**, 941 (1993).
3. F. Amet *et al.*, *Science* **352**, 966 (2016).
4. R. E. Prange, S. M. Girvin, Eds., *The Quantum Hall Effect* (Springer, New York, ed. 2, 1990).
5. M. Katnelson, *Graphene: Carbon in Two-Dimensions* (Cambridge Univ. Press, London, 2012).
6. C. R. Dean *et al.*, *Nat. Nanotechnol.* **5**, 722 (2010).
7. H. Hoppe, U. Zulicke, G. Schon, *Phys. Rev. Lett.* **84**, 1804 (2000).
8. J. A. M. Ostaay, A. R. Akhmerov, C. W. Beenakker, *Phys. Rev. B* **83**, 195441 (2011).
9. P. San-Jose *et al.*, *Phys. Rev. X* **5**, 041042 (2015).
10. D. J. Clarke *et al.*, *Nat. Commun.* **4**, 1348 (2013).
11. R. S. K. Mong *et al.*, *Phys. Rev. X* **4**, 011036 (2014).

10.1126/science.aaf6604

BIOCHEMISTRY

Methane—make it or break it

Biochemical data resolve the controversy over how methanogenic archaea produce methane

By **Thomas J. Lawton** *and* **Amy C. Rosenzweig**

About 500 to 600 million metric tons of methane, a potent greenhouse gas, are emitted annually worldwide; ~69% of this methane is produced biologically by anaerobic archaea known as methanogens (*1*). In some environments, methane emissions are partly offset by anaerobic methane oxidizing archaea (ANME) and aerobic methanotrophic bacteria (*1*). In methanogens, the methyl-coenzyme M reductase (MCR) enzyme uses a nickel-containing cofactor (F_{430}) to catalyze the final step of methane synthesis (see the figure, panels A and B) (*2*, *3*). The MCR reaction is reversible, and MCR may also catalyze the first step of methane oxidation by ANME (*2*). The MCR catalytic cycle begins with F_{430} in the reduced Ni(I) form (*4*), but what happens next has been unclear. On page 953 of this issue, Wongnate *et al.* (*5*) report evidence for a Ni(II)-thiolate intermediate.

MCR is notoriously difficult to study because the active Ni(I) state—termed MCR_{red1}—is extremely sensitive to oxygen (*4*). Various states of MCR with and without substrates and in different oxidation states have been characterized, but it is unclear which are relevant to catalysis (*3*, *6*–*8*). Two possible mechanisms have been investigated (*2*, *3*). The distinguishing feature of the two mechanisms is the nature of the first intermediate. In mechanism I, an organometallic methyl-Ni(III) intermediate is formed (MCR_{Me}); in mechanism II, a Ni(II)-thiolate ($MCR_{ox1\text{-}silent}$) and a methyl radical are produced (see the figure, panel C).

The two Ni species should be distinguishable via spectroscopic methods. In particu-

Departments of Molecular Biosciences and of Chemistry, Northwestern University, Evanston, IL 60208, USA. Email: thomas.j.lawton@gmail.com; amyr@northwestern.edu

GRAPHIC: C. BICKEL/*SCIENCE*

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

lar, MCR_{Me} contains an unpaired electron, which will give rise to an electron paramagnetic resonance (EPR) signal, whereas $MCR_{ox1\text{-}silent}$ is EPR-silent. The problem lies in observing such signals (or the lack thereof) before this first intermediate reacts further. Wongnate *et al.* overcome this obstacle by using a coenzyme B (CoBSH) (see the figure) analog that is one carbon atom shorter (CoB_6SH). This modification slows the reaction rate sufficiently to observe the first intermediate (*9*).

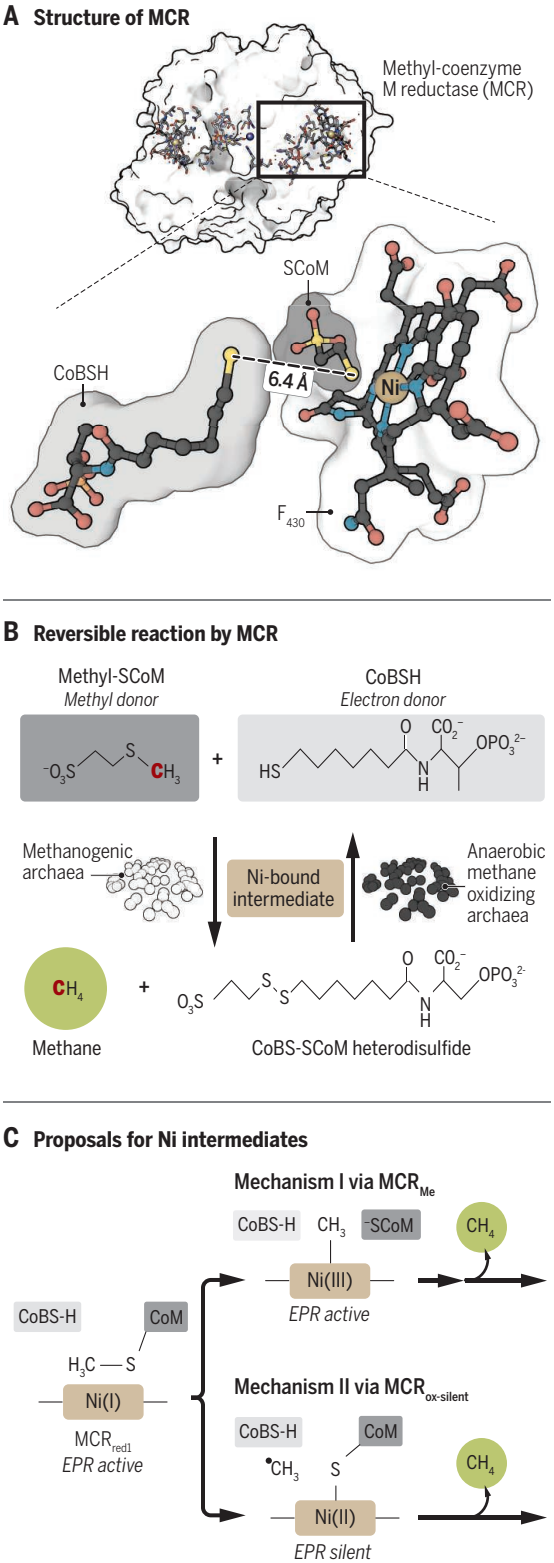
The authors first used stopped-flow and rapid chemical-quench methods to compare the rates of MCR_{red1} disappearance and of methane synthesis. They found that the two processes occur at the same rate. This observation is inconsistent with mechanism I, which involves formation of the comparatively stable methyl-Ni(III) intermediate. Instead, the kinetics support mechanism II, in which the methyl radical would react rapidly to form methane.

With these data in hand, Wongnate *et al.* took advantage of the slower reactivity of CoB_6SH to trap and characterize the elusive intermediate. They generated rapid freeze-quench samples at various time points during the first minute of the reaction and collected EPR spectra. The result is clear: The intensity of the EPR signal from MCR_{red1} decreases by 90% at a rate similar to that of methane synthesis, without the appearance of a new signal attributable to Ni(III). Thus, the intermediate is EPR-silent, consistent with a Ni(II)-thiolate (mechanism II) and not a methyl-Ni(III) (mechanism I) or any other Ni(III) species.

To confirm the intermediate's identity, the authors analyzed the same samples with another spectroscopic technique, magnetic circular dichroism (MCD). The MCD spectra of the starting MCR_{red1} and many other forms of MCR are distinct (*8*, *10*). The spectrum of the intermediate closely resembles that of the $MCR_{ox1\text{-}silent}$ state, in which Ni is coordinated by the thiolate group of coenzyme-M (SCoM) (see the figure, panel C) (*7*, *10*). This direct evidence for a Ni(II)-thiolate species is further bolstered by combined computational and experimental thermodynamic analyses that strongly favor mechanism II.

Identification of the intermediate ends more than two decades of controversy and sets the stage for building a consensus MCR mechanism. Several key questions remain unanswered, however. For example, can radical intermediates in the early stage of the reaction be observed? The EPR data show that small amounts of a radical do form as MCR_{red1} disappears, but it is unclear how this species compares with previously reported radicals (*11*). Other previous observations must also be reconciled with the new results (*11*, *12*). Looking beyond the first step of the reaction, the changes that bring SCoM bound to F_{430} close enough to CoBSH (see the figure, panel B) to react remain unclear. Finally, does anaerobic methane oxidation proceed via the reverse mechanism, and, if so, what factors influence the direction of the reaction?

The answers to these questions will affect efforts to design biomimetic catalysts, engineered enzymes for methane oxidation, and commercially viable synthetic biological pathways (*13*). We are one step closer to knowing how to make methane; will we also be able to break it? ■



How methanogens make methane. MCR is the key enzyme in methane synthesis by methanogens. (**A**) The crystal structure of MCR from *Methanothermobacter marburgensis* (PDB code 1MRO) is in the $MCR_{ox1\text{-}silent}$ state, where the Ni atom in cofactor F_{430} binds to the thiolate group of SCoM. (**B**) In the reaction catalyzed by MCR, methyl-SCoM reacts with CoBSH to form and methane and the CoBS-SCoM heterodisulfide. MCR is believed to also catalyze the reverse reaction in ANME. (**C**) The exact mechanism by which the MCR reaction proceeds has been unclear. Two main mechanisms have been proposed. Mechanism I proceeds via a methyl-Ni(III) intermediate, whereas mechanism II involves a Ni(II)-thiolate intermediate. Wongnate *et al.* now provide convincing evidence for mechanism II.

REFERENCES

1. R. Conrad, *Environ. Microbiol. Rep.* **1**, 285 (2009).
2. S. W. Ragsdale, in *The Metal-Driven Biogeochemistry of Gaseous Compounds in the Environment*, P. M. H. Kroneck, M. E. Sosa Torres, Eds. (Springer Netherlands, Dordrecht, 2014), pp. 125–145.
3. U. Ermler, *Dalton Trans.* **21**, 3451 (2005).
4. M. Goubeaud, G. Schreiner, R. K. Thauer, *Eur. J. Biochem.* **243**, 110 (1997).
5. T. Wongnate *et al.*, *Science* **352**, 953 (2016).
6. T. Wongnate, S. W. Ragsdale, *J. Biol. Chem.* **290**, 9322 (2015).
7. U. Ermler, W. Grabarse, S. Shima, M. Goubeaud, R. K. Thauer, *Science* **278**, 1457 (1997).
8. E. C. Duin *et al.*, *J. Biol. Inorg. Chem.* **9**, 563 (2004).
9. Y.-C. Horng, D. F. Becker, S. W. Ragsdale, *Biochemistry* **40**, 12875 (2001).
10. J. L. Craft, Y.-C. Horng, S. W. Ragsdale, T. C. Brunold, *J. Am. Chem. Soc.* **126**, 4068 (2004).
11. M. Dey, X. Li, R. C. Kunz, S. W. Ragsdale, *Biochemistry* **49**, 10902 (2010).
12. S. Scheller, M. Goenrich, S. Mayr, R. K. Thauer, B. Jaun, *Angew. Chem. Int. Ed. Engl.* **49**, 8112 (2010).
13. C. A. Haynes, R. Gonzalez, *Nat. Chem. Biol.* **10**, 331 (2014).

10.1126/science.aaf7700

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

CELL BIOLOGY

Formin' filaments at a faster CLIP

A microtubule-tracking protein amplifies actin assembly

By Klemens Rottner[1,2]

Cells of all kingdoms form a rigid but dynamic cytoskeleton that is essential for cell shape and growth. Microtubules and actin filaments build the eukaryotic cytoskeleton, and although they serve distinct functions, cooperation and indirect connections between these systems exist (*1*). On page 1004 of this issue, Henty-Ridilla *et al.* (*2*) report how a protein that accumulates at the growing tips of microtubules also elongates actin filaments, revealing an unexpected interaction between the two cytoskeletal systems.

Both microtubules and actin filaments display a polarity of growth, with fast growing "plus" ends. The building blocks, the α/β-tubulin heterodimer and the actin monomer, hydrolyze guanosine 5′-triphosphate and adenosine 5′-triphosphate, respectively, within these filaments as a functional prerequisite for specific modes of regulation. These include interactions with filament regulators or the flux of monomers through filaments called treadmilling. However, the structures of both building blocks and filaments differ. Microtubules are hollow tubes built of 13 protofilaments and have a diameter of 25 nm, whereas the roughly 7-nm-thick actin filaments constitute two helically entwined protofilaments.

In interphase cells, microtubules are frequently anchored with their minus ends in a microtubule organizing center, from which the plus ends emanate (*3*). By contrast, actin filaments assemble into bundles, networks, or a mixture of both (*4*), and frequently associate with plasma- or endomembranes, exerting forces during shape changes and migration or mediating transport processes inside the cell (*4*). "Plus-tip" proteins control the cycles of microtubule plus-end growth and abrupt depolymerization (known as catastrophe) (*5*). Microtubule plus tips include members of the end binding (EB) protein, cytoplasmic linker protein (CLIP), and cytoplasmic linker protein–associated protein (CLASP) families. Henty-Ridilla *et al.* focused on CLIP-170, the first plus tip to be tagged with a green fluorescent protein and visualized in cells as "spot fountains" arising from the microtubule organizing center (*6*).

Unexpectedly, Henty-Ridilla *et al.* observed that CLIP-170 enhances the plus-end growth of actin filaments that is mediated by the formin family of assembly factors. Formins promote filament nucleation, which is the major rate-limiting step of filament assembly in vivo, and processive filament elongation (*7*). For this, they form dimers to surf on the plus ends of actin filaments and add monomers for growth. Henty-Ridilla *et al.* show that, in vitro, this process is accelerated by CLIP-170. Most experiments were performed with a formin called mouse Diaphanous homolog 1 (mDia1), whose ability to elongate actin filaments was boosted by CLIP-170 more than threefold in vitro; however, the authors confirmed that one-third of the 15 mammalian formins had the same effect, albeit to different extents.

Interestingly, acceleration of actin filament growth required a CLIP-170 domain called formin elongation effector domain (FEED), found previously in the yeast formin inhibitor suppressor of myo2-66 (Smy1). Smy1 blocks formin by binding to its actin-binding site [called the formin homology 2 (FH2) domain] (*8*). Activation of actin assembly by CLIP-170 required no other domains; thus, regions that interact with microtubules or the plus-tip protein EB1 were dispensable and thus separable from the CLIP-170 effects on formins. Moreover, elongation of actin filaments by formin is antagonized by the elongation inhibitor capping protein; these proteins engage in a tug-of-war for interacting with the growing actin filament plus end (*9, 10*). Capping protein still counteracted mDia1-mediated filament elongation when CLIP-170 was present, but less efficiently than in the absence of the latter.

Henty-Ridilla *et al.* noted that CLIP-170 and mDia1 mostly associate along microtubule sides, but are biased to the plus ends upon addition of EB1. Most remarkably, the authors co-reconstituted microtubule and actin dynamics simultaneously in vitro, allowing direct exploration of the interactions of all relevant players with both types of filaments. Actin filaments were visualized to nucleate

[1]Zoological Institute, Technische Universität Braunschweig, Spielmannstrasse 7, 38106 Braunschweig, Germany. [2]Helmholtz Centre for Infection Research, Inhoffenstrasse 7, 38124 Braunschweig, Germany. Email: k.rottner@tu-braunschweig.de

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Cytoskeletal cooperation

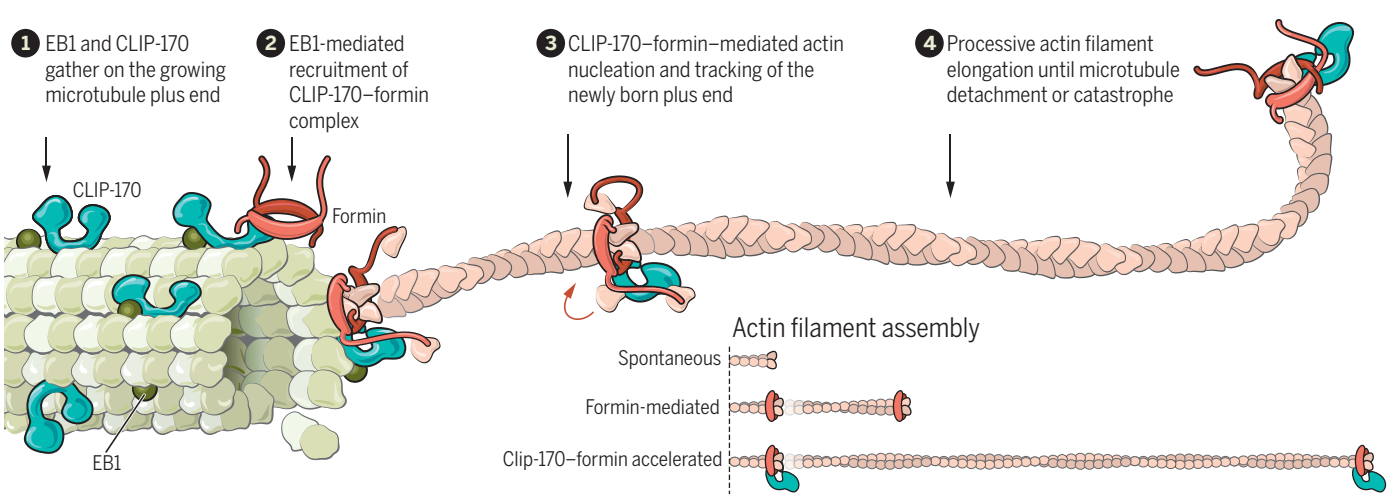


ILLUSTRATION: K. SUTLIFF/*SCIENCE*

Published by AAAS

from microtubule tips, with their minus ends attached to microtubules for dozens of seconds, and CLIP-170 and mDia1 dimers tracking the newly born actin filament plus end (see the figure). Although actin nucleation by the facultative microtubule-end tracking protein adenomatous polyposis coli (APC) (*11*) can also precede formin-dependent elongation, actin nucleation and microtubule binding of APC are mutually exclusive (*12*), thus precluding actin nucleation off microtubules by APC, unlike CLIP-170. Clearly, this reaction remains to be demonstrated in living cells, but the authors also discovered that the previously established contribution of CLIP-170 to the dendritic complexity of rat hippocampal neurons (*13*) requires a functional FEED domain, leading them to conclude that the coordination of microtubule and actin systems by CLIP-170 could promote neuronal dendritic arborization and probably other types of cell shape changes.

It is not yet known whether this CLIP-170–dependent process is mediated by mDia1 or other formins. Also, if actin filaments grow off microtubule tips, then how do their minus ends stay attached to microtubules? It is also not clear how a domain (FEED) that inhibits FH2-domain function in yeast can accelerate FH2-dependent, processive filament elongation (*14, 15*). Finally, does EB1-mediated confinement of CLIP-170–formin complexes to microtubule tips contribute to activation of actin nucleation? We are just beginning to grasp how the authors' observations, and the future studies they inspire, translate into mechanisms that underlie a hitherto unappreciated aspect of microtubule and actin filament cooperation in vivo. ■

REFERENCES AND NOTES

1. S. Etienne-Manneville, *Ann. Rev. Cell Dev. Biol.* **29**, 471 (2013).
2. J. L. Henty-Ridilla, A. Rankova, J. A. Eskin, K. Kenny, B. L. Goode, *Science* **352**, 1004 (2016).
3. A. Akhmanova, M. O. Steinmetz, *Nat. Rev. Mol. Cell Biol.* **16**, 711 (2015).
4. L. Blanchoin, R. Boujemaa-Paterski, C. Sykes, J. Plastino, *Physiol. Rev.* **94**, 235 (2014).
5. G. J. Brouhard, *Mol. Biol. Cell* **26**, 1207 (2015).
6. F. Perez, G. S. Diamantopoulos, R. Stalder, T. E. Kreis, *Cell* **96**, 517 (1999).
7. E. N. Firat-Karalar, M. D. Welch, *Curr. Opin. Cell Biol.* **23**, 4 (2011).
8. J. A. Eskin, A. Rankova, A. B. Johnston, S. L. Alioto, B. L. Goode, *Mol. Biol. Cell* **27**, 828 (2016).
9. J. P. Bombardier *et al.*, *Nat. Commun.* **6**, 8707 (2015).
10. S. Shekhar *et al.*, *Nat. Commun.* **6**, 8730 (2015).
11. D. Breitsprecher *et al.*, *Science* **336**, 1164 (2012).
12. K. Okada *et al.*, *J. Cell Biol.* **189**, 1087 (2010).
13. L. Swiech *et al.*, *J. Neurosci.* **31**, 4555 (2011).
14. D. R. Kovar, E. S. Harris, R. Mahaffy, H. N. Higgs, T. D. Pollard, *Cell* **124**, 423 (2006).
15. S. Romero *et al.*, *Cell* **119**, 419 (2004).

ACKNOWLEDGMENTS

I thank T. Stradal (Helmholtz Centre for Infection Research, Braunschweig) for comments on the manuscript and figure.

10.1126/science.aaf9624

ROBOTICS

Learning from nature how to land aerial robots

Smaller robots can use mechanical intelligence to simplify the task of perching on a target

By **Mirko Kovac**

One of the main challenges for aerial robots is the high-energy consumption of powered flight, which limits flight times to typically only tens of minutes for systems below 2 kg in weight (*1*). This limitation greatly reduces their utility for sensing and inspection tasks, where longer hovering times would be beneficial. Perching onto structures can save energy and maintain a high, stable observation or resting position, but it requires a coordination of flight dynamics and some means of attaching to the structure. Birds and insects have mastered the ability to perch successfully and have inspired perching robots at various sizes. On page 978 of this issue, Graule *et al.* (*2*) describe a perching robotic insect that represents the smallest flying robot platform that can autonomously attach to surfaces. At a mass of only 100 mg, it combines advanced flight control with adaptive mechanical dampers and electroadhesion to perch on a variety of natural and artificial structures.

In nature, many birds and insects are highly skilled in precise perching on complex structures using a variety of techniques (see the top half of the figure). Large birds, for example, use visual feedback to maintain a constant expansion pattern on their retina that allows for a smooth deceleration in air to reach zero velocity just before touching the surface (*3*). This aerial braking is achieved through a complex deep-stall maneuver that uses morphing wings and airflow sensing (*4*). Simultaneously, the legs are extended and the feet attach firmly to the structure on contact through a passive grasping system called the "digital tendon locking mechanism" (*5*).

Smaller animals have a different perching strategy that relies more on mechanical intelligence of their body morphology. Bumblebees first hover near the structure and then use a combination of visual sensing and mechanoreceptors to slowly align their body and attach to the surface (*6*). Houseflies perch on ceilings without even reducing their flight speed before impact, but instead they simply extend their legs as compliant soft structures to passively damp the impact and then perch (*7*). Smaller insects, such as ballooning spiders, spin a silk-thread parachute as an aerial support structure to travel long distances using the wind (*8*). The silk thread can also act as a tensile anchoring element allowing them to perch on suspended surfaces in a completely passive manner (*9*). Similarly, flying seeds can be seen as ultralightweight perching structures that use exclusively passive mechanisms (for example, interlinking bristles into feathers) to attach themselves to animals for long-distance hitchhiking (*10*).

> ***"Perching can be achieved with complex sensing, planning, and dynamic flight maneuvers as used by birds and large aerial robots. When the system is scaled down, the use of a more mechanical perching technique can be beneficial."***

The field of perching aerial robotics has shown a similar pattern; higher levels of reliance on sensing and control is needed for larger vehicles, but more attention is paid to mechanical intelligence for smaller systems (see the bottom half of the figure). For example, the Stanford Univ. Scansorial Unmanned Aerial Vehicle (*11*) uses onboard sensors to detect a wall. It then performs a dynamic flight maneuver to pitch up and perch on the surface using passively aligned microspines on its legs. In many ways, this approach is similar to that of birds because it relies on distance sensing, dynamic flight maneuvers, and locally adaptive claws that attach to the surface via smart mechanical interlinking.

Aerial Robotics Laboratory, Department of Aeronautics, Imperial College London, SW7 2AZ London, UK. Email: m.kovac@imperial.ac.uk

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

As system scale decreases, the dynamic flight accuracy is generally reduced and computational power is limited. Highly robust perching is achieved by relying on passive features enabled by mechanical processes. For example, the École Polytechnique Fédérale de Lausanne (EPFL) microglider (*12*) manages to perch by using a smart attachment mechanism without relying on any sensor feedback. A mechanical trigger releases two elastically preloaded spikes that snap forward and attach to the surface. The mechanism is designed such that the mechanical impulse created by the snapping movement is sufficient to decelerate the robot smoothly to zero velocity just before the pins touch the surface. Its perching technique is similar to that of the housefly that uses its legs to dampen impact and attach to the surface without executing a complex flight pattern.

Perching with tensile structures similar to ballooning spiders has been used in (*13*), albeit for larger vehicles of ~26 g and in combination with a fairly complex force-compensated flight controller. Nevertheless, the same principle of relying on a thread to anchor the vehicle has been used to greatly reduce control requirements and allow for a highly robust perching maneuver.

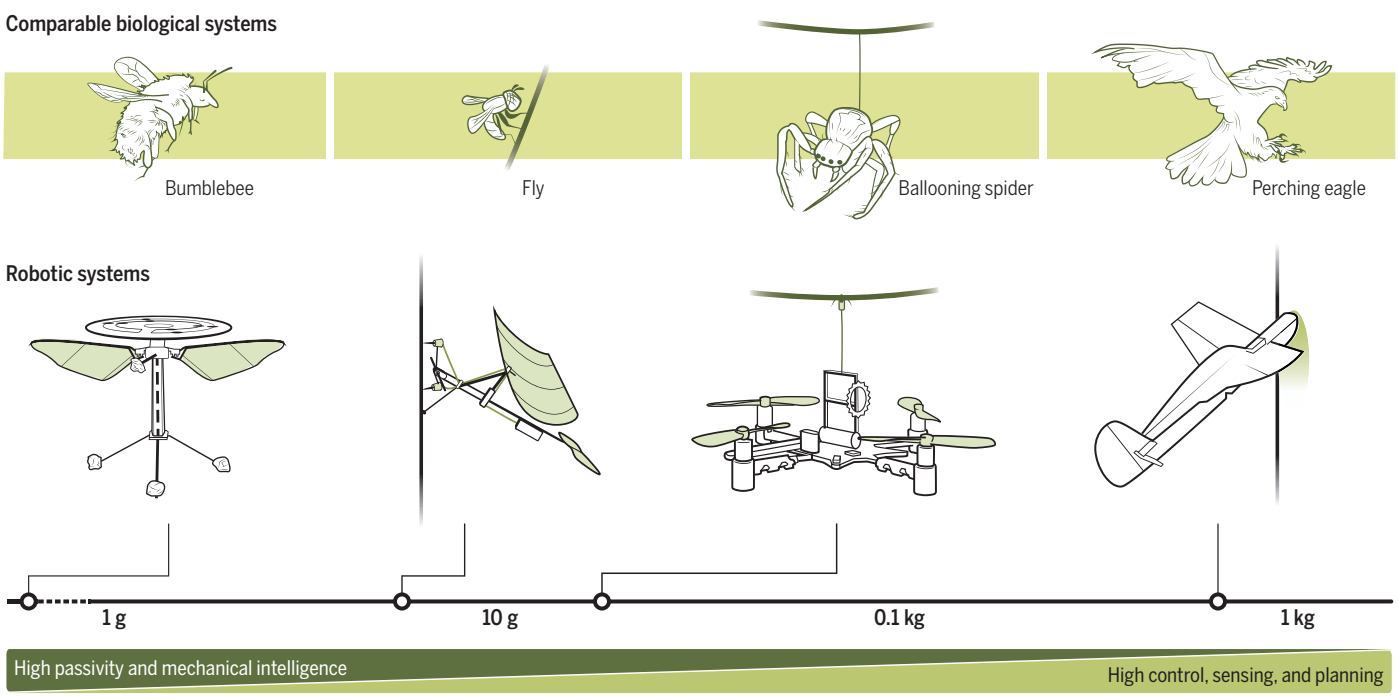
The robotic insect of Graule *et al*. perches by hovering beneath the target structure and initiating a controlled approach to the surface. It aligns its body such that an electrostatic adhesive pad makes contact with the surface and electrically engages the pad for attachment. Electrostatic adhesive is a highly compatible solution for this scale because it is nondirectional and will work on practically any material under most environmental conditions. The robot also uses a passive polyurethane foam mount that adapts to the surface geometry of the perching substrate and balances the vehicle orientation and approach trajectory before attachment. In a similar manner, bees attach to surfaces by slowly approaching the surface and then using local adaptation with mechanical alignment structures to perch.

Perching can be achieved with complex sensing, planning, and dynamic flight maneuvers as used by birds and large aerial robots. When the system is scaled down, the use of a more mechanical perching technique can be beneficial. The mechanical insect of Graule *et al*. has demonstrated that mechanical intelligence can successfully be combined with complex flight control to enable robust perching behaviors. The work is also a prime example that demonstrates how engineering can learn from nature to build the next generation of aerial robots. A future direction for the field could be to selectively use environmental vectors such as wind to travel larger distances or, similar to allochory seeds, avoid aerial locomotion capabilities altogether and travel by perching on animals and larger mobile robots at no energetic cost. ■

REFERENCES

1. D. Floreano, R. J. Wood, *Nature* **521**, 460 (2015).
2. M. A. Graule *et al*., *Science* **352**, 978 (2016).
3. D. N. Lee, M. N. O. Davies, P. R. Green, D. Weel, *J. Exp. Biol*. **180**, 85 (1993).
4. A. C. Carruthers, A. L. R. Thomas, S. M. Walker, G. K. Taylor, *Aeronaut. J.* **114**, 673 (2010).
5. T. H. Quinn, J. J. Baumel, *Zoomorphology* **109**, 281 (1990).
6. C. Evangelista *et al*., *J. Exp. Biol*. **213**, 262 (2010).
7. W. G. Hyzer, *Science* **137**, 609 (1962).
8. J. A. C. Humphrey, *Oecologia* **73**, 469 (1987).
9. J. M. Schneider, J. M. Roos, Y. Lubin, J. R. Henschel, *J. Arachnol*. **29**, 114 (2001).
10. Y. Aoyama, K. Kawakami, S. Chiba, *Biodivers. Cons*. **21**, 2787 (2012).
11. A. Lussier Desbiens, A. T. Asbeck, M. R. Cutkosky, *Int. J. Robotics Res*. **30**, 355 (2011).
12. M. Kovac, J. Germann, C. Hürzeler, R. Y. Siegwart, D. Floreano, *J. Micro-Nano Mechatronics* **5**, 77 (2009).
13. A. Braithwaite, T. Alhinai, M. Haas-Heger, E. McFarlane, M. Kovac, Spider Inspired Construction and Perching with a Swarm of Nano Aerial Vehicles; www.imperial.ac.uk/aerial-robotics/construction/.

10.1126/science.aaf6605



The size and mass dependence of perching. In both nature and robotics, perching for smaller flying bodies relies on the mechanical intelligence of the body morphology, whereas larger bodies use predominantly complex control, sensing, and planning. Mass is indicated on a logarithmic scale. In the top panel, perching in biology is shown with typical masses indicated: Bumblebee (*Bombus ruderatus*); housefly (*Musca domestica*); ballooning spider (genus *Stegodyphus*); and Verreaux's eagle (*Aquila verreauxii*). In the bottom panel, perching of aerial robots is shown from the smallest to largest examples: The robotic insect of Graule *et al*.; the École Polytechnique Fédérale de Lausanne Perching Microglider; the Imperial College String-Based Percher; and the Stanford Univ. Scansorial Unmanned Aerial Vehicle.

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

INTERNATIONAL MIGRATION

International migration under the microscope

Fragmented research and limited data must be addressed

By **Frans Willekens,[1,2] Douglas Massey,[3] James Raymer,[4] Cris Beauchemin[5]**

Although humanitarian crises, such as the ongoing mass exodus from Syria toward Europe, tend to focus global attention on migration, each year millions of people migrate to and from affected countries throughout the world. Progress has been made in understanding drivers of migration, and we have relatively good data on immigrant populations, but we lack information on how many people leave their country each year to settle elsewhere and who these emigrants are. The impact of migration on the individual and on sending and receiving communities and countries is only partly understood. Economic effects can be very different from the impacts on society and culture; some gain from migration, whereas others lose. The lack of knowledge creates systemic risks and uncertainties and frustrates public debate and the formation of effective policies. As high-level leaders convene to discuss such issues at the first United Nations World Humanitarian Summit, we outline priorities for migration data collection, research, and training.

POLICY

CAUSES AND CONSEQUENCES. Contrary to common beliefs, international migration has been low for decades (*1*, *2*). In 2015, about 3.3% of the world's population, ~244 million people, lived in a country other than the country of birth (*3*). This seems surprisingly low given levels of inequality among countries and the proportion of people who desire to emigrate. Wide variations exist in the percentage of a population that is foreign-born. It is <1% in China and India, 14% in the United States and Western Europe, and more than 50% in some countries of the Middle East and the Caribbean. Large countries have generally smaller shares of foreign-born than small countries.

[1]*Max Planck Institute for Demographic Research, Rostock, Germany.* [2]*Netherlands Interdisciplinary Demographic Institute, The Hague, Netherlands.* [3]*Office of Population Research, Princeton University, Princeton, NJ, USA.* [4]*School of Demography, Australia National University, Canberra, Australia.* [5]*Institut National d'Études Démographiques, Paris, France. Email: willekens@demogr.mpg.de*

People have different motives for leaving their country. Some escape poverty and threats. Others seek access to quality education; employment; health care; and amenities; or to join a partner, form a family, or reunite the family. In countries without adequate social protection schemes, families often see migration of family members as a means to diversify risk and sources of income. Transnational social networks may provide social protection not available from the public and private sector at home.

Violent conflicts, natural disasters, and other life-threatening situations trigger peak levels of migration. The United Nations estimates that 60 million people were forced to flee their homes because of war and persecution, compared with 37 million 10 years ago (*4*). The conflict in Syria has been generating an unprecedented magnitude of displaced persons. By mid-February 2016, more than half of the prewar population of 22 million had been forced to flee their homes, with 4.7 million of them international refugees (*5*, *6*).

"...some [communities and countries] gain from migration, whereas others lose."

Many people consider emigration but few leave. A worldwide Gallup survey of 750 thousand adults (15+) in 2005 found that 14% would like to emigrate if they could. Only 8% of them were planning to do so within 12 months (*7*). Migration is costly and risky. People with resources, including skills, have better opportunities elsewhere and are better equipped to take the risk and cover the costs. People may migrate internally and subsequently leave the country.

Migration changes population structure and generates benefits and costs to migrants and their families and to communities and countries. Although many consequences of migration are not well understood, largely because of data problems, the demographic consequences are relatively well understood. Immigration is not a cure for low birth rates or a sustainable solution to aging populations. Immigration increases sociocultural diversity of a population. This can be both an asset (unleashing creativity, competition, and trade) and a liability (perceived threat to cultural identity and social cohesion). Sending countries are particularly affected by migration because of its selectivity.

DATA DEFICITS. Our understanding of migration is handicapped by fragmentation of research and training along disciplinary lines and inadequate measurements. We need a comprehensive approach to migration and a central understanding that transcends disciplinary boundaries.

The main sources of data on international migration are national population censuses, which typically record the place of birth. Several organizations have invested in making these data publicly available (*8*–*10*). The quality of data varies but can be improved by application of the UN Recommendations for population and housing censuses.

Although censuses are an important source of data, several features limit their usefulness as a source for up-to-date data on migration flows. First, immigrants are recorded, but emigrants are not. Second, the age or year of migration is not recorded, so data are ill-suited for analyses of migration trends and of effects on migration of social, economic, or political events and processes, as well as natural disasters. Third, return migrations and frequent migrations go unrecorded. Fourth, censuses come only every 10 years in most countries.

Researchers developed techniques to estimate from census data the number of people that change country of residence at least once within a few years before the census (*1*, *2*). These are not flow data, but they come close. The main sources of flow data are administrative records, such as civil registers and residence permits. A civil register is updated continuously but relies on people's reporting of migration. If people do not report, the event goes unnoticed. Emigration data are particularly weak because departures are often not reported. With arrivals reported and departures not reported, the number of immigrants is overestimated. A further problem is that there is no consistency among administrative sources in the measurement of migration. Ideally, destination countries would communicate their number of arrivals to sending countries.

The lack of accurate flow data and the fact that arrival data receive larger media attention than departure data explains why, in

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

Record numbers of migrants. Syrian migrants enter Hungary at the southern border with Serbia, August 2015.

most countries' opinion polls, people overestimate immigration and the proportion of immigrants in the population. Public attitudes depend on these perceptions, so data inadequacy shapes public debate and, ultimately, the policy response. Another consequence of inadequate reporting of departures is the lack of information on citizens living abroad. Several countries stimulate return migration but lack data for effective policies. The situation improved for OECD countries since the Organization for Economic Cooperation and Development (OECD) started publication of data on immigrants and emigrants (expatriates) in member countries based on census questions on country of birth (*11*).

An emerging data source consists of geolocated Internet Protocol (IP) addresses used to map locations from where users sent email or used social media within a given period. Twitter and Facebook data and Yahoo! email accounts have been used to infer migration flows (*12*). Google search data have been used to infer migration intentions and preferred destinations. Although geo-locators track locations of online connections and not addresses of users or owners, big data may complement traditional data sources, provided that they are available on a regular basis and anonymous and that the selection bias can be resolved. Recent research has shown how migration data from different sources with different measurements may be combined (*13*).

Sample surveys are the main data source to determine who migrates, why, and when. Household, labor force, and demographic surveys collect detailed information on a representative selection of individuals and their activities. They usually include questions on migration, used to develop and test theories of migration, moving from description to explanation and prediction. Sample surveys are typically cross-sectional, so it is not possible to infer mechanisms linking cause and effect. If an unobserved factor influences both the decision to migrate and the outcome of migration, we cannot quantify the effect of migration. In that case, we need an experimental research design or, because that is not usually possible, longitudinal data.

PRIORITIES. Scientists and policy-makers should prioritize three investments to increase our ability to explain and predict migration.

Data communication and modeling. National statistical offices need to cooperate and work with the UN Population Division to standardize measurements according to UN recommendations (*14*) and to improve data availability. Research is needed to develop models to harmonize and estimate missing data, as well as to improve data availability. Recent research on European migration (*13*) has illustrated how data and measurement models can be combined to provide consistent and complete estimates of international migration flows. Extending this work to other countries and regions would provide an invaluable source for improving existing migration flow data and for assessing new data.

Data collection. Although migration modules should be added to established longitudinal surveys, we propose a dedicated World Migration Survey (WMS), echoing a proposal by the International Union for the Scientific Study of Population (*15*, *16*). The WMS should include respondents and partnering institutes in countries of origin, transit, and destination and, ideally, should be longitudinal. Capacity-building should be a component. Necessary expertise and

Downloaded from http://science.sciencemag.org/ on May 19, 2016

PHOTO: REUTERS/BERNADETT SZABO TPX

Published by AAAS

experience have been developed in studies involving research institutes in sending and receiving countries, e.g., the Mexican Migration Project (Princeton University), the "push and pull factors of international migration" project (a joint project of Eurostat and NIDI, Netherlands Interdisciplinary Demographic Institute), and the Migrations between Africa and Europe Project (led by INED, the French National Institute for Demographic Studies).

Training. We propose M.Sc. and Ph.D. programs in migration and population diversity. These should adopt a holistic perspective integrating demography, economics, analytical sociology, geography, cognitive anthropology, political science, and international migration law. The programs should ensure that migration issues are properly treated in the production of statistics and the formulation of policies.

Our recommendations could help expand and improve the evidence base available for public debates and policy formation. Our hope is that these debates will lead to policies to promote full participation of all population groups in society and turn diversity into a valuable asset in a globalizing world. ■

REFERENCES

1. G. J. Abel, N. Sander, *Science* **342**, 1520 (2014).
2. G. J. Abel, "Estimates of global bilateral migration flows by gender between 1960 and 2015" (Working paper 2/2016, Vienna Institute of Demography, Vienna, 2016).
3. UN, International Migration 2015 [wall chart] (UN, New York, 2016).
4. UNHCR, Worldwide displacement hits all-time high as war and persecution increase (2015); www.unhcr.org/558193896.html.
5. UNHCR, Syria regional refugee response [data portal] (2016); https://bit.ly/Regional Response.
6. UNHCR, Total number of Syrian refugees exceeds four million for first time (2016); www.unhcr.org/559d67d46.html.
7. N. Esipova, J. Ray, A. Pugliese, *Gallup World Poll: The Many Faces of Global Migration* (in cooperation with GALLUP, International Organization for Migration, Geneva, 2011).
8. Population Division, International Migration (Population Division, UN, New York, 2016); www.unmigration.org.
9. OECD, OECD migration databases (2016); www.oecd.org/els/mig/oecdmigrationdatabases.htm.
10. World Bank, Migration and Remittances Data (2016); https://bit.ly.MigrationRemittances.
11. J.-C. Dumont, G. Lemaître, *OECD Econ. Stud.* **2005**(3), 49 (2005).
12. P. Gerland, Population Division, Migration, mobility, and big data: An overview (Population Division, United Nations, New York, 2015).
13. J. Raymer, A. Wiśniowski, J. J. Forster, P. W. F. Smith, J. Bijak, *J. Am. Stat. Assoc.* **108**, 801 (2013).
14. United Nations, "Recommendations on statistics of international migration" (Statistical papers series M, no. 58, Department of Economic and Social Affairs, United Nations, New York, rev.1, 1998).
15. C. Beauchemin, Statement prepared for the Informal Hearings for High-level Dialogue on International Migration and Development, 15 July 2013, New York (International Union for the Scientific Study of Population, Paris, 2013).
16. C. Beauchemin, *Int. Migr. Rev.* **48**, 921 (2014).

10.1126/science.aaf6545

SCIENTIFIC COMMUNITY

Preprints for the life sciences

The time is right for biologists to post their research findings onto preprint servers

By **Jeremy M. Berg,[1] Needhi Bhalla,[2] Philip E. Bourne,[3] Martin Chalfie,[4] David G. Drubin,[5] James S. Fraser,[6] Carol W. Greider,[7] Michael Hendricks,[8] Chonnettia Jones,[9] Robert Kiley,[9] Susan King,[10] Marc W. Kirschner,[11] Harlan M. Krumholz,[12] Ruth Lehmann,[13] Maria Leptin,[14] Bernd Pulverer,[14] Brooke Rosenzweig,[15] John E. Spiro,[16] Michael Stebbins,[17] Carly Strasser,[18] Sowmya Swaminathan,[19] Paul Turner,[20] Ronald D. Vale,[21] K. VijayRaghavan,[22] Cynthia Wolberger[23]**

A preprint is a complete scientific manuscript (often one also being submitted to a peer-reviewed journal) that is uploaded by the authors to a public server without formal review. After a brief inspection to ensure that the work is scientific in nature, the posted scientific manuscript can be viewed without charge on the Web. Thus, preprint servers facilitate the direct and open delivery of new knowledge and concepts to the worldwide scientific community before traditional validation through peer review (*1*, *2*). Although the preprint server arXiv.org has been essential for physics, mathematics, and computer sciences for over two decades, preprints are currently used minimally in biology.

POLICY

The ASAPbio meeting (Accelerating Science and Publication in biology) was held on 16 and 17 February 2016 to explore the wider use of preprints for disseminating ideas and results in the life sciences. The ~70 invited participants included junior and senior working scientists; and representatives of public and private funding agencies, industry, databases, and scientific journals. All talks and breakout sessions were streamed over the Internet to encourage community participation, and a full record of the meeting is available (*3*). The meeting goals were to analyze the roles that preprints might play in communicating results in biology and to debate the potential advantages and disadvantages of greater use of preprints for the progress of science, career development, and the integrity of the scientific record. In the three sections below, three classes of attendees—academic scientists, funders, and publishers—provide their perspectives on the meeting and its outcomes.

ONLINE SURVEY

Tell *Science* what you think about preprint servers at http://scim.ag/1T5LkfI

ACADEMICS' PERSPECTIVE

J. M. Berg, N. Bhalla, M. Chalfie, J. S. Fraser, C. W. Greider, M. Hendricks, M. W. Kirschner, R. Lehmann, P. Turner, C. Wolberger

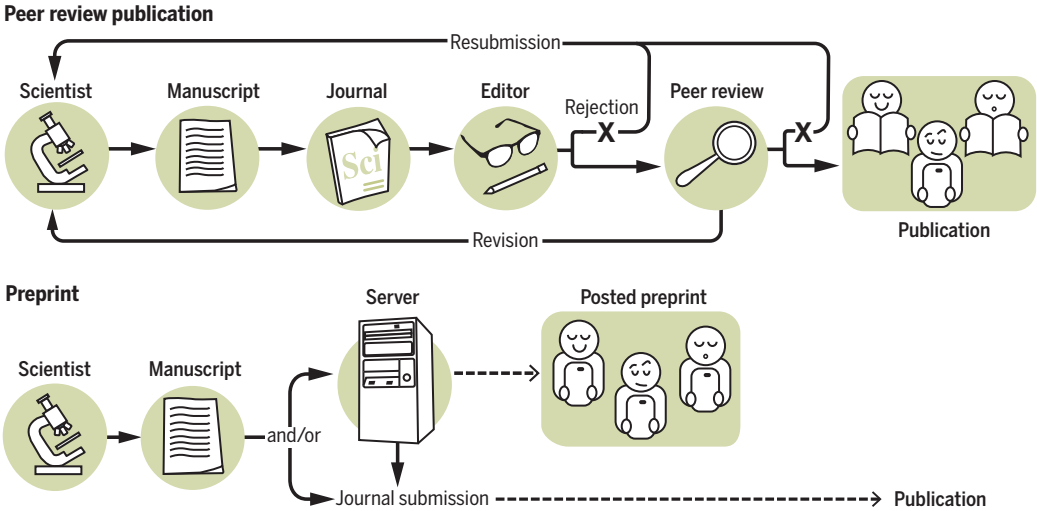
Motivated by frustrations in the slow speed of publishing (*1*), we and other junior and senior life scientists participated in the ASAPbio meeting. Physicists have embraced sharing their work as preprints for 25 years. Paul Ginsparg, founder of arXiv, described how physicists, mathematicians, and computer scientists check arXiv when they wake up each morning to learn about advances in their fields. Even though physicists publish their work later in journals, arXiv has become THE way to communicate new discoveries. Ginsparg also described how preprints empower younger scientists to move their work and careers forward.

Knowledge and opinions of preprints varied among the ASAPbio attendees at the start of the meeting, but many came to appreciate their benefits. Currently, the time between manuscript submission and paper publication is unpredictable and can be long. Depositing a manuscript in a preprint archive makes the work publicly available almost immediately. Posting preprints has the added benefit of democratizing the flow of information and making it available to all investigators across the globe, while allowing journals to make their own judgments of appropriateness and interest after peer review. Publicly available preprints provide an opportunity for authors to obtain feedback beyond the few scientists who see the manuscript during peer review. Finally, preprint archives also document the history of the ideas, as old versions of a manuscript are maintained even after revisions of the work are submitted.

Ginsparg was emphatic that a preprint, because it has a time stamp and is publicly available, plays a key role in establishing priority of discovery. But will this model be widely accepted by biologists? Some suggested that the archive could be flooded with weak papers meant only to assert priority. But decades of experience have demonstrated that scientists do not post poor-quality work on arXiv because of the impact on their reputations; we expect professional biologists to behave similarly. After hearing various points of view, ASAPbio at-

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS



Peer review and preprints. Preprints accelerate dissemination, whereas manuscripts are being improved through peer review.

tendees, in a private and optional poll, voted nearly unanimously in favor of preprints being used for establishing priority (*4*).

If preprints are to help early-career scientists, their use in hiring and promotion is of paramount importance. The ability to cite preprints in grant applications and progress reports would benefit scientists at all career stages. Although not peer-reviewed, a preprint provides tangible evidence of a scientist's most recent work, which is often of greatest interest to review panels. Again, by private ballot, nearly all of voting attendees thought that preprints should be considered as evidence of achievement in evaluations for academic advancement and for funding (*4*).

We also debated the use of preprints in reporting results of clinical studies. In its favor, some argued that clinical research would benefit from more open and timely access to data and noted that papers published in respected medical journals can also be misleading or wrong. Some who were opposed questioned whether research involving human subjects might require additional safeguards in scrutiny by institutional review boards and disclosures of conflicts of interest. As occurred in the physical sciences, different fields in biology and biomedical research may come to embrace preprint archiving at different times and to different degrees.

What would help promote the use of preprints by life scientists? Several steps are essential: broader acceptance of preprint posting by journals (in process and well on its way); the development of search engines for finding and linking preprints to published versions of manuscripts; and the recognition of preprints by grant, hiring, and promotion committees. These steps will likely come. But, motivated by the meeting, many ASAPbio attendees (P. E. Bourne, M. Chalfie, D. A. Colón-Ramos, S. L. Díaz-Muñoz, D. G. Drubin, M. B. Eisen, J. S. Fraser, C. W. Greider, J. K. Polka, R. Schekman, B. Stillman, R. D. Vale, H. Varmus, K. VijayRaghavan, L. B. Vosshall, C. Wolberger) are not waiting: They have taken a step toward embracing a new culture of science communication by posting a preprint this year, most of them for the first time.

FUNDERS' PERSPECTIVE

R. Kiley, C. Jones, P. E. Bourne, K. VijayRaghavan, M. J. Stebbins, J. E. Spiro, C. Strasser, B. Rosenzweig

By the end of the ASAPbio meeting, funders felt that preprints, operating in parallel with peer-reviewed publication, could play a valuable role in communicating research results.

An important issue was whether funders would recognize preprints in grant proposals and/or when reviewing the productivity of a researcher. In grant applications, funders typically ask applicants to identify relevant "peer-reviewed publications" along with the relevant persistent identifiers. Applicants are often invited to detail "other scientific contributions"—which could be used to list preprints—but this section is typically less well-populated and may well be considered less important by external reviewers. Equally, the annual reports from funding agencies focus on peer-reviewed research outputs. However, this somewhat ambivalent approach to preprints is not the consequence of an explicit policy but rather reflects the hitherto limited use of preprints by biologists. From discussions at ASAPbio, our group of funders has identified the following benefits and challenges of preprints.

Benefits. From the perspective of a research funder, we can see substantial benefits from the widespread adoption of preprints. Apart from the obvious benefit that research findings become available more quickly—an important consideration given data showing that the median review time at journals has grown from 85 days to >150 days during the past decade (*5*)—preprints provide funding agencies (and those reviewing funding proposals) with a more current and complete view of a researcher's ideas and progression of work than does a formal, peer-reviewed product of research.

Preprints enable reviewers to assess an applicant's ideas by scrutinizing the research findings, rather than using the journal name (or its impact factor) as a proxy for quality. Funders are keen to uphold the principle that funding decisions should be based on the merit of the research.

Next, preprints provide reviewers with an opportunity to see—in real time—reactions from the community and how the researcher responds to these.

Finally, preprints offer an opportunity for early-career scientists to demonstrate productivity and to provide evidence of independence. Preprints also offer more opportunity for early-career scientists to get peer feedback, especially if they lack the professional networks or the funds to attend conferences.

Challenges. The biggest challenge of using preprints to help inform funding decisions is that they represent work at various stages of development and, by extension, of varying degrees of quality. This puts even more

[1]Institute of Personalized Medicine, University of Pittsburgh School of Medicine. [2]Department of Molecular, Cell and Developmental Biology, University of California, Santa Cruz. [3]Office of the Director, National Institutes of Health. [4]Department of Biological Sciences, Columbia University. [5]Department of Molecular and Cell Biology, University of California, Berkeley. [6]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco. [7]Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine. [8]Department of Biology, McGill University. [9]Wellcome Trust. [10]Rockefeller University Press. [11]Department of Systems Biology, Harvard Medical School. [12]Yale School of Medicine, Yale University. [13]Kimmel Center for Biology and Medicine of the Skirball Institute, Department of Cell Biology, New York University School of Medicine. [14]European Molecular Biology Organization. [15]The Leona M. and Harry B. Helmsley Charitable Trust. [16]Simons Foundation. [17]Laura and John Arnold Foundation. [18]Gordon and Betty Moore Foundation. [19]Nature Research Group. [20]Department of Ecology and Evolutionary Biology, Yale University. [21]Howard Hughes Medical Institute and Department of Cellular and Molecular Pharmacology, University of California San Francisco. [22]Department of Biotechnology, Ministry of Science and Technology, Government of India. [23]Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine. Email: ron.vale@ucsf.edu

ILLUSTRATION: N. CARY/SCIENCE

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

onus on the grant reviewers to read the preprint (as they would a peer-reviewed paper) and judge the quality and whether the conclusions reached are supported by the data presented. Even if the reviewers have all the necessary skills, reviewing preprints will add to their workload at a time when the burden on reviewers is higher than ever (*6*). Consequently, without careful guidance, reviewers may bias their assessment against applicants who cite material that has not yet been published in a peer-reviewed journal.

Other challenges include ensuring that preprints are easily discoverable (e.g., through services like PubMed, Scopus, and Web of Science); and remain permanently accessible. In cases where preprints are subsequently published in peer-reviewed journals, mechanisms must be established to direct the reader from one version to the other.

Recommendations. To move forward on these issues, funders should consider these:

1. Publishing an explicit statement encouraging researchers to make early versions of their manuscripts available through acceptable preprint repositories.
2. Permitting the citation of preprints in acceptable repositories in grant proposals as evidence of productivity, research progress, and/or preliminary work.
3. Providing guidance to reviewers on how to assess preprints in grant proposals.
4. Working with the community to fund a common infrastructure and standards for acceptable preprint repositories to ensure that they are easily discoverable and remain accessible in the long term.

JOURNALS' PERSPECTIVE

B. Pulverer, M. Leptin, D. G. Drubin, S. King, H. M. Krumholz, S. Swaminathan

Representatives of journals and publishers attended ASAPbio. Here we summarize pertinent points of discussion.

Preprints and peer-reviewed journals serve different purposes. The contribution of citable and stably archived preprints to scientific progress was not questioned. Preprints also represent markers of achievement and progress that can be considered in research assessment and in establishing priority of findings. This can be particularly important for young researchers who need to document progress during the early phase of their independent careers.

Some participants argued for completely replacing journals by preprint servers, noting that the extent to which peer review improves manuscripts has not been formally quantified and that peer review still misses errors. However, the majority of attendees felt that peer review orchestrated through journals still plays a valuable role in providing in-depth analysis and scrutiny. Surveys also show that many researchers believe that their own work often improves through the peer-review process (*7*). Following the model of the preprint server arXiv in physics, preprints and formal papers should exist in parallel, synergizing and fulfilling complementary functions. Posting preprints facilitates the rapid communication of scientific findings, whereas peer review provides a more formal certification process that promotes reliability and reproducibility.

Many publishers support preprints, and journals could benefit. Many journals already have policies that support the posting of preprints on recognized servers before submission to their journal (*8*). A draft statement developed by publishers for ASAPbio stating that "Posting on a recognized preprint server does not constitute prior publication or a breach of…embargo policies…, and will prejudice neither the peer review process nor publication…" received broad approval by ASAPbio attendees (*4*). We also feel that comments on the preprint would help to make the subsequent formal review process more efficient and could result in an improved final manuscript. Pipelines that enable direct submission from preprint servers to journals should increase acceptance of this dual mechanism in the community.

Several issues will require further thought.

A preprint creates a "date stamp" in attributing scientific findings to researchers. Premature posting to "stake claims" will need to be self-moderated by the community. Furthermore, establishing priority also will depend on credibility gained through peer review.

Preprints and published research papers represent a continuum in the evolution of a body of work and should be formally linked, such that the scientific paper supersedes the preprint as the version of record that should be cited. "Versioning" is necessary to allow preprints and papers to be adapted, both in response to further progress in the research project and in response to comments from other scientists.

Important questions are (i) whether preprint submissions should be screened for adherence to scientific and ethical standards and (ii) how to handle data that raise ethical concerns or that contravene national policy or guidance. Preprints of clinical data could be very useful, but it is essential that it be made clear, especially to the press and the public, that preprints must be formally recognized only as non–peer-reviewed units of information. Preprint archives may need to give clear guidance to authors about such studies. Thus, citations to preprints could have a different format and be separated from citations to the peer-reviewed literature.

Optimal commenting and discussion formats on preprints remain to be defined and can range from direct communication with authors to public, signed comments.

The raison d'être of a preprint server is the sharing of research findings without delay, but preprints are not necessarily only an alternative or a preliminary step to publishing in formal journals. Posting of units of information smaller than research papers should be encouraged, as long as the data reporting allows others to replicate the work. The value of sharing review articles and commentaries as preprints is less clear.

In summary, if, as is the case for physics preprints, the community engages in constructive and objective discussion of preprints, both the scientific community and journals are likely to benefit directly.

CONCLUSIONS

Preprints could play important roles in accelerating scientific progress; they could serve the needs and foster the careers of scientists; and, in cooperation with existing journals, they could enhance the current system for communicating results and ideas in the life sciences. However, preprints are relatively new to biology, and many questions remain unanswered. Will funding agencies encourage the use of preprint servers? Will all journals accept manuscripts for publication after they have been disseminated as preprints? Will the life sciences community find ways to make biology preprints easily discoverable? And will researchers themselves decide to submit, cite, and evaluate work presented in preprint form? The cooperative spirit displayed by the attendees at ASAPbio gives hope that these complex issues, as well as others that limit the communication of scientific ideas and results, can be addressed in a productive and thoughtful manner. ■

REFERENCES AND NOTES

1. R. D. Vale, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13439 (2015).
2. iBiology, What are preprints? (2016); www.ibiology.org/ibioeducation/what-are-preprints.html.
3. *ASAP*bio, 2016 meeting (2016); http://asapbio.org.
4. *ASAP*bio, Draft document (2016); http://asapbio.org/drafts.
5. K. Powell, *Nature* **530**, 148 (2016).
6. M. Arns, *Nature* **515**, 467 (2014).
7. Sense About Science, *Peer Review Report 2009: Final Report* (Sense About Science, London, 2009).
8. Wikipedia, List of academic journals by preprint policy (2016); http://bit.ly/Journalpreprintpolicy.

ACKNOWLEDGMENTS

We thank D. A. Colón-Ramos, J. K. Polka, R. D. Vale, and H. Varmus for organizing and Howard Hughes Medical Institute for hosting ASAPbio.

10.1126/science.aaf9133

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

Carefully choreographed "waggle" dances inform hivemates about the location of food sources and potential nesting sites.

HISTORY OF SCIENCE

The bee whisperer

A new biography reveals how an ethologist with Jewish heritage earned the Nobel Prize for Nazi-funded research

By **Viviane Callier**

In *The Dancing Bees*, historian of science Tania Munz recounts the fascinating story of Karl von Frisch, the Austrian ethologist who discovered that honeybees communicate the direction and distance of food sources through elaborate dancelike behaviors. The book is a biography of a remarkable and gifted scientist whose work proved foundational for the discipline of comparative physiology, but it is also much more than that. Von Frisch's scientific research occurred during a time period that spanned two world wars and the Cold War. His seminal discovery, which would later earn him a Nobel Prize, was made while he was funded by the Nazi Ministry of Food and Agriculture.

Von Frisch grew up in an intellectually vibrant community in Brunnwinkl, Austria, where his interest in science started very early. Munz carefully describes his first experiments, conducted while von Frisch was an undergraduate studying medicine at the University of Vienna. Working under the mentorship of his uncle, Sigmund Exner, a senior figure in the field of experimental physiology, von Frisch studied how the pigments in crustaceans' eyes change depending on the amount of light in the environment, enabling them to see.

After completing his undergraduate studies, he pursued graduate research in zoology at the Zoological Institute in Munich and then at the Biological Research Station in Vienna. There, he studied how minnows undergo rapid color changes when swimming in different environments. By cutting the sympathetic trunk, a bundle of nerves that run from head to tail, just next to the vertebral column, at successively higher positions along the spinal cord, von Frisch found that the pigment cells in the minnow's skin were controlled by signals from two nerve pathways, one for the anterior and one for the posterior half of the minnow's body. Students of comparative physiology will appreciate Munz's accessible yet detailed descriptions of these experiments that, although not technically complicated, were elegant and ingenious.

In 1925, von Frisch succeeded his graduate mentor, Richard Hertwig, as head of the Institute of Zoology at Munich University. The turmoil in the wake of Germany's defeat in World War I, and uncontrolled unemployment and inflation, would soon lead to the rise of the Nazi regime.

When Adolf Hitler became chancellor in 1933, laws designed to "Aryanize" the university went into effect. Because von Frisch was unable to account for the heritage of his maternal grandmother, the regime declared him to be one-quarter Jewish. Faced with the possibility of losing his professorship, von Frisch called on his scientific colleagues and allies to vouch for him and help him navigate the Nazi bureaucracy. In the end, his efforts were successful. He was even able to secure research funding from the Nazi government.

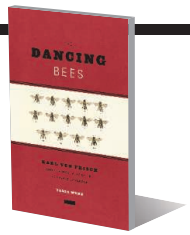
During this period, nosema, a deadly infection caused by a fungal parasite, was devastating the bee population and threatening Europe's crops. Von Frisch astutely shifted his research program to focus on efforts to combat nosema. However, he did manage to pursue a number of basic research projects, including his work deciphering the meaning behind the direction and frequency of the honeybee's dance—work for which he would later earn the Nobel Prize.

Many people were persecuted and suffered unimaginable fates at the hands of the Nazis. Von Frisch, by contrast, escaped relatively unscathed and, in fact, ultimately benefited from the hostile treatment he experienced under the regime, which would later help him build key relationships and gain funding from the American Rockefeller Foundation.

When we think about science under National Socialism, we often think of eugenics and the race to develop an atomic weapon, but what are we to make of the more benign discoveries, like von Frisch's? The stories of scientific discovery that typically make their way into textbooks are clean and elegant, removed from both time and place. In this engaging and accessible book, Munz shows that there is often more to such stories than meets the eye.

The reviewer is at The Ronin Institute for Independent Scholarship, Montclair, NJ 07043, USA. Email: vcallier@gmail.com

The Dancing Bees
Karl von Frisch and the Discovery of the Honeybee Language
Tania Munz
University of Chicago Press, 2016. 286 pp.



10.1126/science.aaf5312

PHOTO: © SCOTT CAMAZINE/ALAMY STOCK PHOTO

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

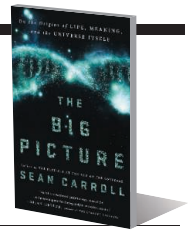
PHYSICS

Philosophy for physicists

A cosmologist explores the philosophical implications of the foundational laws of nature

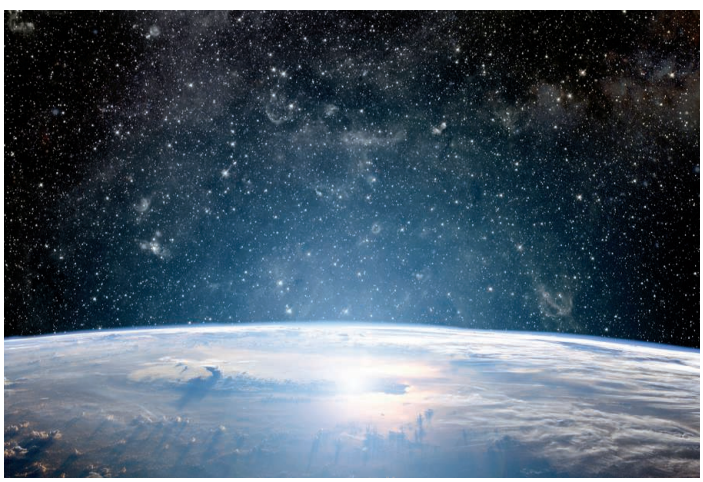
By **Barry Loewer**

The Big Picture
On the Origins of Life, Meaning, and the Universe Itself
Sean Carroll
Dutton, 2016. 480 pp.



The 20th-century philosopher Wilfrid Sellars characterized the aim of philosophy as "to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term." This is also physicist Sean Carroll's aim in his new book, *The Big Picture*. He sets out to show how various phenomena, including thought, choice, consciousness, and value, hang together with the scientific account of reality that has been developed in physics in the past 100 years. He attempts to do all this without relying on specialized jargon from philosophy and physics and succeeds spectacularly in achieving both aims.

Carroll calls his view "poetic naturalism," the latter of which consists of two parts. First, he argues that observation and the scientific method are the only reliable ways of learning about the world. The second component is that the systematic application of the scientific method has taught us that there is one natural world that, at a fundamental level, consists of quantum fields distributed in space-time, the distribution of which is subject to physical laws. Currently, these laws are those of quantum field theory and general relativity. Any future physical theory is expected to subsume these two tenets as correct in their respective domains.



"What is the fundamental nature of reality?" asks Sean Carroll in *The Big Picture*.

The "poetic" part of Carroll's view refers to the idea that there are many true descriptions of the world, including those contained in the special sciences, the languages of psychology, economics, ethics, poetry, and so on. In any particular context, our purposes determine the best way of talking.

For Carroll, attempts to describe the natural world are legitimate as long as they serve the intended purpose and the claims made in them are consistent with the established theories of physics and with each other. He says "consistent," but I think he means something stronger, along the lines of "their claims are entailed by fundamental physical laws and facts." Thus, although generalizations in economics may be the best way of talking about changes in the money supply, true economic generalizations about the money supply must be defined by the laws of physics that describe the motions of elementary particles and the fields that constitute the distribution of money and the rest of the economy.

Carroll observes that the laws of physics are complete within the domain of physics, meaning that any change in the configurations of fields and their particles can be accounted for in terms of earlier configurations and fundamental laws. He argues that it follows from the structure of quantum field theory and the results of particle experiments that all the fields and forces that are responsible for the behavior of macroscopic objects are known. Because any process must be implemented according to fundamental physical laws, it follows that this is sufficient to exclude certain kinds of putative phenomena (e.g., astrological and paranormal influences) from reality. Carroll does not mean that these are a priori impossible but rather that they are inconsistent with established physics and so are ruled out by poetic naturalism.

There are challenges to poetic naturalism that need more discussion than Carroll gives them. One is the origin of "time's arrows," especially causation. Carroll first addressed this issue in a previous book (*1*), and his efforts to fill in the details in *The Big Picture* make for a lively research project in physics and philosophy (*2*).

Another challenge is understanding how thought, consciousness, and free will fit into physical theory. Thoughts are "about" things outside themselves, consciousness possesses a "what it is like" feel, and free choice seems to involve decisions that originate in an agent. It is difficult to see how these are entailed by the motions of fields and particles. Carroll discusses a number of well-known arguments that attempt to show that such phenomena are inconsistent with a purely physical fundamental ontology and rightly concludes that these are not persuasive. But poetic naturalism should not be satisfied until it can include an account of how these elements emerge from fundamental physics or, if such an account is not forthcoming, why they do not involve nonphysical fundamental ontology.

A related challenge has to do with understanding the place of value in the big picture. Carroll follows Hume and noncognitivists in ethics, arguing that value claims are not descriptions of objective aspects of reality but rather that value is invented by human beings. But it is not clear that this view does justice to the way we think about ethics and value.

A last challenge that poetic naturalism should seek to address is why the universe has the laws and parameter values it does and exactly how one is to understand what it is to be a law. Carroll does not answer these questions, but he does argue that theological explanations and metaphors are of no help.

Recently, a number of prominent physicists (Feynman, Weinberg, Hawking, and Krauss) have dissed philosophy, claiming either that it is of no use to physicists or that the problems it addresses have been solved already by physics. *The Big Picture* shows why these claims are misguided.

The reviewer is at the Rutgers Center for Philosophy and the Sciences, Rutgers University, New Brunswick, NJ 08901, USA. Email: loewer@rci.rutgers.edu

REFERENCES

1. S. Carroll, *From Eternity to Here: The Quest for the Ultimate Theory of Time* (Dutton, London, 2010).
2. See, for example, D. Albert, *Time and Chance* (Harvard Univ. Press, Boston, 2000).

10.1126/science.aaf6858

PHOTO: 4MAKSYM/ISTOCKPHOTO.COM

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

CITIES ARE THE FUTURE

Rapid urbanization is overtaxing the planet, but it may not have to

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

By **Nicholas S. Wigginton, Julia Fahrenkamp-Uppenbrink, Brad Wible, and David Malakoff**

Earth has become an urban planet. More than half of the world's people now live in cities, and the proportion is growing. And urban areas are sprawling even faster than they are adding people, swallowing up both farmland and wildlands.

The implications are sobering. The land area needed to provide city residents with food, energy, and materials is expanding; this ecological footprint is often 200 times greater than the area of a city itself. The resulting carbon emissions, added to those from cities themselves, mean that urbanization is now the main driver of climate change.

The rise of cities is not, however, all doom and gloom. By some metrics, consolidating human populations helps shrink our individual environmental footprints, and cities are serving as laboratories for further improvements. Researchers are exploring creative approaches to harvesting urban waste streams, integrating renewable sources of energy, and improving transit.

This collection of Reviews, Perspectives, and News features, as well as interactives you can find at http://scim.ag/1Tb0WdZ, delves deeper into how we came to live in cities and what urbanization means for the future of our planet and ourselves. One message is clear: The urban planet is here to stay, and the decisions we make today about how we build and live in cities will affect generations to come.

INSIDE

NEWS

Rise of the city *p. 906*

Roots of the urban mind *p. 908*

A plague of rats *p. 912*

China rethinks cities *p. 916*

Vancouver's green dream *p. 918*

REVIEWS

City-integrated renewable energy for urban sustainability *p. 922*

Emerging solutions to the water challenges of an urbanizing world *p. 928*

PERSPECTIVES

Transport solutions for cleaner air *p. 934*

The ecological future of cities *p. 936*

Living in cities, naturally *p. 938*

Meta-principles for developing smart, sustainable, and healthy cities *p. 940*

Hidden linkages between urbanization and food systems *p. 943*

Building functional cities *p. 946*

RELATED ITEMS

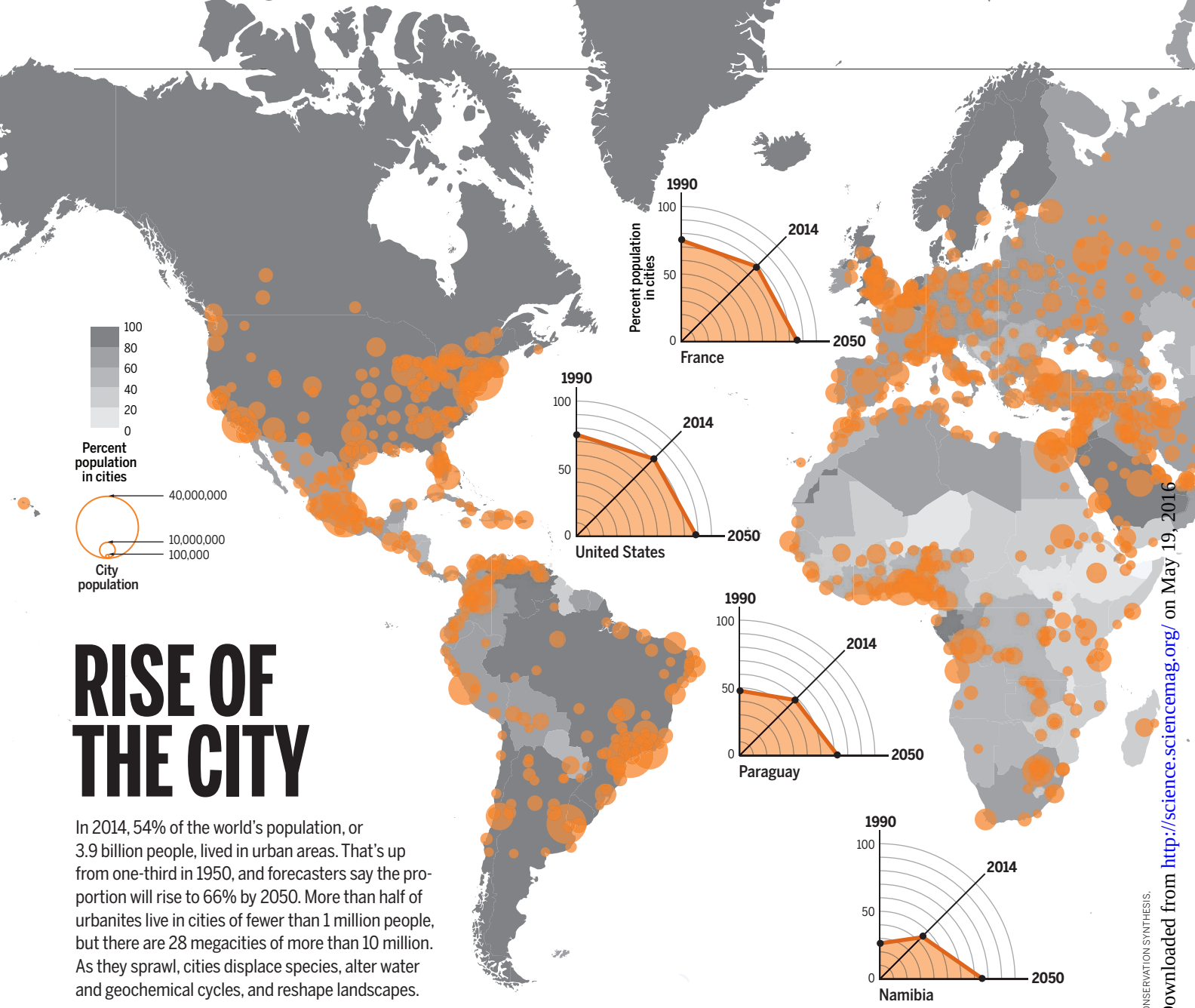
- EDITORIAL P. 873
- LETTERS P. 886
- PODCAST
- VIDEO http://scim.ag/1Tb0WdZ

PHOTO: © CHARLIE JOE/GETTY IMAGES

Cities like Dubai, United Arab Emirates (pictured), must balance the need to provide resources for growing populations with urbanization's effects on the environment.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

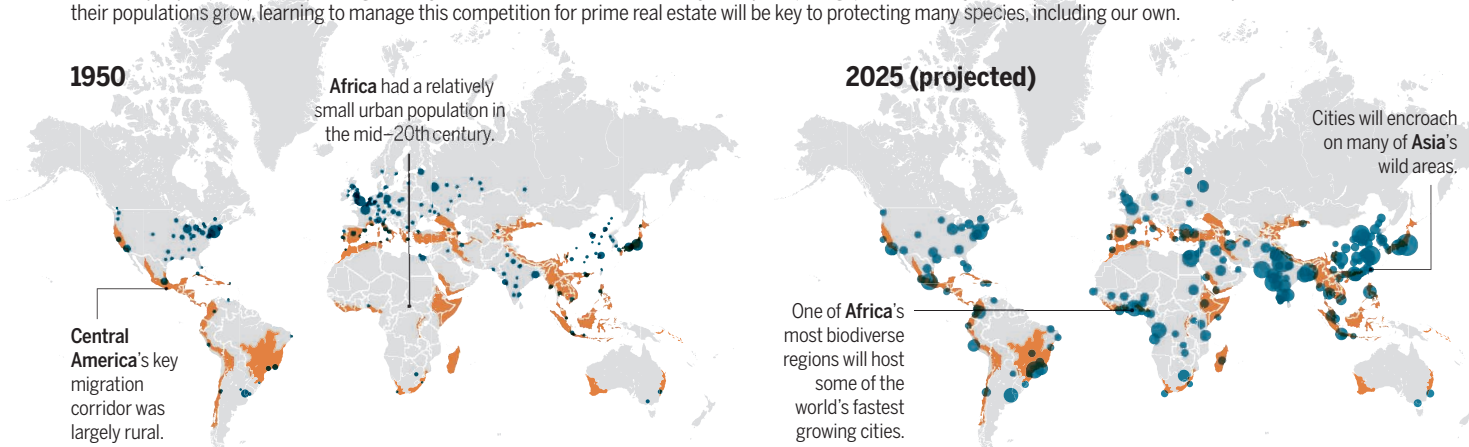


RISE OF THE CITY

In 2014, 54% of the world's population, or 3.9 billion people, lived in urban areas. That's up from one-third in 1950, and forecasters say the proportion will rise to 66% by 2050. More than half of urbanites live in cities of fewer than 1 million people, but there are 28 megacities of more than 10 million. As they sprawl, cities displace species, alter water and geochemical cycles, and reshape landscapes.

Where the wild things were

People like to live in many of the same places favored by wild plants and animals, such as rich coastal plains and river valleys. As a result, cities of more than 200,000 people (blue) are expanding in many of the world's richest biodiversity hot spots (orange). And as many cities sprawl across the landscape faster than their populations grow, learning to manage this competition for prime real estate will be key to protecting many species, including our own.



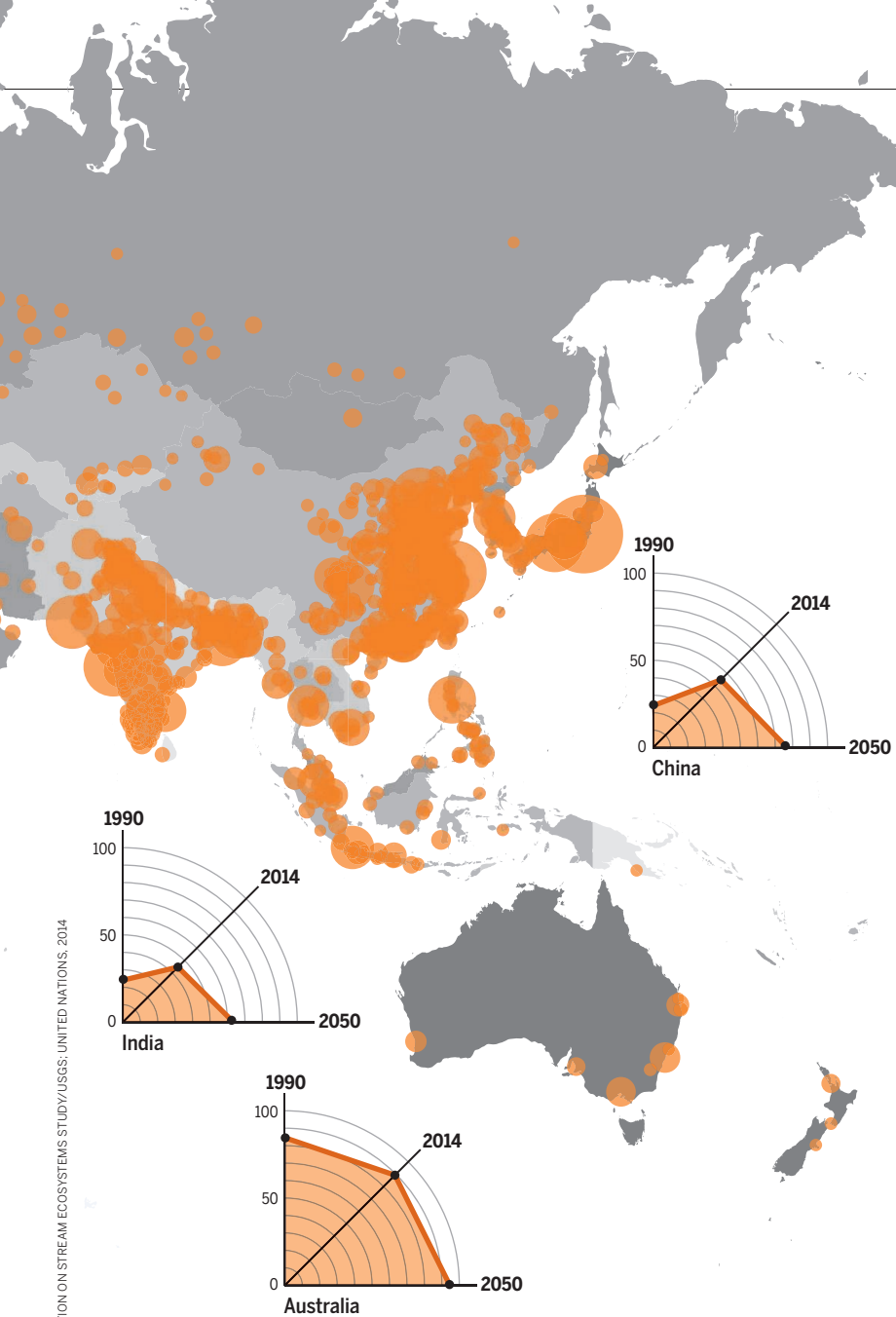
Downloaded from http://science.sciencemag.org/ on May 19, 2016

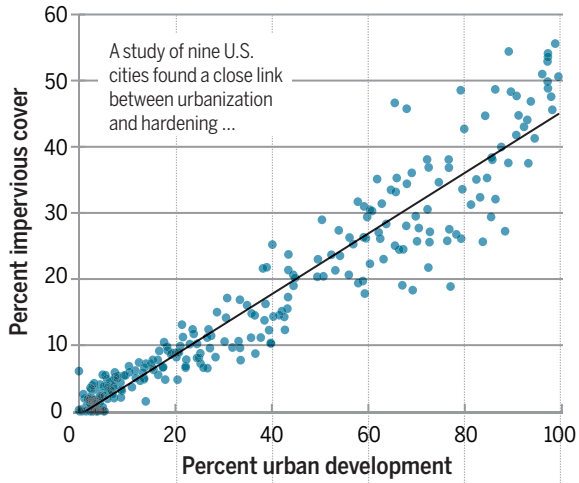
CREDITS: (GRAPHICS) G. GRULLÓN/SCIENCE; (DATA, TOP TO BOTTOM) UNITED NATIONS, 2014; CONSERVATION SYNTHESIS, CENTER FOR APPLIED BIODIVERSITY SCIENCE AT CONSERVATION INTERNATIONAL

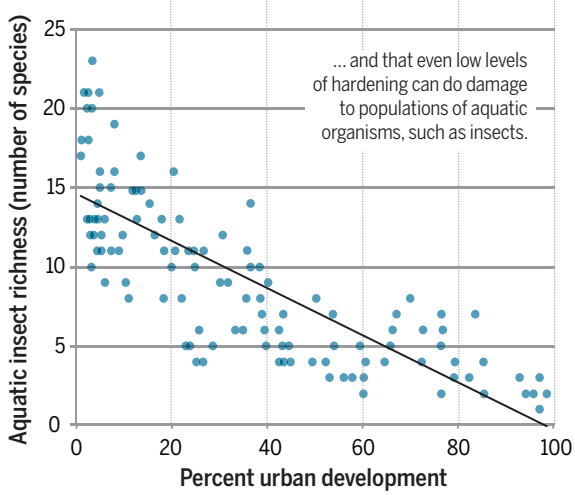
Published by AAAS

Urbanization's hard truth

Cities alter the environment in many ways, from creating air pollution and unusually warm "heat islands" to producing noise and artificial light. One of the most dramatic ecological changes is caused by the spread of impervious surfaces, such as paved roads and roofs. The hardening prevents water from soaking into the soil, promoting flash floods and polluted runoff that damage aquatic ecosystems.

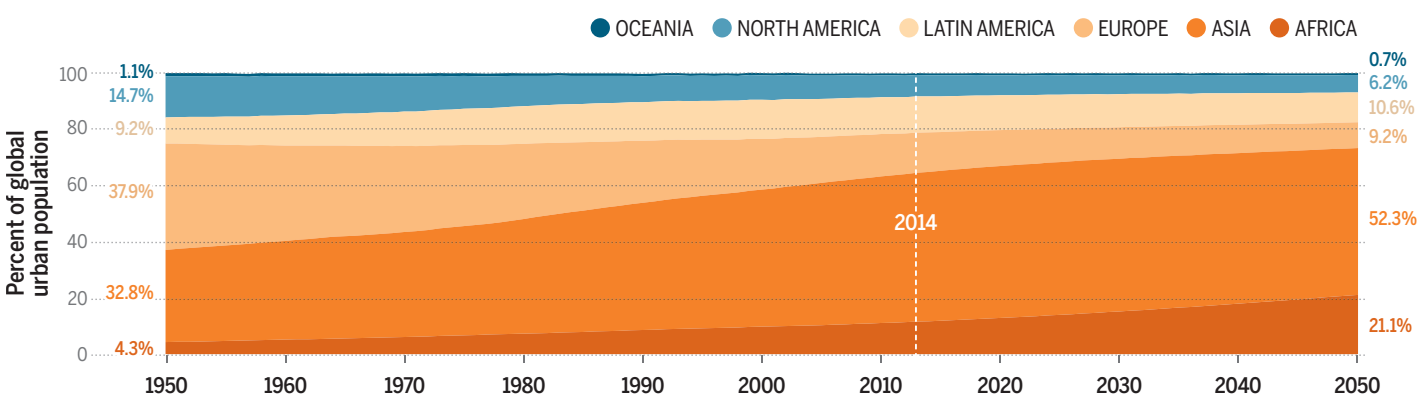






The changing geography of urbanites

In 1950, most of the world's city dwellers were in Europe and the Americas. But Asia and Africa now host the world's largest and fastest growing cities. Cities in just three nations—China, India, and Nigeria—are expected to add nearly 1 billion residents in coming decades, with most of the growth occurring in cities of fewer than 1 million. By 2050, nearly 75% of urbanites will be in Asia and Africa.



CREDITS: (GRAPHICS) G. GRULLÓN/*SCIENCE*; (DATA, TOP TO BOTTOM) EFFECTS OF URBANIZATION ON STREAM ECOSYSTEMS STUDY/USGS; UNITED NATIONS, 2014

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

PHOTO: © TON KOENE

Downloaded from http://science.sciencemag.org/ on May 19, 2016

ROOTS OF THE URBAN MIND

The stress of living with strangers may have spawned cherished aspects of city life

By **Greg Miller**

Stuff a Stone Age human into a time machine and put him on the New York City subway at rush hour and his mind would be blown, if not by the speeding train, bright lights, and strange smells, then by the sheer concentration of people. One crowded subway car might contain more human beings than he'd ever come to know in a lifetime spent hunting and gathering on the grasslands of Africa.

For the vast majority of our history as a species, humans lived in small, mobile communities of dozens to perhaps a couple hundred individuals. Everybody knew everybody else, which made it possible to keep track of who was most likely to help you during a hunt—or kill you in your sleep. Today more than half of us live in cities, surrounded by multitudes of people we'll never meet. It's a radical change, and it happened in an evolutionary eye blink. We navigate our increasingly urban modern world with Paleolithic brains.

In the traditional view, agriculture was the innovation that paved the way for cities. About 12,000 years ago in the Neolithic period, early farmers figured out how to grow enough food to feed many people in one place, and eventually some of these hamlets grew big enough to be called cities. In this view, aspects of city life like fashion and public architecture developed long after agriculture, as a byproduct of life in large groups.

But overcoming food constraints is only part of the story, according to evolutionary psychologist Robin Dunbar of the University of Oxford in the United Kingdom, who has synthesized his own and others' work into a novel hypothesis about the origins of urban life. He argues that in addition to getting enough to eat, the first villagers had to leap another equally significant hurdle: the cognitive demands and social stresses

Published by AAAS

Most people who meet in the Xingu people's village square in Brazil already know each other. But in a large city, people must share space with millions of strangers.

created by living in the midst of strangers. "The stress of living in large communities is very, very intense," Dunbar says. "It's a major problem all primates face. You need mechanisms that will defuse the situation and allow you to stay together."

Dunbar, who is best known for provocative research on social networks, thinks Neolithic people invented, or at least stumbled on, just such mechanisms. For example, architecture that designates certain spaces for certain types of social interactions helps people know what to expect when entering a new building. Material trappings like jewelry and pottery signal the social group and status of strangers. And religion and elaborate group cultural rituals, like singing or dancing together, act as social glue among people in large groups, he says.

"It's a brilliant synthesis," says anthropologist Richard Sosis of the University of Connecticut (UConn), Storrs, of Dunbar's idea. "Like any good theory it unites disparate observable facts and enables you to ask new questions."

Yet Dunbar's ideas have gotten a cool reception from other anthropologists and archaeologists, many of whom wouldn't even comment for this story. They don't see the need to bring in psychology, saying that there's plenty of evidence that agriculture drove larger, more permanent human settlements. Inferring the mental states of prehistoric humans is tricky business, and Dunbar's account will be difficult—but perhaps not impossible—to test.

DUNBAR IS BEST KNOWN for the number that bears his name. "Dunbar's number" is the number of significant social relationships that one individual can maintain. He argues that it's determined by the brain's capacity to store information about individuals and relationships. In primates, the number correlates with the relative size of the neocortex, the folded sheet of tissue on the surface of the brain (*Science*, 5 October 2012, p. 33). For chimps, it averages about 45, according to Dunbar's analysis of published field studies.

For humans it's something like 150. That's probably not everyone you know, but it's the friends you'd most likely invite to your wedding. Dunbar says the number emerges again and again in a handful of direct tests and in data from sources as varied as the number of recipients on Britons' Christmas card lists, the size of Roman infantry divisions, and the number of Facebook "friends" with whom people actually interact.

In a small group, it's possible to remember each individual and their reputation and connections to the rest of the group. In groups larger than Dunbar's number, it's impossible to keep track, and that's incredibly stressful for all social primates, Dunbar says. He cites several lines of evidence, including findings that crowding reduces fertility in captive and wild primates, and work suggesting (though perhaps not conclusively) that violent crime and mental illness are more prevalent in big cities (see paper in this special issue, p. 938).

For most of our history as a species, humans solved this problem by living in fission-fusion societies that break up when they get too big, much as modern hunter-gatherer groups do today. But something changed in the Neolithic, allowing humans to begin living in much bigger communities.

Dunbar began thinking seriously about this question a few years ago, after participating in a multidisciplinary project funded by the British Academy in London called Lucy to Language. He and archaeologists Clive Gamble of the University of Southampton and John Gowlett of the University of Liverpool, both in the United Kingdom, led a team that examined the published archaeological record for clues to the evolution of human social life. If psychological adjustments were crucial to the dawn of bigger settlements, they would have preceded or accompanied them, rather than appearing only later in well-established communities. In reports on Near East Neolithic settlements, the researchers began to see hints that mechanisms for relieving the social stress of living in larger communities emerged early.

One set of strategies utilizes physical objects to make social interactions with strangers more predictable and less scary. Imagine walking into a coffee shop in a strange city. The space is filled with strangers, but the layout is familiar. There's a counter and behind it a guy in an apron. He's the barista, obviously, and you know how to interact with him even though you don't know him. You place your order, hand over money, and step aside so someone else

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

PHOTO: © JOHN VAN HASSELT

can do the same. Everyone knows their role. The thousands of such interactions that happen in a day mostly go smoothly, because they're so predictable.

That's how material culture, including uniforms, helps make life among strangers more manageable today, says archaeologist Fiona Coward of Bournemouth University, Talbot, in the United Kingdom, who also worked on the Lucy project. Its roots go back to when people first started living in large groups, she says. "What we see in the Near East as people start to settle down is an elaboration of material environments." Coward has spent several years combing papers to compile a database that includes roughly 30,000 humanmade objects discovered at nearly 600 Neolithic sites in the Near East. She's been tabulating objects and their styles, from tools like awls to decorations like beads, to track the development and spread of material culture at these sites.

The conventional thinking is that people settled down first and then started accumulating more stuff, but Coward thinks it may have been the other way around. Her preliminary analysis suggests that at many (though not all) sites, rich material culture preceded the growth of settlements. If so, material culture may have enabled larger communities rather than merely being a byproduct of them. The growing diversity of tools, jewelry, and figurines helped delineate social roles, Coward says. Instead of trying to keep track of hundreds or even thousands of individuals, people could lump their neighbors into a much smaller number of social categories. As Gamble, who was Coward's postdoctoral adviser, puts it, "Material culture had to become more complex if we were going to expand without losing our sanity."

Architecture may also have served as a social aid, Dunbar and Coward argue in a chapter in a 2014 book called *Lucy to Language: The Benchmark Papers*. The ways Neolithic people arranged their spaces suggest the development of new forms of social life, Dunbar and Coward argue, drawing support from the work of others, including anthropologist Brian Byrd, who has studied the transition to more permanent settlements in the Near East and U.S. West.

Most of our clues about our hunting and gathering ancestors come from people who live that lifestyle today, explains Byrd, a primary investigator at the Far Western Anthropological Research Group, a cultural resources consulting company, and a research associate at the University of California, Davis. Visit the hunting and gathering !Kung people of southern Africa, for example, and you'll likely find buildings made from brush that offer little privacy. "In their huts, the doors are wide open, they all face the center, and the hearths are out front," Byrd says. "Everybody's business is everybody's business."

The lack of a distinction between public and private space may work for people in small groups who know each other well. But as people transitioned to larger, more permanent settlements, they began creating distinctive spaces for different activities and interactions. For example, at Beidha, an early Neolithic site in Jordan, Byrd and colleagues have found what appear to be dwellings for individual families, often arranged around a central public space. Individual dwellings often had areas for storing grain or other resources—among the first signs of private property—but food preparation and other cooperative activities were done in public. Dividing the space like this would have helped limit and formalize social interactions, Byrd says. "Dunbar's ideas on psychosocial stresses dovetail nicely with my perspective."

The same forces may have been at work at Çatalhöyük, the largest and best-preserved Neolithic settlement, in modern Turkey (*Science*, 20 November 1998, p. 1442). There, starting 9500 years ago, people built enclosed dwellings that sealed off life inside from outside eyes and let inhabitants control who came in. "Çatalhöyük is an extraordinary site because the houses are built close together like cardboard boxes, bang up against each other," says archaeologist Trevor Watkins of the University of Edinburgh, who has worked at the site. "You moved around on the flat roofs and from there you descended by ladder into the building," he says. "It's very deliberate—you couldn't just walk accidentally into someone's doorway."

In a place like Çatalhöyük, which was home to up to 10,000 people, there must have been conventions about whether you knocked on a door or asked to be let in, or whether you approached a certain house at

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

People gathering to pray at a Jakarta mosque (left) may not know each other, but they are united by religion and dress. Similarly, ancient religion may have eased social stress among hunter-gatherers who met at the ancient site of Göbekli Tepe in Turkey.

PHOTO: © VINCENT J. MUSI/NATIONAL GEOGRAPHIC CREATIVE

all given your social connection with the inhabitants, Watkins says. People would have needed new rules to navigate this emerging social world, and Watkins thinks the buildings themselves could have served as a guide. "There are norms of behavior embodied in the architecture and the way the buildings are organized and the imagery they contain."

OTHER INNOVATIONS helped people cope with another potential contributor to social strife: freeloaders. In groups smaller than Dunbar's number, everyone belongs to the same social network, and peer pressure alone keeps people on their good behavior, Dunbar says. In larger groups, more formal rules and policing mechanisms become necessary.

Other researchers have come to similar conclusions, and a growing number have argued that religions with all-seeing, punishing gods arose to take over this policing role in bigger societies (*Science*, 28 August 2015, p. 918). Dunbar says his thesis fits with their research. "Hunter-gatherers don't have moralizing gods," he says. The timing works: Archaeological evidence suggests that the rise of organized religion coincided with or even preceded the emergence of larger, permanent communities. For example, at Göbekli Tepe, an early Neolithic site in modern Turkey, archaeologists suspect hunter-gatherers came together en masse for rituals before they lived together in villages (*Science*, 18 January 2008, p. 278).

Those rituals probably involved music and dancing, practices that Dunbar has argued strengthen social bonds. He thinks these activities act much like social grooming in monkeys and apes, promoting bonds through the release of endorphins, the brain's natural version of opioid painkillers.

In a 2012 study published in the journal *Evolutionary Psychology*, for example, Dunbar and colleagues found that drumming, dancing, and singing in groups increased volunteers' endorphin levels, as measured by testing their pain thresholds with an overinflated blood pressure cuff or an ice pack. Active participation was crucial—pain tolerance did not rise when people merely listened to music.

They also tested the next step—whether music promotes social bonds—using a London-based choir group. Members practice in groups of 20 to 80 and then come together once a year to perform as a "megachoir" of more than 200. (Dunbar notes that these numbers approximate those of fission-fusion societies in which small bands of hunter-gatherers periodically live together for a time.) In both the small and large choir groups, members had higher pain thresholds after singing, and they reported more positive emotions and stronger feelings of inclusion and connectivity. The before-and-after difference in social closeness was more pronounced in the megachoir, the researchers reported last year in the journal *Evolution and Human Behavior*, suggesting that cultural phenomena like national anthems and religious music promote social bonding in large groups of strangers, they said.

"The idea that something like singing can bond larger groups makes a lot of sense," says Sosis of UConn, who notes that this fits with his observations of Micronesian communities, where communal song and dance are reserved for festivals when people gather from other islands.

NOT ALL THE PIECES of Dunbar's broader thesis fit together quite so neatly, however, and some researchers point to gaps in the evidence. Primatologist Joan Silk of Arizona State University, Tempe, for example, isn't convinced that social stress is such a big obstacle to living in large groups. Among chimpanzees, as groups get bigger individuals do spend more time grooming to maintain social ties, but they also have to range over greater distances and expend more energy to find food, Silk says. Thus their group size may be limited by food rather than psychology. "When groups split, it could be because the energetic costs of living in large groups becomes too high or because the quality of social ties has been eroded," Silk says. "It's hard to say which plays a more important role from the available evidence."

The same question applies to Neolithic humans. So far, only patchy evidence suggests that mechanisms for overcoming social stress emerged before humans massed together. That leaves plenty of room for the traditional view that the richness of objects, architecture, and culture arose almost by accident after agriculture freed people from spending all their time securing food. In this view, says Peter Turchin, an evolutionary anthropologist at UConn, "Agriculture happened and then social complexity just sort of bubbled up." But Turchin thinks Dunbar is right to challenge this chronology. "It's a bit controversial, but I actually buy his idea."

Many of the Neolithic innovations that Dunbar credits with aiding the transition to larger communities are conspicuous in our cities today. In fact, they're some of the things we most associate with urban life, from elaborate architecture and cultural institutions to a dazzling selection of material goods. If Dunbar is right, it's not so much that we need cities to produce all these things. It's that we couldn't have cities without them. ■

Greg Miller is a science and technology journalist in Portland, Oregon.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

A rat trapped in Pau da Lima will help researchers understand how the rodents help spread *Leptospira*, a sometimes deadly bacterium.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

A PLAGUE OF RATS

As more people crowd into urban slums,the risks posed by rodent-borne diseases are on the rise

By **Warren Cornwall**, *in Salvador, Brazil; Photography by* **Mauricio Susin**

Rats haunt the slums of Pau da Lima. Their paw prints surround drain pipes. Burrows pock dirt walls. Shriveled black feces speckle patio edges. The rodents even leave their mark in the blood of the people living here in a crowded favela on the edge of this sprawling coastal city, Brazil's third largest. Many residents carry antibodies for *Leptospira*, a bacterium found in rat urine that can be deadly to humans.

"There's so many rats. You can't believe it. Outside, inside," says Carlos Bautista as he sits on the step of his brick shack, looking out over a pile of sodden trash and a makeshift chicken coop.

The haunting is deeply personal for Bautista. Six years ago, his 22-year-old wife died, unexpectedly, from lung damage caused by leptospirosis. Soon after, Bautista sent his son to live in the countryside with his grandparents. "It's better to have him alive there than to have him here" exposed to rats and disease, he says in a voice barely above a whisper.

Rats have long been one of the world's

Published by AAAS

most ubiquitous—and infamous—forms of urban wildlife, synonymous with pestilence and squalor. They've attracted only sporadic attention from scientists, however. Much about the secretive city rat—chiefly the Norway rat, *Rattus norvegicus*—remains a mystery. But as the world's urban population surges and more people crowd into rat-plagued neighborhoods like Pau da Lima, the rodents are getting renewed attention from researchers and public health experts. Over the past decade, scientists in a number of cities have launched efforts to better understand rat behavior and evolution, and the role they play in spreading disease.

One of the most intensive and longest-running investigations into rat-human interactions is occurring here in Pau da Lima, a chaotic jumble of buildings astride a small, hilly swathe of this city of 2.9 million people. For the last 2 decades, researchers have scrutinized the bodies, homes, and habits of favela residents—rat and human alike—while dodging encounters with gun-toting gangs. The goal is to decipher the forces driving leptospirosis, which kills some 60,000 people annually worldwide, and find the best ways to curb a disease that experts warn is an underappreciated threat in the burgeoning slums of a more urban world.

"When we think about the slums in Jakarta or Manila or Cali, Colombia, what you see in Pau da Lima is exactly what you see in those areas, if not worse," says Albert Ko, a physician and infectious disease expert at Yale University, and a founder of the Salvador research project. "We need to find out what solutions can be done immediately that are also generalizable to many of the urban slums."

Researcher Arsinoê Pertile performs a necropsy, collecting tissue samples that might reveal whether rats of a certain age or sex serve as particularly important disease carriers.

KO'S INTEREST in Pau da Lima's rats dates back to 1996, when a surge of deathly ill people, many with failing kidneys, started appearing at the Salvador hospital where he worked. At the time, leptospirosis was considered a rural disease. The corkscrew-shaped spirochete that causes it dwells in the kidneys and urinary tracts of rats and farm animals, and it infects people when their skin or mucus membranes come in contact with water contaminated by the animals' urine. Many people show no sign of infection, or just fever and aches. But a small fraction develop severe kidney damage or massive bleeding in the lungs, although researchers aren't sure why.

Alarmed, Ko and Brazilian colleagues spent much of a year tracking the outbreak. The results, spelled out in a 1999 article in *The Lancet*, were among the first alerting the world that this infection had moved to cities. Over 8 months, they found 326 severe cases, with 50 people dead, and traced the cause to a strain of *Leptospira* found primarily in rats. They noticed that infections surged after intense rains, and that most of the sick came from favelas on the city's outskirts, nearly half of which had open sewers. One was Pau da Lima.

Beginning in 2001, that neighborhood became the focus of an ambitious attempt to merge infectious disease research, urban ecology, and community development. With backing from Brazilian and U.S. funding agencies, the scientists recruited local officials and favela residents to help understand and counter the disease.

The result has been pioneering work on a disease so neglected that it doesn't even make some lists of neglected diseases, says physician Joseph Vinetz of the University of California, San Diego, who studies leptospirosis. "As far as I know, there are no systematic urban studies like the one going on in Salvador."

ON AN APRIL MORNING, a dozen researchers wearing long white lab coats slip through gaps between buildings lining a traffic-clogged street, then descend steep paths into Pau da Lima for a day of data gathering. The coats are sweltering in the damp, tropical heat. But security trumps comfort. The uniforms serve as a white flag of neutrality to the drug-trafficking gangs that control these neighborhoods. Still, the scientists are constantly on guard, watching for police entering the favela, often the prelude to a gun battle.

Ecologist Arsinoê Pertile, a graduate student at Salvador's Federal University of Bahia (UFBA) here, walks past shacks clinging to a hillside of red clay. Brassy rhythms pump from stereo speakers, melding with the whine of saws and the bang of hammers, a testament to the constant construction of improvised dwellings in a place the scientists have dubbed Valley 4. This and several neighboring ravines are home to more than 3000 people squeezed into less than a fifth of a square kilometer—a population density twice that of New York City. Nearly 90% of the residents are squatters. The average person survives on the equivalent of $2.60 a day.

At the valley floor, Pertile enters a cramped, walled courtyard that serves as a tiny open-air market, selling beer and other drinks. There, a cage trap tucked between a washing machine and a stack of empty bottles has captured a new addition to her research: a fist-sized ball of dull brown fur with a pair of shining black eyes.

That afternoon, Pertile will kill and dissect the rat, recording its size and sex and taking tissue samples. Among other things, its urine will be checked for *Leptospira*. "Close to 80% (of the rats) had *Leptospira*" during one recent research

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

sweep, she says, examining the caged animal as the shop owner calmly sweeps the floor nearby. Researchers hope to help pin down the chief reservoirs of the bacteria by correlating rats' age, sex, and capture site with *Leptospira* levels in the urine. Some rats may be more important carriers than others.

The traps are part of a much broader effort to construct a detailed picture of the favela's rats. To identify hot spots of rat activity, for instance, scientists have scattered throughout the valley hundreds of plastic squares the size of dinner plates, coated with a sticky film of soot mixed with methyl alcohol. They record a visual impression of every paw and tail that passes over them.

On this day, Pertile and her research partner, Luciano Lima, a rat exterminator for the city, are particularly interested in how quickly rats are repopulating Valley 4 after a recent extermination effort—and where they are coming from. Genetic studies have found that rat populations in each of the favela's valleys are relatively distinct, suggesting the rodents don't stray too far. Now, with traps at 60 spots in three valleys, they hope to learn whether rat numbers are rebounding because the animals are moving in from neighboring slums, or simply because Valley 4 survivors are reproducing. The answer could help shape future rat control programs.

The market is emblematic of how intimately people and rats coexist here. The building is perched at the edge of a stream of gray, fetid water, fed by trickles from white plastic pipes jutting from nearby buildings. It's the valley's improvised sewer system. Empty plastic bottles and food bags litter the streamside, tossed there rather than carried on the long hike to the top of the valley.

Lima, a 7-year veteran of the rat wars, points out the water, food, and thick vegetation that make this a rat haven. He traces the route the rats can take from the streambank, up through a drain pipe and into the back of the bar. Within 4 months of the last extermination attempt, the valley "was again full of rodents," he says.

DECLARING WAR ON RATS might seem the obvious way to address a rat-borne disease. But the Pau da Lima study has also shown that killing rats isn't always the answer. That's because researchers have realized another major culprit is water, specifically untreated sewage and runoff. It connects everything in the favela—the rats, the bacteria, and the people.

Even in wealthy cities, rats can be rife with *Leptospira* infections, but the number of human cases is low. That's probably because modern infrastructure steers most sewage and rainwater—and *Leptospira*—into pipes and away from people, says Federico Costa, an ecologist at UFBA who now directs the Pau da Lima work. Not so in Valley 4. Open sewers are the norm. The months-long rainy season turns paths to muddy streams and floods homes in the valley bottom.

A few hundred yards downstream from the courtyard market, researchers crowd into the small living room of one of those homes. Jamile da Cruz Nascimento, her husband, and their three children live in the disease's bullseye—in a low spot, near the confluence of two sewage-filled streams.

For nearly 2 decades, the favela of Pau da Lima has a been a focus of urban rat research.

Flood waters can reach almost to her doorstep. Her sons often go barefoot or wear flip-flops rather than closed shoes; their skin is constantly exposed to polluted water.

Asked how serious she thinks leptospirosis is, Nascimento gives it a 10 out of 10. She pegs her risk of getting it as a five out of five. A friend who lived nearby died from the disease several years ago. "We have many cases here," she says. "It's very serious."

A small electric fan wags hypnotically back and forth as a researcher slides a needle into the arm of Eric, Nascimento's lanky, 12-year-old son. Dark blood courses into a vial, to be tested for signs of recent *Leptospira* infections. Another worker questions the boy: Has he had a fever or joint pain in the last year? No. How often has he recently walked in floodwater? Frequently. Did he wear rubber boots? Sometimes.

So far, Nascimento and her children have gotten good news: They've never shown signs of a *Leptospira* infection. But she says her husband, who works as a garbage collector, always has a positive blood test.

Given the abundance of rats and *Leptospira* in the slum, why doesn't everyone here get sick? Ko's team would like to know. Understanding why Nascimento and her kids dodged the bacteria, while her husband didn't, for example, could offer clues to strategies for coping with the dangers.

Overall, researchers have found that approximately 3.2% of the favela's residents are infected every year. One in 30 of those infections leads to mild sickness, and one in roughly 200 causes severe illness, based on the number of cases admitted to the city's hospitals.

A string of studies published in the last 3 years lays out key risk factors. Houses with signs of rat infestation are nearly twice as likely to have a *Leptospira* infection, for instance. Other factors are barometers of poverty. The danger increases the farther downhill someone lives, tracing the slum's economic pecking order, which has pushed the poorest residents into the lowest, wettest areas. One study found the chance of infection fell by half with every additional dollar a day a person earned.

THE LINK between infection and water suggests that killing rats alone won't be enough to protect Pau da Lima's residents. "People are very close to the sewage and very close to the rats," Costa says. "I think if it could be done, a system that collects most of the water … would avoid most of the infections."

But in the favela, that's a big "if." In the early years of the project, scientists and community leaders successfully lobbied the federal government for $36 million to build a road and sewer lines through some of the most flood-prone parts of Valley 4 and a neighboring valley, as well as new housing for 271 households, or about 7% of the population. But today just a fraction of the project is built. A tidy row of apartment buildings sits empty along a freshly paved street at the lowest end of Valley 4. But sewage still flows down an open stream. In the adjoining valley, the only sign of work is a dirt road running down a hillside.

Delays ate into the funding, the researchers say. And in 2015, gangs controlling the other valley shut down the construction there, fearing it would give police easier access. "That project should have been done 10 years ago," Ko laments.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

Researchers question Pau da Lima residents Jamile da Cruz Nascimento and her 12-year-old son Eric as part of their effort to understand disease risks.

GIVEN SUCH DIFFICULTIES, the scientists are looking for cheaper, quicker ways to make inroads against leptospirosis. They want to know how to fine-tune extermination campaigns, and whether things as simple as fencing off certain areas, or giving everyone rubber boots, could help.

The researchers hope to first test such interventions using a computer model that simulates how people, rats, and the bacteria commingle in the favela. Right now, the model—being developed by Costa and scientists at Yale and the University of Liverpool in the United Kingdom—is relatively coarse; researchers can run scenarios only at the level of a whole valley. Eventually, they hope to be able to see patterns at a much finer scale, just tens of meters.

Still, insights from Pau da Lima are influencing city government practices. Fifteen years ago, for instance, Salvador's rat control efforts were haphazard, often centering on affluent neighborhoods where residents and the politically connected complained, Ko says. Today, in contrast, the city focuses on 11 neighborhoods with the highest leptospirosis rates. In 2015, exterminators went house to house in five of the worst areas, leaving poison where they found signs of rats. When a case of leptospirosis is reported anywhere in the city, a team applies the same treatment within 200 meters of the patient's home.

Ko admits that they don't yet have data confirming that such strategies reduce infections. But "my gut feeling," he says, "is that we have to figure out ways to reduce the rat population."

A MORNING spent with the city's rat patrol, however, offers a glimpse into the difficulties of brute-force extermination. An aging Volkswagon van parks on a ridge above Valley 4, and eight people spill out, wearing white and blue polo shirts emblazoned with a logo reading "Centro de Controle de Zoonoses." They are responding to a report that a 12-year-old boy had come down with a fever—perhaps leptospirosis.

The agents fan out into the valley to hunt for signs of rats and distribute poison. Last year, the city had five of these vans and a staff of 120 conducting extensive, neighborhood-wide campaigns, says Maria Gorete Magalhães Rodrigues, who oversees the city's rat program. But with the arrival of the Zika virus, four vans and 80 workers were reassigned to fight mosquitoes. Now, her rat team responds only to reports of infections.

"Leptospirosis is not taken as seriously as it should be," she says through a translator. Then she switches briefly to English: "But I fight for leptospirosis."

After less than 20 minutes of rat hunting, the agents in Valley 4 suddenly retreat up the hill and back to the van. It turns out one had seen a policeman and feared bullets might start flying. Valley 4's rats, it seemed, would be safe for the moment. ■

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

Compact, walkable enclaves such as Guangzhou's Liuyun Xiaoqu are the new model for China's city planners.

Downloaded from http://science.sciencemag.org/ on

CHINA RETHINKS CITIES

After decades of reckless growth, the country revises its vision

By **Dennis Normile**, *in Guangzhou, China*

Just a few steps separate one of the worst examples of China's recent urbanization from one of the best. Tianhe Road, a main artery here in China's third-largest city, is eight lanes wide but snarled with cars. Pedestrians scurry between bland office towers and huge malls. Only the brave ride bicycles.

Just a block away, pedestrians and cyclists rule. Liuyun Xiaoqu is a verdant neighborhood about 5 kilometers from the city core where small streets once clogged by cars have given way to walkways and planters. Boutiques, cafes, and groceries cluster at ground level in the midsize apartment buildings. Even on a drizzly day, the neighborhood is alive with window-shoppers, retirees, and young couples. Residents can stroll to stores, schools, and restaurants, and many work nearby. It's a striking contrast not just with the surrounding city, but with newer residential areas on the fringes of town, where people live in gigantic superblocks, isolated from workplaces and amenities, and rely on cars.

Guangzhou is not unique. China's urban population doubled between 1978 and 2010, while the area covered by cities tripled. The sprawl has driven up car use and helped make China's cities a major source of carbon emissions. Choking pollution and sedentary lifestyles are harming urban health.

Officials have concluded that the country's urbanization has gone awry. This past February, the State Council and the Communist Party's Central Committee—the nation's highest authorities—adopted new guidelines that call for compacter cities with denser networks of streets, more pedestrian and cycling lanes, better public transport, mixed-use zoning, and more green space. "This is a significant turning point for Chinese cities," says Yan Song, an urban planner at the University of North Carolina, Chapel Hill.

Experts warn that transformation will be difficult, and could take decades. "It's going to be evolution, not revolution," predicts Yang Li, an urban planner at the Guangzhou office of the Institute for Transportation & Development Policy (ITDP), which promotes sustainable urban development. But over time, entire cities—not just isolated neighborhoods—may come to resemble Liuyun Xiaoqu.

IN 2014, the World Bank described China's urbanization as "unprecedented in scale." Less than 20% of China's population lived in cities in 1978; by 2014, more than half did. The bank and others noted that China's urbanization helped lift hundreds of millions of people out of poverty. But they have also critiqued the new cities as unsustainable.

Guangzhou's bus rapid transit line (bottom) cut emissions, saved travel time, reduced operating costs, and smoothed out the traffic jams seen before the system was built (top).

PHOTOS: (LEFT TO RIGHT) © WENJUN CHEN; INSTITUTE FOR TRANSPORTATION AND DEVELOPMENT POLICY

Planners initially opted for Soviet-inspired wide streets, huge blocks, and massive buildings, and then added U.S.-style highways and ring roads radiating ever farther from city centers. Zoning laws fragmented cities into single-use residential or office districts. Later, gated compounds with a few high-rise towers in huge, parklike settings leapfrogged far beyond transit lines and other services.

In a vicious circle, the spread of car-dependent superblocks "quickly caused [traffic] congestion, so the government made more and wider streets," says Dongquan He, an environmental scientist at Energy Foundation China's Beijing office. Researchers estimate that superblock residents use 65% to 80% more transit energy than those living in mixed-use, walkable neighborhoods.

A perverse incentive exacerbated sprawl: Many governments depend on revenue gained by incorporating farmland and then selling the building rights. "The revenue from these transactions can amount to 40% or more of a city's budget," Qian Zhang, a geographer at the Chinese Academy of Sciences's (CAS's) Institute of Geographic Sciences and Natural Resources Research in Beijing, and colleagues wrote in 2014 in *Land Use Policy*.

The result: Cities have sprawled faster than populations have grown, so that "average population density … dropped by more than 25%," the World Bank estimates, while air pollution increased. Shanghai's carbon dioxide (CO_2) emissions hit 13.1 tons per capita in 2011, higher than the 2013 CO_2 emissions of 5.7 tons per capita in London and the 8.3 tons per capita in Los Angeles, California, according to a 2015 study by the Lawrence Berkeley National Laboratory in California. "Reducing the urban carbon footprint is a key" to China meeting its Paris Agreement pledge to reduce greenhouse gas emissions, He says.

Public health is also at stake. In big cities, the incidence of chronic metabolic, circulatory, and respiratory diseases was roughly double rural rates in 2008, according to a study by Yong-Guan Zhu, a biogeochemist at CAS's Institute of Urban Environment in Xia-

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

men. Air pollution contributed to 1.2 million premature deaths in China in 2010, according to data extracted from the 2010 Global Burden of Disease Study.

Reconfiguring China's cities along the lines of Liuyun Xiaoqu will be part of the solution. Built in the 1980s, the area began as a strictly residential, gated compound with about 6500 apartments. In 2000, residents gained ownership—and greater control—of their apartments. Entrepreneurs opened shops in first-floor units. The area is near a major stadium, so Guangzhou spruced it up before the 2010 Asian Games with resurfaced walkways, pocket parks, and playgrounds. Gradually, Liuyun Xiaoqu opened its gates and restricted cars. Commercial and residential property values have risen more quickly than in surrounding areas, according to an ITDP study. "The beneficial policies were very simple, but they had a snowball effect," says C. C. Huang, an urban policy analyst at Energy Innovation, a consulting firm in San Francisco, California.

Guangzhou's bus rapid transit (BRT) system offers another lesson. Gaining attention worldwide for their efficiency and low cost, BRT systems have dedicated bus lanes that are typically located in the center of a roadway to avoid curbside jams. Many BRTs have a single platform between two lanes, requiring special buses with driver's side doors. But Guangzhou's has dual platforms, allowing the use of regular buses that operate beyond the BRT corridor.

Opened in 2010, the BRT runs for 23 kilometers along a main east-west artery. During peak periods, 350 buses carry 26,900 passengers an hour, more than most of China's subway lines. The BRT will cut CO_2 emissions by 86,000 tons and particulate emissions by 4 tons annually during its first 10 years, ITDP estimates. Building a BRT costs about 10% of the price of a subway.

China is also looking to its past for a lesson in sustainability. "China used to be the bicycle kingdom," Song says. But in what she calls one of the biggest mistakes of China's urbanization, many cities removed once ubiquitous bike lanes. Now, they are restoring them. Leading the way is Hangzhou, which operates the world's largest bike-sharing program, with 65,000 bikes.

Remaking China's cities won't be easy. Calls to open gated compounds, for instance, have drawn fierce opposition from residents worried about traffic and privacy. People have grown attached to cars. "It will be difficult to drive them back to public transit," Song predicts. Governments, meanwhile, will need to cooperate on transit and other projects that are "complicated technically." He warns. But he is optimistic that China is "moving in a good direction." ■

PHOTOS (LEFT TO RIGHT): VISIBLE LIGHT © BEN NELMS; FALSE COLOR THERMAL IMAGE © INO

VANCOUVER'S GREEN DREAM

The city wants to dramatically shrink its environmental footprint, but obstacles loom

By **Kenneth R. Weiss**, *in Vancouver, Canada*

With a practiced eye, Innes Hood snaps a thermal image of a window-walled condominium tower dozens of floors high. The glass glows white hot, a signature of wasted heat radiating into this port city's chilly air. Concrete and steel slabs also shimmer with the telltale rosy colors of waste. The slabs "are not insulated, so they act like radiating fins," says Hood, a Vancouver, Canada–based sustainability consultant with Stantec Inc.

Hood's camera is taking aim at one of Vancouver's biggest challenges in its quest to become the greenest city on the planet: improving the energy efficiency of buildings. In part because of leaky windows and poor insulation, buildings consume nearly two-thirds of the energy used in the City of Glass—a nickname referring to Vancouver's landmark towers offering stunning ocean and mountain views. Now, city leaders have pledged to cut energy use and greenhouse gas emissions from existing buildings by

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

Red reveals heat leaking from a tower in downtown Vancouver in this image, which combines a visual light photograph with a false color thermal scan. Energy-wasting buildings that radiate heat pose one of the biggest challenges to the city's effort to slash energy use and curb greenhouse gas emissions.

20% by 2020, and to require structures built after 2030 to produce no net emissions. It's part of a long-term goal of fashioning a city that operates entirely on renewable energy and produces little waste.

Vancouver isn't the only metropolis setting formidable targets. Around the world, urban leaders are embarking on an array of efforts to reduce the strain that cities place on the environment. Although cities are often seen as a source of environmental problems, many analysts argue they are also key to solving them; large urban populations, for instance, can use far less energy per capita than rural or suburban residents—if the right policies and infrastructure are in place. But time is short: Experts say that by 2050, about two-thirds of the world's population will live in urban areas, and how cities handle rapid urbanization will be critical to protecting biodiversity and human health, as well as combatting climate change. By one estimate, the world's urban areas could prevent the release of 45 gigatons of carbon dioxide—more than eight times the annual emissions of the United States—if the building boom over the next 15 years produces more efficient infrastructure.

Vancouver has taken the urgency to heart. In 2011, it adopted a Greenest City Action Plan that has made it a prominent pioneer in urban greening, including efforts to transform raw sewage and food waste into energy, and to coax residents to use less water and leave their cars to walk, bike, or ride public transport. Despite some successes, however, researchers say Vancouver's experience illustrates the sobering limits that cities face in acting on their own to reduce environmental impacts, given the complexities of national politics and a globalized economy.

BORDERED BY WATER on three sides, Vancouver squeezes more than 600,000 residents into its city limits—and is the heart of a sprawling metropolis of 2.3 million. Its downtown hosts the headquarters of numerous timber and mining companies, reflecting a region that has relied on resource extraction for jobs and income.

Despite this economic history, a green thread has long run through Vancouver's fabric. In 1971, the city gave birth to Greenpeace, the no-holds-barred environmental group. In 1990, it became one of the first cities to officially recognize the threat of climate change. City Hall has long embraced packing more people into the central city, a strategy once termed "ecodensity," with the aim of using fewer resources per capita. And an ample supply of hydropower from British Columbia's rivers has helped give Vancouver the smallest carbon footprint of any major city in North America.

Still, when Gregor Robertson ran for mayor in 2008, he believed the city could do even more. The provincial legislator and cofounder of the Happy Planet organic juice company pledged to make Vancouver "the greenest city on Earth." It was a winning message for voters. Once elected to office, Robertson assembled a task force that labored for 2 years before issuing the 2011 action plan, which laid out a host of specific, measurable targets.

Much of the plan is far from sexy, dealing with the nitty-gritty of quotidian city life, such as how to improve trash disposal, transit operations, and building codes. Still, the blueprint has helped catalyze real, albeit limited, progress.

WALKING AHEAD of a rumbling garbage truck in an alley in East Vancouver, Jez Figol peers inside a green city waste bin meant for food scraps and other organic waste. Discovering a misplaced milk carton, she pulls out a warning sticker and slaps it on the bin. She adds the address to her clipboard of those violating Vancouver's tough trash-handling rules, which are

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

To cut methane emissions from its landfill, Vancouver strives to remove and compost all organic matter from its waste stream, producing a soil amendment.

designed in part to reduce the amount of methane, a potent warming gas, burping from the city's landfill.

Figol, a no-nonsense woman who used to investigate dog bites, is one of two inspectors who spend their days digging into garbage cans to crack down on careless residents who don't properly separate their waste. The city has expanded its recycling program for paper, glass, and plastic to include other items. It also requires wood removed from pre-1940 houses to be reused, recycled, or burned as an alternative fuel. Vancouver's landfill, like many others, uses extraction wells to capture the methane produced by bacteria breaking down the trash. Most of the collected gas is used to heat nearby greenhouses that grow tomatoes year-round. The surplus is burned in a gas flare.

Despite these efforts, however, officials figured that 40% of landfill methane was escaping into the atmosphere, where it is far more powerful at trapping heat than carbon dioxide. So they took another step, trying to remove all organics from the waste stream. In 2013, volunteer "ambassadors" fanned out across the city, handing out kitchen pails and instructions to toss plate scrapings, uneaten meat, bones, soiled pizza boxes, and yard trimmings into the green bins. All this and more now goes to two composting facilities, which turn it into soil amendment for farmers and landscapers.

Initially, compliance wasn't what city leaders hoped. Then they had a bright idea: switching trash schedules, and picking up the smelly, fruit fly–attracting green bins every week, instead of every 2 weeks. "That did it," says Albert Shamess, the city's director of waste management. "We saw a 40% reduction in garbage and a 60% increase in composting."

Still, Figol needs to remind some residents of the rules. In the alley, she finds a trash bag stuck in the wrong bin. She attaches a large red tag—so the drivers would leave it—and confronts a startled mother, kids in tow, who emerges from the offending address. "It's our job to do face-to-face education," Figol says. And if that doesn't work, next come fines.

WITH THE WAVE OF A KEY FOB, Simon Donner unlocks a silver Fiat 500 parked in a car-share spot below City Hall. "I've always wanted to try one of these," says the geography professor at the University of British Columbia here. The city makes it easy for him to not own a car, a bonus for this climate scientist who likes to walk the talk. It has invested heavily in bike lanes and mass transit, and claims to be the car-sharing capital of North America, thanks to a fast-growing fleet of short-term rentals offered by four companies. City officials estimate that each rent-a-ride removes as many as 11 private vehicles from the road.

Mostly, Donner bikes to work. He gets a thrill when he passes a sign that displays the number of bikers that have pedaled by a bridge leading to downtown. "I'm a scientist so I love data," he says. "I go out of my way to trip the bike counter." Last year, it recorded 1.37 million bike trips. Now, he wonders: "Will we beat that this year?"

Survey data show the city has already met the 2011 action plan's goals of cutting kilometers driven per person by 20% and coaxing residents to make half of their trips in the city by foot, bicycle, or mass transit. But that success has brought strains: The city's bus and subway systems now bulge with commuters during the rush. Already-packed buses often speed past passengers queued up in the rain. This being Canada, the buses display signs that apologetically read: "Sorry, Bus Full."

VANCOUVER is the first city in North America to mine sewage for heat. Kicking up a sweet-smelling mist, a stream of wastewater races under a catwalk in a concrete basement, and then through a giant strainer removing big particles that might gum up the works. "There's a lot of wasted heat going down the drain," says Chris Barber.

PHOTO: © BEN NELMS

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

Barber manages this unusual plant, which is designed to capture the heat and use it to help supply hot water to 6000 residences, including the condominiums built for the 2010 Winter Olympic Games, and some commercial buildings. A mix of sewage, hot shower water, and other wastewater flows through a labyrinth of painted pipes in meticulously clean rooms that feed an enormous heat exchanger bigger than a semi–trailer truck. The harvested heat is returned to the buildings via separate pipes that carry water heated to 65°C–95°C. (On cold days, natural gas boilers add extra heat.) The cooled sewage goes to a treatment plant, and then the ocean.

The plant, modeled after a facility in Oslo, was more expensive to build than a conventional power plant but is far cheaper to operate, Barber said. And city officials say encouraging the construction of more waste-to-heat plants will be key to reaching their goal of cutting overall greenhouse gas emissions by 80% by 2050. So far, they've achieved a 7% cut.

Other sources of renewable energy will drive further progress. The city has adopted policies designed to help utilities cover the heftier up-front cost of clean energy systems by spreading it across many buildings and many years. One proposed project, for example, will enable a utility that provides heat to more than 210 of the largest downtown buildings to invest $100 million to switch from natural gas fuel to "clean wood waste"—leftovers from the timber and construction industries.

NEW POWER PLANTS, however, won't fix the Glass City's millions of square meters of heat-leaking windows and buildings that pop up like burning candles on thermal images. "Hands down, the existing glass towers downtown are going to be the biggest challenge," says Sean Pander, manager of Vancouver's green building program. "Glass curtain-wall buildings are terrible" and expensive to retrofit, he says. "We've got a lot of them."

Some of the older buildings will be torn down and replaced by midcentury. And officials have steadily tightened efficiency rules for new construction, heading toward the 2030 zero-emissions standards. What will happen to the remaining older buildings, however, remains unclear. The City Council has adopted an energy retrofit strategy that aims to identify the leakiest, least-efficient buildings and offer small incentive payments to support voluntary improvements. But it left it to city staff to work out how to handle the massive costs or the disruption of relocating occupants.

ENERGY-WASTING BUILDINGS are just one obstacle standing between Vancouver and its lofty goals. Although its mass transit system is bursting, for instance, last year voters rejected raising a sales tax to pay for improvements, meaning ridership is unlikely to grow further.

Local voters and City Hall have little say, meanwhile, in other decisions that could have a major impact on the city's environmental footprint. National and provincial officials will decide whether energy companies can build a proposed oil pipeline and natural gas export facility near Vancouver's port, which could dramatically increase greenhouse gas emissions from shipping operations. And as much as the city would prefer nothing but zero-emission vehicles on its streets, fuel and emissions standards are in the hands of the provincial and federal governments. Provincial rules have even limited the city's ability to combat the loss of trees: A $10,000 provincial cap on fines for tree removal means regulators have little leverage to deter affluent landowners from deforesting for a better view.

Such issues highlight a broader challenge, analysts say: making sure city efforts mesh with national and international policies. A cautionary example arose during a recent meeting in Paris of the C40 Cities Climate Leadership Group, made up of cities that have set ambitious climate goals. Officials in Nanjing, China, won an award for electrifying the city's buses. But the switch won't do much to reduce overall greenhouse gas emissions, because 70% of China's electricity comes from coal, notes Dabo Guan, a climate change economist at the University of East Anglia in Norwich, U.K. "It doesn't help, it's actually worse," Guan says.

Most of the choices that determine a city's environmental footprint lie beyond the reach of planners, says Jennie Moore, an urban sustainability expert at the British Columbia Institute of Technology, Burnaby. Moore and colleagues calculated that Vancouver requires a land area of some 2.3 million hectares, about 212 times greater than the city itself, to produce the food, energy, and materials it consumes each year. If Vancouver wants to become truly green, she estimates it will need to shrink its footprint by at least two-thirds. And only its residents can make that happen.

Choices made by the city government account for just 40% of Vancouver's footprint, Moore found; the remaining 60% is determined by how residents choose to live, move around the city, and eat. Consuming red meat, for instance, dramatically boosts a person's footprint because of the land, water, and energy required for livestock. Her best guess of what a sustainable Vancouver would look like: Everyone would live in a small, efficient home of about 46 square meters per person, eat a vegetarian diet, and get around on non-fossil-fuel-based transportation. Moore and colleagues have reduced this to a laugh line: "closet-dwelling vegans on bikes."

SITTING in his City Hall office, Mayor Robertson—who bikes to work and keeps a nearly vegetarian diet (no dairy, but seafood)—acknowledges that Vancouver's goal is "big, audacious." And he concedes that as the city has picked off the lower hanging fruit, "the gains get tougher as we get greener." Yet he remains undeterred—and he thinks the struggle will position Vancouver to compete in the new economy of an urbanized world. "Many cities recognize the huge economic upside, being branded strongly with green and renewable," he says.

Even as Vancouver becomes more efficient, he noted, it remains the fastest growing metropolitan economy in Canada. "People want to be here," Robertson says, "and work in low-carbon industries and live in a multicultural, cosmopolitan city where they don't need to own a car." ■

Kenneth R. Weiss, who won the 2007 Pulitzer Prize for Explanatory Reporting, is a journalist in California.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS

REVIEW

City-integrated renewable energy for urban sustainability

Daniel M. Kammen[1,2,3]* and Deborah A. Sunter[1,3]

To prepare for an urban influx of 2.5 billion people by 2050, it is critical to create cities that are low-carbon, resilient, and livable. Cities not only contribute to global climate change by emitting the majority of anthropogenic greenhouse gases but also are particularly vulnerable to the effects of climate change and extreme weather. We explore options for establishing sustainable energy systems by reducing energy consumption, particularly in the buildings and transportation sectors, and providing robust, decentralized, and renewable energy sources. Through technical advancements in power density, city-integrated renewable energy will be better suited to satisfy the high-energy demands of growing urban areas. Several economic, technical, behavioral, and political challenges need to be overcome for innovation to improve urban sustainability.

Since 2007, a greater percentage of the global population has been living in urban areas than in rural areas. Increased urbanization is expected to continue, with two-thirds of the world's population projected to live in urban areas by 2050, a net urban influx of 2.5 billion people (*1*). Cities today are generally not equipped to address dramatic urban growth and strain on existing infrastructure in a sustainable way, especially with respect to their energy systems.

To be sustainable, cities must themselves, or in the resources that they command, become low-carbon, resilient, and livable (*2*). Although there can be considerable variation in methods for evaluating the emissions footprint of cities (*3*), with 54% of the population living in urban areas, it is estimated that cities are currently responsible for 60 to 70% of anthropogenic greenhouse gas emissions (*4*). The two main strategies for transitioning to a low-carbon city are to shift from fossil fuels to cleaner energy sources and to reduce urban energy consumption levels. The low-carbon transition can be accomplished through energy-efficiency measures, behavioral interventions, and incorporating carbon sinks such as urban parks. Cities and their energy systems should also be resilient to natural and human-made threats (*2*). The energy systems of cities are increasingly vulnerable to the effects of climate change and extreme weather, including storms, flooding, and sea-level rise, and also to natural and human-induced disasters. In addition, urban energy systems directly affect the well-being and happiness of urban inhabitants. Health conditions, economic competitiveness, cultural appeal, and social, gender, and racial equality are influenced by high-energy sectors such as transportation, food production, and water quality.

Here we evaluate some of the more promising recent technological advancements that could help urban areas become sustainable cities. Many opportunities exist, but focusing on city-integrated renewable energy—defined as distributed, non–fossil fuel energy generated locally in urban areas—has the potential to help cities meet several sustainability needs. Many of these renewable sources increase regional energy independence and can be redundant with other sources, thus increasing resiliency. Although there are several existing barriers to their adoption, solutions will involve increased power densities of renewable energy technologies, improved infrastructure capable of supporting widespread integrated energy generation systems, and increased urban energy efficiency, particularly in the buildings sector.

[1]Energy and Resources Group, University of California, Berkeley, CA, USA. [2]Goldman School of Public Policy, University of California, Berkeley, CA, USA. [3]Renewable and Appropriate Energy Laboratory, University of California, Berkeley, CA, USA.
*Corresponding author. Email: kammen@berkeley.edu

City-integrated renewable energy

About 75% of power generated globally is consumed in cities (*5*). Generating city-integrated energy at the site of energy use could substantially contribute to the environmental, economic, and social aspects of urban sustainability. Four characteristic advantages of such distributed energy systems include the ability to (i) offer low to zero carbon emissions, (ii) offset capital-intensive investments for network upgrades, (iii) impart local energy independence and network security, and (iv) motivate social capital and cohesion (*6*).

With limited available installation space, renewable energy generation within urban areas poses particular challenges. We use the balance between the high energy demand of cities and the available energy density supplied by renewable sources as a starting point for an analytic framework for decarbonized urban spaces (Fig. 1). However, in the waves of innovation that will be needed, strategies ranging from space-based solar energy to small modular nuclear power systems, deep geothermal systems, and other generation options could transform the energy landscape. In addition to reducing greenhouse gas emissions, these strategies may reduce the consumption of water, air, and other resources.

Solar energy

Recent economic and technical advances have made city-integrated solar technologies increasingly attractive. Since 2010, the installed price of solar energy has dropped by as much as 50% (*7*). Despite substantial economic progress and anticipated cost parity with fossil fuels, renewable energy technologies have often been criticized for their low power densities, making them inappropriate for urban applications. Conservative estimates of the power density of solar photovoltaics are around 10 W/m^2 (*8*), assuming an average direct solar irradiation of 100 W/m^2, which is typical for the United Kingdom, and a photovoltaic efficiency of 10%. However, the solar resource is highly region-dependent, and in some regions, annual direct solar irradiation can exceed 300 W/m^2 (*9*). Many of the regions expected to experience the greatest increase in urbanization are located in solar-rich regions. For example, the majority of the total land area of India experiences an annual direct solar irradiation of over 200 W/m^2 (*10*). Additionally, the efficiency of photovoltaics has increased steadily and has already surpassed 40% in the laboratory, using concentrated multijunction cells (*11*). Hence, under optimal conditions, the power density of photovoltaics could exceed 120 W/m^2.

Several studies have estimated the photovoltaic potential of existing cities. City-integrated photovoltaics have the potential to satisfy 62% of the current electricity needs of Oeiras, Portugal (*12*), and 66% of the electricity needs of Bardejov, Slovakia (*13*). High-efficiency commercially available photovoltaics only on suitable rooftops could satisfy 19.7 to 31.1% of the daily electricity demand and 47.7 to 94.1% of the morning peak electricity demand of Mumbai, India (*10*). With a 20% adoption rate, solar-powered urban microgrids could reduce the grid demand in Cambridge, MA, to almost zero at midday (*14*).

Heating accounts for 40 to 50% of the global energy demand and 75% of the energy demand within the buildings sector (*15*). Urban solar thermal energy, specifically for space and domestic water heating, has been an area of particular research interest. With efficiencies up to 80% (*15*), solar hot-water systems offer a thermal power density up to 240 W_t/m^2 under optimal conditions (W_t, watt-thermal), whereas the global averaged thermal power density of solar heat collectors is 67 W_t/m^2 (*16*). Because domestic solar hot-water heaters are low-cost and compact, one study showed that 84% of urban households in China could install the system on their rooftops (*17*). Solar thermal energy is also used for passive and active space heating. A study of five Australian cities showed that the use of a Trombe wall could offer energy savings up to 17% (*18*). Seasonal solar thermal energy storage is an approach that stores solar thermal energy collected in the summer for heating in the colder months. This technology provides up to 91% of the total energy needs of a large residential building in Richmond, VA (*19*).

The exploitability of the solar resource is highly affected by urban form. Although taller buildings offer higher surface-to-volume ratios, allowing for increased facade-integrated solar technologies, they also increase the risk of vertical obstruction and shading (*20*). Although building facades provide almost triple the area of building roofs, they

Downloaded from http://science.sciencemag.org/ on May 19, 2016

received only 41% of the total irradiation in Karlsruhe, Germany (*21*). However, an optimized urban form could increase the solar irradiation on facades by up to 45% in greater London, whereas an increase of only 9% is possible for rooftops (*22*).

Geothermal energy

Geothermal energy has been harnessed for both electricity and heat. A typical geothermal plant has an electrical power density of 283 W/m^2 (*23*), but some plants are estimated to have a total power density of nearly 800 W/m^2 (*16*). Geothermal plants can be placed on multiple-use lands, sharing space with activities such as farming and skiing. However, the surrounding land can be affected by subsidence, erosion, landslides, and induced seismicity (*16*, *23*). When taking into account all such affected areas, the typical electrical power density of a geothermal power plant is estimated to be 50 to 80 W/m^2 (*16*). Geothermal energy is more efficiently extracted as heat. Iceland's Hellisheidi combined heat and power plant is able to generate hot water at 25,000 W$_t$/m^2 (*16*). Most cities, however, are located in areas with far fewer geothermal resources. Conventional geothermal resources would only produce a mere 0.017 W/m^2 of electricity in the United Kingdom (*8*), but deep (>10 km) geothermal power could transform this baseload resource into a far more substantial (>10%) element of urban energy supply (*24*). The power density of deep enhanced geothermal systems is between 0.59 and 1.19 W/m^2, depending on the available resource temperature (*25*).

Although the use of geothermal energy for electrical generation in cities is limited, more than 60 countries are using geothermal energy for household, commercial, and industrial heat (*16*). The thermal power densities in most regions are relatively high, even at moderate depths. In the United States, the average thermal power density of ground-source heat pumps is 40 W$_t$/m^2 for horizontal closed loops buried just 1 to 2 m deep and 100 W$_t$/m^2 for boreholes at least 150 m deep (*16*). Ground-source heat pumps for heating and cooling in Chinese urban buildings could reduce energy consumption by 10 to 15% in civil buildings and 25 to 30% in public buildings (*26*). Ground-source heat pumps could also meet the heating demands of 58 to 70% of buildings in Westminster, London, and, if a well-organized district heating system was used, all heating demands could be satisfied throughout the urban area (*27*).

Urban areas may be particularly well suited for ground-source heat pumps because of the urban heat-island effect, a phenomenon in which human activities cause cities and metropolitan areas to be warmer than their surrounding rural areas. The increased anthropogenic heat fluxes into the subsurface of a city result in elevated groundwater temperatures, enhancing the geothermal resource. In Karlsruhe and Cologne, Germany, the anthropogenic heat fluxes could sustainably provide 32 and 9% of the annual residential space heating needs, respectively (*28*). The anthropogenic influence on the subsurface temperature is greater in megacities such as Shanghai, where the existing heat content in the urban aquifer is 22 times the annual heating demand of the city (*29*).

Wind energy

Urban wind energy provides opportunities for not only renewable electrical generation but also ventilation, pollution dispersion, and mitigation of the urban heat-island effect. Urban wind energy has not been widely adopted, largely because of challenges and concerns related to installation space, low and turbulent urban wind-speed characteristics, vibration, noise, safety, shadow flicker (periodic shadows cast by the rotating blades of wind turbines), and aesthetics (*30*). Although modern wind farms typically produce 2 to 3 W/m^2 with horizontal-axis wind turbines (HAWTs), counter-rotating vertical-axis wind turbines (VAWTs) can achieve 30 W/m^2 (*31*). In addition to increased power density, VAWTs offer several advantages

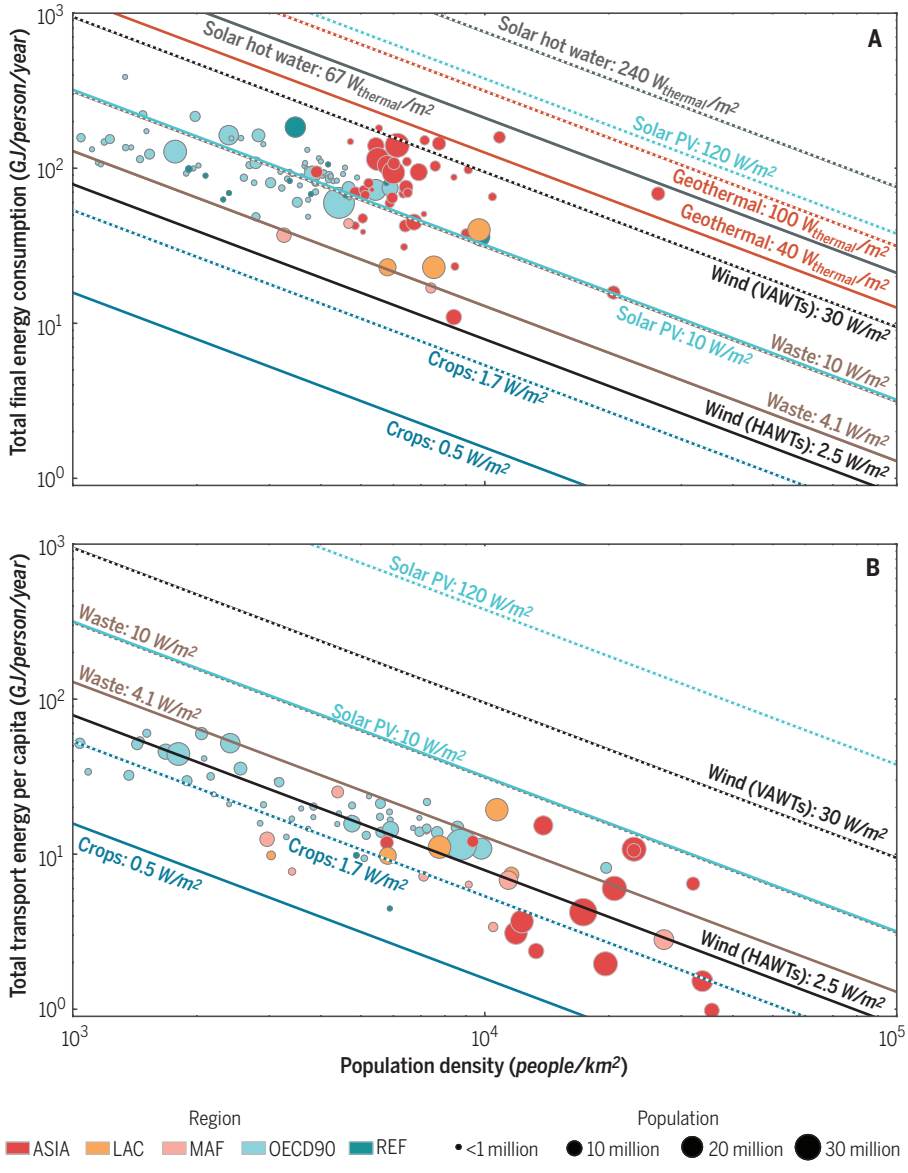


Fig. 1. City-integrated renewable energy potential. (**A**) Potential for renewable sources to satisfy total final urban energy consumption and (**B**) urban transportation energy demands. Solid lines represent typical performance. Dotted lines represent potential performance, based on optimal conditions and technologies currently available in the laboratory. All resources are evaluated based on electric power densities, except geothermal energy and solar hot water. Geothermal energy is evaluated based on thermal power densities for horizontal closed loops buried 1 to 2 m deep at the low end and boreholes at least 150 m deep at the high end. Data on total final energy consumption and region definitions are from (*77*), city population size and density data are from (*78*), and transportation data are from (*79*). "Waste" refers to LFGTE. PV, photovoltaics. Population is indicated by circle sizes, and regions are indicated by circle colors (ASIA, Asia excluding OEDC90 countries; LAC, Latin America and the Caribbean; MAF, the Middle East and Africa; OEDC90, member countries of the Organization for Economic Cooperation and Development as of 1990; REF, Eastern Europe and the former Soviet Union).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

that are particularly relevant to the urban environment. These include lower dependence on wind direction, ability to handle higher turbulence and varied wind speeds, lower manufacturing costs, and decreased impact on birds and aircraft (*30*). Another benefit is that the generator and gear box can be installed at ground level, allowing building-mounted turbines to be more easily serviced (*30*).

Numerous researchers have investigated urban wind-flow characteristics, resulting in estimates of its potential for urban electricity generation. Excluding thermal energy needs, urban wind could provide 33% of residential building electricity needs with HAWTs in urban areas of New Zealand (*32*) and 40% with VAWTs in San Cataldo, Sicily (*33*). There are limited examples of existing urban wind projects. In Bahrain, the World Trade Center twin towers use three vertically arranged HAWTs, providing 11 to 15% of the electrical energy needs (*30*). Although the two VAWTs installed at the Pearl River Tower in Guangzhou, China, provide only 5% of its energy needs (*30*), they offer several design advantages. The curved glass facade of the building funnels air to the VAWTs at speeds of 1.5 to 2.5 times the ambient wind speed, allowing the turbines to generate 15 times more energy than freestanding wind turbines could (*34*). Additionally, by allowing wind to pass through the building, wind-induced forces on the building are reduced, which in turn reduce the quantity of steel and concrete needed to maintain the building's stability (*34*).

Biomass energy

Power densities for biomass energy are highly dependent on the regional climate, because it affects which plants are able to grow locally. Conventional crops have a range of power densities from roughly 0.05 to 1.7 W/m^2; the highest densities come from crops grown in tropical locations with genetic modification, fertilizer, and irrigation (*8*). The ongoing debate over biofuel sustainability and social and environmental justice considerations places this potential energy source in a complex and unsatisfactory position. Direct combustion of urban biomass offers at least a clearer life-cycle path to evaluate than conversion and use of biomass as biofuels. If short-rotation poplar was grown on marginal lands in Boston, for example, it could satisfy 0.6% of the yearly primary energy demand in Massachusetts (*35*).

Given the low power densities, urban agriculture may be better suited for food than for energy. Urban farms help reduce urban heat-island effects, mitigate urban stormwater impacts, and lower the energy needed for food transportation (*36*). A life-cycle analysis of a community farm in South London has shown that urban food supply systems can achieve reductions in greenhouse gas emissions that are potentially larger than those of parks and urban forests (*37*).

Energy from urban waste

Although not entirely renewable, energy from waste could play a key role in sustainable urban energy. Urban residents produce roughly twice the waste of their rural counterparts (*38*). Increased global urbanization is expected to result in increases in municipal solid waste (MSW) (Fig. 2). Typical management strategies include recycling, burning, or landfilling. Although recycling can reduce the life-cycle energy, it may not be economically or energetically realistic for some waste.

Energy from waste may be derived from landfill-gas-to-energy (LFGTE) systems and waste-to-energy (WTE) plants. Landfill gas is generated through the biological degradation of organic materials in MSW, typically consists of methane (50 to 60%) and carbon dioxide (40 to 50%), and can be collected and burned to generate energy. The average LFGTE power density that could be extracted from this gas is estimated to be 4.1 W/m^2, but, if optimized, a theoretical power density of 10 W/m^2 is feasible (*39*). In a WTE plant, MSW is incinerated, generating 0.6 MWh per ton of MSW on average (MWh, megawatt-hour), with the potential to generate up to 1.8 MWh per ton (*40*). Additionally, many WTE plants use cogeneration, providing useful heat. If all of the MSW generated in the United States in 2011 was sent to WTE plants, enough energy would be produced to power and heat 12 and 8% of American households, respectively (*40*). Additionally, carbon-capture systems can be integrated into WTE plants, reducing carbon dioxide emissions by an estimated 90% (*41*).

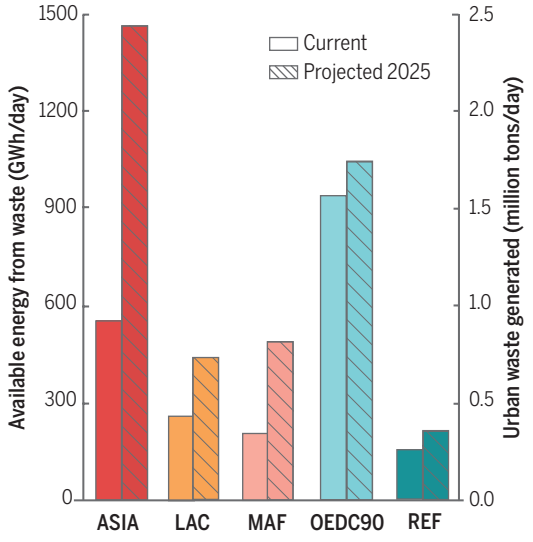


Fig. 2. Estimates of current and projected generation of urban waste. The amount of MSW that is generated daily (solid bars) and the projected daily waste production by 2025 (striped bars) in each region are indicated on the right axis [data are from (*38*)]. If all urban waste were to be burned in WTE plants, the corresponding energy that would be generated, based on an average waste energy value from (*40*) that includes nonrecyclable plastics, is indicated on the left axis.

Opportunities for reducing energy consumption

Two major sectors for reducing energy consumption are buildings and transportation. To limit global climate change to 2°C above preindustrial levels, the greatest global investment is required in the buildings sector (an estimated incremental expenditure of $300 billion/year for 2015–2020) for both retrofitting and constructing new buildings to high energy-efficiency standards (*42*). The next largest investment is required for transportation vehicles, with an estimated additional expenditure of $70 billion/year (*42*).

Building efficiency improvements

Buildings account for 40% of the world's energy consumption and 30% of annual greenhouse gas emissions (*43*). To accommodate the growing urban population, new buildings are needed. Eighty percent of all buildings that will stand in India in 2030 had yet to be constructed as of 2010 (*44*). This new construction creates opportunities not only for energy-efficient and climate-resilient buildings but also for local optimization of urban form. New buildings could be made 70% more efficient than existing buildings through the use of insulated windows, modern gas and oil furnaces, and more efficient air conditioners (*45*). Strategies for zero-energy buildings involve minimizing the energy use of a building, especially for heating and cooling, and adopting renewable energy technologies to meet the remaining minimal energy needs (*46*).

Retrofitting existing buildings saves embodied energy, avoids generating waste from building demolition, and is often more cost-effective than constructing new buildings. Despite these advantages, the rate at which the current building stock is retrofitted is startlingly slow. In the United States, the existing commercial building stock is being retrofitted at a rate of roughly 2.2% a year, with the median energy savings from these retrofits at roughly 11% per building (*47*). Most of these retrofits consist of minimally invasive measures with short payback periods, such as lighting and HVAC (heating, ventilation, and air conditioning system) replacements. Integrated energy-efficiency measures are needed to reach savings as high as 50% (*48*). These measures include upgrades to the building envelope, mechanical systems, lighting and electrical systems, and system controls, as well as changes in tenant behavior (*48*).

Transportation

The five potential metrics for the decarbonization of urban transportation are (i) fuel carbon intensity, (ii) energy intensity, (iii) vehicle occupancy rate, (iv) trip distance, and (v) the number of motorized trips (*49*). These metrics are each discussed below.

Fuel carbon intensity

Fuel carbon intensity is a measure of greenhouse gas emissions per unit energy. To reduce this metric, cleaner energy sources are required and may be achieved through fuel substitution or vehicle electrification. Although fuel substitution with biofuels requires minimal changes to the

Downloaded from http://science.sciencemag.org/ on May 19, 2016

vehicle and fueling infrastructure and reduces global carbon dioxide emissions when evaluated over the life cycle of the fuel, tailpipe carbon dioxide emissions may be comparable to those produced by fossil fuels (*50*). To eliminate tailpipe pollution, vehicles powered by hydrogen or electricity may be better suited for urban transportation. Although roughly 96% of hydrogen production today uses conventional methods with fossil fuels (*51*), there are numerous emerging low-carbon production methods (*52*). However, cost and energy storage remain major obstacles. The electrification of the transportation sector is expected to grow rapidly, with plug-in hybrid vehicles predicted to account for 58% of new light-duty vehicle sales in the United States by 2030 (*53*). This growth requires improvements in low-cost, compact, long-lasting battery technologies and vehicle charging infrastructure. The environmental benefit of electrifying urban transportation will largely depend on the emissions from the electricity generation. The required energy needs for transportation could be achieved, or partially achieved, using city-integrated renewable energy (Fig. 1B).

Energy intensity

Energy intensity is the energy required to move a vehicle one kilometer. Achievable advances in engine technology can improve the fuel economy of automobiles by over 50% and trucks by over 30% (*54*). Although such improvements are possible, the greatest reduction requires a systems approach, taking into account numerous other factors including vehicle lightweighting, accessory load management, powertrain systems optimizations, and aerodynamics. Advances in lightweight materials show particular promise: Passenger vehicle fuel efficiency can be improved by 6 to 8% for each 10% reduction in weight (*55*).

Vehicle occupancy rate

Public transportation and carpooling are commonly used strategies to increase vehicle occupancy rates. A higher urban density increases the attractiveness of public transportation. For reasonable spatial and temporal availability, the urban density threshold for public transportation is estimated to be 5,000 people/km^2 (*56*). Cities with population densities below this threshold are those with the highest percentage of their transportation needs being met by private motorized vehicles (Fig. 3).

Trip distance and number of motorized trips

Trip distances and the numbers of trips taken per year per person depend on the built environment. A city with high density and mixed-use development allows for shorter trips that are more conducive to nonmotorized means, such as walking and biking. The built environment must include safe infrastructure to facilitate nonmotorized transportation. A study of five U.S. cities found that bicycle ridership increased between 21 and 171% within one year of building protected bicycle lanes (*57*). Bicycle use and walking are much more common in other areas of the world. Globally, 37% of trips are nonmotorized, with the greatest use of nonmotorized transport (50%) in the Asia-Pacific region and Africa and the least (8%) in North America (*58*).

Other factors

The urban transportation landscape is changing with increased car-sharing programs and the emergence of self-driving cars. It is still somewhat unclear what effect these will have on energy consumption, because they may result in increased low- or even no-occupancy motorized transportation. One model predicts that shared autonomous vehicles could reduce the number of cars in use by a factor of 10, but the total motorized distance traveled would increase by 11% (*59*).

Challenges of dramatic urban decarbonization

Economic challenges

Some believe that it is too expensive to invest in dramatic decarbonization; however, it may be even more expensive not to. Global infrastructure needs for 2015–2020 are ~$6.7 trillion/year under business-as-usual scenarios, and the incremental costs of low-carbon infrastructure are on the order of −$70 billion/year to $450 billion/year (*60*). Although these are global estimates, port cities with populations over 1 million are particularly vulnerable to infrastructure expenditures related to coastal flooding. Nearly 40 million people and $3 trillion of assets are currently exposed to a 1-in-100-years coastal flood event (*61*). By the 2070s, the exposed population could grow by a factor of 3 and the value of vulnerable assets could increase 10-fold under the combined effects of sea-level rise, subsidence, increased urban populations, and economic growth (*61*). Despite the global need for climate adaptation, investment in adaptation is a small part of the overall urban economy. In a study of ten megacities, investment in climate adaptation was at most 0.33% of a city's gross domestic product (GDPc) and substantially less in developing countries (*62*).

Technical challenges

Some of the main technical challenges for implementing city-integrated renewable energy are the uncertainty and variability in urban energy use and the methods used to account for the associated emissions (*3*). Municipal governments typically measure emissions using a territorial approach, primarily counting emissions that enter the atmosphere within the jurisdiction's geographic boundary (*63*). There is increasing interest in accounting for emissions on a consumption basis, allocating all emissions in global supply chains to the points at which products and services are consumed (*64*). This approach emphasizes the mitigation potential of households and includes a wider range of emission sources, including transportation, energy, food, goods and services, water, waste, and home construction.

Cities differ in both their energy needs and their available energy resources. Hot and cold climates have substantial air conditioning and heating needs, respectively. Cities with more industrial processes typically consume more energy; however, they also have increased potential for district heating and combined heat and power. Urban form also strongly influences transportation energy needs. Available energy resources depend on the status of the electrical grid, location of renewable resources, and socioeconomic conditions. There is considerable variation in energy use not only between cities but also between neighborhoods within the same city. For example, average household carbon footprints in the San Francisco Bay Area vary between neighborhoods according to income, vehicle ownership, household size, home size, carbon intensity of electricity production, population density, and other factors (Fig. 4) (*65*).

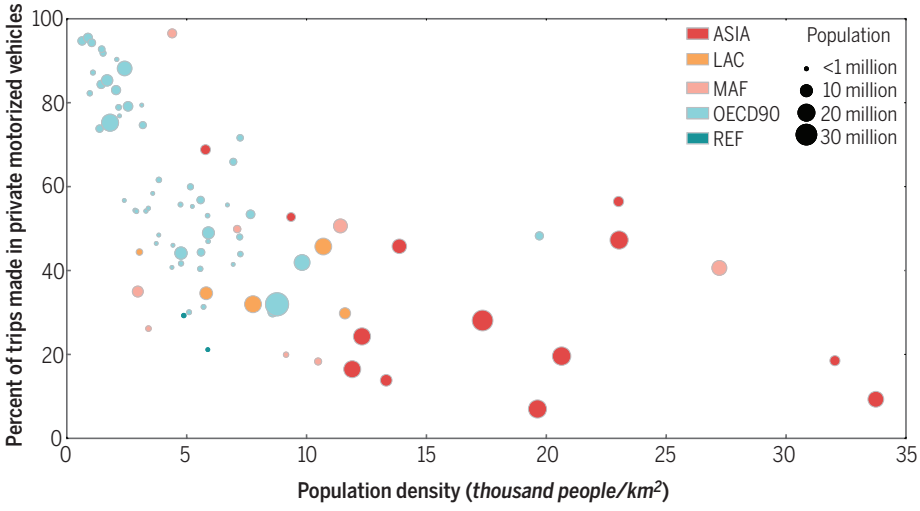


Fig. 3. Relative use of private motorized vehicles. Plotted are the percentages of trips taken using private motorized vehicles, relative to urban population densities. Data are from (*79*).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Given the uncertainty and variability in urban energy systems and urbanization dynamics, a flexible and adaptive solution is necessary. Poor visibility, aging infrastructure, and spatially and temporally variable energy generation from distributed renewable energy sources have made the current electrical grid susceptible to frequent disturbances that can lead to cascading failures. One solution would be a smart grid with integrated energy storage. A smart energy grid should not be limited to electricity; rather, electricity, thermal, and gas grids should be combined and coordinated, emphasizing the role of district heating in future sustainable cities (*66*). Even if a smart gird is well monitored and controlled, the high variability of renewable energy resources requires adequate storage. As the prices of batteries go down and their performances improve, distributed electrical storage shows potential not only in standalone units but also in the electric vehicle fleet. Electric vehicles can provide ancillary services to the grid, such as voltage and frequency regulation, peak power leveraging, and reactive power support to enhance the operational efficiency, secure the electric grid, and reduce power system operating costs (*67*). Electric vehicles couple well with renewable resources. For example, installation of photovoltaics in parking lots in Frauenfeld, Switzerland, could supply 15 to 40% of the future electric vehicle energy demand in that city (*68*), and urban electric vehicle strategies in cities in developing countries could particularly benefit the poor.

Opportunity for thermal storage is growing. Using thermal storage could double the photovoltaic capacity of Shanghai, for example (*69*). Although water remains the most widely used material for sensible heat storage, other methods, such as packed beds and phase change materials, are emerging (*70*). Thermochemical heat storage offers the greatest potential energy density and does not suffer from heat losses during storage, but development efforts are at an early stage.

Behavioral challenges

From the mundane decisions of whether or not to unplug a cellphone charger or take public transportation to more momentous decisions such as installing solar panels on a roof, behavior shapes how we live our lives and the energy choices we make. On an individual level, the effect is minimal, but in aggregate it is substantial. Occupant behavior, for example, can double the energy consumption of a building (*71*). Similarly, driving style can influence a vehicle's fuel consumption by up to 20% (*72*). A lack of information, or selfishness, may lead us to make poor energy choices, even when they are not in our individual and collective best interest (*73*).

Conventional power generation systems, typically located outside of cities and neighborhoods, are out of sight and, therefore, out of the minds of general consumers. This apathy has led to major misunderstandings. For example, only 12% of Americans could pass a basic electricity literacy test (*73*). There is not only a lack of understanding of electricity fundamentals but also of the economic value of energy efficiency and renewable energy generation. For example, most businesses remain uninterested in investing in renewable power, because energy generation is outside their core business goals, despite the increased profit margins that these installations could enable.

A harmful misconception is that freedom and social well-being are best achieved through abundance and excessive consumption. Often, efficiency improvements are outpaced by increased consumption. For example, the floor area of new single-family detached houses in the United States has increased so much since 1978 that single-family housing uses more energy than multifamily housing both per household and per person, despite the efficiency gains achieved through the enforcement of new building codes (*74*). Even WTE programs indirectly encourage increased consumption. Coupled with increased consumption is a strong sense of entitlement. Whereas conventional power plants have typically been located outside of cities and neighborhoods, renewable energy generation is best placed in resource-optimal sites and/or close to the end user. Unfortunately, there has been strong resistance to this, because many people prefer such development to happen elsewhere ("not in my backyard"). Individuals often find their time and comfort to be more important than that of others and choose private vehicle use over public transportation, leading to increased congestion, delays, and inconvenience for the broader community. Not only do individuals need to understand how their behavior affects energy use, society, and the environment but, more importantly, they need to care.

Policy challenges

Although each country suffers its own political challenges, similarities can be found in the treatment of renewable energy and energy-efficiency measures. The research and development funding in these areas is nonexistent in some countries and undersupported in others, slowing innovation. Global research and development funding in renewable energy fell 3% in 2012 and 2% in 2013 (*75*).

Subsidies and incentives are often inconsistent. Unlike those for conventional generators, policies aimed at encouraging renewable power technologies have changed frequently, discouraging widespread adoption of the technologies (*73*). When incentives are removed abruptly, projects can be abandoned before completion and companies can be bankrupted. Additionally, frequently changing subsidies make it increasingly difficult to obtain financing for renewable energy projects. In the United States, policy variation between states deters investment, complicates compliance, discourages interstate cooperation, and encourages tedious and expensive litigation (*76*). An effort to promote renewables has to be sustained, orderly, substantial, predictable, credible, and ramped (*73*).

Policy-makers have focused their efforts on technical challenges. Although there are still

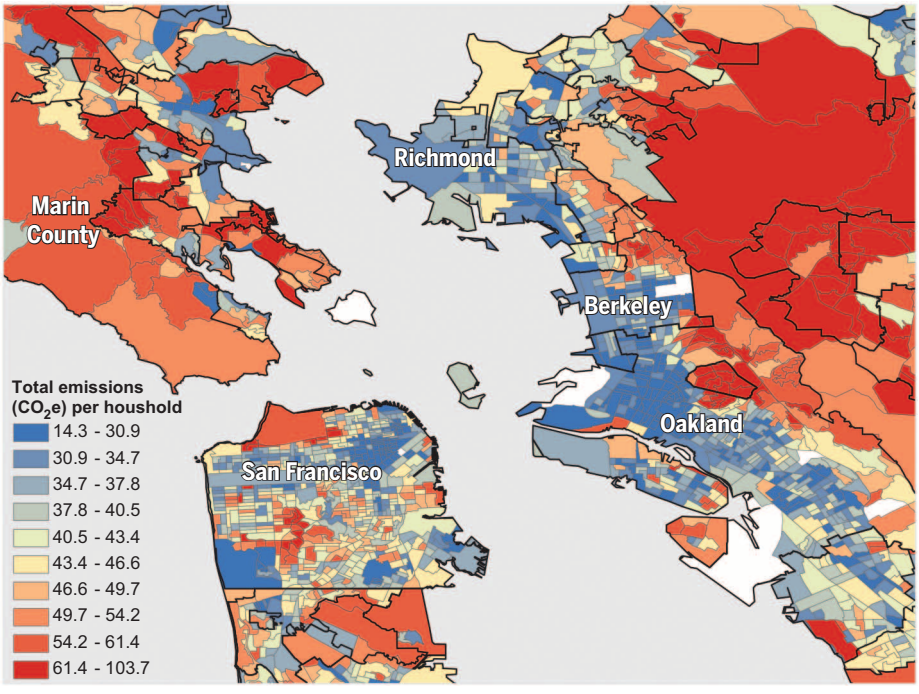


Fig. 4. Neighborhood variation in household carbon footprints. The map shows the average household carbon footprint of census block groups in selected San Francisco Bay Area cities. The color gradient indicates deciles from the lowest to the highest carbon footprint. Data are from the model developed in (*65*) and are available at http://coolclimate.berkeley.edu.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

opportunities for technical improvements, more comprehensive policies are needed to overcome economic and behavioral challenges. Emphasis is needed on government efforts to increase public understanding of energy systems and the environmental impact of behavioral choices.

An innovation agenda for urban sustainability

Achieving a sustainable urban energy system will require a dramatic rethinking of our infrastructure, information systems, and critical social and environmental justice issues. We pose here a number of immediate opportunities to "green" the process of urban evolution, as well as pressing research questions for sustainable cities, both theoretical and practical. We recommend analysis and practice to reach sustainability goals, accompanied by a new suite of data-intensive metrics on which to base planning decisions.

Renewable generation and reductions in consumption

Although city-to-city and regional variations are important to consider, many city governments could immediately (i) encourage energy storage and low-carbon generation at the building level through smart net-metered urban distribution networks; (ii) reclassify electric vehicles as appliances, so that electric vehicle purchases could be amortized into building capital budgets; and (iii) provide intra-city and city-suburban mass transit in the cleanest and most inclusive forms possible. In the near term, the following research questions should be addressed.

Can networked smart buildings themselves become the building blocks of a low-carbon city? Buildings that are designed to create clean energy, store excess generation, and feed this stored energy back into the regional matrix when demand or prices warrant would be key elements of an energy-smart network.

Do cities have an optimal size or density? Urban infrastructure exists at scales that are immediate in terms of buildings and transportation embarkation-disembarkation points but also complex and decentralized in terms of networks that supply food and water to the formal city center and suburbs. Many services that are seen as "citywide" may, in fact, be better suited to regional distribution or even remote management. The trend toward megacities, particularly in Asia, presents opportunities for improvements to infrastructure, such as mass transit, and livable high-density housing. At the same time, walkability and quality of life, as well as the potential for reduced carbon emissions, are all degraded if urbanism forces the mass movement of people, goods, water, and energy. Similarly, although increased population density may reduce transportation energy use, the resulting increased power density demand may not be appropriately met by city-integrated renewable energy, unless multiple sources are combined. These linkages of physical and managerial infrastructure open both theoretical and practical questions of scale in provision and planning.

Overcoming challenges to dramatic decarbonization

Economic, technical, behavioral, and policy challenges have been identified as barriers to dramatic urban decarbonization. However, there are several immediate actions that can be taken to begin to address these: (i) Economically value clean urban environments specifically through the positive environmental justice benefits, and use this valuation to invest in disadvantaged communities. (ii) Standardize carbon and water accounting to improve resource efficiency today and enable a transition to resource and pollution markets over time. (iii) Mix fee-bates (fees associated with polluting vehicles that finance clean vehicle purchases) and congestion pricing to improve urban air quality and reclaim city centers for pedestrians and social spaces. Meanwhile, the following pressing research questions remain.

Can urbanization in emerging economies become a force for sustainability and equality? The unprecedented growth in population and resource demands in large Asian cities has made the urban environment more polluted and more of a resource drain than any other demographic trend over the past four decades. Resource allocation, combined with a new focus on quality of life, should become a means to reverse the trends that have swept Asia. These development trends will ultimately be played out in Africa and elsewhere.

How will we give environmental justice a more central role? Arguably the most central issue in urban sustainability is whether city management can move to a paradigm where environmental justice is not an occasional response to crises of inequality but one where we reap the benefits of proactive and inclusive planning, design, and operation. It is important that renewable energy generation, improved energy-efficiency technologies, and low-carbon transportation are widely accessible, particularly to low-income populations that typically lack ownership of their residential buildings and have the longest commutes. Improvements in walkability, urban parks, and air quality should make cities more livable for all inhabitants.

Ultimately, these and other questions will require a new coordination of technical, social, behavioral, and market innovations. Cities planned around resource demands, or personal automobiles, have been tried and ultimately found lacking in their ability to create sustainable spaces. A wave of innovations in physical form, function, and ideally justice and equity offers a new path toward low-carbon sustainable cities. The challenge is to accelerate innovation and deployment so that cities can substantially reduce greenhouse gas emissions in ways that make them more livable and equitable, not less so.

REFERENCES AND NOTES

1. United Nations Department of Economic and Social Affairs, Population Division, *World Urbanization Prospects: The 2014 Revision, CD-ROM Edition* (United Nations, 2014).
2. J. Simon, *Sustainable Cities: Governing for Urban Innovation* (Palgrave, 2015).
3. A. Chavez, A. Ramaswami, *Energy Policy* **54**, 376–384 (2013).
4. United Nations Human Settlements Programme, *Cities and Climate Change: Policy Directions* (UN, 2011).
5. D. Dodman, *Environ. Urban.* **21**, 185–201 (2009).
6. A. M. Adil, Y. Ko, *Renew. Sustain. Energy Rev.* **57**, 1025–1037 (2016).
7. U. S. Department of Energy, "Solar Power," in *Quadrennial Technology Review 2015* (U.S. Department of Energy, 2015), chap. 4, Technology Assessments.
8. D. J. MacKay, *Sustainable Energy—Without the Hot Air* (UIT Cambridge, 2009).
9. NASA, Surface Meteorology and Solar Energy data set, release 6.0 (NASA, 2014); https://eosweb.larc.nasa.gov/sse/.
10. R. Singh, R. Banerjee, *Sol. Energy* **115**, 589–602 (2015).
11. National Renewable Energy Laboratory, "Best Research-Cell Efficiencies" (National Renewable Energy Laboratory, 2015); www.nrel.gov/ncpv/images/efficiency_chart.jpg.
12. M. Amado, F. Poggi, *Energy Procedia* **48**, 1539–1548 (2014).
13. J. Hofierka, J. Kaňuk, *Renew. Energy* **34**, 2206–2214 (2009).
14. A. Halu, A. Scala, A. Khiyami, M. C. González, *Sci. Adv.* **2**, e1500700 (2016).
15. Z. Wang, W. Yang, F. Qiu, X. Zhang, X. Zhao, *Renew. Sustain. Energy Rev.* **41**, 68–84 (2015).
16. V. Smil, *Power Density: A Key to Understanding Energy Sources and Uses* (The MIT Press, 2015).
17. H. Wei, J. Liu, B. Yang, *Appl. Energy* **126**, 47–55 (2014).
18. C. Castellón, A. Castell, M. Medrano, I. Martorell, L. F. Cabeza, *J. Sol. Energy Eng.* **131**, 041006 (2009).
19. L. T. Terziotti, M. L. Sweet, J. T. McLeskey Jr., *Energy Build.* **45**, 28–31 (2012).
20. S. Freitas, C. Catita, P. Redweik, M. C. Brito, *Renew. Sustain. Energy Rev.* **41**, 915–931 (2015).
21. K. Fath *et al.*, *Sol. Energy* **116**, 357–370 (2015).
22. J. J. Sarralde, D. J. Quinn, D. Wiesmann, K. Steemers, *Renew. Energy* **73**, 10–17 (2015).
23. A. Kagel, D. Bates, K. Gawell, *A Guide to Geothermal Energy and the Environment* (Geothermal Energy Association, 2007).
24. T. J. Reber, K. F. Beckers, J. W. Tester, *Energy Policy* **70**, 30–44 (2014).
25. A. Lopez, B. Roberts, D. Heimiller, N. Blair, G. Porro, *U.S. Renewable Energy Technical Potentials: A GIS-Based Approach* (Technical Report NREL/TP-6A20-51946, National Renewable Energy Laboratory, 2012).
26. J. Zhu *et al.*, *Energy* **93**, 466–483 (2015).
27. Y. Zhang, K. Soga, R. Choudhary, *Géotech. Lett.* **4**, 125–131 (2014).
28. S. A. Benz, P. Bayer, K. Menberg, S. Jung, P. Blum, *Sci. Total Environ.* **524-525**, 427–439 (2015).
29. K. Zhu, P. Blum, G. Ferguson, K.-D. Balke, P. Bayer, *Environ. Res. Lett.* **5**, 044002 (2010).
30. T. F. Ishugah, Y. Li, R. Z. Wang, J. K. Kiplagat, *Renew. Sustain. Energy Rev.* **37**, 613–626 (2014).
31. J. O. Dabiri, *J. Renew. Sustain. Energy* **3**, 043104 (2011).
32. N. Mithraratne, *Energy Build.* **41**, 1013–1018 (2009).
33. A. Gagliano, F. Nocera, F. Patania, A. Capizzi, *Int. J. Energy Environ. Eng.* **4**, 43 (2013).
34. M. A. Kamal, S. Saraswat, *Civ. Eng. Archit.* **2**, 116–120 (2014).
35. M. Saha, M. J. Eckelman, *Appl. Energy* **159**, 540–547 (2015).
36. K. Ackerman *et al.*, *Econ. Soc. Rev.* **45**, 189–206 (2014).
37. M. Kulak, A. Graves, J. Chatterton, *Landsc. Urban Plan.* **111**, 68–78 (2013).
38. D. Hoornweg, P. Bhada-Tata, *What a Waste: A Global Review of Solid Waste Management* (World Bank, 2012).
39. H. R. Amini, D. R. Reinhart, *Waste Manag.* **31**, 2020–2026 (2011).
40. N. J. Themelis, C. Mussche, *2014 Energy and Economic Value of Municipal Solid Waste (MSW), Including Non-Recycled Plastics (NRP), Currently Landfilled in the Fifty States* (Columbia University Earth Engineering Center, 2014); www.americanchemistry.com/Policy/Energy/Energy-Recovery/2014-Update-of-Potential-for-Energy-Recovery-from-Municipal-Solid-Waste-and-Non-Recycled-Plastics.pdf.
41. A. Scott, *Chem. Eng. News* **94**, 18 (2016).
42. International Energy Agency, *Energy Technology Perspectives 2012: Pathways to a Clean Energy System* (International Energy Agency, 2012).
43. United Nations Environment Programme, Sustainable Buildings and Climate Initiative, *Buildings and Climate Change: Summary for Decision-Makers* (United Nations Environment Programme, 2009).
44. McKinsey & Company, *Environmental and Energy Sustainability: An Approach for India* (McKinsey & Company, 2009); www.mckinsey.com/business-functions/sustainability-and-resource-productivity/our-insights/environmental-and-energy-sustainability-an-approach-for-india

Downloaded from http://science.sciencemag.org/ on May 19, 2016

45. K. Kaygusuz, *Renew. Sustain. Energy Rev.* **16**, 1116–1126 (2012).
46. D. H. W. Li, L. Yang, J. C. Lam, *Energy* **54**, 1–10 (2013).
47. V. Olgyay, C. Seruto, in *ASHRAE Transactions* (vol. 116, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, 2010), pp. 244–251.
48. C. Killien, *Deep Green Renovation: Broad Scale Strategies for Achieving Deep Energy Savings in Existing Buildings* (American Institute of Architects, 2011).
49. J. Sager, J. S. Apte, D. M. Lemoine, D. M. Kammen, *Environ. Res. Lett.* **6**, 024018 (2011).
50. T. D. Searchinger, *Environ. Res. Lett.* **5**, 024007 (2010).
51. A. A. AlZaharani, I. Dincer, G. F. Naterer, *Int. J. Hydrogen Energy* **38**, 14505–14511 (2013).
52. I. Dincer, *Int. J. Hydrogen Energy* **37**, 1954–1971 (2012).
53. T. S. Stephens, A. K. Birky, J. Ward, *Vehicle Technologies Program Government Performance and Results Act (GPRA) Report for Fiscal Year 2015* (Argonne National Laboratory, 2014).
54. U.S. Department of Energy, *A Workshop to Identify Research Needs and Impacts in Predictive Simulation for Internal Combustion Engines (PreSICE)* (U.S. Department of Energy, 2011).
55. W. J. Joost, *JOM* **64**, 1032–1038 (2012).
56. A. Grubler, D. Fisk, Eds., *Energizing Sustainable Cities: Assessing Urban Energy* (Routledge, 2013).
57. C. Monsere *et al.*, *Lessons from the Green Lanes: Evaluating Protected Bike Lanes in the U.S.* (NITC-RR-583, National Institute for Transportation and Communities, 2014).
58. A. Aguiléra, J. Grébert, *Int. J. Automot. Technol. Manag.* **14**, 203–216 (2014).
59. D. J. Fagnant, K. M. Kockelman, *Transp. Res. C Emerg. Technol.* **40**, 1–13 (2014).
60. C. Kennedy, J. Corfee-Morlot, *Energy Policy* **59**, 773–783 (2013).
61. S. Hanson *et al.*, *Clim. Change* **104**, 89–111 (2011).
62. L. Georgeson, M. Maslin, M. Poessinouw, S. Howard, *Nat. Clim. Change* 10.1038/nclimate2944 (2016).
63. W. K. Fong *et al.*, *Global Protocol for Community-Scale Greenhouse Gas Emission Inventories: An Accounting and Reporting Standard for Cities* (Greenhouse Gas Protocol, 2014).
64. C. M. Jones, D. M. Kammen, *Environ. Sci. Technol.* **45**, 4088–4095 (2011).
65. C. Jones, D. M. Kammen, *Environ. Sci. Technol.* **48**, 895–902 (2014).
66. H. Lund *et al.*, *Energy* **68**, 1–11 (2014).
67. F. Mwasilu, J. J. Justo, E. K. Kim, T. D. Do, J. W. Jung, *Renew. Sustain. Energy Rev.* **34**, 501–516 (2014).
68. H.-M. Neumann, D. Schär, F. Baumgartner, *Prog. Photovolt. Res. Appl.* **20**, 639–649 (2012).
69. P. Lund, *Energy Convers. Manage.* **63**, 162–172 (2012).
70. P. Tatsidjodoung, N. Le Pierres, L. Luo, *Renew. Sustain. Energy Rev.* **18**, 327–349 (2013).
71. N. Baker, K. Steemers, *Energy and Environment in Architecture: A Technical Design Guide* (E & FN Spon, 2000).
72. J. Gonder, M. Earleywine, W. Sparks, *SAE Int. J. Passeng. Cars Electron. Electr. Syst.*, **5** 450–461 (2012).
73. B. K. Sovacool, *Energy Policy* **37**, 4500–4513 (2009).
74. Y. Ko, *J. Plann. Lit.* **28**, 327–351 (2013).
75. Frankfurt School–United Nations Environment Programme Centre, Bloomberg New Energy Finance, *Global Trends in Renewable Energy Investment 2014* (Frankfurt School of Finance and Management, 2014); http://fs-unep-centre.org/system/files/globaltrendsreport2014.pdf.
76. B. K. Sovacool, C. Cooper, *Electr. J.* **20**, 48–61 (2007).
77. Global Energy Assessment, *Global Energy Assessment - Toward a Sustainable Future* (Cambridge Univ. Press and the International Institute for Applied Systems Analysis, 2012).
78. Demographia, *Demographia World Urban Areas: 11th Annual Edition* (Demographia, 2015).
79. International Association of Public Transport (UITP), Millennium Cities Database for Sustainable Mobility (UITP, 2001).

ACKNOWLEDGMENTS

We thank F. Creutzig, C. Jones, B. Gould, J. Sager, J. Apte, and D. Lemoine for useful discussions. This research was supported by the Karsten Family Foundation and the Zaffaroni Foundation through their support of the Renewable and Appropriate Energy Laboratory (to D.M.K.) and by an Energy Efficiency and Renewable Energy Postdoctoral Research Award from the U.S. Department of Energy (to D.A.S.).

10.1126/science.aad9302

REVIEW

Emerging solutions to the water challenges of an urbanizing world

Tove A. Larsen,[1]* Sabine Hoffmann,[1] Christoph Lüthi,[1] Bernhard Truffer,[1,2] Max Maurer[1,3]

The top priorities for urban water sustainability include the provision of safe drinking water, wastewater handling for public health, and protection against flooding. However, rapidly aging infrastructure, population growth, and increasing urbanization call into question current urban water management strategies, especially in the fast-growing urban areas in Asia and Africa. We review innovative approaches in urban water management with the potential to provide locally adapted, resource-efficient alternative solutions. Promising examples include new concepts for stormwater drainage, increased water productivity, distributed or on-site treatment of wastewater, source separation of human waste, and institutional and organizational reforms. We conclude that there is an urgent need for major transdisciplinary efforts in research, policy, and practice to develop alternatives with implications for cities and aquatic ecosystems alike.

Water has become a challenge of global dimensions (*1*). Many researchers and policy-makers have focused on large water users such as agriculture, the impact of future droughts on food security, and the quality of receiving water, giving little thought to the ability of cities to handle the urban water cycle adequately (*2*). Urban water management (UWM) has recently gained more attention, in part due to the comprehensive Sustainable Development Goal on Water (SDG-6) (*3*). The generally accepted approach to UWM builds on a well-established socio-technical system that, at least in the more affluent part of the world, has solved most of the water and hygiene-related problems afflicting cities at the turn of the 20th century. The core centralized services are the provision of safe drinking water, urban hygiene (for the purpose of public health), and protection against flooding (*4*), complemented by water pollution control.

UWM in high-income countries

The UWM system relies on investment-intensive, usually underground, pipe networks that provide single-quality drinking water and evacuate stormwater and wastewater. In many places, reservoirs and long-distance water conveyance systems compensate for inadequate local water resources. In addition, water and wastewater treatment plants provide an interface to the aquatic environment, treating raw water for drinking-water purposes and wastewater for water pollution control. Indeed, the main components of the UWM system have been considered the most important medical advance since 1840 (*5*) and still serve as the prevailing model for prospering cities worldwide (*6*). An additional important infrastructure—besides water supply and wastewater removal and treatment—is the stormwater drainage system. On a local level, the built environment has a strong influence on the natural hydrological characteristics of a catchment. A substantial part of the global urban area of 658,760 km^2 (*7*) comprises impermeable surfaces. This leads to a higher surface runoff and a faster response time to the rain event (*8*). Without adequate drainage infrastructure, unwanted urban flooding events will occur.

In the process of urban water use, waste is produced in the form of wastewater. However, wastewater also contains important resources, including water, organic matter, heat, and nutrients such as phosphorus and nitrogen (Table 1). For example, the amount of nitrogen passing through the human metabolism on a global scale and therefore potentially ending up in wastewater is on a par with major components of the nitrogen cycle. For a population of 9 billion, nitrogen in wastewater would be of the same order of size as the anthropogenic production of 35 Mt of reactive nitrogen per year (about 25% of the present production) suggested as the upper boundary for a "safe operating space" of humanity (*9*). In view of the large losses of nitrogen in agricultural production (*10*), the world can only be kept within the suggested boundary with a dramatic increase in nitrogen recycling from wastewater.

The current UWM approach has worked so well because it delivers its main services securely at a good quality to a majority of people in a region. Its institutional side is characterized by planning and investment processes traditionally delegated to municipal water authorities. These actors follow well-formulated regulatory codes in their operations and rely primarily on highly specialized technical expertise.

The downsides of the current UWM system are its strong dependence on large quantities of water (Fig. 1), high investment costs, and a need for stable institutions, as well as long planning horizons and inefficient use of resources. Whereas most of these disadvantages have different implications depending on the context, inefficient use of resources is a global issue. Despite the high amounts of energy in wastewater (Table 1), wastewater management is a net consumer of energy, and recycling of nitrogen is only possible to a very

[1]Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland. [2]Faculty of Geosciences, University of Utrecht, Heidelberglaan 2, NL-3584 CS, Utrecht, Netherlands. [3]Institute of Environmental Engineering, ETH Zurich, 8093 Zurich, Switzerland.
*Corresponding author. Email: tove.larsen@eawag.ch

Downloaded from http://science.sciencemag.org/ on May 19, 2016

small degree within the existing system (*11*). Additionally, the substantial investments in infrastructure required to move the large amounts of water in and out of cities and treating the resulting wastewater are of interest beyond the local setting. The most transparent report from the Organization for Economic Cooperation and Development (OECD) (*12*) calculates a global investment need of US$772 billion year^{-1} (or about 1% of gross domestic product) by 2015 for the OECD countries and Brazil, Russia, India, and China. However, other literature estimates are highly variable, from US$190 billion year^{-1} to US$1037 billion year^{-1} (*13*, *14*). Additional US$114 billion year^{-1} [US$71 billion to US$166 billion year^{-1} (*15*)] are required to achieve universal access to safe drinking water and adequate sanitation for all by 2030. Assuming that the investments for water supply and wastewater management are similar in magnitude, the total water infrastructure value for a connected global population of 9 billion people would amount to about US$60 trillion (*16*).

Today, UWM is incurring increasing economic, social, and environmental costs, even in countries with a long tradition of successful practices. This is a consequence of aging built infrastructures, increasing urbanization, emerging contaminants, competitive water uses, and measures to mitigate the effects of climate change (e.g., water-saving measures). Furthermore, public utilities have often missed out on charging full-cost tariffs and are increasingly confronted with a backlog of investments (*13*). These recent developments question whether and in what form the existing UWM system can still be the best solution for the world as it has been since the beginning of the 20th century (*13*).

Limitations for the global diffusion of centralized UWM

Whereas the need for reform in industrialized countries might still be a matter of debate, the proliferation of current UWM practices to the rest of the world is certainly riddled with major problems. Although reliable information on sewers (Fig. 2A) and treatment plants is scarce for Africa and Asia, there is general agreement that connection rates remain very low [an estimated connection rate of 14% for Africa and 18% for Asia in 2000 (*17*)], and the overall treatment of the collected wastewater remains highly insufficient, even in capital cities (*18*). For Latin America, connection rates are higher, but only 15% of municipal wastewaters are treated (*19*).

This backlog is compounded with the current unprecedented global population growth rate. A major part of this growth is projected to take place in the cities of Africa and Asia, including many countries with a low human development index (HDI) as well as pronounced and increasing water scarcity (Fig. 2B). Small and medium-sized towns will bear the brunt of this future urbanization growth (*20*), notably in the provision of access to safe drinking water (Fig. 2C) and sewers (Fig. 2A). High urban growth rates lead to high planning uncertainty, especially in areas with a low HDI and low institutional reliability (*21*). In combination with high discount rates, indicating a strong preference for spending money on immediate benefits rather than on long-term investments (*22*), there is little willingness and ability to embark on large-scale infrastructure projects. A modeling study based on past investment patterns (*17*) estimated that even on the most optimistic assumptions, only 36% of the African population and 44% of the Asian population will be connected to a sewer network by 2050. The implementation of well-functioning, nutrient-eliminating wastewater treatment plants depends on the previous construction of sewers and often involves substantial delays. A case in point illustrating the enormous resources and sector investments needed has become apparent with the Swachh Bharat national campaign by the government of India to achieve a turnaround in India's poorly served cities and towns (*23*).

Apart from the lack of capital, there are also other, more general reasons why conventional UWM is not the best solution for rapidly growing

Table 1. Resources in wastewater. For nutrients and water, global averages are given. No global information is available concerning warm water and organic matter in wastewater. Local loads depend inter alia on nutritional status, household devices, water availability, and habits.

Water (liters person^{-1} day^{-1})		
Domestic	184	Global average (*69*)
Industrial	300	Industrial global average (*69*)
Energy (MJ person^{-1} year^{-1})		
Heat contained in warm water	2800	Typical European country (*11*)
Chemical energy contained in organic matter	540	Typical European country (*11*)
Chemical energy "embedded" in N and P	180	Global average, year 2000 (*11*, *17*)
Nutrients from human metabolism (g person^{-1} day^{-1})		
Nitrogen (N)	10	Global average, year 2000 (*17*)
Phosphorus (P)	2	Global average (*17*)

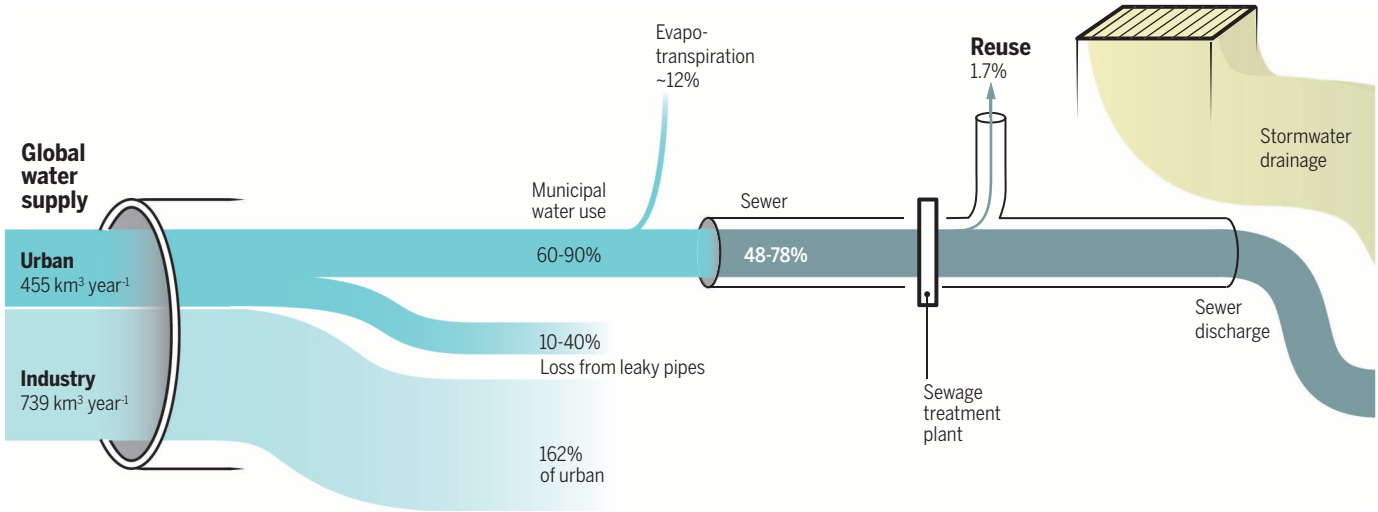


Fig. 1. The global urban water cycle. According to country-specific data from FAO (*69*), the global municipal water withdrawal is estimated to be 454.8 × 10^9 m^3 year^{-1} (184 liters person^{-1} day^{-1}), and 738.8 × 10^9 m^3 year^{-1} (300 liters person^{-1} day^{-1}) for industrial use. This corresponds to 12% and 19%, respectively, of the total global water withdrawal. Shiklomanov (*74*) estimates global urban evapotranspiration to be around 12%. Typical water "losses" due to leaky supply systems are between 10 and 40% (*69*, *75*). Globally, around 1.7% [7.7 × 10^9 m^3 year^{-1}; from (*36*)] of the municipal water supply is reused in this way—mostly for irrigation.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

cities. Network-based infrastructures are designed and built for "final design performance." High growth rates therefore impose large idle capacities during the early life of the infrastructure, with correspondingly high per-user costs (*22*). Furthermore, high planning uncertainty also increases the risk of sunk costs if the expectations of city growth are not fulfilled or if not enough water is available for the correct functioning of the sewers. The lack of stable energy supplies, spare parts, and know-how for reliable operation are additional factors that limit the expansion of centralized systems (*24*). As a special case, the improvement of sanitation conditions in informal settlements in low- and middle-income countries has proved difficult because of disabling institutional environments, a lack of secure tenure and rule of law, which often prevent private or public investments in infrastructure (*25*, *26*). In view of the expected increase in the populations of such informal settlements from today's 1 billion to 2 billion in 2030 (*27*), this is a quantitatively important situation with dramatic consequences not only for the inhabitants themselves, but also for the urban and natural environment.

On the basis of those facts, we conclude that for the areas with the highest rate of urbanization, there is an urgent need to develop more cost-effective and resource-efficient systems that deliver the desired water services of UWM without the prohibiting constraints of the conventional centralized system.

Alternative solutions to conventional UWM

As the currently dominant conventional approach to UWM is unlikely to meet the challenges of an increasingly globalizing world [see also (*28*–*30*)], a shift toward a "new paradigm" is required (*31*). Three of the more salient candidates for new water paradigms that substantially depart from the present strategy are integrated water resources management (IWRM), adaptive management (AM), and ecosystem-based approaches (EBAs) (*31*). A shared feature of these reform agendas is that they give primacy to organizational and institutional reforms in order to orient water management toward providing sustainable water services rather than merely delivering quantities of water. These approaches have gained traction in science and policy-making in recent years (*31*). In particular, they have inspired new policy approaches such as the European Framework Directive, or kindred approaches in countries like Australia, South Africa, and China. However, the impacts on real water systems have been limited (*30*, *31*). One reason is that the routines and practices of water professionals are not directly determined by planning discourses or governmental mission statements. Rather, they are oriented to technical expertise and the professional cultures that have developed over decades in line with the dominant UWM systems.

This debate tells us that it is not enough to hope for technological breakthroughs or to believe in the wisdom of more inclusive governance arrangements alone. Rather, the joint development of new institutional conditions and technological designs is needed to ensure fundamental improvements [so-called transitions in socio-technical regimes (*32*)]. What is at stake, therefore, is a mainstreaming of novel system alternatives in the UWM sector that would respond to the challenges noted above. A number of technological and institutional approaches look promising. They represent potential foci of future innovation efforts. However, they are nonexclusive, and there are many overlaps and potential synergies between them.

Stormwater drainage

Urbanization means that not only the population but also the area in need of drainage increases. Estimates for 2000 to 2030 indicate an enlargement

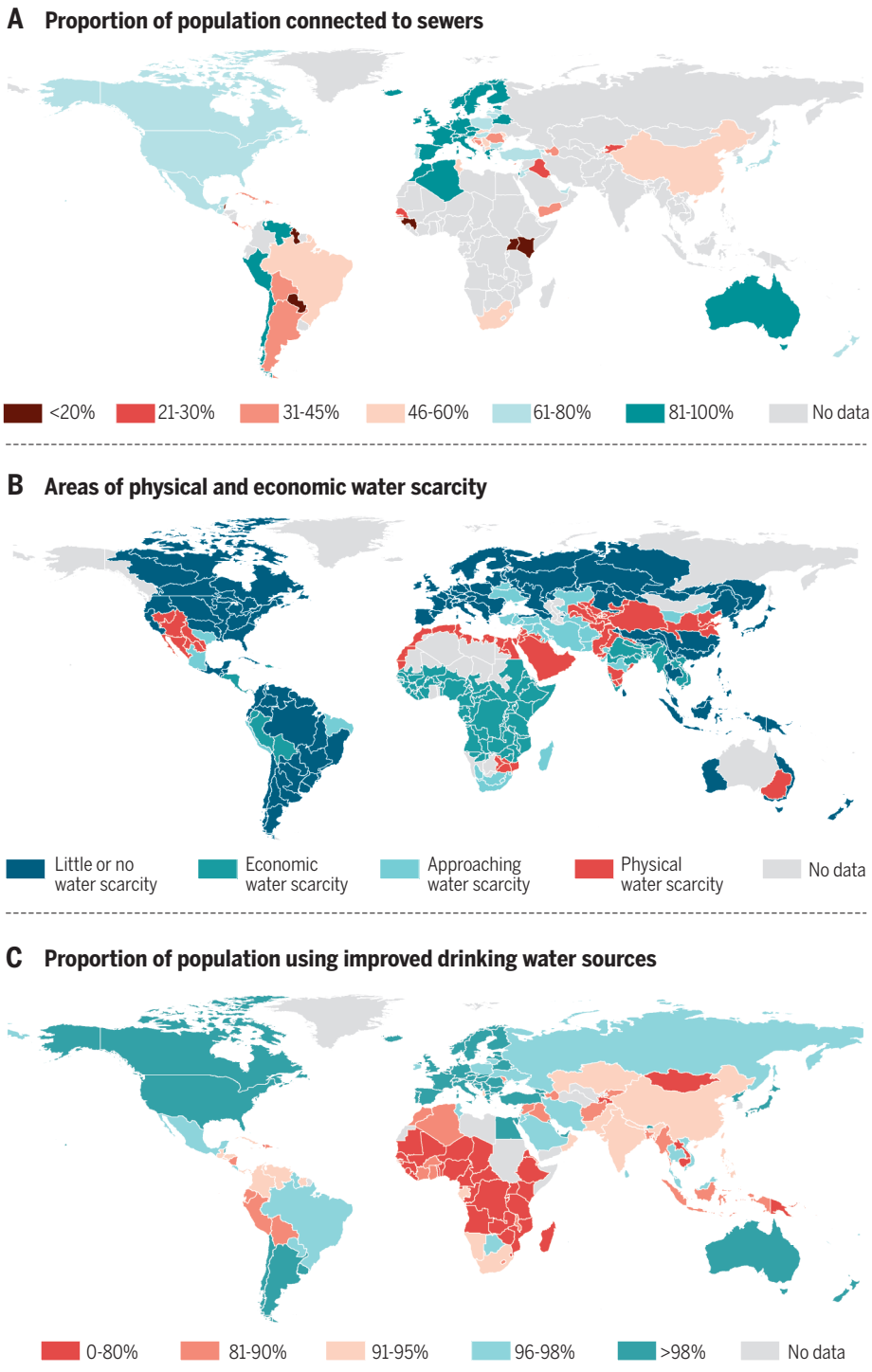


Fig. 2. Important global challenges of UWM. (**A**) Proportion of population connected to sewers (*76*). (**B**) Areas of physical and economic water scarcity [data from International Water Management Institute (IWMI); updated in 2015; map reproduced with permission from IWMI (*77*)]. (**C**) Proportion of population using improved drinking-water sources (*78*).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

of the global urban areas by an additional 60 to 200% (*33*). It is therefore no surprise that the limits of the conventional UWM approach were first recognized in stormwater drainage. Concepts such as sustainable urban drainage systems (SUDS), low-impact urban design and development (LIUDD), water-sensitive urban design (WSUD), and green infrastructures (GI) appeared in the scientific literature toward the end of the 20th century (*34*). The primary goal of these concepts is to maintain or reintroduce a more natural state of the urban hydrological catchment, to reduce the impact of stormwater drainage on the aquatic environment, and to reduce flood risk. All these concepts introduce a strong element of decentralized measures and emphasize the importance of long-term planning.

Increasing water productivity

This practice helps to reduce net water consumption and utilize the available water more efficiently. Three main strategies designed to increase water productivity are reducing water waste, down-cycling or reuse of lower-quality water, and regenerating high-quality water from used water (*35*). In the last two strategies, the collected wastewater is in most cases treated in wastewater treatment plants to the desired quality, making it fit for reuse. Globally, around 1.7% [7.7×10^9 m^3 $year^{-1}$; from (*36*)] of the municipal water supply is reused in this way—mostly for irrigation (Fig. 1). In California, 61% of the reused 0.8×10^9 m^3 $year^{-1}$ of water is applied for irrigation, and the rest is mainly used for recharging the groundwater and for the targeted alimentation of surface water (*36*). There are comparatively few large-scale direct potable reuse schemes, and these compete in terms of energy consumption and costs with desalination technology. The advantage of this approach is its compatibility with conventional network-based UWM. However, it requires additional infrastructure for treatment and redistribution, thus increasing the energetic, financial, and institutional burden. More innovative solutions are found in Table 2.

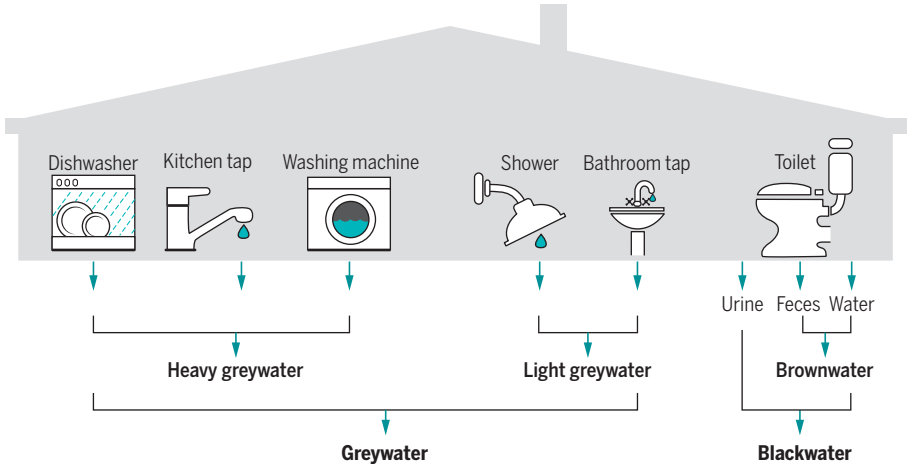


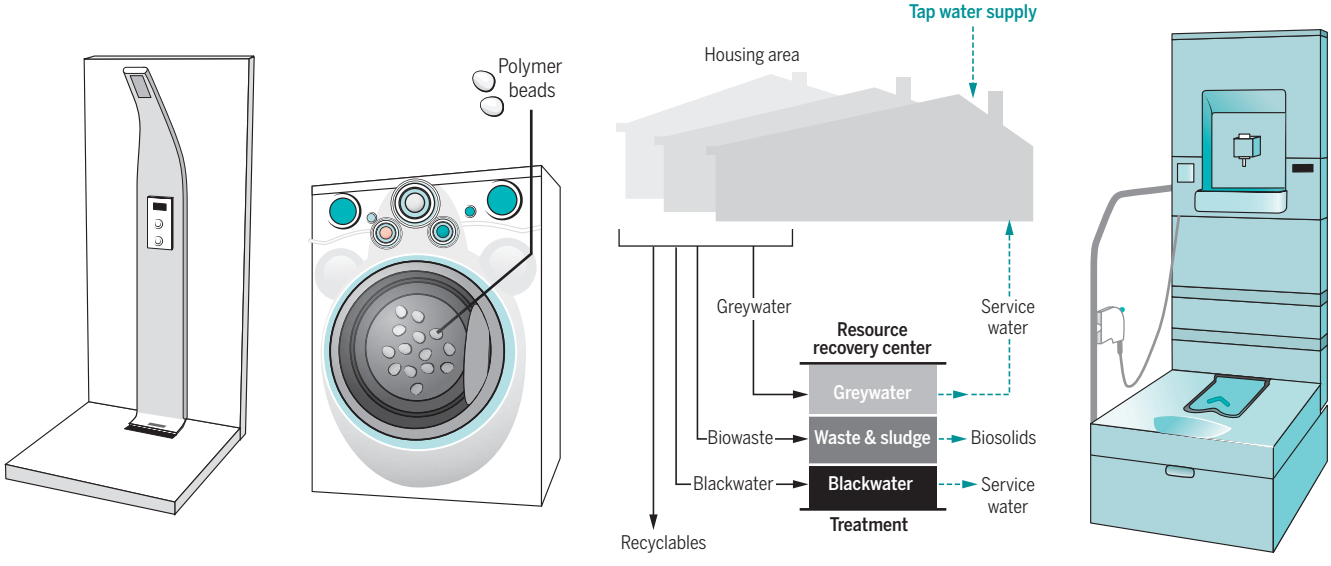
Fig. 3. With source separation of wastewater in the household, new types of wastewater can be constructed for optimal treatment. It is even possible to include treatment and recycling processes in a single device. This offers totally new perspectives for mass-produced, consumer-friendly wastewater treatment technology (for examples, see Table 2).

Source separation of waste

Separating wastewater streams as early as possible alleviates resource recovery and/or facilitates the treatment process. This can take place at the household level, but also at the level of a single household device (Fig. 3). In particular, the separation of greywater promises new ways of reusing water. Compared with wastewater reuse, greywater recovery involves smaller hygienic concerns, has a reduced "yuck" factor, and demands less treatment effort. Especially in arid

Table 2. Examples of emerging solutions to UWM challenges.

Increasing water productivity		Distributed treatment	Source separation of waste
Reuse	Substitution		
Recycling shower (*70*)	Waterless washing machine (*71*)	Distributed treatment of waste at district level (*72*)	Blue Diversion Toilet (*73*)



Downloaded from http://science.sciencemag.org/ on May 19, 2016

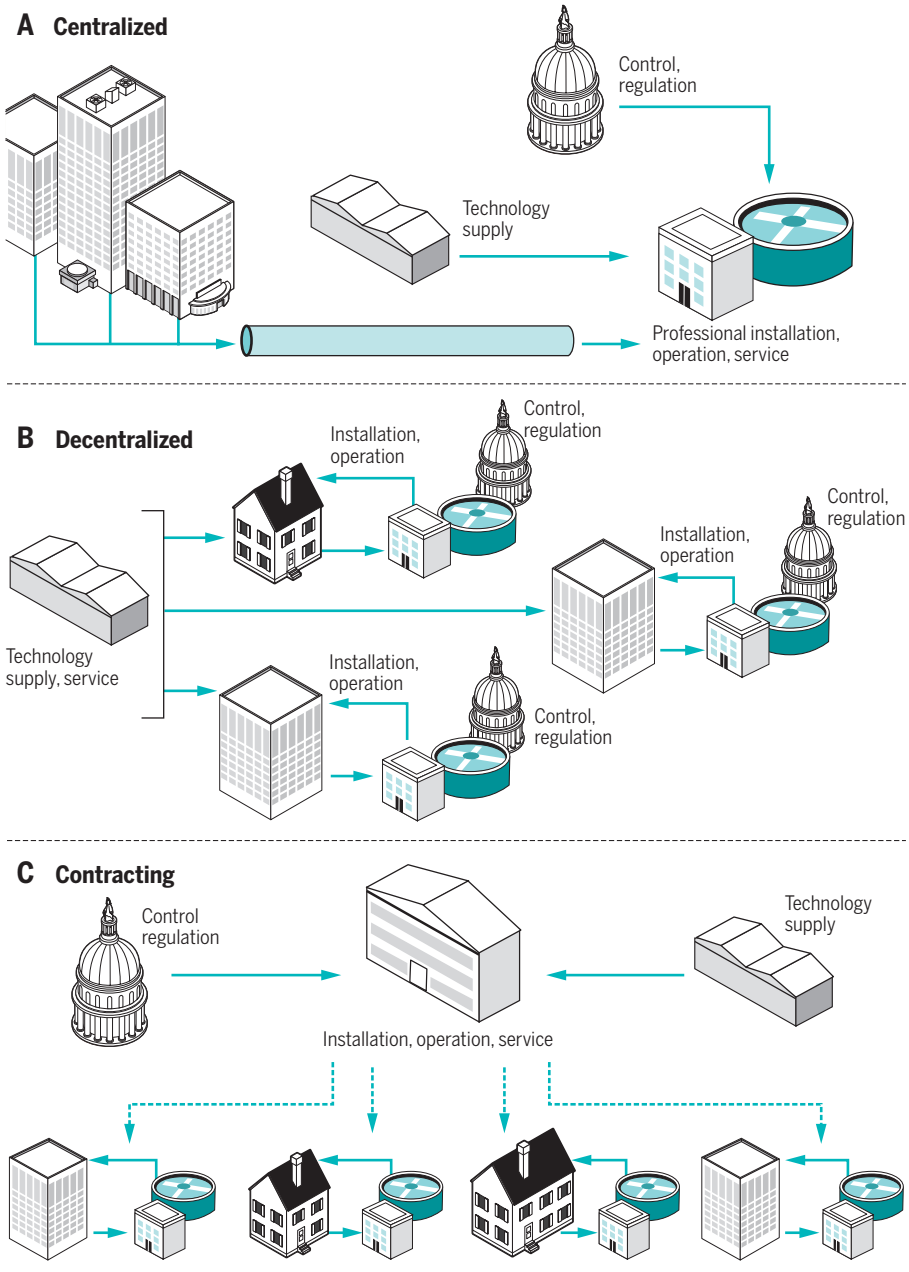


Fig. 4. Alternatives for management of centralized and decentralized wastewater treatment plants. (**A**) Fully centralized, (**B**) fully decentralized, and (**C**) contracting scheme for decentralized technology with centralized operation.

regions, greywater provides greater potential potable water savings than local stormwater capture because it provides a steady source of water during periods with little or no rainfall (*36*, *37*). Not only water, but also energy and nutrients are more easily recovered from source-separated wastewater: energy from hot water (*38*) and from feces (*39*) as well as nutrients from urine (*40*). Some well-documented examples of source separation include the 40 million domestic biogas reactors in China (*41*) or the almost 100,000 urine-diverting dry toilets in peri-urban areas of eThekwini, South Africa (*42*). Although most examples of source separation are still found in areas without existing sewers, there are good reasons to apply this concept more broadly in order to render UWM more resource efficient (*43*).

Distributed or on-site treatments

Decentralized systems have the advantage that they can be installed in the short term when needed, thereby reducing the requirement for large-scale investment in sewers and centralized wastewater treatment plants. Moreover, they allow the local reuse of water and therefore increase water productivity (Table 2). Also, the argument of lower costs of centralized systems due to economies of scale at the treatment plant has become much less persuasive in recent years (*44*). Over the last few decades, large numbers of decentralized wastewater treatment plants have been installed worldwide (*45*). Their performance has been judged as mediocre at best, with some authors stating that their failures are not primarily due to immature technology but rather to weak or unsuitable organizational models and institutional setups (*37*). Whereas this may be partially true (see next section), many small-scale technologies are little more than scaled-down conventional treatment plants, originally developed with a very different set of requirements than we would imagine for small-scale technology.

Institutional and organizational reforms

Multiple efforts on this front have been advocated in the sector since the late 1980s (*46*). Great hopes were originally placed in a stronger involvement by private actors in service delivery and infrastructure investment. However, evidence about the success of these reforms is mixed (*47*–*50*). Research has shown that public and private organizations can be equally effective and efficient in strategic planning (*51*), and that success of reforms depends on how well competences of utilities and institutional context conditions are aligned with the technological characteristics of the sector (*52*, *53*). For the case of distributed systems, this means that innovation in organizational and regulatory models is badly needed. The centralized management approach, combined with centralized treatment technology, has been considered the most cost-efficient organizational and regulatory mode for most of the 20th century (*44*) (Fig. 4A). This is especially true when it is compared to a fully decentralized approach, where end users are responsible for operating their treatment plants and regulators have to oversee innumerable individual installments (Fig. 4B). Recent advances in sensor and communication technology, however, enable new contracting schemes (Fig. 4C) where central operators can monitor large fleets of individual appliances and thereby guarantee very good performance in terms of effluent quality and convenience for the end user (*54*).

Ways forward for policy and research

Overall, any promising approach to solving the urban water challenges requires innovation and development processes in almost all technical, organizational, and institutional dimensions. However, the UWM sector seems to be very poorly prepared to deal with innovations (*55*, *56*). The legacy of the network logic, slow renewal cycles, and high long-term investments lead to risk aversion with respect to novel technologies. Little competence in innovation management has consequently been built up in most water utilities (*53*). Policy-makers and end users have also been reluctant to accept disruptive changes, so that increased efforts are necessary to mainstream innovations [compare direct potable reuse (*57*)]. Furthermore, a major international research and policy effort is needed in the field of sustainable water futures (*1*). This represents a policy challenge on par with other processes of global change.

On-site treatment and source separation, especially in combination, open up the potential for locally adapted water services and the recovery and reuse of valuable resources. A good example

Downloaded from http://science.sciencemag.org/ on May 19, 2016

is the "Reinvent the Toilet Challenge" of the Bill & Melinda Gates Foundation, calling for the next generation of on-site wastewater treatment technology. This call aimed to stimulate the academic community to develop an innovative toilet for the urban poor with no requirement for conventional network infrastructures (water pipes, sewers, electricity networks), while simultaneously promoting maximal resource recovery and zero emissions (*58*). By concentrating on the toilet alone, this initiative essentially broke with the convention of treating domestic wastewater as a single stream and suggested that source separation in on-site settings is an attractive way forward (Table 2). The Blue Diversion Toilet, a urine-diverting dry toilet with a separate water cycle developed within this program, exemplifies the high potential of source separation and resource recovery in a single household device (Fig. 3). By separating human waste from the water cycle, energy-efficient on-site treatment and recycling of the water become possible (*59*), and the separate collection of urine and feces allows for a large number of processes to recover nutrients and energy (*60*). Even on-site treatment of urine and feces may be conceivable, opening the market for these technologies beyond urban slums (*61*). This is, however, still in the development stage only.

The aim of such efforts is not only to solve technical problems, and it must involve more parties than scientists and engineers alone. The complexity, ambiguity, and uncertainty of such radical innovation processes require a approach that transcends the boundaries between different disciplines and bridges knowledge generated by research, policy, and practice (*62*). Longer-term industrial transformation policy programs must be envisaged to meet the challenges of UWM, especially as the sector is characterized by large and long-lived infrastructure. We can learn here from recent experiences in sectors such as agriculture, defense, health, or energy (*63*). First, successful transitions need long-term support for basic research on a broad variety of alternative solutions. For UWM, this means that we should not aim for one single superior alternative to replace the established technological paradigm. Second, public policies should complement and support innovations made by private actors but without curtailing competition between alternative solutions. UWM can learn much from how this complementarity played out in the field of renewable energy (*64*, *65*). Finally, technologies should be tested in a broad variety of experimental settings to ensure robustness, cost-effectiveness, social acceptance, and the wide applicability of alternative solutions.

The present rapid urbanization in areas with water scarcity and/or missing or aging urban water infrastructure is an immense challenge, as well as a formidable chance for developing next-generation technologies and management structures. In Australia, increasing water scarcity has led to large-scale academic efforts to develop the alternative resource of stormwater (*66*), whereas the aging infrastructure of the United States has led to similar efforts in the area of infrastructure management (*67*). Singapore pursues centralized recovery of wastewater to reduce dependence on Malaysian water resources (*36*), and in China, small-scale membrane bioreactors are proliferating to provide enough water for its growing cities (*68*). There will be no one-size-fits-all solution, but with the immense challenges for UWM ahead of us, it will be important to accelerate research efforts and to profit from the lessons learned about successful innovations in other sectors.

REFERENCES AND NOTES

1. C. Pahl-Wostl *et al.*, *Curr. Opin. Environ. Sustain.* **5**, 708–714 (2013).
2. J. Rockström *et al.*, *Ecohydrology* **7**, 1249–1261 (2014).
3. United Nations, Sustainable Development Knowledge Platform, https://sustainabledevelopment.un.org/sdg6.
4. T. A. Larsen, W. Gujer, *Water Sci. Technol.* **35**, 3–10 (1997).
5. A. Ferriman, *BMJ* **334**, 111 (2007).
6. World Bank, *World Development Report 2009: Reshaping Economic Geography*. World Development Reports (World Bank, Washington, DC, 2009).
7. A. Schneider, M. A. Friedl, D. Potere, *Environ. Res. Lett.* **4**, 044003 (2009).
8. A. Mejía, E. Daly, F. Rossel, T. Jovanovic, J. Gironas, *Water Resour. Res.* **50**, 1984–2001 (2014).
9. J. Rockström *et al.*, *Nature* **461**, 472–475 (2009).
10. B. Eickhout, A. F. Bouwman, H. van Zeijts, *Agric. Ecosyst. Environ.* **116**, 4–14 (2006).
11. T. A. Larsen, *Water Res.* **87**, 513–521 (2015).
12. R. Ashley, A. Cashman, in *Infrastructure to 2030: Telecom, Land Transport, Water and Electricity,* OECD, Ed. (OECD, Paris, 2006).
13. OECD, *Water and Cities: Ensuring Sustainable Futures, OECD Studies on Water* (OECD, Paris, 2015).
14. M. Fay, T. Yepes, "Investing in infrastructure: what is needed from 2000 to 2010?" *Policy Research Working Papers* (World Bank, Washington, DC, 2003).
15. G. Hutton, M. Varughese, "The Costs of Meeting the 2030 Sustainable Development Goal Targets on Drinking Water, Sanitation, and Hygiene," *Water and Sanitation Program: Technical Paper* (World Bank, Washington, DC, 2016).
16. M. Maurer, D. Rothenberger, T. A. Larsen, *Water Sci. Technol.* **5**, 145–154 (2006).
17. G. Van Drecht, A. F. Bouwman, J. Harrison, J. M. Knoop, *Global Biogeochem. Cycles* **23**, GB0A03 (2009).
18. A. Peal, B. Evans, I. Blackett, P. Hawkins, C. Heymans, *J. Water Sanit. Hyg. Dev.* **4**, 563–575 (2014).
19. A. Noyola, A. Padilla-Rivera, J. M. Morgan-Sagastume, L. P. Güereca, F. Hernández-Padilla, *Water* **40**, 926–932 (2012).
20. UN-Habitat, *State of the World's Cities 2006/2007* (UN-Habitat, Nairobi, Kenya, 2006).
21. UN-Habitat, *Slums of the World: The face of urban poverty in the new millenium?* Working paper (Nairobi, Kenya, 2003).
22. M. Maurer, in *Source Separation and Decentralization for Wastewater Management,* T. A. Larsen, K. M. Udert, J. Lienert, Eds. (IWA, London, 2013), pp. 85–100.
23. K. Wankhade, *Environ. Urban.* **27**, 555–572 (2015).
24. C. Lüthi, A. Panesar, in *Source Separation and Decentralization for Wastewater Management,* T. A. Larsen, K. M. Udert, J. Lienert, Eds. (IWA, London, 2013), pp. 455–462.
25. M. Jacobsen, M. Webster, K. Vairavamoorthy, "The Future of Water in African Cities: Why Waste Water?" *Directions in Development* (World Bank, Washington, DC, 2012).
26. M. O'Keefe, C. Lüthi, I. K. Tumwebaze, R. Tobias, *Environ. Urban.* **27**, 421–440 (2015).
27. UN-Habitat, *State of the World's Cities 2012/2013* (Routledge, New York, 2013).
28. T. Endo, *Water Policy* **17**, 332–348 (2015).
29. O. Varis, K. Enckell, M. Keskinen, *Int. J. Water Resour. Dev.* **30**, 433–444 (2014).
30. D. R. Marlow, M. Moglia, S. Cook, D. J. Beale, *Water Res.* **47**, 7150–7161 (2013).
31. C. Pahl-Wostl, P. Jeffrey, N. Isendahl, M. Brugnach, *Water Resour. Manage.* **25**, 837–856 (2011).
32. J. Markard, R. Raven, B. Truffer, *Res. Policy* **41**, 955–967 (2012).
33. K. C. Seto, M. Fragkias, B. Güneralp, M. K. Reilly, *PLOS ONE* **6**, e23777 (2011).
34. T. D. Fletcher *et al.*, *Urban Water J.* **12**, 525–542 (2015).
35. S. B. Grant *et al.*, *Science* **337**, 681–686 (2012).
36. B. Jiménez, T. Asano, Eds., *Water Reuse: An International Survey of current practice, issues and needs,* Scientific and Technical Report Series, no. 20 (IWA, London, 2008).
37. National Academies of Sciences, Engineering, and Medicine, "Using Graywater and Stormwater to Enhance Local Water Supplies: An Assessment of Risks, Costs, and Benefits" (2016).
38. F. Meggers, H. Leibundgut, *Energy Build.* **43**, 879–886 (2011).
39. G. Zeeman, K. Kujawa-Roeleveld, *Water Sci. Technol.* **64**, 1987–1992 (2011).
40. T. A. Larsen, A. C. Alder, R. I. L. Eggen, M. Maurer, J. Lienert, *Environ. Sci. Technol.* **43**, 6121–6125 (2009).
41. National Development and Reform Commission, "Medium and long-term development plan for renewable energy in China (translated from Chinese)" (Beijing, 2007).
42. K. M. Udert *et al.*, *Water S.A.* **41**, 212–221 (2015).
43. T. A. Larsen, J. Lienert, K. M. Udert, Eds., *Source Separation and Decentralization for Wastewater Treatment* (IWA, London, 2013).
44. E. Dahlgren, C. Göcmen, K. Lackner, G. Van Ryzin, *Eng. Economist* **58**, 231–264 (2013).
45. N. K. Singh, A. A. Kazmi, M. Starkl, *Water Sci. Technol.* **71**, 468–478 (2015).
46. M. Finger, R. Künneke, *International Handbook of Network Industries: The Liberalization of Infrastructures* (Edward Elgar, Cheltenham, 2011).
47. E. Araral, *Policy Soc.* **27**, 221–228 (2009).
48. S. Kikeri, J. Nellis, *World Bank Res. Obs.* **19**, 87–118 (2004).
49. P. Marin, "Public-private partnerships for urban water utilities: a review of experiences in developing countries," *Trends and policy options No. 2009/8* (The International Bank for Reconstruction and Development, World Bank, Washington, DC, 2009).
50. G. Pierce, *J. Plann. Lit.* **30**, 119–131 (2015).
51. D. Dominguez, H. Worch, J. Markard, B. Truffer, W. Gujer, *Calif. Manage. Rev.* **51**, 30–50 (2009).
52. R. Künneke, J. Groenewegen, C. Ménard, *J. Econ. Behav. Organ.* **75**, 494–505 (2010).
53. E. Lieberherr, B. Truffer, *Environ. Innovation Societal Transitions* **15**, 101–122 (2015).
54. H. Hiessl *et al.*, *AKWA Dahler Feld. Contracting im Bereich der Wasserwirtschaft*, ISI-Schriftenreihe Innovationspotenziale (Fraunhofer Verlag, Karlsruhe, 2010), pp. 145.
55. M. Kiparsky, D. L. Sedlak, B. H. Thompson Jr., B. Truffer, *Environ. Eng. Sci.* **30**, 395–408 (2013).
56. J. Potts, *Policy Pract.* **11**, 34–43 (2009).
57. S. R. Harris-Lovett, C. Binz, D. L. Sedlak, M. Kiparsky, B. Truffer, *Environ. Sci. Technol.* **49**, 7552–7561 (2015).
58. D. Niemeier, H. Gombachika, R. Richards-Kortum, *Science* **345**, 1287–1290 (2014).
59. R. Künzle, W. Pronk, E. Morgenroth, T. A. Larsen, *J. Water Sanit. Hyg. Dev.* **5**, 448–455 (2015).
60. J. R. McConville, R. Künzle, U. Messmer, K. M. Udert, T. A. Larsen, *Environ. Sci. Technol.* **48**, 12238–12246 (2014).
61. D. G. Randall, M. Krähenbühl, I. Köpping, T. A. Larsen, K. M. Udert, *Water Res.* **95**, 361–369 (2016).
62. T. Jahn, in *Transdisziplinäre Forschung. Integrative Forschungsprozesse verstehen und bewerten*, M. Bergmann, E. Schramm, Eds. (Campus, Frankfurt, 2008), pp. 21–38.
63. D. Foray, D. C. Mowery, R. R. Nelson, *Res. Policy* **41**, 1697–1702 (2012).
64. R. Garud, P. Karnoe, *Res. Policy* **32**, 277–300 (2003).
65. S. Jacobsson, K. Karltorp, *Energy Policy* **44**, 374–384 (2012).
66. B. C. Ferguson, N. Frantzeskaki, R. R. Brown, *Landsc. Urban Plan.* **117**, 32–45 (2013).
67. R. G. Luthy, D. L. Sedlak, *Daedalus* **144**, 72–82 (2015).
68. C. Binz, B. Truffer, L. Coenen, *Econ. Geogr.* **92**, 172–200 (2016).
69. FAO, AQUASTAT Main Database, Food and Agriculture Organization of the United Nations (FAO) (2015).
70. www.recyclingshower.com.au
71. www.xeroscleaning.com
72. www.semizentral.de/en/home
73. T. A. Larsen *et al.*, *J. Water Sanit. Hyg. Dev.* **5**, 64–71 (2015).
74. I. A. Shiklomanov, World Water Resources and their Use a joint SHI/UNESCO product (1999); http://webworld.unesco.org/water/ihp/db/shiklomanov/.
75. D. N. Lerner, *Hydrogeol. J.* **10**, 143–152 (2002).
76. United Nations Statistics Division (UNSD), http://unstats.un.org/unsd/environment/wastewater.htm.
77. International Water Management Institute (IWMI), "Water for Food, Water for Life: A Comprehensive Assessment of Water Management in Agriculture" (Earthscan, London, 2007).
78. World Health Organization (WHO), UNICEF, "Progress on sanitation and drinking water—2015 update and MDG assessment" (WHO/UNICEF, Geneva, 2015); www.unicef.org/publications/files/Progress_on_Sanitation_and_Drinking_Water_2015_Update_.pdf.

ACKNOWLEDGMENTS

We thank R. Siber from Eawag for help in preparing Fig. 2. The authors declare that the publication of this paper presents no conflicts of interest. All present affiliations of all authors are indicated.

10.1126/science.aad8641

Downloaded from http://science.sciencemag.org/ on May 19, 2016

PERSPECTIVE

Transport solutions for cleaner air

Frank J. Kelly[1]* and Tong Zhu[2]*

In cities across the globe, road transport remains an important source of air pollutants that are linked with acute and chronic health effects. Decreasing vehicle emissions—while maintaining or increasing commuter journeys—remains a major challenge for city administrators. In London, congestion-charging and a citywide low-emission zone failed to bring nitrogen dioxide concentrations under control. n Beijing, controls on the purchase and use of cars have not decreased transport emissions to a sufficient extent. As cities continue to grow, not even zero-emission vehicles are the solution. Moving increasingly large numbers of people efficiently around a city can only be achieved by expanding mass transit systems.

By facilitating the movement of people and goods and providing employment, road transport supports economic growth and plays a considerable role in global urbanization. The consequent dependence on fossil fuels (increasingly diesel in Europe) has, however, created specific air pollution challenges. In particular, road transport remains an important source of particulate matter (PM) and nitrogen dioxide (NO_2), which can cause or aggravate asthma and impair lung development in children (*1*). Here, we ask whether the policy interventions already introduced, or under consideration, in two of the world's largest and most economically powerful cities—London, England, and Beijing, China—are good exemplars for other cities to follow (Figs. 1 and 2). We consider some of the changes required so that urban areas do not grind to a standstill because of continued growth and transport demand.

A tale of two cities—London

In view of widespread public concern about the health effects of air pollution, the mayor of London launched an air quality strategy in 2002, entitled Cleaning London's Air (*2*). The main aim of the strategy was the reduction of pollution from road traffic, which was the largest source of the pollutants of concern—PM2.5 and NO_2—in the city. PM2.5 refers to particles that pass through a size-selective inlet with a 50% efficiency cutoff at 2.5 μm aerodynamic diameter, which means that they are small enough to penetrate deep into the lungs. To this end, the goals were to reduce the number of vehicles on the road and to lower emissions from individual vehicles through modernization of the fleet vehicle stock.

To help achieve the first goal, a congestion-charging scheme (CCS) was introduced in central London in February 2003 (*3*). The CCS initially reduced the number of private vehicles entering central London by 18%, but the scheme's effectiveness has since been eroded; since 2012, congestion has broadly been back at pre-2002 levels.

The London-wide LEZ, introduced in February 2008, aimed to tackle the second goal by restricting the entry of the oldest and most polluting vehicles (such as diesel engine heavy goods vehicles, buses and coaches, and larger vans and minibuses) across greater London (*4*). However, although black carbon pollution has gradually decreased on many roads in London since 2008, NO_2 concentrations have either not changed or have gone up (*5*). As a consequence, London does not comply with European Union limit values for NO_2 and is unlikely to do so until 2025 at the earliest. The main reason for this is the increased share of new diesel cars on London's roads, up from 1 in 10 cars in 2000 to 6 in 10 in 2014. Real-world NO_2 emissions from most of these vehicles do not comply with European exhaust emission standards, which were introduced in 1993 to improve air quality (*6*).

A tale of two cities—Beijing

In contrast to the UK's long industrial heritage, China has undergone rapid industrialization over the past few decades, adding thousands of kilometers of urban road and hundreds of millions of vehicles. PM2.5 emissions from traffic have contributed to increasingly poor air quality in Beijing (*7*), threatening public health (*8*). The severity of air pollution in Beijing was first acknowledged when China was bidding for the Olympic Games in the 1990s. To address the problem, in 1998 the Beijing municipal government introduced the first of 16 stages of air pollution control measures. Initial measures mainly focused on shutting down small industrial and domestic coal burning stoves and controlling dust from road and construction sites, as well as phasing out old vehicles that did not meet emission standards. These standards were, however, much less strict than those in place today.

Given fast urbanization and social and economic development in Beijing, later measures were more comprehensive because they needed to deal with the complicated nature of the air pollution sources in Beijing. Long-term and strategic measures included replacing coal with natural gas, creating a public transport system with 554 km of subway, and introducing current European emission standards. During the Beijing Olympics in 2008 (*9*) and the Asia-Pacific Economic Cooperation meeting in Beijing in 2014, aggressive measures were taken to reduce the number of vehicles on the road by

PHOTO: MILES ERTMAN/ROBERT HARDING

Fig. 1. Tower Bridge, the Shard, and City Hall. London continues to experience road congestion as it struggles to deal with an ever-increasing population.

[1]National Institute for Health Research (NIHR), Heath Protection Research Unit in Health Impact of Environmental Hazards, MRC-PHE Centre for Environment and Health, King's College London, London SE1 9NH, UK. [2]State Key Laboratory for Environmental Simulation and Pollution Control, College of Environmental Sciences and Engineering, and Beijing Innovation Center for Engineering Science and Advanced Technology (BIC-ESAT), Peking University, Beijing 100871, China.
*Corresponding author. Email: frank.kelly@kcl.ac.uk (F.J.K.); tzhu@pku.edu.cn (T.Z.)

Downloaded from http://science.sciencemag.org/ on May 19, 2016

allowing cars to be used on only certain days of the week, determined by the last digit of the car license number. This measure resulted in substantial temporary reductions in congestion and pollution concentrations in Beijing and was considered so successful that it was made a routine rule immediately after the Olympics. Today, every car is allowed to drive 4 days per 5 weekdays.

These efforts to control air pollution in Beijing have shown some success. PM2.5 concentrations have started to fall in recent years. However, these benefits have been compromised to some extent by the rapid growth in vehicle numbers from 1.5 million in 2000 to 5.6 million in 2015. To start dealing with this problem, the Beijing government introduced a range of vehicle emission policies, including a lottery system for new car license plates and further modernization of the fleet.

Given that the vehicle fleet in China, especially in Beijing, now has the most advanced engine technologies (*10*), further control measures to cut vehicle emissions will necessitate further restrictions on car movements. Importantly, the Chinese government has not provided incentives for increasing the proportion of private diesel vehicles (*11*). As a result—in line with other countries, such as Japan—the problems of diesel NO_2 emissions from private vehicles contributing to Beijing's poor air quality have been avoided.

Looking to the future

The experiences in London and Beijing clearly show that more fundamental changes in urban transportation systems are required for air quality improvements to be achieved and for city streets to become more pleasant environments. In London, the introduction of an ultralow emission zone (uLEZ) is planned for 2020 (*12*). However, banning all but the cleanest vehicles and incentivizing the use of zero-emission cars and taxis in a small area of central London will be insufficient to achieve compliance with NO_2 standards until at least 2025 (*13*). Beijing will grind to a standstill through excessive congestion, and air quality will suffer further if car ownership levels are allowed to move toward U.S. levels (840 cars per 1000 people). Car ownership (currently 280 per 1000 people) must therefore continue to be strictly controlled.

In many European cities, attitudes toward car ownership appear to be changing. Many urban youth of today, it seems, would rather use taxis and car share than own their own car. Recent polls in London associated with the mayoral elections in May 2016 put air quality issues high on the political agenda and suggest that there is popular support for a further tightening of vehicle use in the city. Cars promote a sedentary lifestyle, and a shift to cycling and walking has health benefits for urban dwellers (*14*).

Upon recognizing the negative impacts that congestion, when entering cities, has on the economy and on air quality, several states in the northeastern United States have introduced electronic toll collection technology, which has helped to reduce delays at toll plazas. Improved traffic flows near these toll plazas were linked to more healthy infants born to mothers living nearby (*15*). Other interventions, such as the Hoy No Circula (HNC) program in Mexico City, have proved less successful. Introduced in 1989 to address the severe air pollution problems in Mexico City, the HNC banned most drivers from using their vehicle 1 weekday per week, determined by the last license plate digit (*16*). Although the program was largely successful in removing ~20% of vehicles from the roads, no evidence could be found that it led to increased use of public transportation or that it improved air quality (*16*).

The car industry has recognized the need for change, and companies such as BMW and Ford have launched car-share schemes. Car-sharing has obvious benefits to cities. Zipcar estimates that every shared vehicle replaces up to 20 private cars, thus reducing total vehicle miles and land devoted to parking. The car industry is also pushing forward the development of city-friendly plug-in electric vehicles. Worries over battery life and the lack of coordinated charging networks have, however, held back vehicle sales, even with generous government subsidies. In the UK, the electric vehicle fleet has grown in the past 2 years from 3500 to more than 50,000 vehicles, suggesting that the tide is turning in favor of zero-emission vehicles. In contrast, goals for fuel cell (hydrogen)–powered cars have not been achieved because of limited fueling networks and high initial vehicle costs.

As cities continue to grow in size and population, even zero-emission vehicles are not the solution. Moving increasing numbers of people efficiently around a city can only be achieved by expanded mass transit systems. The increased use of bus networks and traditional subways are providing city dwellers with more flexible, cheaper, and less polluting options. For instance, China is building 87 new mass transit rail lines over the next 5 years. If construction continues at this pace, China's cities will have half of the world's metro tracks by 2050 (*17*).

Rapid transit systems offer only a partial solution to city congestion and poor air quality, however, because many users may not live within easy walking distance to the transit points. To be effective, rapid transit systems must be linked with other transport options at the start and end of the journey. The creation of Autonomous Mobility-on-Demand networks may solve this problem (*17*). Modeled after bicycle-share programs, users would have access to a network of lightweight electronic vehicles (LEVs) distributed at charging stations throughout the city. Customers could summon a vehicle through a smartphone app and abandon it upon reaching their destination, from where the vehicle would find its way back to a charging station or new user. Trials of this technology are

Fig. 2. Galaxy Soho. Beijing's increasing private car ownership is contributing to air pollution challenges.

"As cities continue to grow in size and population, even zero emission vehicles are not the solution. Moving increasing numbers of people efficiently around a city can only be achieved by expanded mass transit systems."

PHOTO: © HUFTON AND CROW/VIEW/CORBIS

Downloaded from http://science.sciencemag.org/ on May 19, 2016

planned in California and in Milton Keynes, UK (*18*, *19*).

In the city of the future, mass public transport systems may lead to many more car-free roads, transforming the landscape and enhancing the urban experience. Moving toward such an environment depends, however, on continued and exacting scientific research, the translation of such research into realistic and effective policies, and successfully encouraging members of the public to use public transport, value exercise, and not drive for short journeys (*20*). The reward will be improved health and quality of life for the growing urban populations around the world.

REFERENCES

1. RCP/RCPCH, Every breath we take: The lifelong impact of air pollution (2016); available at www.rcplondon.ac.uk/projects/outputs/every-breath-we-take-lifelong-impact-air-pollution.
2. Greater London Authority (GLA), *The Mayor's Air Quality Strategy, Cleaning London's Air* (GLA, 2002).
3. F. J. Kelly *et al.*, *Res. Rep. Health Eff. Inst.* **155**, 5–71 (2011).
4. Transport for London (TfL), "Proposed London Low Emission Zone. Greater London Low Emission Zone charging order 2006. Report to the Mayor following stakeholders, businesses, other organisations and the public" (2006); available at http://content.tfl.gov.uk/lez-mayors-report-chapter-1-5.pdf.
5. A. Font, G. W. Fuller, Roadside air quality trends in London—Identifying the outliers Part 1 (ERG, King's College London, 2015); available at http://content.tfl.gov.uk/roadside-air-quality-trends-in-london-identifying-outliers-part-1.pdf.
6. D. C. Carslaw, G. Rhys-Tyler, *Atmos. Environ.* **81**, 339–347 (2013).
7. Y. Chen *et al.*, *Environ. Pollut.* **212**, 269–278 (2016).
8. G. Yang *et al.*, *Lancet* **381**, 1987–2015 (2013).
9. M. Wang *et al.*, *Atmos. Chem. Phys.* **9**, 8247–8263 (2009).
10. P. Hodges, "China's new car market goes ex-growth as used car sales take off," *Financial Times* (2015); available at http://blogs.ft.com/beyond-brics/2015/06/30/chinas-new-car-market-goes-ex-growth-as-used-car-sales-take-off.
11. M. Carnes, E. Helmers, *Environ. Sci. Eur.* **25**, 15 (2013).
12. GLA, Ultra Low Emission Zone Update to the London Assembly (2014); available at www.london.gov.uk/moderngov/documents/b9960/Minutes%20-%20Appendix%202%20-%20ULEZ%20Update%20Thursday%2006-Feb-2014%2010.00%20Environment%20Committee.pdf?T=9.
13. R. Howard, S. Beevers, D. Dajnak, "Up in the air: How to solve London's air quality crisis: Part 2" (King's College London and Policy Exchange, 2016); available at www.policyexchange.org.uk/publications/category/item/up-in-the-air-how-to-solve-london-s-air-quality-crisis-part-2.
14. NCD Risk Factor Collaboration (NCD-RisC), *Lancet* **387**, 1377–1396 (2016).
15. J. Currie, R. Walker, *Am. Econ. J. Appl. Econ.* **3**, 65–90 (2011).
16. L. W. Davis, *J. Polit. Econ.* **116**, 38–81 (2008).
17. R. Chin, Solving transport headaches in the cities of 2050 (BBC, 2013); available at www.bbc.com/future/story/20130617-moving-around-in-the-megacity.
18. BBC News, "Driverless cars to be introduced in Milton Keynes" (2013); available at www.bbc.co.uk/news/uk-england-beds-bucks-herts-24849948.
19. ExtremeTech, "Google's new self-driving car: Electric, no steering wheel, and incredibly cute" (2014); available at www.extremetech.com/extreme/183174-googles-new-self-driving-car-electric-no-steering-wheel-and-incredibly-cute.
20. F. J. Kelly, J. C. Fussell, *Environ. Geochem. Health* **37**, 631–649 (2015).

10.1126/science.aaf3420

PERSPECTIVE

The ecological future of cities

Mark J. McDonnell[1]* and Ian MacGregor-Fors[2]*

The discipline of urban ecology arose in the 1990s, primarily motivated by a widespread interest in documenting the distribution and abundance of animals and plants in cities. Today, urban ecologists have greatly expanded their scope of study to include ecological and socioeconomic processes, urban management, planning, and design, with the goal of addressing issues of sustainability, environmental quality, and human well-being within cities and towns. As the global pace of urbanization continues to intensify, urban ecology provides the ecological and social data, as well as the principles, concepts and tools, to create livable cities.

For the past half century, the prevailing global economies have created employment opportunities in urban areas, driving massive migrations and overcrowding cities. In 2014, the global urban population reached nearly 4 billion people; it is predicted to gain an additional 2.5 billion people by 2050, mostly in African and Asian cities (*1*). The 2011 United Nations Global Report on Human Settlements warned that future synergistic interactions between urbanization and global climate change will threaten the stability of human societies. Urban ecology has developed the ecological and social data, principles, concepts, and tools required to create livable cities in this increasingly urbanized world.

The discipline of urban ecology arose in the 1990s out of a need to increase our knowledge and understanding of the ecological and human dimensions of urban ecosystems (*2*, *3*). Urban ecology studies have mostly focused on biodiversity, ecosystem processes, and ecosystem service attributes of cities. Increasingly, urban ecologists are integrating these components with socioeconomic processes, urban management, planning, and design (*2*). Furthermore, the direct and indirect socioecological effects of urban environments extend far beyond the physical borders of cities. Some of the most alarming global environmental problems are the result of urbanization, including land-use and land-cover change, alteration of biogeochemical cycles, climate change, biodiversity loss, and biological invasions (*4*). Thus, the scope of urban ecology research extends well beyond city limits.

An evolving discipline

Scientists use a number of different approaches to study urban systems. Some studies focus on describing small-scale physical and ecological elements and patterns within cities. In this vein, early urban ecology studies commonly described the distribution and abundance of organisms residing in remnant woodlands and wetlands within cities. Other studies aim to elucidate the interacting physical, biological, and social components at city- or ecosystem-wide scales (*5*). For example, Hope *et al.* (*6*) discovered a positive relationship between plant diversity and human wealth in metropolitan Phoenix. They found higher plant diversity in wealthier neighborhoods, which adds to growing evidence that the quality of the urban social environments is strongly related to socioeconomic status in developed regions. More recent studies explicitly address issues of sustainability, environmental quality, and human well-being in urban ecosystems. For instance, researchers involved in the Baltimore Ecosystem Study, one of the first in this emerging research direction, have been working to better connect neighborhood urban design with fundamental ecological theory and sound environmental practices.

Ecological understanding of cities

Our ecological understanding of cities has lagged far behind that of nonurban areas, mainly because of the widely held notion that humans disrupt the natural ecological conditions and processes that scientists are attempting to understand (*3*). Thus, nature within cities was long considered unworthy of scientific study, except when it involved solving environmental problems that threatened human well-being.

To address the critical environmental issues facing our cities, ecologists have expanded their traditional knowledge base. This has been necessary because urban systems are highly complex and heterogeneous. Cities often contain large human populations, as well as high proportions of exotic species, distinctive disturbance regimes, few top predators, and elevated levels of nutrients and pollutants (*2*).

Not only has the growth of ecological knowledge about the structure and function of cities improved our understanding of basic ecological principles, it has also provided us with valuable information for creating livable, healthy, and resilient urban environments. Today, cities serve as important living laboratories that present opportunities to test fundamental social and ecological questions, such as the effect of landscape fragmentation on the distribution of organisms, the effect of night lighting on the circadian rhythms of humans and other organisms, and the role of urban green spaces in providing ecosystem services to urban dwellers. Cites also provide opportunities

[1]Australian Research Centre for Urban Ecology, Royal Botanic Gardens Victoria and School of BioSciences, University of Melbourne, Melbourne, Victoria 3010, Australia. [2]Red de Ambiente y Sustentabilidad, Instituto de Ecología, A.C. (INECOL), Carretera Antigua a Coatepec 351, El Haya, Xalapa 91070, Veracruz, Mexico.
***Corresponding author. Email: markmc@unimelb.edu.au (M.J.M.); ian.macgregor@inecol.mx (I.M.-F.)**

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Fig. 1. View of the central business district and Kings Domain gardens in Melbourne, Australia. Like other cities around the world, Melbourne is developing urban ecology action plans to improve quality of life for both humans and other species in the region.

PHOTO: ANDREW BERTULEIT PHOTOGRAPHY/CORBIS; IMAGE 42-72639141

for learning about people's knowledge, attitudes, and values in regard to nature. Increasingly, ecologists and landscape architects work together to develop urban design experiments that deliver both cultural amenities and important ecological information (*7*). For example, in design experiments initiated to increase urban tree cover, ecologists are active participants in the design process as well as the follow-up assessments.

Due to the many physical, biological, and social variations between different cities, urban areas also provide unprecedented opportunities to study the evolutionary and adaptive processes affecting urban biota. For example, in a recent global comparative study of more than 800 species of urban birds, Sol *et al.* found that species loss was primarily due to the birds' inability to adapt to urban conditions, especially with regard to exploiting novel resources and avoiding new risks (*8*). However, other studies have shown that some species currently thriving in cities have adapted to urban environments via behavioral, physiological, morphological, and evolutionary responses (*9*). Examples include great tits (*Parus major*) modifying their song to communicate at noisy locations (*10*) and plants living in highly fragmented urban environments shifting to localized seed dispersal strategies (*11*).

Global ecological assessments on the distribution of plants and animals in urban environments have documented the biological richness of cities and offered insights into maintaining this richness. For example, Aronson *et al.* (*12*) found that the presence of remnant and restored natural vegetation is important for maintaining bird and plant populations in urban areas. Hahs *et al.* (*13*) reported that fewer local plant extinctions occurred in cities that maintained at least 30% native vegetation cover. Unfortunately, few studies have addressed the ecological effects of urban growth in developing countries, where levels of biodiversity are highest and urbanization is progressing rapidly (*14*).

Urban ecology in action

Urban ecology research commonly has an applied outlook, with a focus on identifying mitigation and adaptation strategies to reduce the negative effects of human activities (*3*). Urban ecologists in Melbourne, Australia, have begun to develop action plans to improve the health, livability, resilience, and sustainability of their city (Fig. 1). For example, an urban forest strategy aims to ensure that the city will have a resilient, healthy, and diverse urban forest in the future. Other cities around the world—including London (United Kingdom), Chicago (United States), Singapore, Durban (South Africa), and Portland (United States)—are incorporating similar strategies and plans.

Local governments have been participating in research partnerships focused on diverse themes such as understanding, protecting, and enhancing biodiversity; creating and managing ecological linkages; creating biodiversity-friendly environments by minimizing negative effects on animals, plants, soils, and ecosystems; and assisting the public in experiencing and valuing biodiversity and urban ecology in general. The Chicago Wilderness regional alliance is one such partnership. This group aims to improve the quality of life of both humans and other species in the region. Initially, researchers, government and nongovernment agencies, and the public came together to restore and enhance the region's biodiversity. Their mission has expanded to include issues related to climate change, green infrastructure, and connecting people with nature.

Increasingly, urban ecologists, local communities, and governments are also working together in developing countries. For instance, Chilean researchers conducted detailed studies of urban wetlands in Valdivia (southern Chile). Their acquired environmental and social knowledge encouraged the local government, together with the

Downloaded from http://science.sciencemag.org/ on May 19, 2016

interested community, to develop a land management and planning tool for the protection and future sustainable use of the region's wetlands.

Another example is that of one of the remaining perennial springs (La Carbonera) surrounding the city of Querétaro, located in a semi-arid region of central Mexico. Because the area's wetlands have been studied for several years, it was possible to rescue the spring, together with strong local stakeholder support and action. This is clearly a successful case in which locals worked closely with academics to achieve a common goal. This peri-urban spring has been managed, and the area is currently used to conduct recreational and educational activities while also contributing to the restoration of several critical ecological processes.

Toward livable cities

The discipline of urban ecology has made great strides over the past three decades. To further our knowledge of urbanization's effects on people, biodiversity, and ecosystem processes, urban ecologists must shift from studying patterns to untangling the emerging mechanistic processes behind the reported patterns (*15*). However, it is also crucial to create the tools and procedures for transforming scientific knowledge into action. The need for these advances is pressing, as there is a growing discontent among urban dwellers worldwide, related to the erosion of their quality of life. Many urban dwellers are now calling for the creation of green, sustainable cities that are also healthy and resilient (*16*). Incorporating urban ecology principles into the design, construction, and management of cities will require the cooperation, alliance, and synergy of all stakeholders, thus reforming the way we conceive and prepare land to fulfill the needs of modern human agglomerations.

REFERENCES

1. United Nations Department of Economic and Social Affairs, Population Division, *World Urbanization Prospects: The 2014 Revision, Highlights* (United Nations, 2014).
2. I. Douglas, P. James, *Urban Ecology: An Introduction* (Routledge, 2015).
3. M. J. McDonnell, in *Urban Ecology: Patterns, Processes, and Applications*, J. Niemelä *et al.*, Eds. (Oxford Univ. Press, 2011), pp. 5–13.
4. M. von der Lippe, I. Säumel, I. Kowarik, *Die Erde* **136**, 123–143 (2005).
5. S. T. Pickett *et al.*, *J. Environ. Manage.* **92**, 331–362 (2011).
6. D. Hope *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8788–8792 (2003).
7. A. J. Felson, M. A. Bradford, T. M. Terway, *Front. Ecol. Environ* **11**, 362–367 (2013).
8. D. Sol, C. González-Lagos, D. Moreira, J. Maspons, O. Lapiedra, *Ecol. Lett.* **17**, 942–950 (2014).
9. M. J. McDonnell, A. K. Hahs, *Annu. Rev. Ecol. Evol. Syst.* **46**, 261–280 (2015).
10. H. Slabbekoorn, M. Peet, *Nature* **424**, 267 (2003).
11. P. O. Cheptou, O. Carrue, S. Rouifed, A. Cantarel, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3796–3799 (2008).
12. M. F. J. Aronson *et al.*, *Proc. Biol. Sci.* **281**, 20133330 (2014).
13. A. K. Hahs *et al.*, *Ecol. Lett.* **12**, 1165–1173 (2009).
14. I. MacGregor-Fors, R. Ortega-Álvarez, Eds., Ecología urbana: Experiencias en América Latina (2013); http://www1.inecol.edu.mx/libro_ecologia_urbana.
15. M. J. McDonnell, A. K. Hahs, *Urban Ecosyst.* **16**, 397–409 (2013).
16. United Nations Human Settlements Programme (UN-Habitat), *State of the World's Cities 2008/2009: Harmonious Cities* (Earthscan, 2008).

10.1126/science.aaf3630

PERSPECTIVE

Living in cities, naturally

Terry Hartig[1] and Peter H. Kahn Jr.[2]

Natural features, settings, and processes in urban areas can help to reduce stress associated with urban life. In this and other ways, public health benefits from, street trees, green roofs, community gardens, parks and open spaces, and extensive connective pathways for walking and biking. Such urban design provisions can also yield ecological benefits, not only directly but also through the role they play in shaping attitudes toward the environment and environmental protection. Knowledge of the psychological benefits of nature experience supports efforts to better integrate nature into the architecture, infrastructure, and public spaces of urban areas.

Crowding, noise, and other stressful urban conditions increase the risk of mental disorders such as anxiety and depression (*1*). However, urban areas also have environmental assets that support mental health. For example, parks, green spaces, street trees, and community gardens can facilitate physical activity, social contacts, and stress reduction (*1*, *2*). How can psychological benefits from encounters with natural features and processes offset the psychological costs of other urban living conditions? Answers to this question will help improve the quality of life of today's growing urban populations (*2*, *3*) (Fig. 1).

In his classic work on urban psychology, Stanley Milgram opened on a positive note, arguing that "cities have great appeal because of their variety, eventfulness, possibility of choice, and the stimulation of an intense atmosphere that many individuals find a desirable background to their lives" (*4*). He also saw that cities offer "unparalleled possibilities" for face-to-face contact and communication (*4*). Yet, when considering how psychology can contribute to understanding the experience of living in a city, he turned to the negative; he highlighted overload as a psychological concept useful for linking objective urban circumstances such as high population density to observable behaviors such as incivility. Such contrasting perspectives continue to inform psychological research on urban life.

Fig. 1. Together with ecological benefits such as climate change mitigation and the protection of biological diversity, the renaturing of cities opens opportunities for people to engage with features and processes of the natural world; for example, when tending plants in a community garden.

PHOTO: © BOHEMIAN NOMAD PICTUREMAKERS/CORBIS

Urban-rural differences in mental disorders

Numerous studies have suggested that urban living conditions undermine mental health, whereas conditions in rural areas support health. Peen *et al.* (*5*) performed a meta-analysis of 21 studies, done after 1985, that investigated the prevalence of mood disorders among non-institutionalized

[1]Institute for Housing and Urban Research and Department of Psychology, Uppsala University, Uppsala, Sweden. [2]Department of Psychology and School of Environmental and Forest Sciences, University of Washington, Seattle WA, USA.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

adults in relatively affluent countries. They found that the odds of mood disorder were 28% higher in the urban areas than in the rural areas; however, the results were highly heterogeneous, perhaps because "rural" and "urban" were defined differently in different studies. They recognized the problem with integrating results when urban settings in some studies resemble rural settings in others.

Judd *et al.* (*6*) have argued that binary urban and rural categories are insufficient. Numerous environmental factors can affect the stress that people experience in urban and rural settings and, thus, the prevalence of disorders such as anxiety and depression. These include, for example, residential density, housing quality, air quality, transportation options, access to health and welfare services, and access to parks and green spaces. All of these factors might play a role in whether a given individual develops a disorder, but they may do so in different ways in different combinations. Together with individual vulnerabilities and broader contextual characteristics, research must consider the independent and combined effects of these factors in order to elucidate how specific urban conditions may undermine or support mental health.

PHOTO: © MONALYN GRACIA/CORBIS

Fig. 2. People living in cities can encounter nature in the context of many different activities. Their experiences can enhance their development, health, and well-being, and they can shape their attitudes toward the environment and environmental protection.

Findings of urban-rural differences in mental disorders might be taken as warnings about the consequences of urbanization, but there is a risk of confusing cause and effect. The social, political, and economic forces that drive urbanization in a particular society may also generate mental illness through other mechanisms. For example, in some countries, many people have left rural areas for urban ones under threat of violence; this appears to have boosted the prevalence of major depressive disorder (*7*). Poor conditions in the spontaneous settlements they create may exacerbate the effects of victimization, disruption of social relations, and loss of traditional occupations, but urbanization itself is not the initial cause of disorder.

Psychological benefits of nature experience

Research on the experience of nature also suggests that urban living conditions can undermine mental health, whereas relatively natural conditions can support it. The term "nature" means different things in different contexts. Recognizing personal and cultural aspects of the experience of nature, psychological research considers how different people encounter nature in diverse contexts, from viewing indoor plants in urban offices to walking in wilderness areas.

A significant portion of this research concerns the restorative effects of nature experience, such as regaining the ability to concentrate and reducing blood pressure after intensive mental work (*3*, *8*). Theories about restoration describe how encounters with nature involve psychological distance from everyday demands and interested engagement with environmental features and processes. These components of restorative experience have counterparts in many analyses of why people engage in outdoor recreation, done in the United States and elsewhere since the early 1960s (*8*). Complementing what people have long said they seek through outdoor recreation and substantiating theoretical claims, laboratory and field experiments have repeatedly shown that spending time in natural environments or viewing scenes of nature can quickly help people to lift their mood, improve their ability to direct attention, and reduce physiological arousal to a greater degree than do urban streets and other comparison conditions (*2*, *8*).

As in studies of urban-rural differences in mental disorders, comparisons of single examples of urban and natural environments have been criticized for neglecting relevant forms of variation (*8*). Experimental evidence regarding plausible mechanisms has nonetheless encouraged large observational studies on urban green space and health. Such studies consider plausible cumulative effects of green space and greenery near an urban home. Most such studies focus on quantity, but some have also used quality indicators (*9*). Most of the observational studies have used a cross-sectional design, but recent longitudinal studies have enabled stronger inferences regarding beneficial effects of access or proximity to nature. For example, using 5 consecutive years of data for each of 1064 participants in the British Household Panel Survey, Alcock *et al.* (*10*) showed that those who relocated from a less green (58% local area coverage) to a more green (74%) urban area showed improved mental health over the next 3 years. In contrast, those who moved from more (74%) to less (59%) green urban areas showed a decline in mental health before the move, followed by a return to the pre-move baseline.

Much research on nature experience assumes that too few generations have passed for natural selection to shift the adaptedness of affective and cognitive functioning from the conditions of hominid evolution (*8*), and that humans are therefore poorly adapted to living in urban environments, broadly defined. This natural-urban antithesis neglects the fact that urban environments include settings that are supportive of human functioning; noisy, polluted, car-filled streets lined by anonymous towers are not representative of all urban possibilities. This false antithesis also neglects reasons why people gathered in cities millennia ago, and the consequences of that move for the interplay of natural selection and sociocultural development. Here too, the environmental categories and the urbanization process need closer examination.

The design of cities and human-nature relations

Psychological research has substantiated longstanding reasoning about parks and green spaces as health resources for urban populations (*8*, *11*, *12*). This reasoning has guided the creation of parks in many cities, such as Central Park in New York and Mount Royal Park in Montreal. Extending such precedents, researchers, design professionals, citizen groups, and others are working together to create sustainable urban fabrics in our increasingly urbanized world. These efforts—under banners such as green urbanism, green infrastructure, biophilic design, and renaturing—seek a better synthesis of natural processes and ecosystem functions with architecture and urban infrastructure through acts of creation, preservation, and ecological restoration. Such efforts are needed for psychological as well as ecological purposes. The evidence mentioned above and more like it warn against assuming that people can simply adapt to increasing urban density and its concomitants without negative consequences for health and well-being.

By providing opportunities for people to experience nature in cities and to experience cities

Downloaded from http://science.sciencemag.org/ on May 19, 2016

as natural, such efforts can shape attitudes toward the environment (Fig. 2). People in increasingly large and dense urban areas may have few or no contacts with the natural world in everyday life. Environmental generational amnesia refers to the psychological process whereby each generation constructs a conception of what is environmentally normal based on the natural world encountered in childhood (*13*). A problem arises insofar as the amount of environmental degradation increases across generations, but each generation tends to take that degraded condition as the nondegraded condition: the normal experience. This helps to explain inaction on environmental problems; people do not feel the urgency or magnitude of problems because the experiential baseline has shifted. Providing opportunities for people to experience more robust, healthy, and even wilder forms of nature in cities offers an important solution to this collective loss of memory and can counter the shifting baseline (*14*). Such opportunities include, for example, large green spaces and parks, rivers restored to some former free-flowing condition, expansive views over water and land, and extensive connective pathways for walking and biking.

Thus, cities designed well, with nature in mind and at hand, can be understood as natural, supportive of both ecosystem integrity and public health. Further psychological studies can describe how specific improvements in available opportunities for nature experience come to affect mental health and environmental attitudes (*15*). How will they change if car-clogged spaces give way to natural places where children can play wildly and others reflect quietly?

REFERENCES

1. American Psychological Association Task Force on Urban Psychology, *Toward an Urban Psychology: Research, Action, and Policy* (American Psychological Association, Washington, DC, 2005).
2. T. Hartig, R. Mitchell, S. de Vries, H. Frumkin, *Annu. Rev. Public Health* **35**, 207–228 (2014).
3. European Commission Directorate-General for Research and Innovation, *Towards an EU Research and Innovation Policy Agenda for Nature-Based Solutions and Re-Naturing Cities: Final Report of the Horizon 2020 Expert Group on 'Nature-Based Solutions and Re-Naturing Cities'* (European Commission, Brussels, 2015).
4. S. Milgram, *Science* **167**, 1461–1468 (1970).
5. J. Peen, R. A. Schoevers, A. T. Beekman, J. Dekker, *Acta Psychiatr. Scand.* **121**, 84–93 (2010).
6. F. K. Judd *et al.*, *Aust. N.Z. J. Psychiatry* **36**, 104–113 (2002).
7. A. J. Ferrari *et al.*, *Psychol. Med.* **43**, 471–481 (2013).
8. T. Hartig *et al.*, in *Forests, Trees and Human Health*, K. Nilsson *et al.*, Eds. (Springer, Dordrecht, 2011), pp. 127–168.
9. S. M. E. van Dillen, S. de Vries, P. P. Groenewegen, P. Spreeuwenberg, *J. Epidemiol. Community Health* **66**, e8 (2012).
10. I. Alcock, M. P. White, B. W. Wheeler, L. E. Fleming, M. H. Depledge, *Environ. Sci. Technol.* **48**, 1247–1255 (2014).
11. T. S. Eisenman, *J. Plann. Hist.* **12**, 287–311 (2013).
12. D. C. Rouse, I. F. Bunster-Ossa, *Green Infrastructure: A Landscape Approach* (Planning Advisory Service Report 571, American Planning Association, Chicago, 2013).
13. P. H. Kahn Jr., *Technological Nature: Adaptation and the Future of Human Life* (MIT Press, Cambridge, MA, 2011).
14. P. H. Kahn Jr., P. H. Hasbach, Eds., *The Rediscovery of the Wild* (MIT Press, Cambridge, MA, 2013).
15. F. G. Kaiser, T. Hartig, A. Brugger, C. Duvier, *Environ. Behav.* **45**, 369–398 (2013).

10.1126/science.aaf3759

PERSPECTIVE

Meta-principles for developing smart, sustainable, and healthy cities

Anu Ramaswami,[1]* Armistead G. Russell,[2] Patricia J. Culligan,[3] Karnamadakala Rahul Sharma,[1] Emani Kumar[4]

Policy directives in several nations are focusing on the development of smart cities, linking innovations in the data sciences with the goal of advancing human well-being and sustainability on a highly urbanized planet. To achieve this goal, smart initiatives must move beyond city-level data to a higher-order understanding of cities as transboundary, multisectoral, multiscalar, social-ecological-infrastructural systems with diverse actors, priorities, and solutions. We identify five key dimensions of cities and present eight principles to focus attention on the systems-level decisions that society faces to transition toward a smart, sustainable, and healthy urban future.

By the year 2050, the number of people living in cities is expected to increase by about 2.5 billion (*1*). It is estimated that over 60% of the urban areas that will exist by 2050 have yet to be built, indicating that there will be massive new infrastructure requirements, particularly in Asia and Africa (*2*). Simultaneously, existing cities worldwide are aging and much in need of infrastructure replacement.

Infrastructures—defined broadly as the systems that provide water, energy, food, shelter, transportation and communication, waste management, and public spaces (*3*)—are essential to support human well-being and economic development. However, aggregated globally, these seven infrastructure sectors currently place a large burden on the environment and have a considerable impact on human health (Fig. 1). Urban demands dominate these effects; for example, ~70% of global greenhouse gas (GHG) emissions are attributable to cities (*1*). Because physical infrastructures have life spans of 30 to 50 years, the large imminent global requirement for new urban infrastructure presents a historic opportunity for change. The question is, how can urban infrastructure transformations in the 21st century advance the environmental sustainability and human well-being of our cities by taking advantage of the enormous potential offered by data science and technology?

Although information and communication technologies are important for developing smart, sustainable, healthy cities (*4*), we argue that a larger understanding of urban infrastructure systems is necessary to move from data to information to knowledge and, ultimately, to action for urban sustainability and human well-being. With infrastructure as the focus, we identify five key dimensions of cities and present eight principles to help guide urban transformations toward sustainability and health, drawing on examples from the United States, China, and India.

[1]Hubert H. Humphrey School of Public Affairs, University of Minnesota, Minneapolis, MN 55455, USA. [2]School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0355, USA. [3]Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY 10027, USA. [4]ICLEI–Local Governments for Sustainability, South Asia Secretariat, New Delhi 110020, India.
*Corresponding author. Email: anu@umn.edu

Key dimensions

Economic opportunity is a key driver for urbanization, and infrastructure is a prime enabler. Multi-city data sets are emerging that describe scaling relationships among urban population growth, gross domestic product (GDP), household incomes, and infrastructure-related parameters such as financial investments, energy and water use, and land and road expansions (*5*). Cities with different economic structures (e.g., highly industrial, highly commercial, or mixed economy) are known to exhibit different socio-spatial patterns of development (i.e., urban form) that affect infrastructure design. Yet basic city-level data on urban GDP, sectoral employment, and household incomes are sparse in many developing nations and in smaller cities and towns, where much urban growth is projected to occur.

Urban form or morphology describes the evolving interaction between physical space and human activity in cities. Numerous data sets, from census data to aerial and satellite photographs and remote sensing information, are being integrated to enable planners to characterize urban form. Urban complexity science is advancing new measures (*4*) that focus not only on population density, connectivity, proximity to jobs and services, and diversity and intensity of urban activities but also on understanding self-similarity across scales (from blocks to neighborhoods to cities) and patterns of social segregation (e.g., of migrant and informal populations in a city). Urban form represents the foundation upon which infrastructure develops, shaping energy and material use; access to and contiguity of water bodies, green space, and other critical ecosystems; and urban equity and well-being.

Infrastructure design and socio-spatial disparities within cities are emerging as critical determinants of human health and well-being.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Cities are grappling with multiple and multiscalar health risks pertaining to infrastructure, such as food and water insecurity in households, neighborhood designs that inhibit active living, regional air and water pollution, and extreme heat, cold, and flooding that may be exacerbated by climate change (*6*). Socioeconomic disparities often shape exposure to the various risk factors and mediate and modulate the health outcomes. Addressing these diverse social, environmental, and infrastructural risk factors represents a new paradigm for urban public health. The World Health Organization (*7*) and the Centers for Disease Control and Prevention in the United States (*8*) recommend community-based participatory health planning that connects local capacities with infrastructure, using advanced information processing systems and ambient sensing. Making these connections is challenging, requiring overarching frameworks that connect diverse data and processes across scales to support action (*9*).

The environmental impacts of cities are numerous and multiscalar, driven by the transboundary nature of their infrastructure. Cities produce less than 10% of their food and rely on water, energy, fuel, and construction materials from external sources. Sustainable urban system studies are advancing urban supply-chain footprinting techniques that capture infrastructure use within the city (shaped by economic activity and urban form) in combination with transboundary infrastructure supply chains and trade networks. Such transboundary footprints (*10*) characterize the broader environmental impact of all seven urban infrastructure sectors on various resources (e.g., energy, water, and nutrients) and on pollution. Transboundary analysis considers how cities can reduce environmental impacts within their boundaries and across their supply chains and what the risks are to cities from supply disruptions. Cities are beginning to track their transboundary greenhouse gas footprints (*11*) while incorporating supply disruptions (e.g., power cuts, water scarcity, and food and fuel disruptions) into urban quality-of-life metrics (*12*).

"...a larger understanding of urban infrastructure systems is necessary to move from data to information to knowledge and, ultimately, to action for urban sustainability and human well-being."

The intertwined outcomes of environmental sustainability and human well-being require understanding interactions among all seven sectors within a city in terms of local-scale quality-of-life impacts; at the same time, it is also necessary to connect the transboundary infrastructures with regional and global environmental and health impacts, engaging multiple actors and institutions across scales (Fig. 2). Understanding and enhancing the capacity of social, policy, and governance networks therefore holds the key to change. Analyses of social norming and social networks are yielding insights about peer learning with respect to environment and health in cities; mapping of policy actors reveals how information is shared across sectors and scales (*13*). The five key dimensions outlined here—economic opportunity, urban form, social-infrastructural disparities and human well-being, transboundary infrastructure-environment dynamics, and cross-scale multisector governance—serve as the framework (Fig. 2) from which we draw principles for action.

Basic principles for transforming cities

1) Focus on providing and innovating basic infrastructure for all. Basic and affordable water, energy, sanitation, and transportation have long been recognized as important for all cities but have been difficult to attain in some cases, often because of rapid in-migration, unplanned urban expansions, and challenges in infrastructure financing. With 30 to 40% of the population in several cities in Asia and Africa living in slums (*2*), a healthy city must prioritize basic infrastructure for all. Many smart-city discussions focus on high technology, overlooking more basic, yet innovative, equitable solutions that are emerging, such as fit-for-purpose point-of-use household water treatment in Chinese cities (*14*), water "ATMs" in Indian cities, and prioritization to support nonmotorized transportation in compact mixed-use urban neighborhoods.

2) Pursue dynamic multisector and multiscalar urban health improvements, with attention to inequities. Cities must strive to address health priorities, which vary widely across cities, within cities, and over time, by considering infrastructural, environmental, and sociocultural determinants at different scales (*9*). Such an approach could yield, for example, regional weather and air pollution forecasts that provide customized messaging to vulnerable populations, neighborhood-level health interventions, more equitable access to nutritious food and green spaces, and greater attention to sociocultural

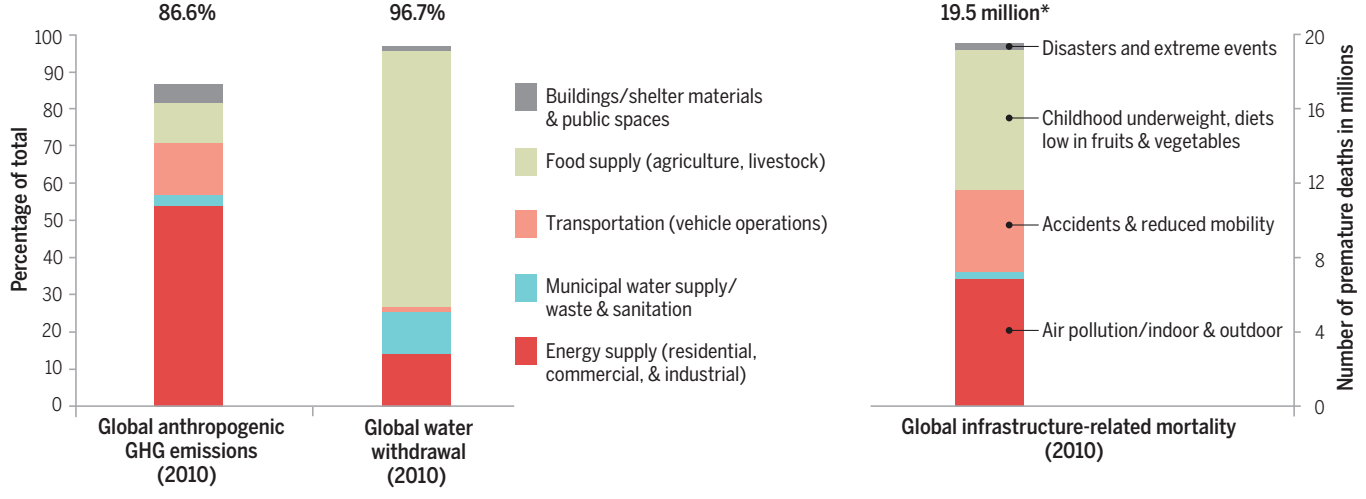


Fig. 1. Impacts of key infrastructure sectors. Shown are the impacts of urban infrastructure sectors on global anthropogenic GHG emissions (*20*), global water withdrawals (*21*, *22*), and global disease burden (*6*). GHG and water impacts associated with buildings and shelter materials include those of producing cement and steel, disaggregated by the authors based on various literature sources. Non-energy GHG emissions are shown for waste and sanitation. Transportation-related premature mortality is from accidents (*23*) and reduced mobility (*6*). The asterisk indicates that total deaths in the right column are non-additive because of overlap.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

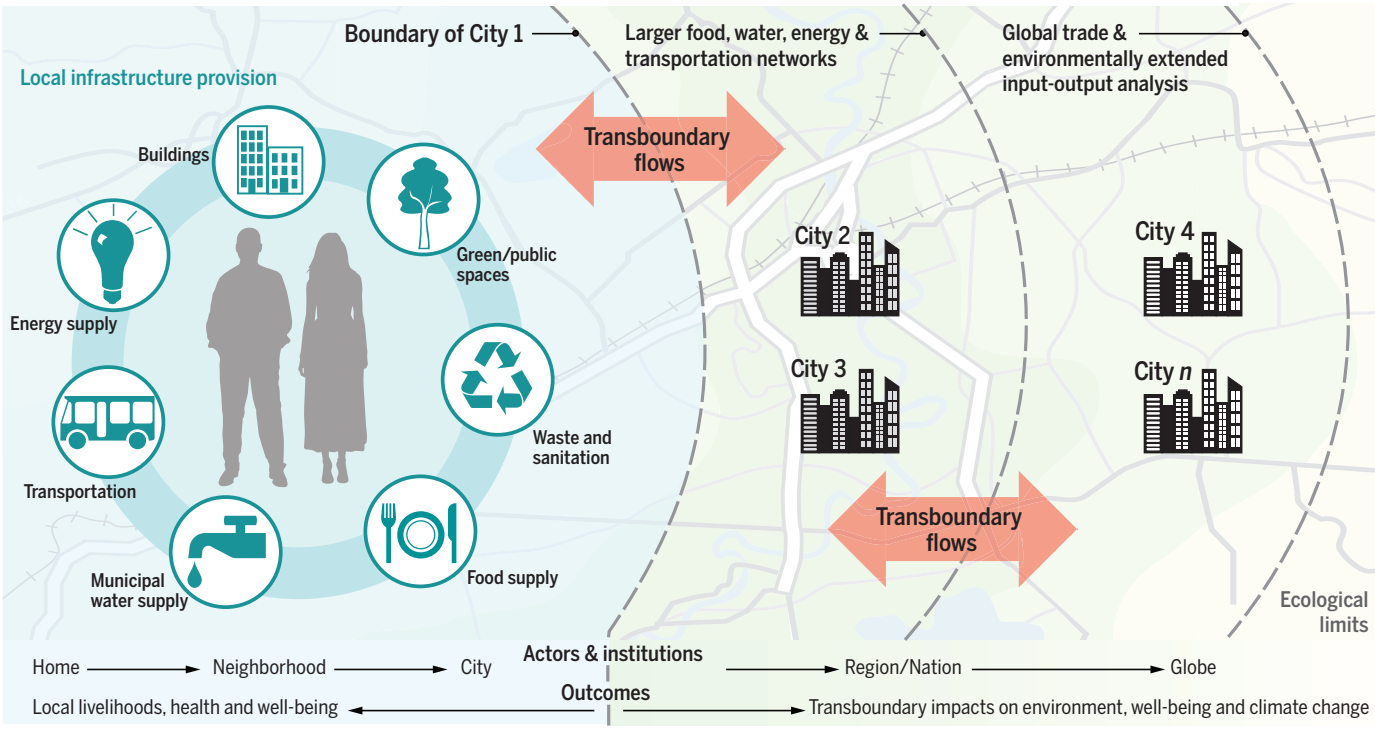


Fig. 2. Intersection of human activities and seven infrastructure sectors within a city, linked to natural ecosystems through transboundary infrastructures across scales. Actors and outcomes (health and sustainability) are also intertwined across scales.

assets that enhance quality of life and human well-being.

3) Focus on urban form and multisector synergies for resource efficiency. As populations urbanize, they become wealthier, increasing material consumption and environmental impact. To counter these effects, urban areas must increase resource efficiency, not by a few percentage points but by factors of 4 to 10. Such efficiency gains cannot come from single-sector interventions in a diffuse urban morphology. Research suggests that an optimally dense urban form, with a high intensity of diverse co-located activities, creates opportunities for systemic multisector infrastructure interventions, yielding the highest-efficiency gains. Advanced district energy systems that use energy cascading, exchange, and storage across industries, power plants, buildings, transportation, water, solid waste management, and renewable energy production offer tremendous potential (*15*). Knowledge of urban morphology, combined with temporal and spatial cross-sectoral infrastructure data, is essential.

4) Recognize diverse strategies for resource efficiency in different city types. A technology-oriented view of smart cities can result in translating high-efficiency solutions from one country or culture to another, where they may not work as well. For example, although tightly insulated, highly instrumented, all-day centrally cooled and heated buildings may be energy-efficient for the United States and the European Union, the same approach may not translate to the more vernacular architecture and informal user practices of Chinese apartments, which tend to be spot-cooled over short periods of time, greatly reducing resource intensity (*16*).

5) Integrate high- and vernacular technologies. Cities should seek local knowledge and systems-level understanding of different solution configurations. For example, municipal plants that convert solid waste to energy are not as effective in developing world cities. The waste streams have lower calorific value, having been sifted through by the informal sector of waste pickers who recycle more than 200 types of waste paper and plastics, which creates greater systems efficiency in terms of material cycling while also promoting local livelihoods. Formalizing and integrating the expertise of waste pickers with state-of-the-art information and waste-to-energy technologies can create hybrid solutions, illustrated, for example, by India's recently revised solid waste management regulations (*17*).

6) Apply transboundary systems analysis to inform decisions about localized versus larger-scale infrastructure. Driven by goals of local self-reliance, efficacy, and anticipated health and well-being benefits, cities are increasingly focusing on more localized infrastructures, such as rooftop and community solar installations, community-supported urban farms, and apartment-scaled wastewater treatment plants. Improved information about transboundary environmental footprints and local well-being impacts are critical to clarify synergies and trade-offs between local versus larger-scale infrastructure networks.

7) Recognize coevolution of infrastructures and institutions (*15*). Matching the scale of engineered infrastructures with that of the institutions with which they must operate is key. For example, neighborhood-scale urban farms, solar gardens, and waste management systems will require new levels of coordination among homes, neighborhood associations, businesses, and city- and state-level governments. At the same time, technology can change institutions; for example, widespread deployment of sensors is enabling remote surveillance of distributed water and wastewater systems. Awareness of the need for new and evolving institutions to manage emerging smart infrastructure can help ease these transitions.

8) Create capacity and transparent infrastructure governance across sectors and scales. Cities need capacity—analytic, administrative, and political—to implement high-impact, cross-sector, cross-scale solutions. Some cities have created sustainability offices that are empowered to convene multiple city departments, and many are leveraging multilevel and cross-national policy-exchange networks (*18*). With the smart-city agenda requiring high-technology expertise, greater involvement of the private sector in infrastructure delivery is inevitable. Many cities are initiating public-private partnerships and/or special financing for smart-city development. These arrangements raise questions about ownership, equity, and governance (*19*). It is equally important to ask where all the information that enables a smart, sustainable, and healthy city will reside. Transparent and adaptive governance arrangements that are open to public input and scientific study will empower cities, and the world, to learn by doing.

These eight principles focus attention on higher-order, systems-level decisions that society must

Downloaded from http://science.sciencemag.org/ on May 19, 2016

make to transition toward a smart, sustainable, and healthy urban future. To achieve the full potential of smart cities, discussions must move beyond data to envision cities as multisectoral, multiscalar, social-ecological-infrastructural systems with diverse actors, priorities, and solutions.

REFERENCES AND NOTES

1. K. C. Seto *et al.*, in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, O. Edenhofer *et al.*, Eds. (Cambridge Univ. Press, 2014), pp. 923–1000.
2. M. Swilling, B. Robinson, S. Marvin, M. Hodson, *City-Level Decoupling: Urban Resource Flows and the Governance of Infrastructure Transitions* (United Nations Environmental Program, 2013).
3. A. Ramaswami *et al.*, *J. Ind. Ecol.* **16**, 801–813 (2012).
4. L. Bourdic, S. Salat, C. Nowacki, *Build. Res. Inform.* **40**, 592–605 (2012).
5. L. M. A. Bettencourt, *Science* **340**, 1438–1441 (2013).
6. S. S. Lim *et al.*, *Lancet* **380**, 2224–2260 (2012).
7. World Health Organization (WHO), "Healthy cities," 2016; www.euro.who.int/en/health-topics/environment-and-health/urban-health/activities/healthy-cities.
8. Centers for Disease Control and Prevention, "Healthy Communities Program (2008–2012)," 2016; www.cdc.gov/nccdphp/dch/programs/healthycommunitiesprogram/.
9. Y. Rydin *et al.*, *Lancet* **379**, 2079–2108 (2012).
10. T. Hillman, A. Ramaswami, *Environ. Sci. Technol.* **44**, 1902–1910 (2010).
11. ICLEI-USA, *U.S. Community Protocol for Accounting and Reporting of Greenhouse Gas Emissions* (ICLEI-USA, 2016); http://icleiusa.org/publications/us-community-protocol/.
12. International Standards Organization, "Sustainable development of communities—Indicators for city services and quality of life," 15 May 2014; www.iso.org/iso/catalogue_detail?csnumber=62436.
13. T. Lee, S. van de Meene, *Policy Sci.* **45**, 199–220 (2012).
14. T. Tao, K. Xin, *Nature* **511**, 527–528 (2014).
15. H. Lund *et al.*, *Energy* **68**, 1–11 (2014).
16. International Energy Agency and Tsinghua University, *Building Energy Use in China: Transforming Construction and Influencing Consumption to 2050* (International Energy Agency, 2015).
17. Government of India, Ministry of Environment and Forests, press release (5 April 2016); http://pib.nic.in/newsite/PrintRelease.aspx?relid=138591.
18. H. Bulkeley, M. Betsill, *Env. Polit.* **14**, 42–63 (2005).
19. R. G. Hollands, *City* **12**, 303–320 (2008).
20. Intergovernmental Panel on Climate Change, *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, O. Edenhofer *et al.*, Eds. (Cambridge Univ. Press, 2014).
21. AQUASTAT, Food and Agriculture Organization, "Water uses," 2016; www.fao.org/nr/water/aquastat/water_use/index.stm).
22. International Energy Agency, "Water for energy," 2016; www.worldenergyoutlook.org/resources/water-energynexus/.
23. WHO, *Global Status Report on Road Safety 2013: Supporting a Decade of Action* (WHO, 2013); www.who.int/violence_injury_prevention/road_safety_status/2013/en/.

ACKNOWLEDGMENTS

The authors are grateful for support from NSF (Partnership for International Research and Education award 1243535 and Sustainability Research Networks award 1444745) and from the U.S. Agency for International Development and the National Academy of Sciences (Partnership for Enhanced Engagement in Research subgrant 2000002841). The principles outlined here were discussed at a workshop convened by ICLEI–South Asia and the University of Minnesota in January 2016, which was attended by more than 40 city officials and policy-makers from the United States, China, and India.

10.1126/science.aaf7160

PERSPECTIVE

Hidden linkages between urbanization and food systems

Karen C. Seto[1]* and Navin Ramankutty[2]*

Global societies are becoming increasingly urban. This shift toward urban living is changing our relationship with food, including how we shop and what we buy, as well as ideas about sanitation and freshness. Achieving food security in an era of rapid urbanization will require considerably more understanding about how urban and food systems are intertwined. Here we discuss some potential understudied linkages that are ripe for further examination.

The rise of urbanization is transforming food systems in many areas, including production on the farm, processing and packaging, distribution and retail, and consumption at the table. Cities concentrate people who live and work in close proximity and facilitate social interaction, which results in a greater variety of available food choices and creates new habits and tastes. Though a few links between urbanization and food systems are attracting increased attention from researchers, we discuss many poorly understood linkages that require further study if we are to achieve food security in an urban era.

Urbanization is a complex phenomenon that involves change in multiple dimensions, including a growing percentage of people who live in urban areas; the expansion of built environments; and changing norms, cultures, and ways of living. Most of what we know about how urbanization affects food systems is focused on two issues: the physical expansion of urban areas (food supply side) and changes in diet (food demand side). A widely expressed concern about urbanization is that expanding cities will result in widespread loss of croplands. Between 1970 and 2000, our global urban footprint expanded by 58,000 km^2 (*1*). By 2030, built-up areas are forecasted to nearly triple in size over the 2000 area, an increase of 1.2 million km^2 (*2*). However, croplands, which account for 12% of the world's ice-free land cover, exceed urban areas, which take up less than 3% (*3*). Though this difference in total area suggests that expansion of urban areas will have a minimal effect on Earth's total cropland, two qualifications are important.

First, because built-up areas are growing faster than urban populations in most parts of the world (*1*, *4*), cropland loss is likely to be acute in countries where urban population growth rates are high and the economy is largely agrarian. Thus, the loss of cropland due to urban expansion is likely to be more of a regional problem than global one. For the affected regions in particular, strategies are needed to manage urban expansion in ways that reduce pressure on farmland, inefficient land-use patterns, and leapfrog development. Toward these goals, the United Nations Human Settlements Programme (UN-Habitat) is supporting efforts to minimize urban sprawl by encouraging densification and more compact cities in countries such as Brazil, Ecuador, and Egypt. An alternate strategy is to safeguard agricultural land from development. The Canadian province of British Columbia established the Agricultural Land Reserve, a land-use zone intended to protect fertile agricultural land from development.

Second, given that cities historically developed in fertile agricultural areas, future expansion of built-up areas will probably encroach on productive agricultural land. This has already occurred in India, Vietnam, China, Turkey, and the United States, where urban expansion has resulted in the loss of prime agricultural land. Loss of the most productive agricultural land leads to decreased average cropland productivity (*5*), which is not a promising trend at a time when closing yield gaps is considered important (*6*, *7*). Furthermore, the loss of agricultural land may result in the need to expand crop production into marginal and fragile lands, as well as other ecosystems. In the tropics, deforestation due to agricultural land expansion is estimated at 12 million ha per year (*8*). Notwithstanding these qualifications, the effect of urban expansion on farmland loss is unlikely to be the most important influence of urbanization on food systems.

On the demand side, it is well documented that the diets of urban and higher-income societies require costly expenditures of land, water, and energy (*9*, *10*). With few exceptions, highly urbanized countries consume more animal protein—in the form of pork, poultry, beef, and dairy products—than the world average. Globally, the average meat consumption per capita is 36.9 kg per year, but there is huge variation between countries, with higher-income countries consuming 81.8 kg per year per capita compared with 17.2 kg per year in lower-income countries (*9*). In the United States, where 81% of the population lives in urban areas, per-capita meat consumption is 89.7 kg per year. In China, per-capita

[1]Yale School of Forestry and Environmental Studies, Yale University, New Haven, CT 06511, USA. [2]Liu Institute for Global Issues and Institute for Resources, Environment and Sustainability, The University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada.
***Corresponding author. Email: karen.seto@yale.edu (K.C.S.); navin.ramankutty@ubc.ca (N.R.)**

Downloaded from http://science.sciencemag.org/ on May 19, 2016

meat consumption quadrupled between 1970 and 2015, a period when the urban population increased from 17.4 to 55.6%. Other countries experiencing high rates of urbanization are undergoing similar dietary transitions (*11*, *12*). It is important to point out that changes in diet are not limited to increased meat consumption but also include more nuanced shifts in quality, composition, and sources of caloric intake. In general, changes in diet can be characterized as a shift from complex carbohydrates, grains, vegetables, and fruits to a higher proportion of animal proteins, refined fats, refined sugars, alcohols, and oils (*13*).

The exact mechanisms that undergird these dietary shifts are not well understood, and many studies conflate urbanization, rises in income, and westernization. More affluent households can afford higher-quality diets with greater variety, the means to physically reach grocery stores that have better selection, and nutritional knowledge. Disentangling the income or economic development effect from the urban effect on diets will require substantially more research that goes beyond comparison of urban diets to rural ones. Given that urbanization is a major global trend, we need research that helps us isolate the urban effects on diet. For example, how does urban spatial form affect food demand? How are our food choices shaped by where we live and work, their relative locations, and our travel behavior? How does living in large mega-urban regions affect our exposure and access to foods, as well as our food preferences and choices? How does urban living influence our exposure to "foodstyles," such as veganism, organic, and local foods, as well as our purchasing habits and diets? We need to move beyond understanding urbanization's direct effects on diet and toward examining its indirect effects on resource use and the environment, such as embodied energy for food production, transport, packaging, cold storage, and the rest of the entire chain from farm to fork.

> ***"We need to move beyond understanding urbanization's direct effects on diet and toward examining its indirect effects on resource use and the environment, such as embodied energy for food production, transport, packaging, cold storage, and the rest of the entire chain from farm to fork."***

Urbanization could drive profound changes in food systems in ways unrelated to rises in income, such as through shifts in norms and attitudes about food; increasing opportunity cost of time and the necessity of convenience; growing availability and access to modern energy; and changes in the structure of the economy, household dynamics, the workforce, and urban form and infrastructure (Fig. 1). The pace of life in cities is fast, and with the onset of the 24-hour workday and more women joining the workforce, there is a growing need to reduce meal cooking time. This creates a demand for time-saving food preparation, from prepackaged items to reduced shopping time; meals away from home; and convenience foods on the go (*14*, *15*). Around the world, more urbanized societies eat more meals away from home (Fig. 2). Urban and food systems coevolve: Increased convenience allows people to work longer, furthering the demand for convenient foods. Although these foods reduce meal preparation time, they also increase the need for preprocessing, packaging, and refrigeration. In 2005, food container and packaging waste accounted for nearly one-third of total municipal solid waste in the United States (*16*).

In addition to packaging waste, food losses and waste place enormous burdens on the environment by contributing to greenhouse gas emissions, eutrophication, and the waste of resources used to produce food that is not eaten

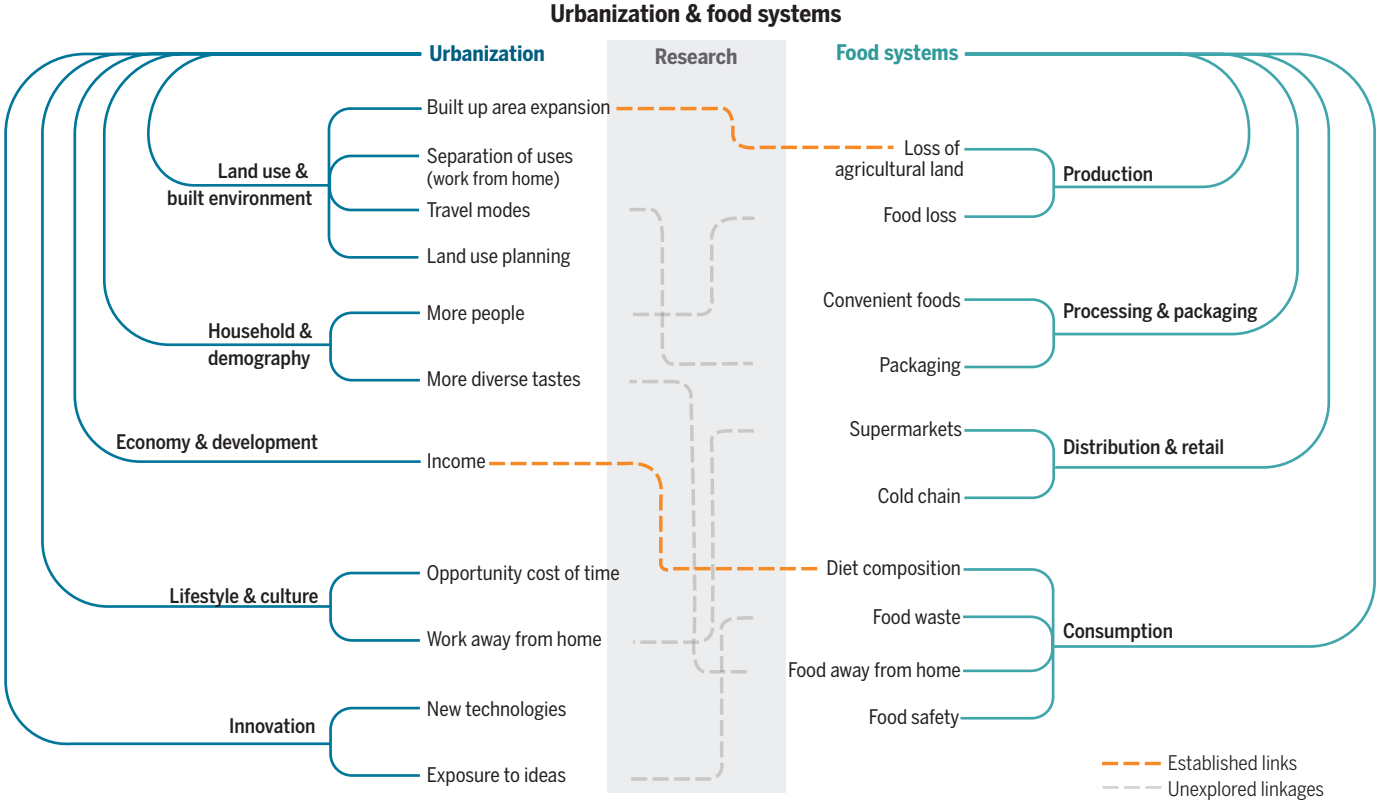


Fig. 1. Established and underexplored linkages between urbanization and food systems. The underexplored linkages are illustrative and not exhaustive.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

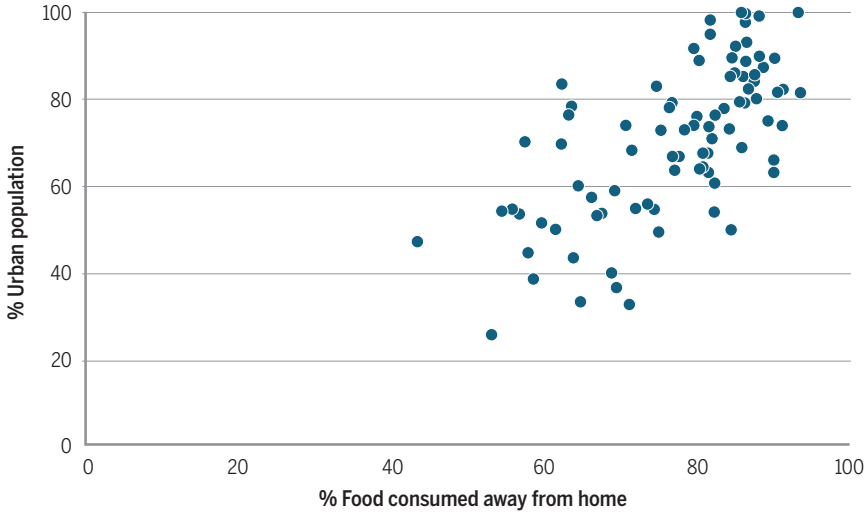


Fig. 2. Percentage of the national population living in urban areas versus percentage of consumer expenditures on food consumed away from home for 86 countries in 2014. Urban population estimates are from the World Bank; estimates for shares of food consumed at home are from the Economic Research Service of the U.S. Department of Agriculture (we inferred the percentage of food consumed away from home to be 100% minus the percentage of food consumed at home).

(*17*). Food loss and waste within the food system depend on many factors, but the level of economic development is a strong predictor: Preconsumer food losses within the supply chain are more prevalent in developing countries, whereas postconsumer food waste is a bigger issue in high-income countries (*18*). Urbanization may actually be beneficial for reducing food losses through improvements in technology, such as cold storage and transportation. There are many reasons why people in the developed world waste so much food: We overbuy and then overprepare and discard excess food. This is overwhelmingly a problem of the affluent, where food is a small percentage of the total household budget, thus reducing the cost of discarding food. Again, as with diets, disentangling the income effect from the urban effect on food loss and waste will require far more research. For example, how much does food consumed away from home, which is strongly associated with urban living, influence food waste? How do mixed-use urban forms, with close access to grocery stores, allow people to purchase food when they need it and reduce household food waste?

With greater numbers of people and activities, urban areas also require more institutions to maintain law and order, sanitation, and public hygiene. More rules and changes in attitudes come together to create new norms and regulations about food safety. Though some of these regulations (such as cold-storage requirements) may appear benign, together they constitute a substantial shift in where and how food is stored, packaged, and sold. In many Asian cities, wet markets are giving way to more sanitized farmers markets and multinational supermarkets, partly due to concerns about the avian flu but also because of changing attitudes around freshness and modernity.

Where they exist, urban planning and land-use zoning regulations greatly affect food retail diversity and accessibility. Transport planning, modes of mobility, street connectivity, and land-use mix affect how often people shop for food, what they buy, and where they eat. In cities where street connectivity is high—characterized by smaller blocks, shorter distances, and many intersections—there is a positive correlation with walking and, thus, more convenience, higher accessibility, and lower energy costs of food procurement. Land-use regulations are central to creating retail "foodscapes" that are diverse by either enabling small-scale, independent food shops and restaurants or limiting food commerce to stores or chains that can afford higher rents and fill large retail spaces.

Urban areas are also hubs of innovation, where new technologies and ideas about food systems emerge, such as vertical farming, entomophagy, the slow food movement, farmers markets, and community-supported agriculture. Urban dwellers are increasingly disconnected from the agricultural land base but are finding creative ways to remain connected to the food system. Changes in urban consumer preferences can have far-reaching consequences. For example, European preferences for non–genetically modified foods have contributed to the rapid expansion of South American farming and the development of entirely new supply chains (*19*). The rapid growth of organic farming is driven by increasing demand from urban consumers in North America and Europe, whereas the majority of organic farmers are located in developing countries (*20*).

Of all the myriad ways in which urbanization affects different parts of the food system, the effects on the people who produce food are perhaps some of the least studied. There are an estimated 500 million smallholder farms in developing countries providing livelihoods for almost 2 billion people (*21*). Urban expansion may displace these farmers into marginal lands, and the increase in land value may encourage them to sell their farms and seek alternate means of livelihood, possibly in urban areas, leading to even greater rural-to-urban migration. Further, as cities grow larger, the food retail market becomes increasingly dominated by large supermarket chains, and this has secondary effects on traditional retailers, small-scale producers, traditional food brokers, and the entire supply chain (*22*).

Most discussions of food security focus on how to sustainably provide for a world of 9 to 10 billion people in 2050. Much less attention has been paid to the fact that ~6.5 billion of these people will live in urban areas and that the transformation toward a more urbanized society will have many consequences for food systems. The overwhelming effects will be indirect. To achieve food security in the urban century, we urgently need more research that unravels how these two systems are linked.

REFERENCES

1. K. C. Seto, M. Fragkias, B. Güneralp, M. K. Reilly, *PLOS ONE* **6**, e23777 (2011).
2. K. C. Seto, B. Güneralp, L. R. Hutyra, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16083–16088 (2012).
3. N. Ramankutty, A. T. Evan, C. Monfreda, J. A. Foley, *Global Biogeochem. Cycles* **22**, GB1003 (2008).
4. S. Angel, J. Parent, D. L. Civco, A. Blei, D. Potere, *Prog. Plann.* **75**, 53–107 (2011).
5. F. Martellozzo *et al.*, *Reg. Environ. Change* **15**, 881–893 (2015).
6. H. C. J. Godfray *et al.*, *Philos. Trans. R. Soc. London Ser. B* **365**, 2769–2777 (2010).
7. J. A. Foley *et al.*, *Nature* **478**, 337–342 (2011).
8. E. B. Barbier, *Am. J. Agric. Econ.* **86**, 1347–1353 (2004).
9. A. J. McMichael, J. W. Powles, C. D. Butler, R. Uauy, *Lancet* **370**, 1253–1263 (2007).
10. B. M. Popkin, *World Dev.* **27**, 1905–1916 (1999).
11. P. Pingali, *Food Policy* **32**, 281–298 (2007).
12. B. M. Popkin, S. Du, *J. Nutr.* **133** (suppl. 2), 3898S–3906S (2003).
13. D. Tilman, M. Clark, *Nature* **515**, 518–522 (2014).
14. J. F. Guthrie, B.-H. Lin, E. Frazao, *J. Nutr. Educ. Behav.* **34**, 140–150 (2002).
15. H. Ma, J. Huang, F. Fuller, S. Rozelle, *Can. J. Agric. Econ.* **54**, 101–119 (2006).
16. K. Marsh, B. Bugusu, *J. Food Sci.* **72**, R39–R55 (2007).
17. M. Kummu *et al.*, *Sci. Total Environ.* **438**, 477–489 (2012).
18. J. Aschemann-Witzel, *Science* **352**, 408–409 (2016).
19. R. D. Garrett, X. Rueda, E. F. Lambin, *Environ. Res. Lett.* **8**, 044055 (2013).
20. H. Willer, L. Kilcher, Eds., "The World of Organic Agriculture: Statistics and Emerging Trends 2011" [International Foundation for Organic Agriculture Movements and Research Institute of Organic Agriculture (FiBL), 2011]; http://orgprints.org/19310/1/world-of-organic-agriculture-2011.pdf.
21. High Level Panel of Experts on Food Security and Nutrition (HLPE), "Investing in smallholder agriculture for food security. A report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security" (Food and Agriculture Organization of the United Nations, 2013); www.fao.org/fileadmin/user_upload/hlpe/hlpe_documents/HLPE_Reports/HLPE-Report-6_Investing_in_smallholder_agriculture.pdf.
22. T. Reardon, C. P. Timmer, C. B. Barrett, J. Berdegue, *Am. J. Agric. Econ.* **85**, 1140–1146 (2003).

10.1126/science.aaf7439

Downloaded from http://science.sciencemag.org/ on May 19, 2016

PERSPECTIVE

Building functional cities

J. Vernon Henderson,[1]* Anthony J. Venables,[1,2] Tanner Regan,[1] Ilia Samsonov[1]

The literature views many African cities as dysfunctional with a hodgepodge of land uses and poor "connectivity." One driver of inefficient land uses is construction decisions for highly durable buildings made under weak institutions. In a novel approach, we model the dynamics of urban land use with both formal and slum dwellings and ongoing urban redevelopment to higher building heights in the formal sector as a city grows. We analyze the evolution of Nairobi using a unique high–spatial resolution data set. The analysis suggests insufficient building volume through most of the city and large slum areas with low housing volumes near the center, where corrupted institutions deter conversion to formal sector usage.

Most cities in the developed world use land in an orderly pattern that allows cities to achieve high productivity. For example, businesses mainly reside in a central business district (CBD), and residential neighborhoods have regular layouts with high densities near the center and lower densities further out (*1*). In contrast, many cities in developing nations have office towers bordered by slums, scattered fringe developments, and a consequent lack of connectivity between firms, workers, and consumers. Such cities are viewed as nonfunctional (*2*), with large numbers of people in informal settlements [62% of the African urban population according to (*3*)], poor transport infrastructure and limited ability to commute (*4*), and low worker productivity (*5*).

Here, we explore factors that may underlie nonfunctionality of many cities in the developing world. We analyze how construction decisions made under weak and corrupt institutions can be a driver of nonfunctionality. The built environment resulting from these decisions accounts for two-thirds of produced capital in developing countries (*6*) and is long lived. As such, weak institutions undermine the competitiveness of cities, and thus, bad decisions made today have effects that last for generations. We first discuss recent model results and then use Nairobi, Kenya—a city of about 5 million people that is growing at a rate of 3 to 4% per year—to map out how the built environment has changed and to explore ways in which it appears to deviate from an efficient pattern, with insufficient building volume through most of the city.

In a recent study (hereafter referred to as HRV) (*7*), we developed a general model of the dynamics of economically efficient urban land use and of key elements that impede efficient urban development. To do so, we adapted a standard urban model to a growth context and the circumstances of developing countries. The model captures rapid population growth and two types of housing technology: Formal housing, in which capital is sunk, buildings are long-lived, and construction decisions (such as building height) are based on expectations of future rents; and informal, or slum settlement, where construction is flexible or adjustable over time (e.g., through use of corrugated iron sheets), building a single story is cheap, but building high is very expensive. This distinction is illustrated in Nairobi, where 57% of slum dwellings are made of sheet metal and 15% of mud and wood, whereas 90% of formal residences are made of stone, brick, or cement block (*8*).

In the efficient outcome in the model, slums form at the edge of the city, where land is cheap. As the city grows, old slums are converted to formal settlement and new ones form on the expanding edge. Formal sector development is subject to periodic demolition and reconstruction, and structures become successively taller and denser as the city grows and land values increase. If slum housing is inherently of lower quality, then slums will eventually be phased out entirely as incomes grow, just as 19th-century tenements and shacks in London and New York disappeared decades ago. Our model (*7*) analyzes two main sources of inefficiency in the dynamics of city development. One arises from the difficulty of forming expectations; for example, pessimism about future city growth undermines willingness to invest and leads to a lower, more sprawling city. The other is institutional obstacles in the process of converting slum developments to formal sector usage.

There are many such institutional obstacles. Formal sector development requires financing and enforcement of contracts, which in turn requires land ownership rights to be formalized to mitigate the risk of expropriation. Land rights are often unclear because of coexisting systems of private ownership (some illegal or quasi-legal), communal ownership, and government ownership. Competing claims may result in lengthy court cases. Slum areas are particularly complex, with "planning or regulatory powers... split between a galaxy of private sector actors, landlords, chiefs and bureaucrats, and gangs" (*9*). Land administration is subject to corruption. The Kenyan elite has been guilty of land-grabbing, with a government inquiry alleging that the land allocation process has been subject to corrupt and fraudulent practices and "outright plunder" (*10*). As a result, the cost and feasibility of conversion to legal formal usage varies depending on a plot's history; plots with high conversion costs remain informal much longer. A spatial jumble of land rights and conversion costs results in a hodgepodge of uses, land-use intensities, and stages of redevelopment throughout the city, including close to the center.

Studying these inefficiencies requires data on individual buildings and the ability to track them through time to quantify the potential loss of building space. Such data are generally difficult to obtain. For our Nairobi study, we used a building footprint data set based on extremely high-resolution aerial photos (well under 50-cm

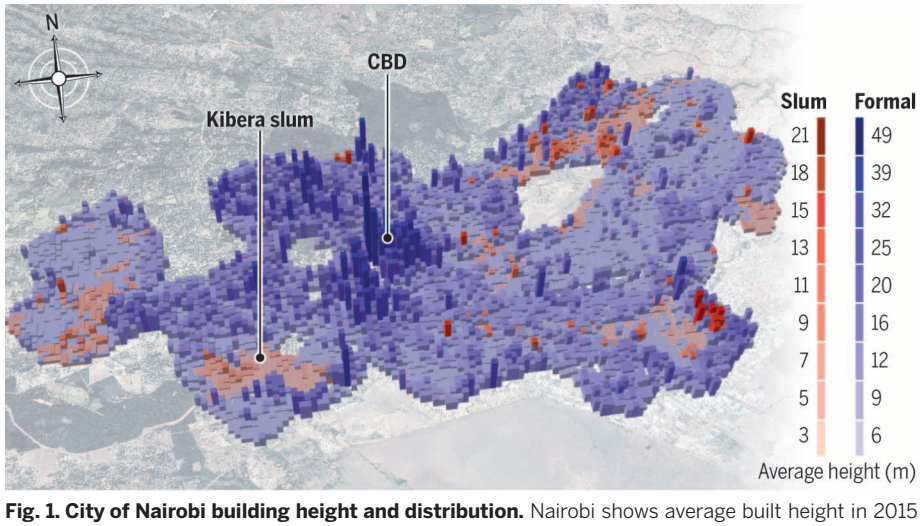


Fig. 1. City of Nairobi building height and distribution. Nairobi shows average built height in 2015 as 150-m by 150-m cells split across the formal and slum sectors. The compass (top left) points north. The location of the Kibera slum and the CBD are marked. The boundary of the city spans about 22 km east to west and 11 km north to south; the map tilt may distort the appearance of distances. Modified from HRV. [Background imagery Airbus Defense and Space 2016, taken from the SPOT5 satellite 20 September 2004].

[1]London School of Economics and Political Science, London WC2A 2AE, UK. [2]Department of Economics, Oxford University, Oxford OX1 3UQ, UK.
*Corresponding author. Email: j.v.henderson@lse.ac.uk

Downloaded from http://science.sciencemag.org/ on May 19, 2016

resolution), which allows clear demarcation even of buildings in slums, satellite data to derive road coverage, and LIDAR data for building heights (*7*). We used city studies that mapped slums in 2003–04 (*11*) and slums and ownership or land rights in 2012 (*12*). For the later time period, house and land prices are available from surveys or scraping the Web. We developed novel methods to integrate and analyze these data, including overlaying building footprints of the city at different points in time to define infill, reconstruction, demolition, and no change.

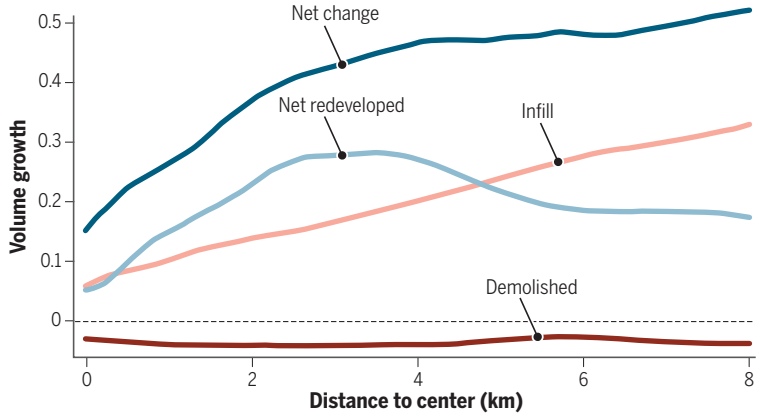


Fig. 2. City of Nairobi volume change by sector since 2004. Formal sector volume change from 2004 to 2015 as a fraction of initial volume by distance. The ratio of net volume to initial volume is broken down into change due to infill redevelopment and demolition as defined in HRV. This sample excludes cells that had no buildings in both 2015 and 2004. Modified from HRV.

As can be seen in the three-dimensional map (Fig. 1), building heights vary widely throughout Nairobi, reflecting formal and informal housing. The city is monocentric, constrained by national parks to the north and south; undefined spaces within the city include an airport, golf course, and the President's complex. Slums include the 1000-acre slum of Kibera, to the south-west of the city center that we discuss below.

Land prices, as recorded in 2015, decline sharply with distance from the CBD. There are no slums in the CBD, and the proportion of developed land occupied by slums peaks at 45% at a distance of 5 km out from the CBD. In Nairobi, slums are not concentrated at the edge (which, according to the model, would be economically efficient), although city mapping may underrepresent emerging slums at the fringe. More to the point, slums appear in a scattered fashion throughout, even near the CBD, which indicates potentially important land market frictions. Building heights in the formal sector average about 23 m in the CBD, in contrast to expectations that there would be overall less height in Nairobi. Heights fall to about 6 to 7 m at a distance of 10 km from the CBD. Slums have similar height throughout the city and, at about 5 m, are less tall than formal buildings, as modeled. Despite the lack of roads and green space and the intense crowding of buildings in slums, height in the formal sector trumps intense footprint coverage in slums, so that building volume (height × footprint) per unit land in slums is always lower than in formal developments. An implication is that the presence of slums near the CBD has a large impact on building volume. At 2, 3, and 4 km from the center, conversion of slums to formal usage would increase building volumes in those slum areas by 148%, 95%, and 53%, respectively, one indication of potentially inefficient land use.

Turning to dynamics, we determined the volume of infill (new buildings where there were none in 2004), net redevelopment (new building where there had been an earlier structure), and demolition (buildings demolished and not yet replaced) as a fraction of initial volume (Fig. 2). In the 0- to 1-km ring at the CBD, use is locked in by roads, colonial buildings, and tall complexes built over the last 40 years. Total road area declines sharply with distance from the center, as modeled by Solow and Vickrey (*13*). Between 1 and 5 km, as the model suggests, there is substantial net redevelopment in the face of escalating land prices, with new buildings taller than their older neighbors. The volume of net redevelopment peaks at 4 km out, where it amounts to just under 30% of old volume. Beyond 5 km, volume changes are dominated by infill.

What about slums? Up to about 2 km from the CBD, slums are demolished and redeveloped. Beyond that there is less redevelopment than might be expected. Why? In HRV, we argue that remaining slums nearer the CBD, like Kibera, have high costs of conversion to formal usage. Slum land near the center, including Kibera, is government-owned (*12*)—a code word for conflicting private claims, with the government having seized ownership but not responsibility. Slum landlords there make high profits and much of the land is controlled by political figures with a vested interest not to develop the land; redevelopment would take away their profits on land to which they have no legal claim. Nearer the fringe, land ownership in slums becomes increasingly private.

The constraint on slum redevelopment nearer the center has significant welfare costs: There is lost volume of space due to not building high as we described above, and the quality of the built space and unit rents are low compared with those of the formal sector. We hypothesize that slum landlords have invested little in land improvements, such as infrastructure and regularized lay-out near the center, because they cannot capture those returns when housing spaces are redeveloped. A simple calculation (reported in HVR) suggests that lack of redevelopment reduces land values in Kibera by the order of about $1 billion. Such a magnitude of potential gain from redevelopment indicates the potential for a political solution: to buyout the actors inhibiting redevelopment and to help relocate tenants.

In summary, Nairobi has many of the features of a "normal" city: High buildings in the CBD and declining heights and land prices away from the center. Yet there is substantial evidence of inefficient land-use. The low–volume intensity of slums and the persistence of slums relatively close to the center lead to a substantial loss of housing capacity. We argue that such persistence is due to the myriad of institutional and political obstacles to redevelopment.

Although the data are specific to Nairobi, the modeling and analysis are more general. Weak and corrupt land market institutions are common throughout much of Africa, which suggests that aspects of our findings for Nairobi have more general applicability. Whereas the focus has been the built environment and much of data is a view from the sky, our continuing work combines this with economic and population censuses and surveys, in order to give detail for what is happening to people and firms on the ground. Combining data sources and institutional details of specific cities will help inform urban policy to improve functionality, as the African urban population trebles over the next three decades.

REFERENCES

1. G. Duranton, D. Puga, in *Handbook of Regional and Urban Economics*, G. Duranton, J. V. Henderson, W. Strange, Eds. (Elsevier, Philadelphia, 2015), vol. 5, pp. 467–560.
2. S. Coulibaly *et al.*, "World development report 2009: Reshaping economic geography" (Report 43738, World Bank Group, Washington, DC, 2008).
3. United Nations Human Settlements Programme, *State of the Worlds' Cities 2012/2013: Prosperity of Cities* (UN-HABITAT, Nairobi, Kenya, 2012).
4. Trans-Africa Consortium, *Public Transport in Sub-Saharan Africa: Major Trends and Case Studies* [International Association of Public Transport (UITP), Brussels, 2010].
5. M. Fay, C. Opal, "Urbanization without growth: a not-so-uncommon phenomenon" (Policy Research Working paper series 2412, The World Bank, 2010).
6. World Bank Group, *Where Is the Wealth of Nations?: Measuring Capital for the 21st Century* (Report 34855, World Bank Group, Washington, DC, 2005).
7. J. V. Henderson, T. Regan, A. Venables, "Building the city: Sunk capital and sequencing" (Discussion paper 11211, Centre for Economic Policy Research, London, 2016).
8. C. Zinnes *et al.*, Kenya Urban Program Baseline Study (NORC, University of Chicago, Chicago, 2012); www.norc.org/Research/Projects/Pages/kenya-urban-program-baseline-study.aspx.
9. B. Marx, T. Stoker, T. Suri, *J. Econ. Perspect.* **27**, 187–210 (2013).
10. R. Southall, *Rev. Afr. Polit. Econ.* **32**, 142–151 (2005).
11. S. Williams, E. Marcello, J. M. Klopp, *Ann. Assoc. Am. Geogr.* **104**, 114–130 (2014).
12. World Bank, "Resettlement Action Plan report for KCC (Embakasi) Informal Settlement in Nairobi County" (Report RP1059, World Bank Group, Washington, DC, 2014).
13. R. M. Solow, W. S. Vickrey, *J. Econ. Theory* **3**, 430–447 (1971).

18 March 2016; accepted 26 April 2016
10.1126/science.aaf7150

Downloaded from http://science.sciencemag.org/ on May 19, 2016

RESEARCH ARTICLE SUMMARY

CANCER IMMUNOTHERAPY

The tumor microenvironment underlies acquired resistance to CSF-1R inhibition in gliomas

Daniela F. Quail, Robert L. Bowman, Leila Akkari, Marsha L. Quick, Alberto J. Schuhmacher, Jason T. Huse, Eric C. Holland, James C. Sutton, Johanna A. Joyce*

INTRODUCTION: Therapies targeted against the tumor microenvironment (TME) represent a promising approach for treating cancer. This appeal arises in part from the decreased likelihood of acquired resistance through mutations in target TME cells, as is frequently observed with cancer cell–targeted therapies. Although classical mechanisms of tumor cell–intrinsic resistance to cytotoxic and targeted agents have been well-defined—including aberrant drug metabolism and transport, drug target mutation, and activation of alternative survival pathways—it still remains unclear whether resistance to TME-directed therapies follows similar principles. Given that TME-targeted agents are increasingly being evaluated in the clinic, it is becoming critical to mechanistically define how resistance may evolve in response to these therapies in order to provide long-term disease management for patients.

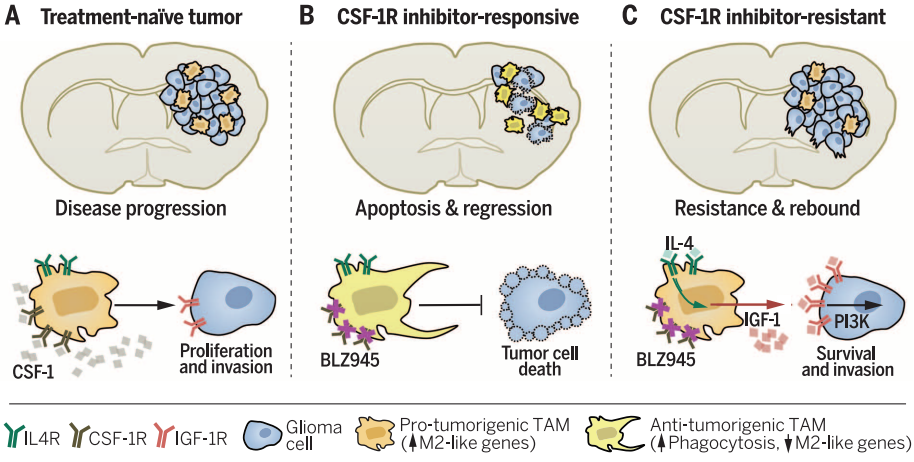
RATIONALE: Macrophages and microglia are of the most abundant noncancerous cell types in glioblastoma multiforme (GBM), in some cases accounting for up to 30% of the total tumor composition. Macrophages accumulate with GBM progression and can be acutely targeted via inhibition of colony-stimulating factor–1 receptor (CSF-1R) to regress high-grade gliomas in animal models. However, it is currently unknown whether and how resistance emerges in response to sustained CSF-1R blockade in GBM. Despite this, multiple clinical trials are currently underway testing the efficacy of CSF-1R inhibition in glioma patients. Therefore, determining whether long-term CSF-1R inhibition can stably regress GBM by using animal models is an important and timely question to address.

RESULTS: Using genetic mouse models of GBM, we show that although overall survival is significantly prolonged in response to CSF-1R inhibition, tumors recur eventually in >50% of mice. Upon isolation and transplantation of recurrent tumor cells into naïve animals, gliomas reestablish sensitivity to CSF-1R inhibition, indicating that resistance is microenvironment-driven. Through RNA-sequencing of glioma cells and macrophages purified from treated tumors and ex vivo cell culture assays, we found elevated phosphatidylinositol 3-kinase (PI3K) pathway activity in recurrent GBM after CSF-1R inhibition, driven by macrophage-derived insulin-like growth factor–1 (IGF-1) and tumor cell IGF-1 receptor (IGF-1R). Consequently, combining IGF-1R or PI3K blockade with continuous CSF-1R inhibition in recurrent tumors significantly prolonged overall survival. In contrast, monotherapy with IGF-1R or PI3K inhibitors in rebound or treatment-naïve tumors was less effective, indicating the necessity of combination therapy to expose PI3K signaling dependency in recurrent disease. Mechanistically, we found that activation of macrophages in recurrent tumors by IL4 led to elevated Stat6 and nuclear factor of activated T cells (NFAT) signaling upstream of Igf1, and inhibition of either of these pathways in vivo was sufficient to significantly extend survival.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aad3018

CONCLUSION: We have identified a mechanism of drug resistance that can circumvent therapeutic response to a TME-targeted therapy and promote disease recurrence in the absence of tumor cell–intrinsic alterations. Specifically, we have uncovered a heterotypic paracrine signaling interaction that is initiated by the TME and drives resistance to CSF-1R inhibition through IGF-1R/PI3K signaling. Given that PI3K signaling is aberrantly activated in a substantial proportion of GBM patients, and that recent clinical trial results show limited efficacy in recurrent (albeit very advanced) GBM, it is possible that this pathway could similarly contribute to intrinsic resistance to CSF-1R inhibition. Our findings underscore the importance of bidirectional feedback between cancer cells and their microenvironment and support the notion that although stromal cells are less susceptible to genetic mutation than are cancer cells, a tumor can nonetheless acquire a resistant phenotype by exploiting its extracellular environment. ■



Resistance to CSF-1R inhibition in glioma. (**A**) Macrophages contribute to GBM progression by creating a protumorigenic niche associated with M2-like gene expression. CSF-1R is a critical receptor for macrophage biology and is under clinical evaluation as a therapeutic target in glioma. (**B**) Targeting CSF-1R early in gliomagenesis significantly prolongs survival in mouse models. CSF-1R inhibition reprograms macrophages to become antitumorigenic by down-regulating M2-like genes and enhancing phagocytosis. Tumor-derived survival factors sustain macrophage viability despite CSF-1R blockade. (**C**) After prolonged treatment, a subset of GBMs acquire resistance to CSF-1R inhibition, and tumors recur. This is driven by elevated macrophage-derived IGF-1 and high IGF-1R on tumor cells, resulting in PI3K pathway activation and enhanced glioma cell survival and invasion. Blocking this pathway and CSF-1R in preclinical trials yields a substantial survival benefit.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: johanna@joycelab.org
Cite this article as D. F. Quail *et al.*, *Science* **352**, aad3018 (2016). DOI: 10.1126/science.aad3018

Downloaded from http://science.sciencemag.org/ on May 19, 2016

RESEARCH ARTICLE

CANCER IMMUNOTHERAPY

The tumor microenvironment underlies acquired resistance to CSF-1R inhibition in gliomas

Daniela F. Quail,[1] Robert L. Bowman,[1] Leila Akkari,[1,2,3] Marsha L. Quick,[1] Alberto J. Schuhmacher,[1] Jason T. Huse,[4] Eric C. Holland,[5] James C. Sutton,[6] Johanna A. Joyce[1,2,3]*

Macrophages accumulate with glioblastoma multiforme (GBM) progression and can be targeted via inhibition of colony-stimulating factor–1 receptor (CSF-1R) to regress high-grade tumors in animal models of this cancer. However, whether and how resistance emerges in response to sustained CSF-1R blockade is unknown. We show that although overall survival is significantly prolonged, tumors recur in >50% of mice. Gliomas reestablish sensitivity to CSF-1R inhibition upon transplantation, indicating that resistance is tumor microenvironment–driven. Phosphatidylinositol 3-kinase (PI3K) pathway activity was elevated in recurrent GBM, driven by macrophage-derived insulin-like growth factor–1 (IGF-1) and tumor cell IGF-1 receptor (IGF-1R). Combining IGF-1R or PI3K blockade with CSF-1R inhibition in recurrent tumors significantly prolonged overall survival. Our findings thus reveal a potential therapeutic approach for treating resistance to CSF-1R inhibitors.

Therapies targeted against the tumor microenvironment (TME) represent a promising approach for treating cancer. This appeal arises in part from the decreased likelihood of acquired resistance through mutations in target TME cells, as is frequently observed with cancer cell–targeted therapies. Because multiple TME-directed therapies are currently advancing through different clinical trials (*1*, *2*), this necessitates an understanding of potential mechanisms of intrinsic or acquired resistance. We have focused on addressing this issue by investigating whether resistance to a macrophage-targeted therapy emerges during the course of long-term trials in various preclinical models of high-grade glioma [glioblastoma multiforme (GBM)].

GBM is the most common and aggressive adult primary brain tumor, and survival is only minimally prolonged by current standard-of-care treatment, including surgery, radiation, and temozolomide chemotherapy (*3*). Accordingly, targeting the glioma TME is emerging as a promising alternative therapeutic strategy. In GBM, tumor-associated macrophages and microglia (TAMs) comprise up to 30% of the bulk tumor mass (*4*). In many cancers, including glioma, elevated TAM numbers are associated with high-grade and poor patient prognosis (*4*–*7*). As such, targeting TAMs in GBM represents an attractive therapeutic approach.

Macrophages critically depend on colony-stimulating factor–1 (CSF-1) for multiple functions; consequently, strategies to target TAMs often include CSF-1 receptor (CSF-1R) blockade (*8*–*10*). In clinical trials, several approaches to inhibit CSF-1R are currently being used, including antibodies and small molecules (*7*, *11*, *12*). However, the long-term effects of these agents on clinical outcome are still under evaluation, and thus, gaining insight into potential mechanisms of drug resistance and/or inefficacy is now critical.

We used a potent and highly selective small-molecule CSF-1R inhibitor, BLZ945. We have shown that BLZ945 blocks early gliomagenesis, and that short-term treatment of advanced, high-grade glioma causes robust tumor debulking after just 7 days (*8*). CSF-1R inhibition has no direct effect on glioma cell viability because these cells do not express CSF-1R in the models we have used. Instead, glioma TAMs remain abundant and become antitumorigenic in response to treatment, by down-regulating markers of M2-like macrophage polarization/alternative activation and adopting a pronounced phagocytic phenotype (*8*). We address in this Research Article the unanswered question of whether long-term CSF-1R inhibition in aggressive late-stage GBM has a sustainable antitumorigenic effect or instead leads to acquired resistance.

A subset of GBMs develop resistance to CSF-1R inhibition in long-term preclinical trials

We first analyzed the kinetics of GBM response to continuous long-term BLZ945 treatment using a transgenic platelet-derived growth factor–driven glioma (PDG) model (RCAS-hPDGF-B/Nestin-Tv-a;*Ink4a/Arf*$^{-/-}$) (Fig. 1A) (*8*, *13*). Two weeks into the trial, we observed maximal tumor regression, with an average volume reduction of 62% (Fig. 1, B and C). At this time point, 8% of animals showed no evidence of residual tumor with magnetic resonance imaging (MRI). In contrast, vehicle-treated tumors exhibited a 2522% increase in volume over the same period (Fig. 1C).

After this regression phase, all BLZ945-treated tumors entered a dormancy phase, which lasted for ≥4 weeks (Fig. 1B). Of treated animals, 44% remained symptom-free and survived to the trial endpoint of 26 weeks ($P < 5 \times 10^{-17}$) (Fig. 1D) with minimal or, in some cases, no evidence of residual tumor as determined by MRI and histology (Fig. 1E). This is in stark contrast to vehicle-treated GBMs, which were purposely selected to be smaller in size upon treatment initiation (fig. S1A), yet median survival was only 15 days after treatment initiation (versus 93 days for BLZ945), and no animals survived beyond 6 weeks (Fig. 1D). After the dormancy phase observed in all BLZ945-treated animals, however, 56% eventually developed resistance, and tumors rebounded, despite effective, continued inhibition of CSF-1R phosphorylation in TAMs (Fig. 1, B and D, and fig. S1B).

We next focused on understanding how resistance to the CSF-1R inhibitor emerged and chose several time points throughout the long-term trial for comparison, including Veh [vehicle; 20% Captisol (Captisol, La Jolla, CA) until symptomatic], 7 days (BLZ945-responsive, regressing), 28 days (BLZ945-responsive, dormant), Reb (BLZ945-resistant, actively rebounding), and EP (26-week endpoint, stably regressed) (Fig. 1B). Histological analysis showed that after 7 days of BLZ945, tumor grade was substantially reduced. At 28 days and EP, histological grade remained low, with 33 and 50% of mice, respectively, showing no evidence of tumor via histology, and the remainder of animals exhibiting either residual disease or grade II tumors (fig. S1, C and D). By comparison, the majority of rebound tumors were grade III or IV and similar in size to Veh tumors at death (fig. S1, C to E). Both Veh and Reb tumors also exhibited a high proliferation:apoptosis index (Ki67:CC3), indicating a state of rapid growth (fig. S1F).

Glioma cells resistant to CSF-1R inhibition in vivo exhibit elevated phosphatidylinositol 3-kinase (PI3K) signaling

To determine the mechanism by which tumor cells acquire resistance, we first performed array comparative genomic hybridization (aCGH) analyses and found no copy number alterations in primary

[1]Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center (MSKCC), New York, NY 10065, USA. [2]Department of Oncology, University of Lausanne, CH-1066, Lausanne, Switzerland. [3]Ludwig Institute for Cancer Research, University of Lausanne, CH-1066, Lausanne, Switzerland. [4]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [5]Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, WA 98109, USA. [6]Novartis Institutes for Biomedical Research, Emeryville, CA 94608, USA.
*Corresponding author. Email: johanna@joycelab.org

Downloaded from http://science.sciencemag.org/ on May 19, 2016

rebound glioma tumorsphere lines (fig. S2, passage 1). To then assess which signaling pathways are altered specifically in recurrent tumors, we first purified glioma cells (PDGFRα$^+$) from Veh, EP, and Reb lesions using fluorescence-activated cell sorting (FACS), and performed RNA-sequencing (RNA-seq). Glioma cells were isolated from EP lesions that were stably regressed but still detectable with MRI. Gene ontology analysis demonstrated that Veh and Reb tumor cells showed an enrichment of cell cycle–related genes, compared with EP tumor cells (fig. S3A), corroborating the observed changes in Ki67 levels (fig. S1F) and supporting the notion that EP tumors were in a state of cell-cycle dormancy. To interrogate which pathways were differentially regulated between the three groups, we used gene set variation analysis (GSVA) (*14*) for each pair-wise comparison. Nine gene sets in total were significantly enriched in Reb tumor cells as compared with EP (fig. S3B), including a PI3K gene set (Fig. 2A), potentially explaining the robust differences in proliferation given the importance of PI3K signaling in cell-cycle regulation. In accordance with this result, we found elevated phosphorylated (p)–Akt (a PI3K substrate) in Reb tissues as compared with that in Veh and EP, using immunofluorescence (IF) staining and Western blotting (Fig. 2B and fig. S3, C and D).

To investigate whether PI3K signaling is functionally important in driving recurrence, we performed a preclinical intervention trial. We treated PDG mice bearing high-grade gliomas with BLZ945 and imaged via MRI until they showed tumor rebound, at which point we intervened with BKM120 treatment (Fig. 2C, trial design 1, and fig. S3E) at an appropriate dose so as to avoid reported off-target effects (*15*). BKM120 was chosen because it is a brain-penetrant pan–Class 1 PI3K inhibitor that is currently being clinically evaluated in GBM patients with recurrent disease after standard therapy. Animals with rebound tumors treated with continued BLZ945 monotherapy led to a median survival of 13 days post-recurrence, whereas rebound tumors treated with BLZ945+BKM120 extended median survival to 51 days (Fig. 2D) and blocked tumor progression after 2 weeks of treatment (fig. S3F). In contrast, BKM120 monotherapy in rebound tumors (discontinued BLZ945) led to a median survival of 10 days, which was indistinguishable from the vehicle control (Fig. 2D). Moreover, BKM120 was only modestly effective in treatment-naïve tumors (fig. S3, G and H). Collectively, these results indicate that continued CSF-1R inhibition is necessary to expose PI3K signaling dependency in rebound tumors and, consequently, a heightened sensitivity to pathway inhibition.

To determine whether recurrence could be prevented by earlier PI3K inhibition, during the initial dormancy phase we treated GBM-bearing PDG mice with BLZ945 alone for 28 days, at which point we added BKM120 until the trial endpoint (Fig. 2C, trial design 2). With this early combination, the percentage of animals that survived to endpoint increased substantially (91% BLZ945+BKM120) as compared with those given single-agent treatments (44% BLZ945 alone and 0% BKM120 alone) (Fig. 2E). Taken together, our results demonstrate that PI3K signaling is engaged during the acquisition of resistance to CSF-1R inhibition in the context of continued BLZ945 treatment.

Our finding that PI3K activation underlies resistance to CSF-1R inhibition was intriguing in light of the high frequency of mutations in the PI3K pathway in glioma patients (*16*). Therefore, we also investigated whether genetic mutations in PI3K/phosphatase and tensin homolog (PTEN) would similarly confer a resistance-like phenotype in mouse models, which could potentially be informative in the clinical setting. To address this question, we compared BLZ945 efficacy in two additional RCAS-hPDGF-B/Nestin-Tv-a GBM models harboring distinct clinically relevant oncogenic alterations besides *Ink4a/Arf* loss, including *Pten* deletion (*Pten* KO model) or *p53* knockdown (*p53* KD model) (materials and methods). After 2 weeks of treatment, we found that BLZ945 efficacy in the *p53* KD model (56% reduction of tumor volume) was comparable with that of the PDG model (62% reduction). However, CSF-1R inhibition was less potent in the *Pten* KO model over the same time period (3% reduction) (Fig. 2F). Furthermore, although we did eventually observe a significant reduction in tumor volume in the *Pten* KO model after a prolonged treatment period of 4 weeks (11% reduction), this did not meet the response evaluation criteria in solid tumors (RECIST) standard for a partial response (Fig. 2G) (*17*). These results suggest that treatment efficacy of CSF-1R

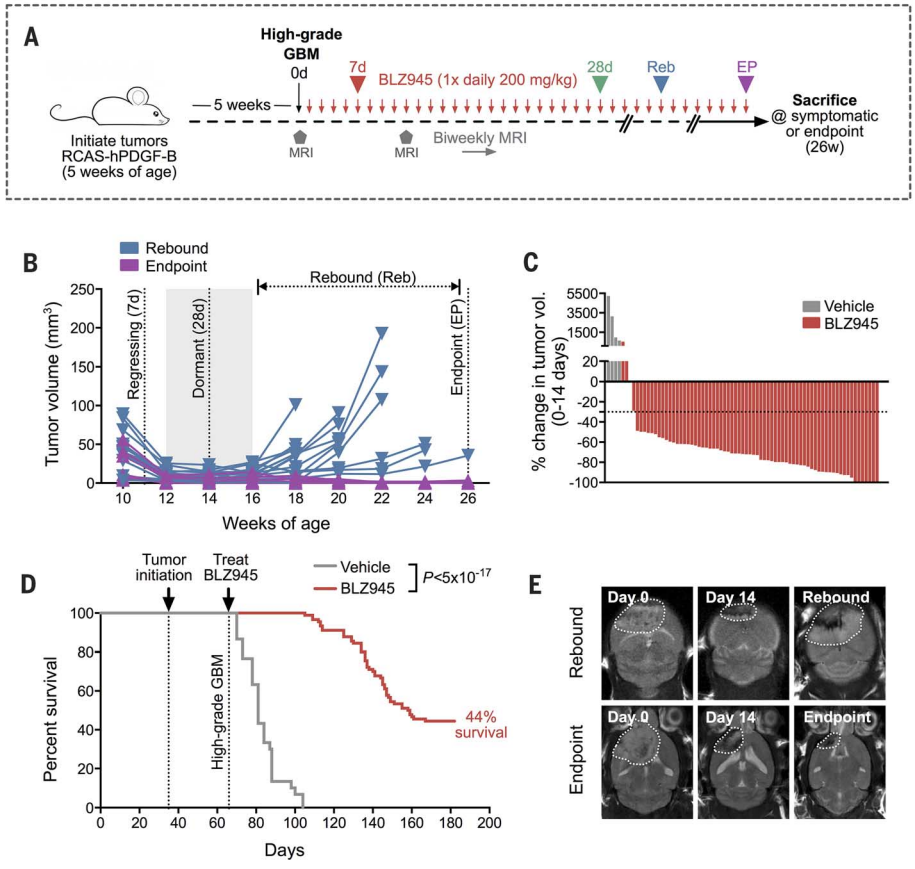


Fig. 1. Fifty-six percent of GBMs develop resistance to CSF-1R inhibition in long-term preclinical trials. (**A**) Long-term trial design for testing BLZ945 efficacy in the PDG model. High-grade PDG tumors were treated with BLZ945 (200 mg/kg/day) or vehicle (20% Captisol) and monitored with biweekly MRI for up to 26 weeks (defined endpoint; materials and methods) or until symptomatic. (**B**) Tumor volume curves from biweekly MRIs of long-term BLZ945 trials (n = 90 animals treated, 23 representative curves are shown). Four key phases are indicated, including 7 days (regressing tumor), 28 days (dormant tumor), Reb (recurrent tumor, evaluated on a mouse-to-mouse basis with MRI), and EP (stable regression at the 26-week endpoint). (**C**) Waterfall plots showing percent change in tumor volume between 0 and 14 days in a representative subset of animals from (D) (BLZ945, n = 71 mice; vehicle, n = 4 mice). (**D**) Kaplan-Meier of BLZ945-treated (n = 90) versus vehicle-treated (n = 30) mice bearing high-grade PDG tumors (Log-rank Mantel-Cox test, $P < 5 \times 10^{-17}$). Median survival for vehicle-treated animals was 15 days after treatment initiation, whereas median survival for BLZ945-treated animals was 93 days. (**E**) Representative MRI images over time of one mouse with a rebounded tumor (top row) and another mouse that had stable disease until EP (bottom row).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

inhibitors may be blunted in patients with existing genetic alterations in the PTEN/PI3K pathway.

Resistance to CSF-1R inhibition is mediated by the microenvironment

We next investigated how PI3K was activated in rebound tumors and first explored whether resistance to BLZ945 was tumor cell–intrinsic or –extrinsic. We previously established that BLZ945 does not directly affect glioma cell lines in culture (*8*) and demonstrate here that CSF-1R inhibition also has no direct effect on viability of a panel of primary cell lines derived from rebound tumors (fig. S4A). We designed an intracranial tumor transplantation model using early-passage Reb cells in order to address the following hypotheses: (i) Resistance is tumor cell–intrinsic, therefore transplanted tumors will not respond to BLZ945, or (ii) resistance is mediated by the treatment-altered microenvironment, therefore transplanted tumors will reestablish sensitivity to CSF-1R inhibition in naïve animals. Transplanted rebound tumors responded to BLZ945 treatment in the naïve setting (Fig. 3A and fig. S4, B to F), indicating that resistance is likely mediated by the TME.

We analyzed the TME in recurrent disease to determine how resistance to CSF-1R inhibition develops. We found that rebounding tumors always emerged adjacent to regions of glial scarring, characterized by reactive astrocytes, calcium deposition, and relatively low vascularity associated with elevated hypoxia (Fig. 3, B to D, and fig. S4, G to I). In contrast, scarring was infrequently observed in the 28-day and EP tumors (fig. S4I). The scar tissue architecture was reminiscent of gliosis in response to neurodegeneration or physical injury (*18*). Given the parallels between a wound-associated microenvironment and tumorigenesis in epithelial tissues (*19*), we hypothesized that this brain injury response may likewise be contributing to a microenvironment that is potentially triggering recurrent disease.

Rebound TAMs adopt a wound-associated signature driven by enhanced interleukin-4 (IL-4) signaling

During gliosis, activated macrophages play a central role in providing growth factors and signaling molecules to nearby astrocytes and neurons to form a reactive barrier that limits the extent of tissue damage in the brain (*18*, *20*). Given that BLZ945 is a macrophage-targeted drug, we analyzed TAM numbers and phenotype in rebound tumors. We have previously shown that TAMs are not depleted in the glioma TME in a 1-week trial with BLZ945, but rather down-regulate expression of M2-like genes and increase phagocytosis of tumor cells (*8*, *21*). Consistently, we show here that TAMs are still present in 7-day, 28-day, EP, or Reb tumors (Fig. 3E and fig. S5A). When we used flow cytometry to discriminate between $CD45^{lo}CD11b^{+}$ cells (putative microglia) versus $CD45^{hi}CD11b^{+}$ cells [putative bone marrow–derived macrophages (BMDMs)] (*22*–*24*) in Veh, EP, and Reb tumors (materials and methods) (fig. S5B), we found that long-term BLZ945 treatment enriched for $CD45^{lo}CD11b^{+}$ TAMs (fig. S5, C and D). This is potentially either a consequence of phenotypic mimicry between the macrophage populations or the result of one macrophage population responding differently to CSF-1R inhibition than the other. Costaining of CD68 or CD206 macrophage markers in combination with Ki67 demonstrated that a subset of remaining TAMs in rebound tumors were proliferating (fig. S5, E to

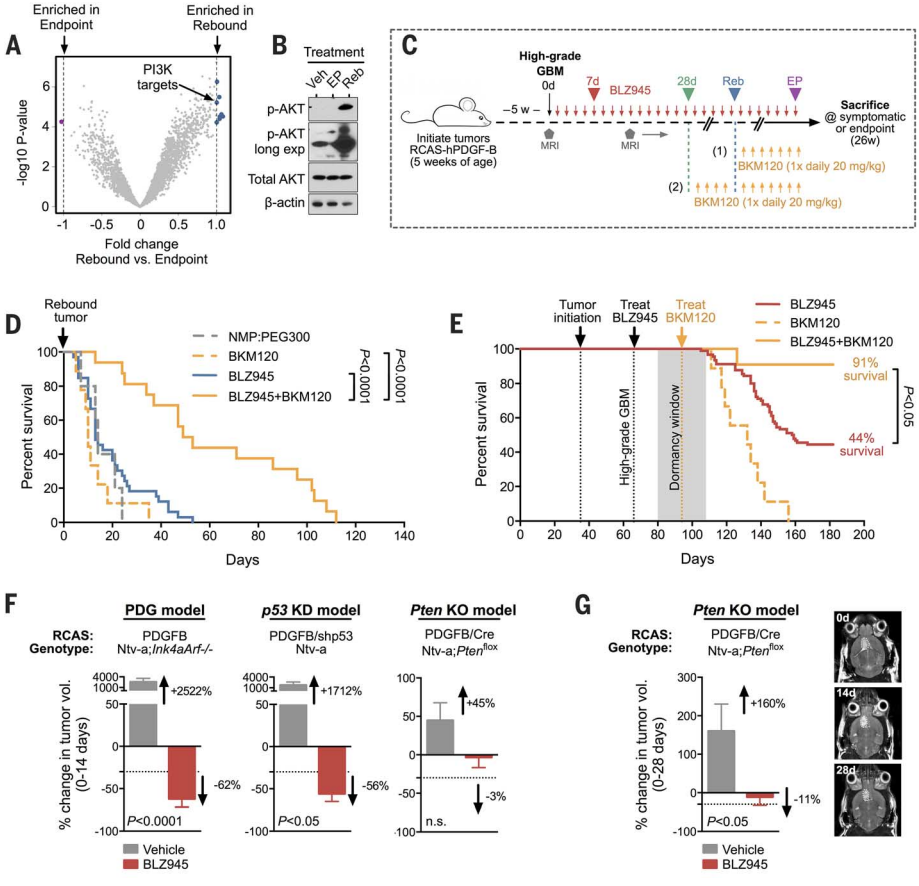


Fig. 2. Combined CSF-1R and PI3K inhibition improves survival in the PDG model. (**A**) Gene set variation analysis based on RNA-seq from FACS-purified EP and Reb tumor cells ($PDGFR\alpha^{+}$) (fig. S3B). Blue circles indicate gene sets significantly enriched in Reb tumor cells, and purple circles identify those enriched in EP tumor cells. The PI3K gene set is indicated with an arrow. Fold change is graphed as log2(fold). Vertical dotted lines indicate fold cutoff for significance (n = 5 or 6 samples per group). (**B**) Immunoblot from snap-frozen Veh, EP, and Reb tumors demonstrating elevated phospho (p)–AKT in Reb tumors compared with Veh and EP (n = 3 experiments, one representative blot is shown). (**C**) Long-term trial design for evaluating BLZ945 and BKM120 combination therapy on PDG tumors. High-grade tumors were treated with BLZ945 until recurrent tumors developed (trial design 1) or until dormancy (28 days, trial design 2), whereupon BKM120 was either added (with continuous BLZ945 treatment) or switched (discontinued BLZ945). (**D**) Survival of animals with recurrent tumors treated either with BLZ945 alone (n = 33 mice), BKM120 alone (n = 9 mice), or BLZ945 in combination with BKM120 (n = 16 mice) (Fig. 2C, trial design 1). Combination of BLZ945+BKM120 led to an increase in overall survival (Log-rank Mantel-Cox test, $P < 0.0001$), and in median survival (51 days) after recurrence compared with BLZ945 (13 days) or BKM120 (10 days) monotherapy. (**E**) Survival of animals with 28-day dormant tumors treated either with BLZ945 alone (n = 90 mice; same cohort as presented in Fig. 1D), BKM120 alone (n = 9 mice), or BLZ945 in combination with BKM120 (n = 11 mice) (Fig. 2C, trial design 2). Combination of BLZ945+BKM120 led to increased overall survival compared with either monotherapy. Log-rank Mantel-Cox test was used to calculate significance. (**F**) Average percent change in tumor volume (0 to 14 days) between vehicle- or BLZ945-treated tumors, in three different RCAS-PDGFB-HA *Nestin-Tv-a* GBM models (termed PDG, *p53* KD, and *Pten* KO) (materials and methods). BLZ945 efficacy in the *p53* KD model (BLZ945, n = 8 mice; vehicle, n = 6 mice) was comparable with that of the PDG model (BLZ945, n = 71 mice; vehicle, n = 4 mice) after 2 weeks (56 and 62% volume reduction, respectively); however, BLZ945 efficacy was less pronounced in the *Pten* KO model (n = 5 mice per treatment group; 3% volume reduction). (**G**) Average percent change in tumor volume (0 to 28 days) between vehicle- or BLZ945-treated *Pten* KO tumors (n = 5 mice per group). BLZ945 caused 11% volume reduction after 4 weeks. Data were analyzed with Student's t test unless indicated otherwise.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

G). Although these results can only be formally confirmed with lineage tracing experiments, they at least suggest that rebound TAMs (enriched for $CD45^{lo}CD11b^{+}$ cells) may undergo a low level of replication, presumably as a means to compensate for the duress caused by prolonged CSF-1R blockade.

To assess potential differences in activation states, we FACS-purified TAMs from Veh, EP, and Reb tumors and performed RNA-seq. Principal component analysis confirmed distinct global gene expression profiles for Veh, EP, and Reb TAMs (fig. S6A), and differential expression analysis revealed large numbers of differentially expressed genes between the three groups (table S1). We first focused on a subset of M2-like genes previously identified as altered by CSF-1R inhibition (*8*, *25*). We found that compared with Veh TAMs, alternative activation was suppressed in EP TAMs, whereas a subset of these genes were highly expressed in Reb TAMs (Fig. 3F). In accordance with previous findings (*8*), no inverse relationship was observed for M1-like markers such as tumor necrosis factor–α (TNFα) across the different treatment groups (*25*). Together, these findings support our hypothesis that the Reb TME is protumorigenic.

Given the similarities in alternative activation between Veh and Reb TAMs according to the M1/M2-like paradigm (*26*), yet clear differences in drug response between treatment-naïve and rebound tumors, we next used a more fine-tuned approach to defining macrophage phenotype. We computationally interrogated a spectrum model of macrophage activation, defined by gene sets that are altered in response to different stimuli, including interferon-γ (IFN-γ), IL4, TNFα, transforming growth factor–β1 (TGFβ1), IL1β, and two Toll-like receptor (TLR) agonists specific for TLR2 [macrophage-activating lipopeptide 2 (MALP2)] and TLR9 [unmethylated CpG-containing oligonucleotide (CPG)] (*27*, *28*). We determined that IL4- and TGFβ1-targeted gene sets were significantly enriched in Reb TAMs compared with Veh TAMs (Fig. 4A) and also found a significant enrichment of these same gene sets in Reb TAMs versus EP TAMs (Fig. 4B). Given that IL4 is a known mediator of alternative activation associated with a wound-healing phenotype in macrophages (*20*, *29*), and the role of TGFβ1 during wound-healing and tissue turnover in multiple contexts (*30*, *31*), these results were consistent with our observation of glial scarring in association with rebound tumors. Indeed, a number of M2-like genes expressed by macrophages involved in wound repair and resolving inflammation (*Retnla*, *Chil3*, and *Ccl17*) were enriched in Reb TAMs (Fig. 3F). In contrast, RNA-seq analyses of an independent set of Reb and 28-day TAM samples revealed no significant differences in expression across this same gene set (fig. S6B), suggesting that a subset of 28-day tumors may be rebound precursors, and that induction of M2-like gene

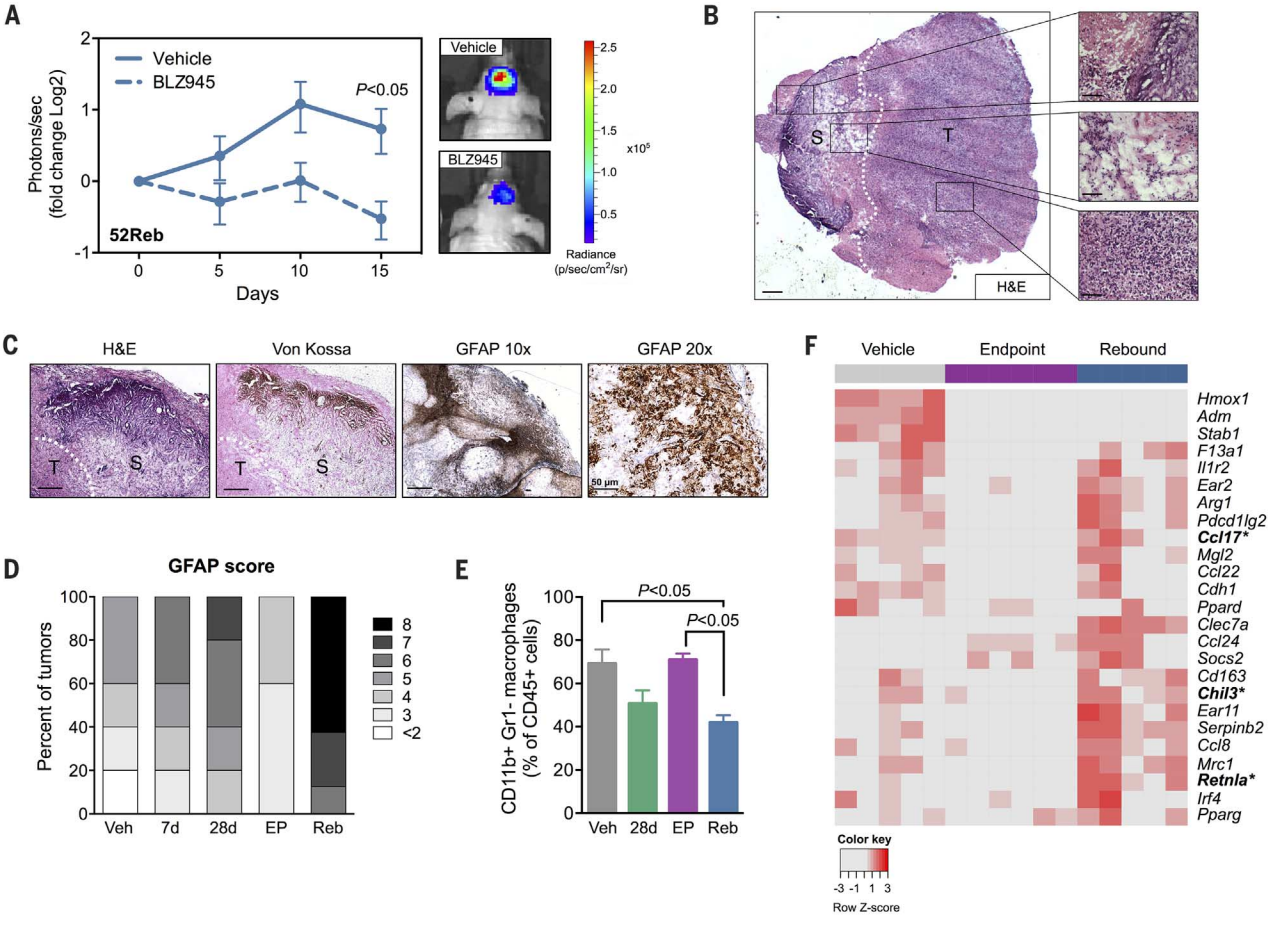


Fig. 3. Resistance to CSF-1R inhibition is mediated by the microenvironment and rebound TAMs are alternatively activated. (**A**) Quantification of BLI from intracranially transplanted 52Reb cells into naïve mice. Results show that 52Reb cells, isolated from a recurrent PDG tumor that developed resistance to BLZ945 treatment in vivo, reestablish sensitivity to BLZ945 in the naïve setting (Student's *t* test day 15, $P < 0.05$, $n = 10$ mice). Representative BLI images at day 15 are shown. (**B**) (Left) H&E (hematoxylin and eosin) of a rebound tumor (T) adjacent to glial scarring (S). Scale bar, 500 μm. (Right) Representative regions of calcification (top), reactive astrocytic barrier (middle), and recurrent tumor (bottom). Scale bars, 50 μm. (**C**) H&E, Von Kossa, and GFAP (glial fibrillary acidic protein) staining on rebound tissue. Scale bars, 200 μm; except GFAP, 20× = 50 μm, as indicated. (**D**) Allred score for the astrocyte marker GFAP, showing a higher intensity and proportion of $GFAP^{+}$ staining in Reb tissues ($n = 8$ mice) compared with other treatment groups ($n = 5$ mice per group). (**E**) Flow cytometry of TAMs ($CD45^{+}CD11b^{+}Gr1^{-}$) in Veh, 28-day, EP, and Reb tumors ($n = 5$ to 7 tumors). Data were analyzed with a one-way ANOVA and Tukey's multiple comparisons test. (**F**) Heatmap showing RNA-seq expression changes of M2-like associated genes in Veh, EP, and Reb TAMs ($n = 5$ or 6 per group). Wound-associated genes are indicated in bold with asterisks.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

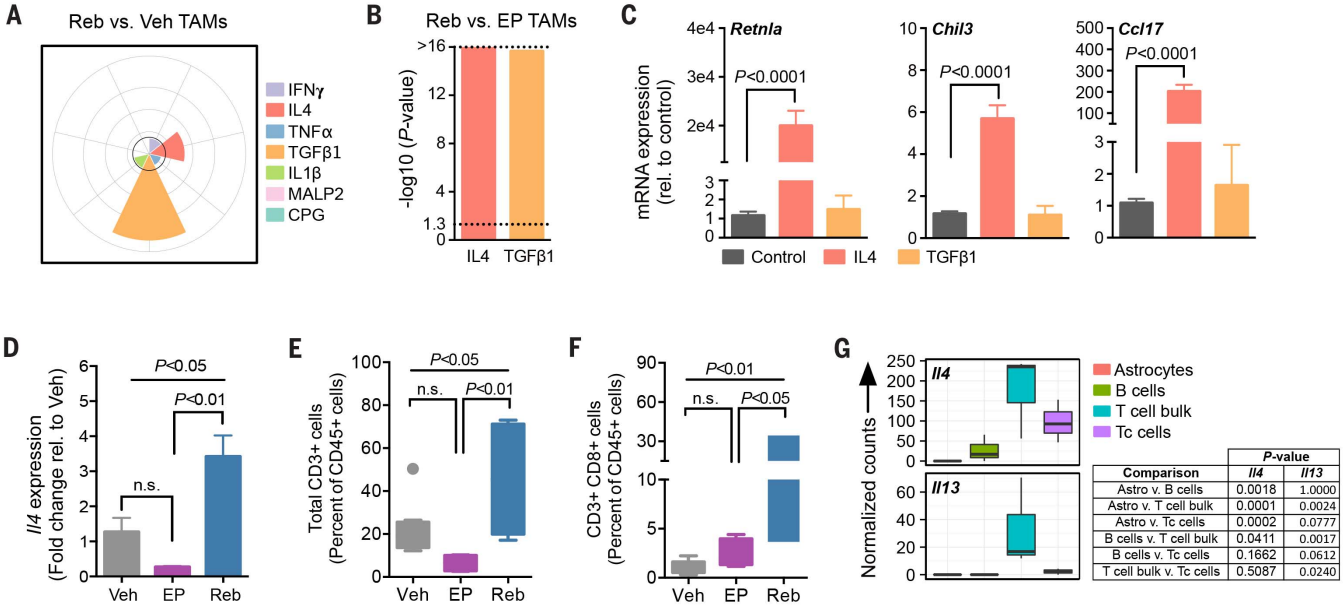
expression precedes recurrence. When we investigated the ability of IL4 or TGFβ1 to regulate expression of these wound-associated genes in BMDMs in vitro, we found that IL4 was a potent inducer, whereas TGFβ1 was not (Fig. 4C and fig. S6, C and D), indicating the specificity of IL4 for regulation of this particular gene set. Corroborating these findings, we confirmed increased expression of *Il4* in rebound tumors by means of quantitative reverse transcription polymerase chain reaction (RT-PCR) (Fig. 4D) and elevated canonical downstream signaling by means of immunofluorescent staining for phosphorylated Stat6 (p-Stat6) in PDG tissue samples from the long-term trials (fig. S6, E and F). These results do not exclude the possibility that TGFβ may be playing a role in rebound tumors that is distinct from, or indirectly connected to, alternative activation of TAMs.

To determine the cellular source of IL4 in our model, we used multicolor flow cytometry to immunoprofile Veh, EP, and Reb tumors using a panel of myeloid [CD11b, Gr1, Ly6G, Ly6C, CD11c, Tie2, and major histocompatibility complex (MHC) II] and lymphoid (CD19, CD3, CD4, CD8, and FoxP3) cell markers. Although we found few cell types that were specifically enriched in rebound tumors (fig. S7, A to G), there was a significant increase in the proportion of $CD3^+$ T cells in rebound tumors, driven by the $CD8^+$ fraction (Fig. 4, E and F). FACS-purification of these cells from rebound tumors along with other putative contributors of IL4 [including astrocytes (Fig. 3D), B cells (*32*–*34*), and bulk T cells] revealed that *Il4* expression was enriched in both bulk T and $CD8^+$ T cell fractions (Fig. 4G and fig. S7H). In comparison, expression of *Il13*, a closely related cytokine that shares the IL4Rα subunit in its heterotypic receptor to initiate canonical Stat6 signaling, was enriched in the bulk T cell fraction (Fig. 4G). Expression of *IL4* was assessed in a panel of human cell types and detected in $CD8^+$ T cells as expected, in addition to monocytes, eosinophils, astrocytes, and brain microvascular cells (fig. S7I). Although additional analyses are needed to establish the precise cellular source of these cytokines in brain malignancies in patients, our data are consistent with IL4 being produced by multiple cell types.

The IGF-1/IGF-1R signaling axis is induced in rebound gliomas

We next investigated how IL4 and wound-associated gene expression might be connected to PI3K signaling in rebound tumors. Differential gene expression analysis revealed that TAM-derived *Igf1* was one of the most significantly up-regulated genes in Reb TAMs compared with EP TAMs, which we confirmed in comparisons with Veh TAMs or 28-day TAMs (Fig. 5, A and B, fig. S8A, and table S1). This was particularly interesting because *Igf1* is an IL4 target gene in macrophages (fig. S8B) (*35*–*37*), it is a known mediator of tissue repair and neuroprotection (*38*–*41*), and one of its canonical downstream signaling pathways is PI3K/Akt (*42*). Congruent *Igf1r* up-regulation was identified in glioma cells purified from rebound tumors (Fig. 5C and table S1), elevated p-IGF-1R was found in rebound tumors by means of immunostaining and Western blotting (Fig. 5D and fig. S8, C and D), and *Igf1* up-regulation was found in snap-frozen rebound tissue samples by means of quantitative RT-PCR (fig. S8E). Additionally, levels of *Igf1* expression were substantially higher in Reb TAMs than tumor cells, whereas *Igf1r* expression was enriched in tumor cells as compared with TAMs (fig. S8F). Together, these data demonstrate elevated IGF-1 signaling in recurrent disease.

To assay the effects of IGF-1 specific to the rebound setting, we used multiple approaches: First, through Western blot analysis, we confirmed higher baseline levels of p-IGF-1R in early-passage rebound tumor cell lines as compared with cell lines that we were able to propagate from dormant tumors in culture (materials and methods) (fig. S8G). We determined that phosphorylation and downstream signaling could be reduced in rebound cell lines by using an inhibitor of IGF-1R (fig. S8G) and that early-passage rebound cells were more sensitive



Comparison	P-value Il4	Il13
Astro v. B cells	0.0018	1.0000
Astro v. T cell bulk	0.0001	0.0024
Astro v. Tc cells	0.0002	0.0777
B cells v. T cell bulk	0.0411	0.0017
B cells v. Tc cells	0.1662	0.0612
T cell bulk v. Tc cells	0.5087	0.0240

Fig. 4. IL4 activates wound-associated genes in rebound TAMs and is produced by T cells in the rebound tumor microenvironment. (**A**) Spectrum model of macrophage activation by using GSEA of transcriptional programs regulated by the indicated cytokines (materials and methods). Reb TAMs exhibit activation of programs driven by TGFβ1 and IL4 compared with Veh TAMs ($n = 5$ mice per group). The −log10 (P value) is plotted for each gene set (0, 4, 8, 12, and >16, radially outward in gray concentric circles). The black circle indicates a −log10 (P value) cutoff of 3. (**B**) GSEA as in (A), confirming that Reb TAMs exhibit activation of programs driven by IL4 and TGFβ1 compared with EP TAMs ($n = 5$ or 6 mice per group). Dotted line demarcates $P < 0.05$. (**C**) Quantitative RT-PCR analysis demonstrates that IL4 induces expression of the wound-associated genes *Retnla*, *Chil3*, and *Ccl17* in BMDMs in culture, whereas recombinant TGFβ1 treatment does not. Bone marrow isolate was obtained from five or more independent WT mice for replicate experiments. (**D**) Quantitative RT-PCR analysis of *Il4* expression in snap-frozen whole-tumor samples from Veh, EP, or Reb treatment groups, demonstrating elevated expression in the Reb setting ($n = 4$ tumors) compared with both Veh ($n = 4$ tumors) and EP ($n = 4$ tumors). (**E**) Flow cytometry of total $CD3^+$ T cells ($n = 5$ to 10 tumors) and (**F**) $CD3^+$ $CD8^+$ cytotoxic T cells ($n = 5$ to 10 tumors) in Veh, EP, and Reb tumors. (**G**) RNA-seq data from a panel of cell types [$GFAP^+$ astrocytes, $CD19^+$ B cells, $CD3^+$ $CD8^+$ cytotoxic T cells (Tc), and remaining $CD3^+$ $CD8^-$ bulk T cells] isolated from Reb tumors by means of FACS ($n = 3$ tumors). Results show elevated *Il4* expression in bulk T and Tc cells, and *Il13* expression in bulk T cells. Significance values for (C) to (F) were calculated with one-way ANOVA and Tukey's multiple comparisons test.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

to IGF-1R blockade than were naïve glioma cells in vitro by using multiple pharmacological inhibitors (Fig. 5E and fig. S8H).

Next, to model the effects of macrophage-derived IGF-1 on rebound tumor cells, we designed an ex vivo culture system using primary glioma microenvironment cultures (GMECs). GMECs contain multiple cell types from the glioma TME when harvested at early passage, including macrophages, T cells, and astrocytes, among others (fig. S8I) (*8*). We hypothesized that rebound GMECs would be able to stimulate production of IGF-1 by macrophages and subsequent growth of tumor cells. To test this, we collected conditioned media (CM) from rebound GMECs and applied it to wild-type (WT) BMDMs for 24 hours. After this treatment, we collected CM from the GMEC-stimulated BMDMs (Stim CM)

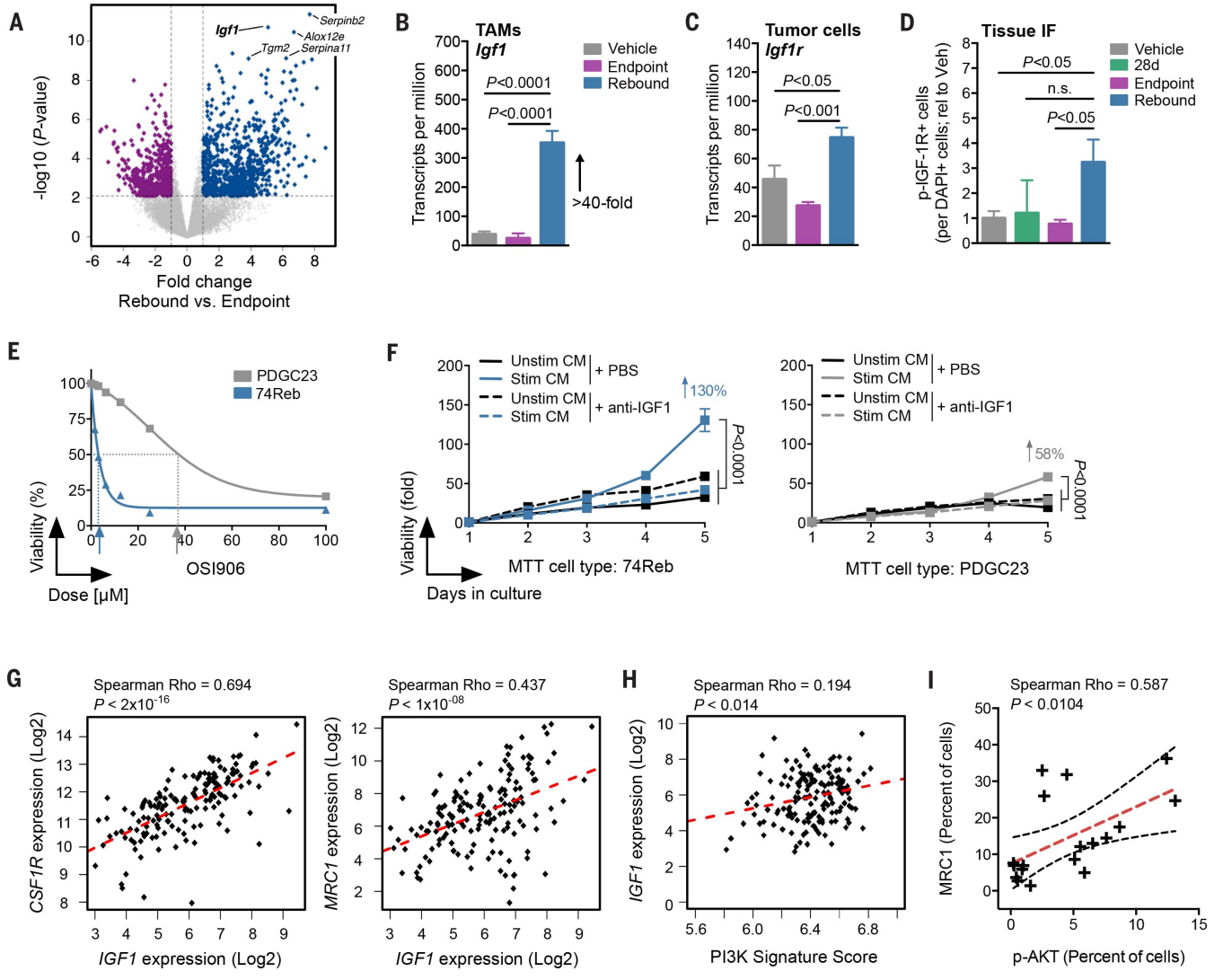


Fig. 5. The IGF-1/IGF-1R axis is induced in rebound gliomas. (**A**) Volcano plot showing the fold change [log2(fold)] between Reb (n = 5 tumors) and EP (n = 6 tumors) TAMs on the x axis and the significance [−log10(P value)] on the y axis. Blue dots indicate genes up-regulated in Reb TAMs, whereas purple dots indicate genes down-regulated in Reb TAMs. *Igf1* is labeled in the top right quadrant. (**B**) RNA-seq barplot depicting *Igf1* transcripts per million (TPM) in Veh, EP, and Reb TAMs (n = 5 or 6 tumors per group; one-way ANOVA and Tukey's multiple comparisons). (**C**) RNA-seq barplot depicting *Igf1r* TPM in Veh, EP, and Reb tumor cells (n = 5 or 6 tumors per group; one-way ANOVA and Tukey's multiple comparisons). (**D**) Quantification of immunofluorescent staining of phospho (p)–IGF-1R in Veh, 28-day, EP, and Reb tumor tissues. Results show pIGF-1R is elevated in rebound tumors compared with all other groups (n = 5 to 8 tumors per group; one-way ANOVA and Dunnett's multiple comparisons to Reb). (**E**) MTT assay demonstrating higher sensitivity of an early-passage primary rebound PDG cell line (74Reb, blue) to IGF-1R inhibition with OSI906 compared with a primary treatment-naïve PDG cell line (PDGC23, gray) (n = 3 independent replicates, one representative experiment shown). IC_{50} values are indicated with arrows. (**F**) MTT proliferation assays of 74Reb cells (left) compared with PDGC23 cells (right), treated with conditioned media (CM) from BMDMs that were stimulated with rebound glioma microenvironment culture CM (fig. S8, I and J). Stimulated BMDM CM (Stim CM) induced growth of 74Reb cells more than PDGC23 cells (130 versus 58%, respectively), and this effect was blocked with an anti–IGF-1 neutralizing antibody (n = 4 replicate experiments). A one-way ANOVA and Dunnett's multiple comparisons with Stim CM + PBS was used to calculate differences at 5 days (P < 0.0001 for all). (**G**) Correlation between *IGF1* and *CSF1R* or *MRC1* expression from TCGA-GBM data. (**H**) Single sample GSEA for a hallmark PI3K signature was used to assign a pathway activity score (materials and methods) across patients from the TCGA-GBM data set. PI3K signature scores were then correlated with *IGF1* expression levels as shown. (**I**) Linear regression analysis of immunohistochemical staining for phospho (p)–AKT and MRC1 in serial sections from GBM patient tissue (n = 18 patients). A significant correlation between MRC1 protein levels and AKT signaling was observed. For correlational analyses [(G) to (I)], a Spearman coefficient was used to assess significance, and a line of best fit is shown (dashed red). The 95% confidence band (dashed black) is also shown in (I).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

and applied it to either rebound or naïve tumor cell lines in an MTT assay, plus or minus a neutralizing antibody against IGF-1 (experimental design is provided in fig. S8J). We found that Stim CM induced proliferation of rebound cell lines more effectively than naïve cell lines, and this effect was blocked by IGF-1 neutralization (Fig. 5F). These results suggest that the cells within a recurrent tumor are capable of stimulating production of IGF-1 by macrophages, which in turn gives a proliferative advantage to rebound tumor cells.

To assess the relevance of the IGF-1/IGF-1R axis in human malignancy, we evaluated *IGF1* expression in publicly available human GBM gene expression data sets. We found that *IGF1* expression was significantly correlated with macrophage-specific genes (*CSF1R*, *CD68*, and *AIF1*) and with genes associated with an M2-like phenotype (*CD163* and *MRC1*) in The Cancer Genome Atlas (TCGA) GBM samples (Fig. 5G and fig. S9A). No such correlations were found for genes enriched in astrocytes (*GFAP* and *ALDH1L1*) (fig. S9A), another key cell type in the glial scar phenotype (*18*). We also confirmed that *IGF1* expression was significantly correlated with a PI3K signature score (Fig. 5H) generated from single sample gene set enrichment analysis (GSEA) for hallmarks of PI3K signaling (*43*). Consistently, immunohistochemistry (IHC) quantitation on an independent set of human GBM tissue samples revealed a significant association between p-AKT and the M2-associated protein MRC1 (Fig. 5I, fig. S9B, and table S2). Together, these data support our hypothesis that high IGF-1 levels translate to elevated PI3K signaling in patients, and that this axis is associated with M2-like gene expression.

To determine whether macrophages are the predominant source of *IGF1* in humans, we used quantitative RT-PCR to show that macrophages express high levels of *IGF1* compared with that in different immune cell types, astrocytes, endothelial cells, and glioma cells (fig. S9C). Consistently, we found that *IGF1* expression was enriched in TAMs compared with the tumor bulk in GBM (fig. S9D) (*44*), and in mesenchymal GBM compared with other molecular subtypes (fig. S9E) in which high macrophage content is a hallmark histological feature. These data corroborate our findings in the PDG model by suggesting that macrophages are an important source of *IGF1* in human malignancy.

Last, we used publicly available data sets to assess survival correlations in patients. We first generated Kaplan Meier curves using a median cutoff for *IGF1* expression levels and found no significant difference in overall survival between

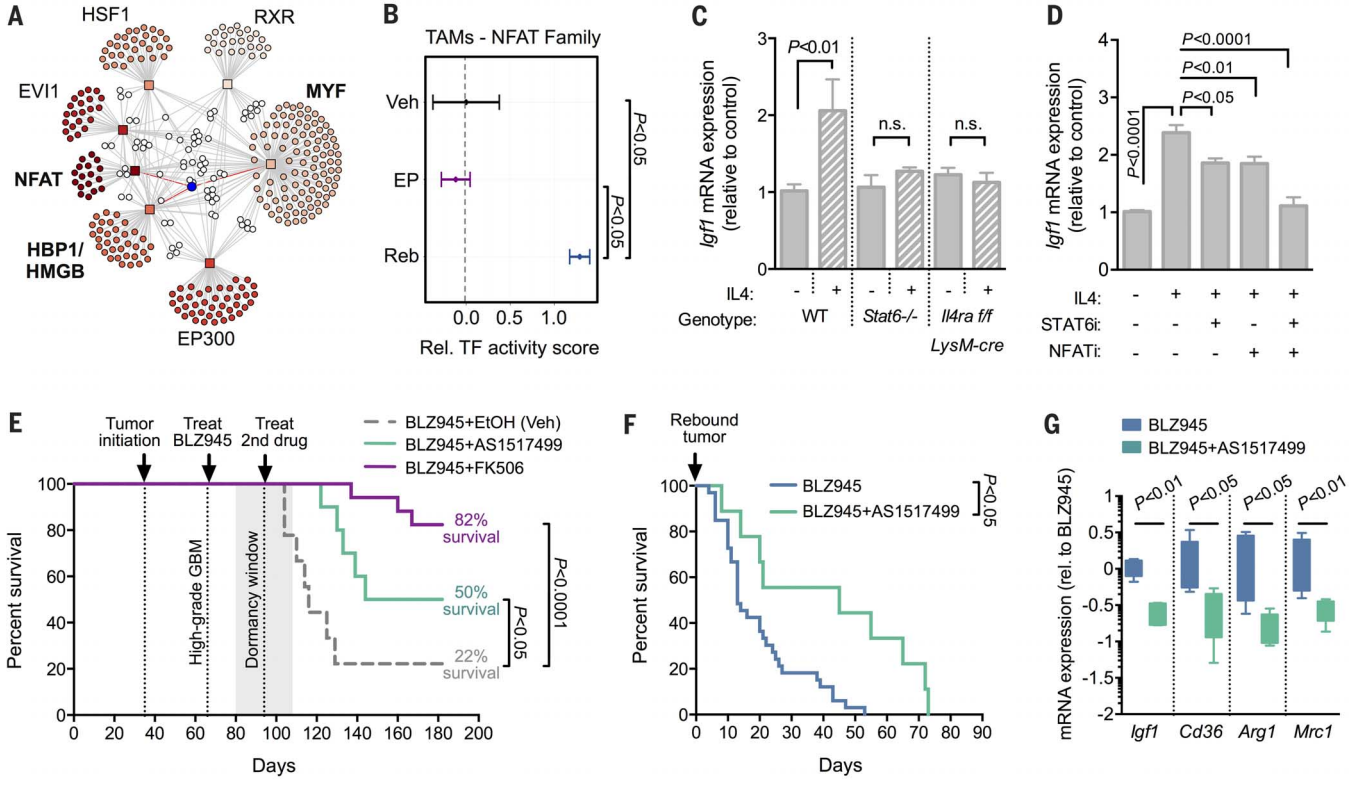


Fig. 6. NFAT and Stat6 cooperate to regulate *Igf1* expression in rebound TAMs. (**A**) Transcription factor (TF) network analysis from RNA-seq data showing enriched TF families (squares) connected with a line to target genes (circles). White circles indicate genes targeted by multiple TFs. *Igf1* is shown as a blue circle connected by red lines to three regulating TF families. (**B**) Predicted NFAT TF activity in Veh, EP, and Reb TAMs, showing a high score specifically in Reb TAMs (n = 5 or 6 tumors per group). (**C**) Quantitative RT-PCR analysis of *Igf1* in BMDMs derived from WT, *Stat6*$^{-/-}$, or *Il4ra flox; LysM-cre* mice, treated ± recombinant mouse IL4 (10 ng/ml; n = 5 independent experiments). Student's t test was used for pairwise comparisons within each genotype. (**D**) Quantitative RT-PCR analysis of *Igf1* in BMDMs derived from WT mice, treated ± recombinant mouse IL4 (10 ng/ml), a Stat6 inhibitor (AS1517499, 50 nM), and/or an NFAT inhibitor (INCA-6, 40 μM; n = 6 independent experiments). A one-way ANOVA and Dunnett's multiple comparisons with the +IL4 condition was used to calculate significance. (**E**) Survival of PDG animals with high-grade tumors treated first with BLZ945 alone until dormancy (28 days) and then enrolled on combination therapy with either AS1517499 (n = 10 mice), FK506 (NFAT inhibitor; n = 17 mice), or EtOH vehicle control (n = 9 mice). Combination therapy with either inhibitor led to a significant increase in overall survival (FK506, 82% survival, $P < 0.0001$; AS1517499, 50% survival, $P < 0.05$), compared with vehicle control (22% survival). Log-rank Mantel-Cox test was used to calculate significance. (**F**) Survival curve representing animals with recurrent tumors treated either with continuous BLZ945 alone (n = 33 mice) or BLZ945+AS1517499 (n = 9 mice). Combination therapy led to a significant increase in overall survival (Log-rank Mantel-Cox test, $P < 0.05$), and in median survival after recurrence (45 days) compared with BLZ945 monotherapy (13 days). (**G**) Quantitative RT-PCR analysis of *Igf1*, *CD36*, *Arg1*, and *Mrc1* levels in a subset of animals from (F). Results show a significant reduction of *Igf1* expression in rebound tumors after Stat6 inhibition (n = 5 mice per group, $P < 0.01$), and a reduction of known IL4-Stat6 transcriptional targets (*CD36*, $P < 0.05$; *Arg1*, $P<0.05$; *Mrc1*, $P < 0.01$), confirming drug efficacy in the brain (Mann-Whitney test).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

IGF1^{high} and *IGF1*^{low} patients (fig. S9F). Given that baseline IGF-1 signaling is critical during normal homeostasis in the brain (*41*), and also the extremely rapid progression of GBM in patients, we also surveyed the top 10% of *IGF1*^{high} patients (versus all remaining) and found a clear decrease in overall survival (fig. S9F). When we used this same stringent top-10% cutoff for a macrophage marker (*AIF1*; also known as *Iba1* in mouse) that correlates significantly with *IGF1* expression levels (fig. S9A), there was no separation of survival curves (fig. S9F), suggesting that differences in survival are not simply due to differences in macrophage abundance, but rather to differences in degree of *IGF1* expression.

NFAT and Stat6 cooperate to regulate *Igf1* expression in rebound TAMs

We next used transcription factor (TF) activity analysis to identify putative transcriptional networks regulating *Igf1* expression in rebound tumors. Seven TF families showed enriched activity in Reb TAMs compared with EP TAMs (Fig. 6A, fig. S10A, and table S3). Three of these were found to regulate *Igf1* [nuclear factor of activated T cells (NFAT), MYF, and HMGB families] (Fig. 6A), of which the NFAT family showed enriched activity specifically in Reb TAMs compared with both EP and Veh TAMs (Fig. 6B). These results were particularly interesting given the cooperative relationship between NFAT and Stat6 (canonical IL4 pathway) in transcriptional regulation (*45*). Corroborating these results, we found that IL4, but not TGFβ1, induced *Igf1* expression in BMDMs in vitro, which was reduced by an NFAT inhibitor (fig. S10B).

To further characterize the role of IL4-induced NFAT and/or Stat6 signaling in *Igf1* regulation, we performed a series of in vitro and in vivo experiments. First, we confirmed that IL4 strongly induced expression of *Igf1* and the three representative alternative activation/wound-associated genes (*Retnla*, *Chil3*, and *Ccl17*) in WT BMDMs, and that this capacity was reduced in BMDMs from either *Stat6*$^{-/-}$ or *Il4ra flox; LysM-cre* mice (Fig. 6C and fig. S10C). Furthermore, whereas pharmacological inhibition of either Stat6 or NFAT partially reversed the effects of IL4 on *Igf1*, *Retnla*, *Chil3*, and *Ccl17* expression, dual inhibition of these pathways in WT BMDMs completely blocked the effects of IL4 on this gene set (Fig. 6D and fig. S10D). To validate the importance of these pathways in vivo, PDG mice with high-grade GBMs were treated continuously with BLZ945 alone until 28 days, at which point FK506 (a NFAT-calcineurin inhibitor) or AS15171499 (a Stat6 inhibitor) was added until the trial endpoint. With addition of either of these inhibitors, the percentage of animals that survived to endpoint was significantly increased (22% BLZ945+Veh, 50% BLZ945+AS1517499, and 82% BLZ945+FK506) (Fig. 6E). In accordance with these results, when we treated animals with AS1517499 in combination

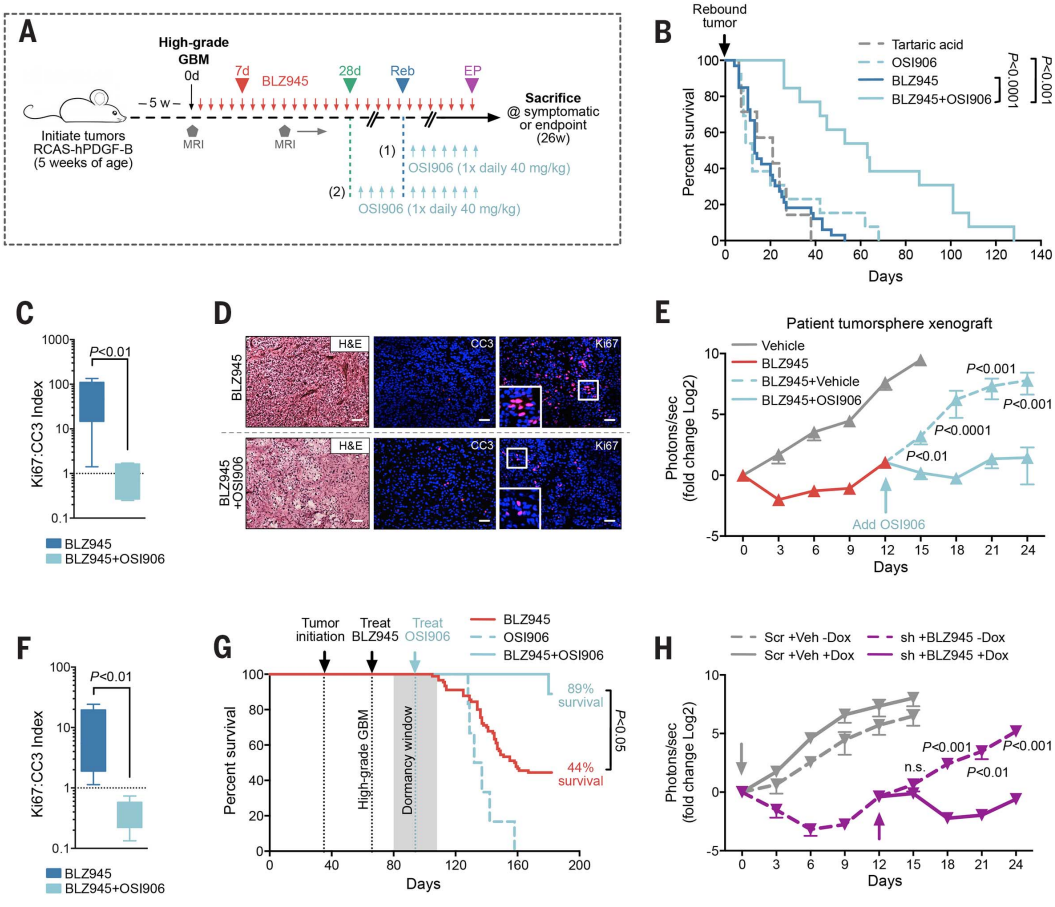


Fig. 7. Combination of CSF-1R inhibition and IGF-1R inhibition significantly improves outcome in preclinical models. (**A**) Long-term trial design for testing BLZ945 and OSI906 combination therapy on PDG tumors. High-grade tumors were treated with BLZ945 until recurrent tumors developed (trial design 1) or until dormancy (28 days, trial design 2), when OSI906 was either added (with continuous BLZ945) or switched (discontinued BLZ945). (**B**) Survival of animals with recurrent tumors treated either with BLZ945 alone (n = 33 mice), OSI906 alone (n = 13 mice), or BLZ945 in combination with OSI906 (n = 13 mice) (Fig. 7A, trial design 1). Combination therapy of BLZ945+OSI906 led to an increase in overall survival (Log-rank Mantel-Cox test, $P < 0.001$), and in median survival after recurrence (63 days) compared with BLZ945 (13 days) or OSI906 (12 days) monotherapy. (**C**) Ki67:CC3 (cleaved caspase 3) proliferation:apoptosis index from immunofluorescent staining of recurrent tumors treated with BLZ945 alone versus BLZ945+OSI906 for 2 weeks (Mann-Whitney test, $P < 0.01$, n = 7 or 8 mice). (**D**) Representative H&E and immunofluorescent images corresponding to data in (C). Scale bars, 50 μm. (**E**) BLI from orthotopically xenografted patient-derived tumorspheres (TS573) that were subject to 24 days of treatment with BLZ945 (red) versus vehicle control (gray). Results demonstrate that treatment with BLZ945+OSI906 blunts outgrowth of rebound tumors compared with BLZ945+vehicle. Mann-Whitney test was used to calculate P values for each time point (n = 5 to 20 mice). (**F**) Ki67:CC3 index from immunofluorescent staining of TS573 orthotopic xenograft tissues (Mann-Whitney test, $P < 0.01$, n = 5 mice per group). (**G**) Kaplan-Meier analysis of PDG animals treated either with BLZ945 alone (n = 90 mice; same cohort as presented in Fig. 1D), OSI906 alone (n = 6 mice), or with BLZ945 in combination with OSI906 during dormancy (n = 9 mice) (Fig. 7A, trial design 2). Combination therapy of BLZ945+OSI906 extended overall survival compared with either monotherapy. Log-rank Mantel-Cox test was used to calculate significance. (**H**) BLI of orthotopically xenografted U251 cells genetically engineered to express an *IGF1R*-targeted dox-inducible shRNA (sh; n = 15 or 16 mice; purple lines) or a scrambled control vector (Scr; n = 4 or 5 mice; gray lines). Colored arrows indicate respective administration of dox. Graph shows two combined shRNAs; data for individual hairpins are shown in fig. S13E. Mann-Whitney test was used to calculate significance.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

with continued BLZ945 treatment at a later time point, during the rebound phase we were able to extend survival (Fig. 6F), and quantitative RT-PCR analysis of these tumors confirmed reduced levels of *Igf1* expression along with additional known targets of IL4-Stat6 signaling (*CD36*, *Arg1*, and *Mrc1*) (Fig. 6G) (*27*). Collectively, these data suggest that both NFAT and/or Stat6 signaling contribute to macrophage activation and IGF-1 regulation in rebound tumors, and that pharmacological blockade of either of these pathways is sufficient to reduce the incidence of disease recurrence. However, these models are limited by their inability to distinguish cell type–specific pathway blockade. Furthermore, although FK506 is used in multiple clinical settings, targeted specificity of the Stat6 inhibitor in vivo is not fully characterized. In future studies, elucidating the effects of T cell–specific IL4 expression, or macrophage IL4R, on BLZ945 efficacy through the use of genetic models would provide additional insight into mechanistic communication between different TME cell types.

Combination of CSF-1R and IGF-1R inhibition improves outcome

To formally test our hypothesis that the IGF-1/IGF-1R axis may underlie resistance to CSF-1R inhibition, we targeted IGF-1R in vivo using both pharmacological and genetic approaches. First, we designed similar preclinical intervention trials as for BKM120, except with an inhibitor of IGF-1R (OSI906/Linsitinib). We treated PDG mice with BLZ945 until they showed signs of rebound as seen with MRI, at which point we intervened with OSI906 (Fig. 7A, trial design 1). OSI906 was chosen because it is currently being clinically evaluated for multiple cancer types, its effect on BMDM viability in vitro was minimal compared with other IGF-1R inhibitors tested (fig. S11A), and we confirmed that it is brain-penetrant in rebound tumors by showing reduced p-IGF-1R immunostaining (fig. S11B). In concordance with the BLZ945+BKM120 trial results, we found that rebound tumors treated with OSI906 and continuous BLZ945 significantly extended median survival to 63 days (versus 13 days after recurrence for continuous BLZ945 monotherapy) (Fig. 7B), and markedly reduced tumor progression and proliferation:apoptosis ratios after 2 weeks of treatment (Fig. 7, C and D, and fig. S11C). In contrast, OSI906 monotherapy in rebound tumors (discontinued BLZ945) led to a median survival of just 12 days (Fig. 7B) and was only modestly effective in treatment-naïve tumors (fig. S11, D and E). Together, these results mirror those from the BKM120 trials and suggest that continued CSF-1R inhibition is necessary to drive IGF1R/PI3K signaling dependency, rendering recurrent tumors sensitive to pathway inhibition. Similar combination treatment efficacy and mechanistic commonalities (including IF quantification of Ki67:CC3 ratios, CD206, and p-IGF-1R) were observed in orthotopic xenograft trials performed with patient-derived proneural tumorspheres and with established human glioma cell lines (Fig. 7, E and F, and fig. S12, A to F).

To determine whether tumor outgrowth could be prevented by earlier IGF-1R inhibition, PDG mice with high-grade GBMs were treated continuously with BLZ945 alone until 28 days, at which point OSI906 was added until the trial endpoint (Fig. 7A, trial design 2). With early combination treatment, we indeed extended overall survival and increased the percentage of animals that survived to endpoint (89% BLZ945 +OSI906 versus 44% BLZ945 alone or 0% OSI906 alone) (Fig. 7G). Together, our results suggest that targeting either IGF-1R or PI3K signaling in GBMs resistant to CSF-1R inhibition may interfere with disease progression and improve overall survival.

Given that pharmacological inhibition of IGF-1R using OSI906 cannot confirm whether tumor cell–specific blockade is sufficient to reduce recurrent disease, we used a genetic approach to target this receptor in glioma cells. U251 glioma cell lines were genetically engineered to express a doxycycline (dox)–inducible short hairpin RNA (shRNA) against *IGF1R*, and orthotopic xenograft experiments were performed (trial design is provided in fig. S13A). Two independent shRNAs targeting *IGF1R* were used, and a scramble-sequence shRNA was used as a control (fig. S13, B to D). We found that in both cases, dox-induction of the shRNAs during the rebound phase of the trial (day 12) mitigated tumor progression compared with no-dox control animals (Fig. 7H and fig. S13E). These results support our hypothesis that tumor cell–specific IGF-1R contributes to BLZ945 resistance and disease recurrence.

Discussion

Here, we describe the development of acquired resistance to CSF-1R inhibition in mouse models of GBM. Although initial therapeutic response to CSF-1R inhibition is robust, rapid, and completely penetrant, we found that approximately half of the animals eventually develop resistance, with rapidly progressing rebound tumors. In light of recent results from ongoing patient studies with CSF-1R inhibitors in glioma (*46*) and other cancers, our findings suggest the need to prepare for the emergence of therapeutic resistance to CSF-1R inhibitors in GBM in the clinical setting and determine whether brain malignancies other than GBM will respond similarly to CSF-1R inhibition. Although classical mechanisms of tumor cell–intrinsic resistance to cytotoxic and targeted agents have been well-defined—including aberrant drug metabolism and transport, drug target mutation, and activation of alternative survival pathways (*47*)—it still remains unclear whether resistance to TME-directed therapies will follow similar principles. Given that TME-targeted agents are increasingly being evaluated in the clinic (*1*, *2*), it will be critical to mechanistically define how resistance evolves in response to these therapies in order to provide long-term disease management for patients.

In light of this problem, we have now identified a mechanism of drug resistance that can circumvent therapeutic response to a TME-targeted therapy and promote disease progression in the absence of tumor cell–intrinsic alterations. Specifically, we have uncovered a heterotypic paracrine signaling interaction that is initiated by the TME and drives resistance to CSF-1R inhibition. In rebound tumors, we found that IGF-1 is up-regulated in TAMs in response to IL4 (possibly supplied by T cells or alternative cell types), in part via NFAT activation. IGF-1 secretion into the extracellular environment results in activation of IGF-1R and PI3K signaling in glioma cells, supporting tumor growth and malignancy (Fig. 8, model). Multiple nodes in this signaling loop can be targeted therapeutically by agents that are currently used clinically—including OSI906, BKM120, or FK506—resulting in a substantial improvement in survival in the preclinical setting when combined with CSF-1R inhibition. Indeed, given that PI3K signaling is aberrantly activated in a substantial proportion of GBM patients (*16*), and that recent clinical trial results show limited efficacy

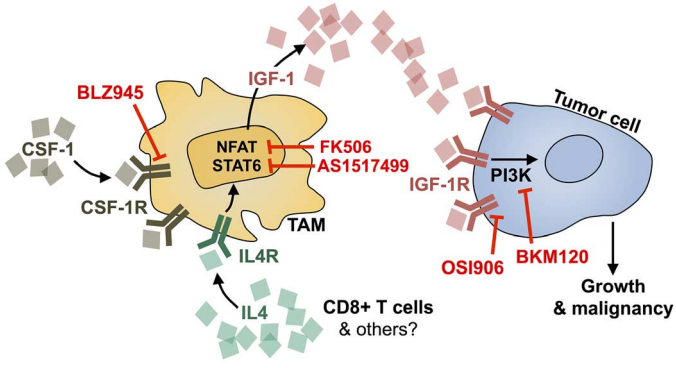


Fig. 8. Working model for mechanism of resistance to CSF-1R inhibition in glioma. IGF-1 is up-regulated in TAMs in response to long-term CSF-1R inhibition in GBM. IGF-1 secretion into the extracellular environment results in activation of IGF-1R on tumor cells and downstream PI3K signaling to support tumor regrowth during continuous BLZ945 treatment. Upstream of IGF-1 in TAMs, NFAT and/or Stat6 transcriptional activity regulate its expression. This is thought to be initiated in response to IL4/IL4R pathway activation, feeding in from other cell types in the TME, including T cells (and possibly others). Multiple nodes in this signaling loop can be targeted therapeutically—including OSI906, BKM120, AS1517499, or FK506—resulting in a substantial improvement in survival in preclinical trials when combined with CSF-1R inhibition.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

in recurrent (albeit very advanced) GBM (46), it is possible that this pathway could similarly contribute to intrinsic resistance to CSF-1R inhibition, and consequently, those patients may benefit from combinatorial inhibition of these pathways from the outset.

Our findings underscore the importance of bidirectional feedback between cancer cells and their microenvironment and support the notion that although stromal cells are less susceptible to genetic mutation than are cancer cells, a tumor can nonetheless persist by exploiting its extracellular environment to acquire a resistant phenotype. Thus, we propose that an integrated analysis of cancer cells with their microenvironment will be critical to understanding both their parallel evolution during tumor progression and their capacity for adaptation in the context of therapeutic intervention and the development of resistance.

Materials and methods

Study design

The overall objective of our study was to understand how resistance to CSF-1R inhibitors develops in high-grade glioma by using various in vivo and in vitro models. Within the animal studies, mice were randomly assigned to different therapy groups, which included treatment with BLZ945 (a CSF-1R inhibitor) in combination with inhibitors of putative resistance pathways (such as IGF-1R or PI3K), versus single-agent or vehicle controls. Survival and disease progression were monitored by using a combination of MRI, histology, flow cytometry, and gene expression analyses throughout all trials. To power our studies, sample size was predefined as at least $n = 3$ independent experiments, replicates, or samples for in vitro and in vivo experiments, and up to a maximum of $n = 90$ mice for survival analyses. Replicate values are indicated for each experiment in corresponding figure legends. All analyses were calculated in a blinded manner through numerical coding of samples. For all long-term survival trials, 26 weeks was selected as a predefined endpoint because mice in the *Ink4a/Arf*$^{-/-}$ background develop spontaneous lymphomas and sarcomas beginning at ~30 weeks of age. For in vivo trials, mice were considered outliers if they developed (i) spontaneous lymphomas (given this disposition in the *Ink4a/Arf*$^{-/-}$ background), (ii) early symptoms of tumor burden >1 week before trial enrollment, or (iii) development of hydrocephalus before trial enrollment.

Biologicals and pharmaceuticals

BLZ945 (8) was obtained from Novartis (Basel, Switzerland) for both in vitro and in vivo use; 6700 nM BLZ945 was used for all in vitro experiments in tumor cells, versus an equal percent dimethyl sulfoxide (DMSO) as a vehicle control. This concentration represents 100 times the median inhibitory concentration (IC_{50}) for BLZ945 in macrophages (8). For in vivo experiments, BLZ945 was obtained preformulated at 12.5 mg/ml. BLZ945 was administered orally 1× daily at 200 mg/kg, and 20% Captisol was used as a vehicle control. BKM120 (Buparlisib; 1 μM unless indicated otherwise), AEW541, ADW742, BMS754807, and OSI906 (Linsitinib; 10 μM unless indicated otherwise) were purchased from Selleckchem (Houston, TX) for in vitro use and were used up to 100 μM for dose response assays versus an equal percent DMSO as a vehicle control. BKM120 and OSI906 were purchased from ChemieTek (Indianapolis, IN) for in vivo use. BKM120 was formulated by dissolving 52 mg into 500 ml NMP, boiling, and then adding 9.5 ml of PEG300. BKM120 was administered orally 1× daily at 20 mg/kg, and NMP:PEG300 (1:19) was used as a vehicle control. Animals were dosed at 20 mg/kg so as to avoid off-target effects, because BKM120 binds tubulin at concentrations above 50 mg/kg in subcutaneous tumor models, but not below 40 mg/kg (15). OSI906 was formulated daily at 4 mg/ml in 25 mM tartaric acid with shaking and sonication for ~15 min. OSI906 was administered orally once daily at 40 mg/kg, and 25 mM tartaric acid was used as a vehicle control. We chose to treat at 40 mg/kg because the maximum tolerated dose for OSI906, 75 mg/kg (48), was found to be toxic within 4 days in our studies. For all combination trials, BLZ945 was administered in the morning, and BKM120 or OSI906 was administered in the evening. The NFAT inhibitor, INCA-6, was purchased from Tocris (Bristol, UK) and was used at a concentration of 40 μM for in vitro use (49). For in vivo inhibition of NFAT signaling, FK506 was used to inhibit the activating interaction between calcineurin and NFAT, at a dose of 10 mg/kg (administered intraperitoneally every 3 days) (50). The Stat6 inhibitor, AS1517499, was purchased from Axon Medchem (Reston, VA) and was used at a concentration of 50 nM for in vitro use and dosed at 10 mg/kg for in vivo use (administered intraperitoneally 2× weekly) (51). The vehicle control for FK506 and AS1517499 in vivo was 10% ethanol (EtOH) and 1% Tween-80 in phosphate-buffered saline (PBS). For in vitro PCR assays, recombinant mouse IL4 (R&D Systems, Minneapolis, MN) was used at a concentration of 10 ng/ml, recombinant mouse TGFβ1 (R&D Systems) was used at a concentration of 50 ng/ml, and the TGFβ1 type 1 receptor inhibitor SB431542 (Tocris) was used at a concentration of 10 μM. For ex vivo GMEC assays, a neutralization antibody against IGF-1 (R&D Systems) was used at a concentration of 0.5 μg/ml. For culture of macrophages in vitro, recombinant mouse CSF-1 was used at a concentration of 10 ng/ml. For in vitro assays using BMDMs, CSF-1 supplementation was excluded from all experimental conditions.

Animals

Crl:NU(NCr)-Foxn1^{nu} immunodeficient mice were purchased from Charles River Laboratories (Wilmington, MA) for orthotopic transplantation studies. NOD/CB17-Prkdc^{scid} immunodeficient mice were purchased from The Jackson Laboratory (Bar Harbor, ME) for orthotopic implantation of human cells. Three different transgenic mouse models expressing the avian tv-a receptor under the control of the nestin (N) promoter in either mixed strain or BL6 backgrounds were used (Ntv-a;*Ink4a/Arf*$^{-/-}$, Ntv-a, and Ntv-a;*Pten*^{flox}), all previously described (13, 52–57) and bred within the MSKCC animal facility. *Stat6*$^{-/-}$, *Il4ra*^{flox};*LysM-cre*, and WT C57BL/6 (BL6) mice were also bred within the MSKCC animal facility and were used for bone marrow isolations. All animal studies were approved by the Institutional Animal Care and Use Committee of MSKCC.

PDG mouse model

The initiation of PDGF-driven gliomas (PDGs) with RCAS-hPDGF-B-HA in adult mice has been previously described (8, 13, 58). Briefly, Ntv-a; *Ink4a/Arf*$^{-/-}$ mice were fully anesthetized with ketamine/xylazine before surgery. Pain management included a 50-μl subcutaneous injection of bupivacaine (0.25%) at the surgical site before surgery, and an intraperitoneal injection of buprenorphine immediately after surgery. Mice were intracranially injected with DF-1:RCAS-hPDGF-B-HA cells (200,000 cells/1 μl) between 5 and 6 weeks of age by using a fixed stereotactic apparatus (Stoelting, Wood Dale, IL). Injections were made into the right frontal cortex, ~1.5 mm lateral and 1 mm caudal from bregma, and at a depth of 2 mm into the subventricular zone (SVZ). In this model, injection into the SVZ induces tumors with low latency (4 to 5 weeks), 100% penetrance, and histological features characteristic of patient GBM including microvascular proliferation and pseudopalisading necrosis (13). The incision was sealed by using Vetbond tissue adhesive (3M, Maplewood, MN). Tumors were imaged with MRI after 5 weeks, and drug intervention was initiated for tumors ≥2 mm^3. A total of 90 animals were treated in long-term experiments with BLZ945 alone, which represented five independent cohorts. These data are compiled and presented in Figs. 1D, 2E, and 7G (red lines).

p53 *KD* mouse model

Injections were performed as described for the PDG mouse model above, except Ntv-a mice were used (WT *Ink4a/Arf*). Mice were intracranially injected with a 1:1 ratio of DF-1:RCAS-hPDGF-B-HA cells and DF-1:RCAS-shp53 cells (total of 300,000 cells/2 μl) between 5 and 6 weeks of age. Injection into the SVZ in the *p53* knockdown (KD) model induces high-grade tumors with low latency (6 to 7 weeks), 100% penetrance, and histological features of human GBM (13, 54, 55). Tumors were detected by MRI after 6 to 7 weeks, at which point drug intervention with BLZ945 was initiated (Fig. 2F).

Pten *KO* mouse model

Injections were performed as described for the PDG mouse model above, except Ntv-a;*Pten*^{flox} mice were used (WT *Ink4a/Arf*). Mice were intracranially injected with a 1:1 ratio of DF-1:RCAS-hPDGF-B-HA cells and DF-1:RCAS-Cre cells (total of 300,000 cells/2 μl) between 5 and 6 weeks of age. Injection into the SVZ in the *Pten* KO model induces tumors with moderate latency (8 to 12 weeks) and penetrance (~20 to 30%). Tumors that form harbor key characteristics of human GBM, including highly infiltrative histology (13, 54, 56, 57). Tumors were detected with

Downloaded from http://science.sciencemag.org/ on May 19, 2016

MRI after 8 to 12 weeks, at which point drug intervention with BLZ945 was initiated (Fig. 2, F and G).

Derivation of mouse primary glioma cell lines from PDG tumors

To derive rebound or dormant cell lines from BLZ945-treated PDG tumors, MRI was used to confirm whether a particular tumor was in rebound or dormancy phase. Macrodissected rebound or dormant lesions from the BLZ945-treated PDG mouse model were manually dissociated and filtered through a 40-μm-mesh filter. The cell suspension was washed twice with PBS and cultured in Mouse Neural Stem Cell (mNSC) Basal Media (Stem Cell Technologies, Vancouver, Canada) containing mNSC proliferation supplement, 1 mg/ml Heparin (Stem Cell Technologies), 10 ng/ml recombinant epidermal growth factor (rEGF; Invitrogen, Carlsbad, CA), and 20 ng/ml recombinant basic-fibroblast growth factor (rbFGF; Sigma, Bandai, Japan). To generate cell lines in monolayer, tumorsphere cultures were expanded, dissociated, and transferred to culture with Dulbecco's minimum essential medium (DMEM) + 10% fetal bovine serum (FBS) (*59*). In total, five rebound cell lines were derived (89AReb, 89BReb, 74Reb, 48Reb, and 52Reb), and one cell line was derived from the 28-day dormant time point. The 28-day dormant cells (fig. S8G) took several weeks before starting to proliferate in culture, and upon transplantation into naïve animals, the cells did not give rise to growing tumors [bioluminescent imaging (BLI) signal remained stable and was monitored up to 22 days]. Derivation of the PDGC23 primary glioma cell line from an untreated/naïve mixed-background PDG mouse was described previously (*8*).

Human cell lines

Human umbilical vein endothelial cells (HUVECs) were purchased from ATCC (Manassass, VA). Human brain microvascular endothelial cells (HBMECs) and human astrocytes were purchased from Sciencell (Carlsbad, CA). Astrocytes were cultured on poly-L-lysine-coated plates, and both HUVECs and HBMECs were cultured on gelatin-coated plates with endothelial cell media (ECM; Sciencell) + 10% FBS + an endothelial cell growth factor supplement. The U251 (commercially available) and TS573 (patient-derived) cell lines were selected on the basis of our previously published work, which showed efficacy in response to BLZ945 in orthotopic xenograft trials (*8*). The patient-derived TS573 glioma tumorsphere line was derived from a consenting patient under Institutional Review Board (IRB)–approved protocols for the banking of excess tumor tissue during routine surgical resection [MSKCC IRB nos. 99-125A(2) and 06-107], as previously described (*8*, *60*). Tumorspheres were maintained in Human Neural Stem Cell (hNSC) Basal Media (Stem Cell Technologies) containing hNSC proliferation supplement, 1 mg/ml Heparin (Stem Cell Technologies), 10 ng/ml rEGF (Invitrogen), and 20 ng/ml rbFGF (Sigma). Tumorspheres were passaged by dissociation with Accutase cell detachment solution (Millipore, Billerica, MA). Characterization and molecular subtyping by Sequenom (San Diego, CA) and NanoString (Seattle, WA) and aCGH were performed as previously described (*8*). Briefly, aCGH on primary spheroids showed a high-level amplification of *PDGFRA* and *CDK6* loci and a regional chromosome 5 loss. NanoString analysis confirmed *PDGFRA* overexpression, and Sequenom analyses were negative for *IDH1/2* mutations.

Isolation of BMDMs

To generate macrophages from bone marrow, femurs and tibiae from *Stat6*$^{-/-}$, *Il4ra*^{flox}*;LysM-cre*, or WT BL6 mice were flushed and cells harvested under sterile conditions. The isolate was filtered through a 40-μm-mesh filter and cultured in 30-ml Teflon bags (PermaLife PL-30) for 5 to 7 days in DMEM + 10% FBS + 10 ng/ml recombinant mouse CSF-1 (R&D Systems). Media were changed every other day.

TGL infections

Cell lines were labeled with a triple-imaging vector [TK-GFP-Luc (TGL)] (*61*) for use in orthotopic in vivo experiments. The TGL vector was developed to enable noninvasive in vivo imaging of tumor growth over time. A standard protocol for retroviral infection was used. Briefly, GP2-293T cells were transfected with the TGL construct and pCL-Ampho at a 1:1 ratio, using Fugene (Promega, Madison, WI) and OptiMEM (Gibco, Thermo Fisher). Twelve hours later, media was replaced with complete antibiotic-free DMEM and collected for 3 consecutive days for infection of target cells.

Orthotopic transplantation experiments

TGL-labeled cells were resuspended in antibiotic-free, serum-free DMEM for all orthotopic injections. Mice were fully anesthetized with ketamine/xylazine before surgery. Pain management included a 50-μl subcutaneous injection of bupivacaine (0.25%) at the surgical site before surgery, and an intraperitoneal injection of buprenorphine immediately after surgery. Mice were intracranially injected with glioma cells between 5 and 6 weeks of age by using a fixed stereotactic apparatus (Stoelting). The number of cells injected was as follows: mouse glioma lines, 2.5×10^4 to 5×10^4 cells/2 μl; human U251 cells, 2.5×10^5 cells/2 μl; and patient-derived TS573 cells, 5×10^4 cells/2 μl. Injections were made to the right frontal cortex, ~1.5 mm lateral and 1 mm caudal from bregma and at a depth of 2 mm. Hydrogen peroxide was used to clean the hole made by the surgical drill, and bone wax was used to close the hole. The incision was sealed by using Vetbond tissue adhesive (3M). One week after injections, mice were randomly assigned to vehicle or BLZ945 treatment groups and dosed 1× daily by means of oral gavage. BLI (Xenogen IVIS-200 Optical In Vivo Imaging System) was performed every 3 to 5 days over the course of the experiment in order to monitor tumor progression and response to therapy. Once BLZ945-treated tumors reached three times the volume of their lowest BLI measurement, tumors were considered "resistant," and mice were randomly assigned to combination treatment with BLZ945 + OSI906, or BLZ945 + vehicle. For dox-inducible *IGF1R* shRNA experiments with U251 cells, mice were additionally assigned to either ± dox groups (dox hyclate diet formulated at 2500 mg/kg; Envigo, Cambridgeshire, UK).

Animal sacrifice and tissue harvest

Mice were euthanized at the defined 26-week endpoint, or when symptomatic (poor grooming, lethargy, weight loss, hunching, macrocephaly/hydrocephalus, or seizures). We selected 26 weeks because mice in the *Ink4a/Arf*$^{-/-}$ background develop spontaneous lymphomas and sarcomas beginning at ~30 weeks of age (*62*). Euthanasia was performed by means of either carbon dioxide asphyxiation or anesthesia (avertin; 2,2,2-tribromoethanol; Sigma) followed by cervical dislocation. For snap freezing of whole-tumor samples, mice were euthanized 1 hour after the last treatment dose, and tissues were collected, frozen immediately in liquid nitrogen, and stored at –80°C for subsequent applications (such as RNA isolation and protein extraction). For isolation of whole tissues for histology, mice were fully anesthetized with avertin, transcardially perfused with 10 ml of PBS, followed by 10 ml of paraformaldehyde (PFA; 4% in PBS). Tissues were incubated in PFA overnight at 4°C, rinsed in PBS, and then transferred to sucrose (30%) for 2 to 3 days at 4°C. For hypoxia analysis, mice were injected intraperitoneally with 60 mg/kg of pimonidazole [hypoxyprobe-1 (HPI)] ~20 min before sacrifice and tissue collection. All tissues were embedded and frozen in Optimal Cutting Temperature (OCT) compound (Tissue-Tek), and 10-μm cryostat tissue sections were used for all subsequent staining and analyses.

IF staining

For IF staining, 10-μm frozen sections were thawed and dried at room temperature and rinsed in PBS. Tissue sections were blocked in 0.5% Blocking Reagent (PerkinElmer, Boston, MA) (1 hour at room temperature or overnight at 4°C), followed by incubation in primary antibody (2 hours at room temperature or overnight at 4°C). All primary antibodies and dilution factors are listed in table S4. Sections were then washed in PBS and incubated with the appropriate fluorophore-conjugated secondary antibody (Molecular Probes, Eugene, OR) at a dilution of 1:500 in 0.5% PNB (1 hour at room temperature). 4′,6-diamidino-2-phenylindole (DAPI) was used as a counterstain before mounting with fluorescent mounting media (Dako, Carpinteria, CA).

IHC and Von Kossa staining

For manual IHC, tissue sections were first subjected to citrate buffer–based antigen retrieval by submerging in antigen unmasking solution (0.94% v/v in distilled water; Vector Laboratories, Burlingame, CA) and microwaving for 10 min, followed by cooling to room temperature for ≥30 min. Endogenous peroxidases were blocked for 10 min with Dual Endogenous Enzyme Block (Dako). Slides were incubated with serum-free protein block (Dako) for 1 hour at room temperature

Downloaded from http://science.sciencemag.org/ on May 19, 2016

and then incubated with primary antibody in a humidity chamber overnight at 4°C. The following day, slides were washed and incubated with horseradish peroxidase (HRP)–conjugated secondary antibodies (Jackson Immunoresearch, West Grove, PA) for 1 hour at room temperature, and positive staining was detected by using diaminobenzidine (DAB) substrate-chromogen. Haematoxylin was used as a counterstain, and slides were mounted with VECTASHIELD mounting media (Vector Laboratories). As an alternative to manual staining, a Ventana autostainer was used for staining of mouse glial fibrillary acidic protein (GFAP), human phospho (p)–AKT, and human MRC1, which included automated deparaffinization, citrate buffer–based antigen retrieval, nonspecific protein and endogenous peroxidase block, antibody incubation, and DAB detection. All primary antibodies and dilution ratios for both manual staining and autostaining are listed in table S4. For visualization of calcium deposition in tissue sections, a Von Kossa staining kit was used as per manufacturer's instructions (Abcam, Cambridge, UK). Briefly, tissues are treated with a silver nitrate solution, which is deposited by replacing calcium reduced by ultraviolet light. Nuclear fast red was used as a counterstain (Vector Laboratories).

Histology and grading

For analysis of tissue histology and grading of tumor malignancy, hematoxylin and eosin (H&E) staining was performed by using a Tissue-Tek automated slide stainer, and slides were mounted with VECTASHIELD mounting media (Vector). Tissues were blindly graded by a neuropathologist (J. Huse, MSKCC) according to standard World Health Organization criteria (*63*). For central nervous system tumors, this grading system is based on a malignancy scale, in which tumors that are minimally proliferative and infiltrative are considered grade I, whereas the most histologically aggressive, infiltrative, and incurable tumors (glioblastomas) are grade IV (*63*). GFAP quantification was performed by use of the Allred scoring method, which is the sum of a proportion score (0 to 5) and an intensity score (0 to 3) for a given marker (*64*).

Image analysis

Stained tissue sections were visualized under a Carl Zeiss (Oberkochen, Germany) Axioimager Z1 microscope equipped with an ApoTome.2, or a Carl Zeiss Axioimager M1 epifluorescence and brightfield microscope. Staining analyses were performed by using a TissueGnostics (Vienna, Austria) slide scanning platform and TissueQuest analysis software. This automated analysis platform was used to quantify number of positive counts, area of positive staining, and/or microvascular density within stained tissue sections in an unbiased manner (*8*).

Patient GBM tissue samples

Patient GBM tissue samples used for staining and analysis (Fig. 5I and fig. S9B) were obtained from the Brain Tumor Center, MSKCC. Eighteen patient tissue samples were used in total. IHC staining for MRC1 and p-AKT was performed as described above ("IHC and Von Kossa staining," above), and quantitation was performed by using the Axioimager Z1 scanning microscope ("Image analysis," above). Correlational analysis across patients was performed by using GraphPad Prism 6.0 ("Data presentation and statistical analysis," below). Patient information for all samples can be found in table S2.

Protein isolation and immunoblotting

To evaluate phospho-protein status of IGF-1R and AKT in culture, cells were seeded at 70 to 80% confluence, serum-starved for ~12 hours, and then treated with OSI906 for 30 min. To evaluate phosphorylation of IGF-1R or AKT in snap-frozen tumor samples, tissues were dissociated with a glass homogenizer on ice in the presence of lysis buffer. All protein lysates were prepared in radioimmunoprecipitation assay lysis buffer supplemented with Halt protease inhibitor (1:100; Thermo Scientific, Waltham, MA) and Halt phosphatase inhibitor (1:100; Thermo Scientific), and protein was quantified by using a BCA protein assay kit (Pierce). Equal amounts of protein (20 μg/lane for cell lines, and 100 μg/lane for tissue samples) were loaded onto SDS–polyacrylamide gel electrophoresis precast gels (Invitrogen) and transferred to polyvinylidene difluoride membranes for immunoblotting. Membranes were blocked in 5% milk, incubated with primary antibodies (table S4) for 1 hour at room temperature or overnight at 4°C, washed with 0.1% Tris-buffered saline and Tween 20 (TBS-T), and incubated with HRP-conjugated secondary antibodies (Jackson Immunoresearch) for 1 hour at room temperature. SuperSignal West Femto or Pico chemiluminescent substrate and CL-XPosure Film (Pierce) were used for signal detection.

RNA isolation, reverse transcription, and quantitative RT-PCR

RNA was isolated with TRIzol and deoxyribonuclease-treated, and 1 μg of RNA was used for cDNA synthesis by using a High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, CA). Mouse taqman probes (Applied Biosystems) were used for quantifying expression of *Igf1* (Mm00439560_m1), *Chil3* (Mm00657889_mH), *Ccl17* (Mm01244826_g1), *Retnla* (Mm00445109_m1), *Il4* (Mm00445260_m1), *Arg1* (Mm00475988_m1), *Cd36* (Mm00432403_m1), *Mrc1* (Mm01329362_m1), *Ubc* (Mm02525934_g1; housekeeping), and *Hprt* (Mm01545339_m1; housekeeping). Human taqman probes (Applied Biosystems) were used for quantifying expression of *IL4* (Hs00174122_m1), *IL13* (Hs00174379_m1), *IGF1* (Hs01547656_m1), and *HPRT1* (Hs02800695_m1; housekeeping).

MTT assays

Cell growth rate was determined by using a MTT cell proliferation kit (Roche, Basel, Switzerland). Briefly, cells were plated in ≥triplicate in 96-well plates; 1×10^3 cells/well were plated for mouse glioma cell lines, and 5×10^3 cells/well were plated for BMDMs. For BLZ945 time course experiments, cells were grown in the presence of 6700 nM of BLZ945 versus an equal percent DMSO, media was changed every 48 hours, and viability measurements were taken every 24 hours. For dose response experiments, cells were grown in the presence of an IGF-1R inhibitor (AEW541, ADW742, BMS754807, or OSI906) versus an equal percent DMSO at the doses indicated ("Biologicals and pharmaceuticals," above), and viability measurements were taken after 24 hours. Reduction of the MTT substrate was detected with colorimetric analysis by using a plate reader as per the manufacturer's protocol. Ten microliters of MTT labeling reagent was added to each well and then incubated for 4 hours at 37°C, followed by the addition of 100 μl of MTT solubilization reagent overnight. The mixture was gently resuspended, and absorbance was measured at 595 and 750 nm on a spectraMax 340pc plate reader (Molecular Devices, Sunnyvale, CA).

Ex vivo GMEC assays

GMECs are early-passage heterotypic cell cultures harvested directly from PDG mouse primary tumors. Flow cytometry characterization of these cultures revealed that passage 1 GMECs contain a high abundance of tumor cells, astrocytes, and macrophages, as well as smaller proportions of myeloid progenitors, T cells, and B cells (fig. S8I). In our ex vivo assays, we derived CM from passage 1 rebound GMECs and used this CM to treat naïve BMDMs in vitro (experimental design is provded in fig. S8J). We collected CM from these GMEC-stimulated BMDMs (Stim CM) and applied it to either rebound glioma cell lines (highly sensitive to IGF-1R inhibition) or naïve glioma cell lines (less sensitive to IGF-1R inhibition), plus or minus a neutralizing antibody for IGF-1. An MTT assay was used to assess changes in growth in response to each treatment condition, as described above.

Flow cytometry and FACS on primary mouse tissues

Mice were fully anesthetized with avertin and transcardially perfused with 20 ml of PBS. The brain was then isolated, and the tumor was macrodissected from the surrounding normal tissue. Tissues were mechanically dissociated and filtered into a single-cell suspension. For flow cytometry, cells were counted and incubated with Fc block for 1 hour ($1:100/10^6$ cells; BD Biosciences, East Rutherford, NJ), followed by a 30-min incubation with LIVE/DEAD fixable dead cell kit (Invitrogen) and then a 1-hour incubation with conjugated antibodies for extracellular markers. For FACS, cells were counted and incubated with Fc block for 1 hour, followed by a 1-hour incubation with conjugated antibodies, and then stained with DAPI for dead-cell exclusion. All antibodies and dilution ratios used for these experiments are listed in table S4. OneComp eBeads (eBioscience, San Diego, CA), ArC Amine Reactive Compensation Beads (Invitrogen), and/or cell suspensions from spleen were used for compensation controls. A BD Biosciences LSR-Fortessa was used for flow cytometry, and a

Downloaded from http://science.sciencemag.org/ on May 19, 2016

BD Biosciences FACSAria III was used for cell sorting.

Distinguishing between putative BMDMs and microglia by flow cytometry

Flow cytometry was used to evaluate the proportions of peripherally derived BMDMs versus resident microglia in PDG tumors across different treatment groups, according to published methods (*22*–*24*). Briefly, after gating on live cells we used cell-surface expression of CD45 and CD11b to distinguish between the two populations, in which $CD45^{lo}$ $CD11b^{+}$ defined putative microglia, whereas $CD45^{hi}$ $CD11b^{+}$ defined putative BMDMs. In using this method, we acknowledge its limitations and recognize that these two populations cannot be definitively distinguished without lineage-tracing experiments that specifically label yolk sac–derived microglia or peripherally recruited BMDMs.

FACS purification of human peripheral immune cell types

Human buffy coats from three consenting healthy donors were obtained from the New York Blood Center. For isolation of neutrophils and eosinophils, buffy coats were directly red blood cell lysed (BD Biosciences PharmLyse) for 15 min at room temperature. All other cell types were isolated from the top layer of a Ficoll gradient separation (HistoPaque, Sigma). Cells were pelleted for 10 min at 300*g* and washed twice with FACS buffer (PBS + 2% fetal bovine serum) and Fc blocked (Biolegend TruStain FcX). Cells were incubated with the appropriate antibodies for 15 min (table S4). Cells were FACS-purified on an Aria III (BD Biosciences). For human macrophage differentiation, PBMCs were isolated from buffy coats following a Ficoll gradient. Monocytes were further purified from the interphase of a 70/30% Percoll gradient. Monocytes were then washed twice with PBS and cultured in Teflon bags (Origin) for 7 days in DMEM + 2% human serum + recombinant human CSF-1 (10 ng/ml; R&D Systems). CSF-1 and media were replaced every 48 hours.

aCGH

All sequencing and quality control was performed at the Integrated Genomics Operation, MSKCC. DNA was isolated from passage 1 PDG neurospheres from rebound tumors or corresponding liver tissue by using TRIzol as per manufacturer instructions (Invitrogen). Three micrograms of DNA was used with an Agilent (Santa Clara, CA) standard cy5/cy3 labeling protocol. Briefly, Agilent Mouse CGH 180k arrays were hybridized at 65°C and 20 revolutions per min for 40 hours. Slides were then scanned by using the Agilent scanner according to the manufacturer's instructions. The raw data were extracted with Feature Extraction by using Agilent default analysis settings. Subsequent analyses were performed in R v3.1.0 by using the "DNAcopy" package (*65*).

RNA-seq

Three RNA-seq experiments were performed in total: (i) FACS-purified tumor cells ($CD45^{-}PDGFR\alpha^{+}$) and TAMs ($CD45^{+}CD11b^{+}Gr1^{-}$) from Veh, EP, and Reb tumors; (ii) FACS-purified astrocytes ($CD45^{-}GLAST^{+}$), B cells ($CD45^{+}CD19^{+}$), Tc cells ($CD45^{+}CD3^{+}CD8^{+}$), and bulk T cells ($CD45^{+}CD3^{+}CD8^{-}$) from Reb tumors; and (iii) FACS-purified TAMs ($CD45^{+}CD11b^{+}Gr1^{-}$) from 28-day and Reb tumors. RNA-sequencing and quality control was performed at the Integrated Genomics Operation, MSKCC, or Genewiz (Plainfield, NJ). In all cases, RNA was isolated by using TRIzol as per manufacturer instructions (Invitrogen), and RNA integrity was assessed with an Agilent Bioanalyzer 2100. RNA-seq libraries were prepared by using the SMART-Seq library preparation kit, and 2 × 50 or 2 × 100 base pair sequencing was performed on an Illumina HISEq.2000. Sequencing quality was assessed with FASTQC (www.bioinformatics.babraham.ac.uk/projects/fastqc). Reads were mapped to the mouse genome (mm10) by using STAR 2.3.0e (*66*) with the default parameters, a minimum intron length of 70 base pairs, and a maximum of 100,000 base pairs. BAM files were generated and sorted, and duplicate reads were then removed by using SAMTOOLS (*67*). Read counts were tabulated with HT-Seq by using "union" mode and the iGenomes GFF file as a reference (Illumina) (*68*).

Gene expression analyses

Raw count data from HT-Seq was imported into R (v3.1.0) and normalized by using limma voom (*69*). Principal component analysis was completed by using the princomp function on mean centered data. A log2-fold change cutoff of 1 and a false discovery rate of 10% were applied for all differential gene expression analyses (table S1). Significantly up-regulated genes from these lists were used in gene ontology analyses by using DAVID (*70*). GSVA was performed on RNA-seq data from FACS-purified $PDGFR\alpha^{+}$ tumor cells, using the GSVA package (*14*) with gene sets from the C2 group from the Molecular Signatures Database (MSigDB) (*71*). A log2-fold change cutoff of 1 and a false discovery rate cutoff of 10% were used to determine differentially enriched gene sets (fig. S3B). The spectrum model of macrophage activation was assessed with GSEA, using the gsea function from the phenoTest package in R (http://rpackages.ianhowson.com/bioc/phenoTest). A minimum P value of $<1 \times 10^{-16}$ was used to represent significance values that were reported as 0.0 and outside of the determined distribution. Gene sets were adapted from a previous study in which mouse macrophages were stimulated with IFN-γ, IL-4, TNF-α, TGF-β, IL-1β, MALP2, or CPG (*27*). A literature-derived IL-4 responsive gene set was generated through the use of QIAGEN's Ingenuity iReport (www.qiagen.com/ingenuity).

TF activity analysis

TF activity analysis was performed as an adaptation of previously published methods: ISMARA (*72*) and RegulatorInference (*73*). Briefly, transcription start sites and Motevo predictions (*74*) of binding sites were downloaded from the SwissRegulon (http://swissregulon.unibas.ch/fcgi/sr/downloads). These were used to determine the number of predicted binding sites for 185 TF families across all mouse promoters. Promoters were designated as 2 kb upstream and downstream of transcription start sites. This tabulated matrix was then used in a ridge regression to model log2 gene expression values generated by means of voom. Ridge regression was performed with the glmnet function in R (*75*). The regularization parameter, λ, was identified for each sample through 10-fold cross validation. The coefficients for each TF family were z-scored and used as relative TF activity scores in subsequent analyses. Differentially enriched TFs were identified by using the z-scored values in limma with a log2-fold change cutoff of 1 and a false discovery rate of 10% (table S3).

External data set analysis

RNA-seq expression data from the TCGA glioblastoma patient data set was downloaded by using TCGA-assembler (*76*). For survival analyses, these data were filtered for patients with updated clinical information from the Broad Firehose. Correlations between *IGF1* and macrophage markers (*CD163*, *MRC1*, *CSF1R*, *CD68*, and *AIF1*) or astrocyte markers (*GFAP* and *ALDH1L1*) were assessed by using a Spearman correlation coefficient. Normalized gene expression data for human bulk tumor versus tumor-associated macrophage fraction were downloaded from the Gene Expression Ominbus (GEO) under accession no. GSE16119 (*44*). Subtype calls for patients were obtained from GlioVis (http://gliovis.bioinfo.cnio.es). *PI3K* signature scores were tabulated with a single-sample gene set enrichment, as used for macrophage activity analysis. The gene set was from the Hallmark collection from MsigDB, systematic name M5923 (*43*).

Data availability

All sequencing data has been deposited to the GEO under the accession no. GSE69104 and aCGH data under the accession no. GSE80399. All code used in this project can be found at https://bitbucket.org/bowmanr/joycelab-brain-tme.

Data presentation and statistical analysis

GraphPad Prism 6.0 or R Studio was used for all data analysis. Parametric data are presented as mean ± standard error (SEM) and were analyzed by means of an unpaired two-tailed Student's t test. For multiple comparisons, a one-way analysis of variance (ANOVA) with Tukey's or Dennett's correction was used as noted in the figure legends. Nonparametric data were analyzed by means of a Mann-Whitney test on ranks. For survival curves, P values were obtained by using the Log Rank (Mantel-Cox) test. Fisher's exact test was used for histological tumor grading. A pairwise Spearman correlation test was used for correlational analyses. $P < 0.05$ was considered as statistically significant in all cases. Principal component analyses, correlation plots, Volcano plots, heatmaps, and network plots were plotted in R Studio using

Downloaded from http://science.sciencemag.org/ on May 19, 2016

base graphics (www.R-project.org), rgl (http://CRAN.R-project.org/package=rgl), gplots (http://CRAN.R-project.org/package=gplots), ggplot2 (*77*), and qgraph packages (*78*).

REFERENCES AND NOTES

1. D. F. Quail, J. A. Joyce, Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* **19**, 1423–1437 (2013). doi: 10.1038/nm.3394; pmid: 24202395
2. M. R. Junttila, F. J. de Sauvage, Influence of tumour microenvironment heterogeneity on therapeutic response. *Nature* **501**, 346–354 (2013). doi: 10.1038/nature12626; pmid: 24048067
3. R. Stupp *et al*., Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352**, 987–996 (2005). doi: 10.1056/NEJMoa043330; pmid: 15758009
4. Y. Komohara, K. Ohnishi, J. Kuratsu, M. Takeya, Possible involvement of the M2 anti-inflammatory macrophage phenotype in growth of human gliomas. *J. Pathol.* **216**, 15–24 (2008). doi: 10.1002/path.2370; pmid: 18553315
5. L. Bingle, N. J. Brown, C. E. Lewis, The role of tumour-associated macrophages in tumour progression: Implications for new anticancer therapies. *J. Pathol.* **196**, 254–265 (2002). doi: 10.1002/path.1027; pmid: 11857487
6. S. F. Hussain *et al*., The role of human glioma-infiltrating microglia/macrophages in mediating antitumor immune responses. *Neuro-oncol.* **8**, 261–279 (2006). doi: 10.1215/15228517-2006-008; pmid: 16775224
7. B. Ruffell, L. M. Coussens, Macrophages and therapeutic resistance in cancer. *Cancer Cell* **27**, 462–472 (2015). doi: 10.1016/j.ccell.2015.02.015; pmid: 25858805
8. S. M. Pyonteck *et al*., CSF-1R inhibition alters macrophage polarization and blocks glioma progression. *Nat. Med.* **19**, 1264–1272 (2013). doi: 10.1038/nm.3337; pmid: 24056773
9. S. J. Coniglio *et al*., Microglial stimulation of glioblastoma invasion involves epidermal growth factor receptor (EGFR) and colony stimulating factor 1 receptor (CSF-1R) signaling. *Mol. Med.* **18**, 519–527 (2012). doi: 10.2119/molmed.2011.00217; pmid: 22294205
10. S. Patel, M. R. Player, Colony-stimulating factor-1 receptor inhibitors for the treatment of cancer and inflammatory disease. *Curr. Top. Med. Chem.* **9**, 599–610 (2009). doi: 10.2174/156802609789007327; pmid: 19689368
11. C. H. Ries *et al*., Targeting tumor-associated macrophages with anti-CSF-1R antibody reveals a strategy for cancer therapy. *Cancer Cell* **25**, 846–859 (2014). doi: 10.1016/j.ccr.2014.05.016; pmid: 24898549
12. W. D. Tap *et al*., Structure-guided blockade of CSF1R kinase in tenosynovial giant-cell tumor. *N. Engl. J. Med.* **373**, 428–437 (2015). doi: 10.1056/NEJMoa1411366; pmid: 26222558
13. D. Hambardzumyan, N. M. Amankulor, K. Y. Helmy, O. J. Becher, E. C. Holland, Modeling adult gliomas using RCAS/t-va technology. *Transl. Oncol.* **2**, 89–95 (2009). doi: 10.1593/tlo.09100; pmid: 19412424
14. S. Hänzelmann, R. Castelo, J. Guinney, GSVA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013). doi: 10.1186/1471-2105-14-7; pmid: 23323831
15. S. M. Brachmann *et al*., Characterization of the mechanism of action of the pan class I PI3K inhibitor NVP-BKM120 across a broad range of concentrations. *Mol. Cancer Ther.* **11**, 1747–1757 (2012). doi: 10.1158/1535-7163.MCT-11-1021; pmid: 22653967
16. C. W. Brennan *et al*., The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013). doi: 10.1016/j.cell.2013.09.034; pmid: 24120142
17. P. Therasse *et al*., New guidelines to evaluate the response to treatment in solid tumors. *J. Natl. Cancer Inst.* **92**, 205–216 (2000). doi: 10.1093/jnci/92.3.205; pmid: 10655437
18. J. E. Burda, M. V. Sofroniew, Reactive gliosis and the multicellular response to CNS damage and disease. *Neuron* **81**, 229–248 (2014). doi: 10.1016/j.neuron.2013.12.034; pmid: 24462092
19. E. N. Arwert, E. Hoste, F. M. Watt, Epithelial stem cells, wound healing and cancer. *Nat. Rev. Cancer* **12**, 170–180 (2012). doi: 10.1038/nrc3217; pmid: 22362215
20. F. Ginhoux, J. L. Schultze, P. J. Murray, J. Ochando, S. K. Biswas, New insights into the multidimensional concept of macrophage ontogeny, activation and function. *Nat. Immunol.* **17**, 34–40 (2016). doi: 10.1038/ni.3324; pmid: 26681460
21. R. L. Bowman, J. A. Joyce, Therapeutic targeting of tumor-associated macrophages and microglia in glioblastoma. *Immunotherapy* **6**, 663–666 (2014). doi: 10.2217/imt.14.48; pmid: 25041027
22. G. Locatelli *et al*., Primary oligodendrocyte death does not elicit anti-CNS immunity. *Nat. Neurosci.* **15**, 543–550 (2012). doi: 10.1038/nn.3062; pmid: 22366759
23. K. Gabrusiewicz *et al*., Characteristics of the alternative phenotype of microglia/macrophages and its modulation in experimental gliomas. *PLOS ONE* **6**, e23902 (2011). doi: 10.1371/journal.pone.0023902; pmid: 21901144
24. J. D. Sedgwick *et al*., Isolation and direct characterization of resident microglial cells from the normal and inflamed central nervous system. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7438–7442 (1991). doi: 10.1073/pnas.88.16.7438; pmid: 1651506
25. F. Kratochvill *et al*., TNF counterbalances the emergence of M2 tumor macrophages. *Cell Reports* **12**, 1902–1914 (2015). doi: 10.1016/j.celrep.2015.08.033; pmid: 26365184
26. S. K. Biswas, A. Mantovani, Macrophage plasticity and interaction with lymphocyte subsets: Cancer as a paradigm. *Nat. Immunol.* **11**, 889–896 (2010). doi: 10.1038/ni.1937; pmid: 20856220
27. R. Ostuni *et al*., Latent enhancers activated by stimulation in differentiated cells. *Cell* **152**, 157–171 (2013). doi: 10.1016/j.cell.2012.12.018; pmid: 23332752
28. D. M. Mosser, J. P. Edwards, Exploring the full spectrum of macrophage activation. *Nat. Rev. Immunol.* **8**, 958–969 (2008). doi: 10.1038/nri2448; pmid: 19029990
29. A. Mantovani, S. K. Biswas, M. R. Galdiero, A. Sica, M. Locati, Macrophage plasticity and polarization in tissue repair and remodelling. *J. Pathol.* **229**, 176–185 (2013). doi: 10.1002/path.4133; pmid: 23096265
30. B. J. Faler, R. A. Macsata, D. Plummer, L. Mishra, A. N. Sidawy, Transforming growth factor-β and wound healing. *Perspect. Vasc. Surg. Endovasc. Ther.* **18**, 55–62 (2006). doi: 10.1177/153100350601800123; pmid: 16628336
31. J. Silver, J. H. Miller, Regeneration beyond the glial scar. *Nat. Rev. Neurosci.* **5**, 146–156 (2004). doi: 10.1038/nrn1326; pmid: 14735117
32. F. E. Lund, B. A. Garvy, T. D. Randall, D. P. Harris, Regulatory roles for cytokine-producing B cells in infection and autoimmune disease. *Curr. Dir. Autoimmun.* **8**, 25–54 (2005). doi: 10.1159/000082086; pmid: 15564716
33. D. P. Harris *et al*., Reciprocal regulation of polarized cytokine production by effector B and T cells. *Nat. Immunol.* **1**, 475–482 (2000). doi: 10.1038/82717; pmid: 11101868
34. B. Johansson-Lindbom, C. A. Borrebaeck, Germinal center B cells constitute a predominant physiological source of IL-4: Implication for Th2 development in vivo. *J. Immunol.* **168**, 3165–3172 (2002). doi: 10.4049/jimmunol.168.7.3165; pmid: 11907068
35. O. Butovsky, A. E. Talpalar, K. Ben-Yaakov, M. Schwartz, Activation of microglia by aggregated beta-amyloid or lipopolysaccharide impairs MHC-II expression and renders them cytotoxic whereas IFN-γ and IL-4 render them protective. *Mol. Cell. Neurosci.* **29**, 381–393 (2005). doi: 10.1016/j.mcn.2005.03.005; pmid: 15890528
36. S. J. Forbes, N. Rosenthal, Preparing the ground for tissue regeneration: From mechanism to therapy. *Nat. Med.* **20**, 857–869 (2014). doi: 10.1038/nm.3653; pmid: 25100531
37. M. W. Wynes, D. W. Riches, Induction of macrophage insulin-like growth factor-I expression by the Th2 cytokines IL-4 and IL-13. *J. Immunol.* **171**, 3550–3559 (2003). doi: 10.4049/jimmunol.171.7.3550; pmid: 14500651
38. B. K. Kaspar, J. Lladó, N. Sherkat, J. D. Rothstein, F. H. Gage, Retrograde viral delivery of IGF-1 prolongs survival in a mouse ALS model. *Science* **301**, 839–842 (2003). doi: 10.1126/science.1086137; pmid: 12907804
39. W. H. Zheng, S. Kar, S. Doré, R. Quirion, Insulin-like growth factor-1 (IGF-1): A neuroprotective trophic factor acting via the Akt kinase pathway. *J. Neural Transm. Suppl.* **2000**, 261–272 (2000).pmid: 11205145
40. S. Dupraz *et al*., The insulin-like growth factor 1 receptor is essential for axonal regeneration in adult central nervous system neurons. *PLOS ONE* **8**, e54462 (2013). doi: 10.1371/journal.pone.0054462; pmid: 23349896
41. A. M. Fernandez, I. Torres-Alemán, The many faces of insulin-like peptide signalling in the brain. *Nat. Rev. Neurosci.* **13**, 225–239 (2012). doi: 10.1038/nrn3209; pmid: 22430016
42. M. Pollak, Insulin and insulin-like growth factor signalling in neoplasia. *Nat. Rev. Cancer* **8**, 915–928 (2008). doi: 10.1038/nrc2536; pmid: 19029956
43. A. Liberzon *et al*., The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015). doi: 10.1016/j.cels.2015.12.004; pmid: 26771021
44. A. Murat *et al*., Modulation of angiogenic and inflammatory response in glioblastoma by hypoxia. *PLOS ONE* **4**, e5947 (2009). doi: 10.1371/journal.pone.0005947; pmid: 19536297
45. N. Hermann-Kleiter, G. Baier, NFAT pulls the strings during CD4$^+$ T helper cell effector functions. *Blood* **115**, 2989–2997 (2010). doi: 10.1182/blood-2009-10-233585; pmid: 20103781
46. N. Butowski *et al*., Orally administered colony stimulating factor 1 receptor inhibitor PLX3397 in recurrent glioblastoma: An Ivy Foundation Early Phase Clinical Trials Consortium phase II study. *Neuro-oncol.* **18**, 557–564 (2016). doi: 10.1093/neuonc/nov245; pmid: 26449250
47. C. Holohan, S. Van Schaeybroeck, D. B. Longley, P. G. Johnston, Cancer drug resistance: An evolving paradigm. *Nat. Rev. Cancer* **13**, 714–726 (2013). doi: 10.1038/nrc3599; pmid: 24060863
48. M. J. Mulvihill *et al*., Discovery of OSI-906: A selective and orally efficacious dual inhibitor of the IGF-1 receptor and insulin receptor. *Future Med. Chem.* **1**, 1153–1171 (2009). doi: 10.4155/fmc.09.89; pmid: 21425998
49. M. H. Roehrl *et al*., Selective inhibition of calcineurin-NFAT signaling by blocking protein-protein interaction with small organic molecules. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7554–7559 (2004). doi: 10.1073/pnas.0401835101; pmid: 15131267
50. K. Loser *et al*., FK506 controls CD40L-induced systemic autoimmunity in mice. *J. Invest. Dermatol.* **126**, 1307–1315 (2006). doi: 10.1038/sj.jid.5700185; pmid: 16470176
51. Y. Chiba, M. Todoroki, Y. Nishida, M. Tanabe, M. Misawa, A novel STAT6 inhibitor AS1517499 ameliorates antigen-induced bronchial hypercontractility in mice. *Am. J. Respir. Cell Mol. Biol.* **41**, 516–524 (2009). doi: 10.1165/rcmb.2008-0163OC; pmid: 19202006
52. C. Dai *et al*., PDGF autocrine stimulation dedifferentiates cultured astrocytes and induces oligodendrogliomas and oligoastrocytomas from neural progenitors and astrocytes in vivo. *Genes Dev.* **15**, 1913–1925 (2001). doi: 10.1101/gad.903001; pmid: 11485986
53. E. Tchougounova *et al*., Loss of Arf causes tumor progression of PDGFB-induced oligodendroglioma. *Oncogene* **26**, 6289–6296 (2007). doi: 10.1038/sj.onc.1210455; pmid: 17438529
54. T. Ozawa *et al*., Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. *Cancer Cell* **26**, 288–300 (2014). doi: 10.1016/j.ccr.2014.06.005; pmid: 25117714
55. M. Squatrito *et al*., Loss of ATM/Chk2/p53 pathway components accelerates tumor development and contributes to radiation resistance in gliomas. *Cancer Cell* **18**, 619–629 (2010). doi: 10.1016/j.ccr.2010.10.034; pmid: 21156285
56. X. Hu *et al*., mTOR promotes survival and astrocytic characteristics induced by Pten/AKT signaling in glioblastoma. *Neoplasia* **7**, 356–368 (2005). doi: 10.1593/neo.04595; pmid: 15967113
57. J. T. Huse *et al*., The PTEN-regulating microRNA miR-26a is amplified in high-grade glioma and facilitates gliomagenesis in vivo. *Genes Dev.* **23**, 1327–1337 (2009). doi: 10.1101/gad.1777409; pmid: 19487573
58. A. H. Shih *et al*., Dose-dependent effects of platelet-derived growth factor-B on glial tumorigenesis. *Cancer Res.* **64**, 4783–4789 (2004). doi: 10.1158/0008-5472.CAN-03-3831; pmid: 15256447
59. J. B. Jensen, M. Parmar, Strengths and limitations of the neurosphere culture system. *Mol. Neurobiol.* **34**, 153–161 (2006). doi: 10.1385/MN:34:3:153; pmid: 17308349
60. N. J. Szerlip *et al*., Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3041–3046 (2012). doi: 10.1073/pnas.1114033109; pmid: 22323597
61. V. Ponomarev *et al*., A novel triple-modality reporter gene for whole-body fluorescent, bioluminescent, and nuclear noninvasive imaging. *Eur. J. Nucl. Med. Mol. Imaging* **31**, 740–751 (2004). doi: 10.1007/s00259-003-1441-5; pmid: 15014901
62. M. Serrano *et al*., Role of the INK4a locus in tumor suppression and cell mortality. *Cell* **85**, 27–37 (1996). doi: 10.1016/S0092-8674(00)81079-X; pmid: 8620534
63. D. N. Louis *et al*., The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* **114**,

Downloaded from http://science.sciencemag.org/ on May 19, 2016

97–109 (2007). doi: 10.1007/s00401-007-0243-4; pmid: 17618441
64. D. C. Allred, J. M. Harvey, M. Berardo, G. M. Clark, Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod. Pathol.* **11**, 155–168 (1998). pmid: 9504686
65. V. E. Seshan, A. Olshen, DNAcopy: DNA copy number data analysis. R package version 1.44.0.
66. A. Dobin *et al*., STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). doi: 10.1093/bioinformatics/bts635; pmid: 23104886
67. H. Li *et al*., The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). doi: 10.1093/bioinformatics/btp352; pmid: 19505943
68. S. Anders, P. T. Pyl, W. Huber, HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015). doi: 10.1093/bioinformatics/btu638; pmid: 25260700
69. C. W. Law, Y. Chen, W. Shi, G. K. Smyth, voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014). doi: 10.1186/gb-2014-15-2-r29; pmid: 24485249
70. G. Dennis Jr. *et al*., DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, 3 (2003). doi: 10.1186/gb-2003-4-5-p3 pmid: 12734009
71. A. Subramanian *et al*., Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005). doi: 10.1073/pnas.0506580102; pmid: 16199517
72. P. J. Balwierz *et al*., ISMARA: Automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res.* **24**, 869–884 (2014). doi: 10.1101/gr.169508.113; pmid: 24515121
73. M. Setty *et al*., Inferring transcriptional and microRNA-mediated regulatory programs in glioblastoma. *Mol. Syst. Biol.* **8**, 605 (2012). doi: 10.1038/msb.2012.37; pmid: 22929615
74. M. Pachkov, I. Erb, N. Molina, E. van Nimwegen, SwissRegulon: A database of genome-wide annotations of regulatory sites. *Nucleic Acids Res.* **35** (Database), D127–D131 (2007). doi: 10.1093/nar/gkl857; pmid: 17130146
75. J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010). doi: 10.18637/jss.v033.i01; pmid: 20808728
76. Y. Zhu, P. Qiu, Y. Ji, TCGA-assembler: Open-source software for retrieving and processing TCGA data. *Nat. Methods* **11**, 599–600 (2014). doi: 10.1038/nmeth.2956; pmid: 24874569
77. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
78. S. Epskamp, A. O. J. Cramer, L. J. Waldorp, V. D. Schmittmann, D. Borsboom, qgraph: Network visualizations of relationships in psychometric data. *J. Stat. Softw.* **48**, 1–18 (2012). doi: 10.18637/jss.v048.i04

ACKNOWLEDGMENTS

We thank X. Chen and N. Shah for excellent technical support. We are grateful to C. Brennan for generously providing the TS573 patient tumorsphere line. We thank R. Benezra for critically reading the manuscript and O. Olson, P. Castel, and S. Schwartz for experimental advice. We thank the Animal Imaging Core, the Antitumor Assessment Core, Flow Cytometry Core, and Integrated Genomics Operation at MSKCC for assistance. The data presented in this manuscript are tabulated in the main paper and in the supplementary materials. All sequencing data has been deposited to the GEO under accession no. GSE69104, and aCGH data under accession number GSE80399. BLZ945 is available from Novartis under a material transfer agreement with J.C.S., Novartis. D.F.Q., R.L.B., and J.A.J. are inventors on a patent application filed by MSKCC related to the use of combination therapies with CSF-1R inhibitors for gliomas. This research was supported by the U.S. National Cancer Institute (NCI) (R01CA148967, J.A.J.), the Ludwig Institute for Cancer Research (J.A.J.), fellowships from Canadian Institutes of Health Research (D.F.Q.), NCI (F31CA167863, R.L.B.), American Brain Tumor Association (L.A.), Obra Social Ibercaja (A.J.S.), and the MSKCC Support Grant from the NCI (P30 CA008748).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/aad3018/suppl/DC1
Supplementary Text
Figs. S1 to S13
Tables S1 to S4

18 September 2015; accepted 8 April 2016
10.1126/science.aad3018

Downloaded from http://science.sciencemag.org/ on May 19, 2016

RESEARCH ARTICLE

ENZYME MECHANISMS

The radical mechanism of biological methane synthesis by methyl-coenzyme M reductase

Thanyaporn Wongnate,[1] Dariusz Sliwa,[1]* Bojana Ginovska,[2] Dayle Smith,[2]† Matthew W. Wolf,[3] Nicolai Lehnert,[3] Simone Raugei,[2] Stephen W. Ragsdale[1]‡

Methyl-coenzyme M reductase, the rate-limiting enzyme in methanogenesis and anaerobic methane oxidation, is responsible for the biological production of more than 1 billion tons of methane per year. The mechanism of methane synthesis is thought to involve either methyl-nickel(III) or methyl radical/Ni(II)-thiolate intermediates. We employed transient kinetic, spectroscopic, and computational approaches to study the reaction between the active Ni(I) enzyme and substrates. Consistent with the methyl radical–based mechanism, there was no evidence for a methyl-Ni(III) species; furthermore, magnetic circular dichroism spectroscopy identified the Ni(II)-thiolate intermediate. Temperature-dependent transient kinetics also closely matched density functional theory predictions of the methyl radical mechanism. Identifying the key intermediate in methanogenesis provides fundamental insights to develop better catalysts for producing and activating an important fuel and potent greenhouse gas.

Methanogenic archaea produce more than 90% of Earth's atmospheric methane (*1*), totaling more than 1 billion tons of methane per year globally (*2*). Furthermore, methanogens living in microbial communities containing sulfate- or nitrate-reducing bacteria are responsible for the annual anaerobic oxidation of 0.1 billion tons of methane (*3*–*6*). The enzyme that catalyzes the chemical step of methane synthesis (Eq. 1) or oxidation (the reverse of Eq. 1) is methyl-coenzyme M reductase (MCR), which contains a nickel hydrocorphinate F_{430} at its active site (*4*, *7*–*9*). This reaction involves conversion of the methyl donor, methyl-coenzyme M (methyl-SCoM), and the electron donor, coenzyme B (CoBSH, N-7-mercaptoheptanoylthreonine phosphate) (*10*), to methane and the mixed disulfide CoBS-SCoM (*11*) (Eq. 1). The substrates bind inside a deep substrate channel with CoBSH nearer to the surface, stretching toward methyl-SCoM, which is close to F_{430} (*12*).

$$\text{Methyl-SCoM} + \text{CoBSH} \rightarrow \text{CH}_4 + \text{CoBS-SCoM} \qquad \Delta G^{0'} = -30\ \text{kJ/mol} \qquad (1)$$

The mechanism of methane formation is not fully resolved, mainly because intermediates in the catalytic cycle have not been identified. Uncovering the MCR mechanism is critical because of the important biogeochemical and environmental roles of this enzyme in generating (and metabolizing) a Janus-like compound that serves as a key energy source and is a potent greenhouse gas. Furthermore, the chemical principles underlying both synthesis and activation of methane may inform the development of catalysts that mimic the structure and/or function of the key enzymatic intermediate(s) or transition state(s). The two proposed mechanisms for how methane is generated differ in whether the first step involves an organometallic methyl-Ni(III) [mechanism I (*13*–*15*)] or a methyl radical intermediate [mechanism II (*16*)] (Fig. 1). In both mechanisms, the nickel center of F_{430} must be in the Ni(I) oxidation state for the enzyme to initiate catalysis (*17*, *18*).

Support for mechanism I is based on experiments using F_{430} model complexes (*19*, *20*), enzymatic studies involving isotope exchange (*21*), and the reaction of the active form of MCR (MCR_{red1}) with various activated alkyl donors such as alkyl halides (*22*–*24*). These substrate analogs react with Ni(I) to generate alkyl-Ni(III) species that undergo reduction to the alkane (as in the forward direction of Eq. 1) or conversion to thioethers—e.g., methyl-SCoM upon reaction with organic thiolates like CoM (as in the reverse reaction) (*22*–*24*). Mechanism II is supported by density functional theory (DFT) computations in which it was argued that formation of the methyl-Ni(III) intermediate is not energetically feasible, being endoergic by 91 kJ/mol (with an activation free energy of 94 kJ/mol), whereas the formation of a methyl radical and Ni(II)-thiolate is exoergic by 10 kJ/mol (with an activation free energy of 63 kJ/mol) (*16*, *25*–*27*).

A third mechanism is also possible in which nucleophilic attack of Ni(I) on methyl-SCoM generates a Ni(III)-SCoM species and, formally, an anionic methyl group that undergoes protonation to generate methane (mechanism III) (Fig. 1). A Ni(III)-thiolate species known as MCR_{ox1} is otherwise formed when growing cells are exposed to sodium sulfide (*18*) or to an oxidizing gas mixture (80% N_2/20% CO_2) (*28*). MCR_{ox1} is also called the "ready" state of the enzyme because it can be activated to the active MCR_{red1} state (*17*, *18*). Both the methyl-Ni(III) (*23*, *29*) and the Ni(III)-thiolate (MCR_{ox1}) (*30*) states have been generated in high yield, are relatively stable, and exhibit distinctive electron paramagnetic resonance (EPR) spectra. Actually, spectroscopic and computational studies indicate that MCR_{ox1} is best described as a high-spin Ni(II)-thiyl radical in resonance with a Ni(III)-thiolate species (*30*, *31*). Conversely, the Ni(II)-$MCR_{ox1\text{-}silent}$ state is EPR-silent. The MCR_{ox1}, $MCR_{ox1\text{-}silent}$, and MCR_{red1} states also display distinct magnetic circular dichroism (MCD) spectra (*31*, *32*). Thus, performing rapid mixing experiments and monitoring the accumulation of an intermediate exhibiting the spectroscopic features of the methyl-Ni(III), $MCR_{ox1\text{-}silent}$, or MCR_{ox1} states associated with decay of MCR_{red1} should provide unambiguous evidence supporting one of the three mechanisms. However, only minor spectroscopic changes are observed when MCR reacts with methyl-SCoM and the natural substrate CoBSH (*33*).

We performed transient kinetic, spectroscopic [ultraviolet-visible (UV-Vis), EPR, and MCD], and computational studies of the first step in the MCR catalytic mechanism to trap and identify the key intermediates that differ between mechanisms I and II. MCR contained a sufficiently high amount (70 to 80%) of the active Ni(I)-MCR_{red1} state to monitor changes in its spectroscopic properties during the reaction and identify intermediates. We rapidly mixed MCR with methyl-SCoM and CoB_6SH, containing a hexanoyl instead of heptanoyl side chain, which sufficiently slows down the first step in the MCR reaction (*34*, *35*) to allow accumulation and detection of the first intermediate in the MCR mechanism.

Rapid kinetic studies rule out methyl-Ni(III) and trap the $MCR_{ox1\text{-}silent}$ intermediate

We performed stopped-flow studies by rapidly mixing a solution containing MCR and methyl-SCoM with the slow substrate CoB_6SH (Fig. 2A). We tracked the reaction at 385 nm to follow Ni(I) decay and at 420 nm to measure the rate at which the Ni(II) or Ni(III) intermediate forms. Although the steady-state and presteady-state rate constants are slower by factors of 1000 and 440 with CoB_6SH than with CoBSH, no spectroscopic changes are observed upon addition of methyl-SCoM alone; in fact, even for a single turnover, both substrates must be present before any reaction can occur (*34*). This strongly suggests that with the slow (CoB_6SH) substrate, MCR employs the same strict ternary-complex mechanism as with the native (CoBSH) substrate (*33*, *34*). The spectroscopic features at both 385 and 420 nm

[1]Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109-0606, USA. [2]Physical Sciences Division, Pacific Northwest National Laboratory, Post Office Box 999, K1-83, Richland, WA 99352, USA. [3]Department of Chemistry and Department of Biophysics, University of Michigan, Ann Arbor, MI 48109-1055, USA.
*Present address: Applied Photophysics Inc., 100 Cummings Center, Suite 440C, Beverly, MA 01915, USA. †Present address: Intel Corporation, 2111 NW 25th Avenue, JF5-202, Hillsboro, OR 97124, USA. ‡Corresponding author. Email: sragsdal@umich.edu

Downloaded from http://science.sciencemag.org/ on May 19, 2016

exhibited monophasic kinetics, with 60% of the starting MCR_{red1} state undergoing conversion at a rate constant of 0.35 ± 0.01 s^{-1}. Two additional slow phases [observed rate constant 2 (k_{obs2}) = 0.05 ± 0.01 s^{-1} and k_{obs3} = 0.008 ± 0.001 s^{-1}] with greatly reduced amplitudes are observed that also occur in control reactions lacking substrate, indicating that these phases correspond to the noncatalytic Ni(I) oxidation and are not relevant to the catalytic mechanism. Over a longer time frame, the spectrum of active MCR_{red1} returns (k_{obs} = 0.011 ± 0.001 min^{-1}), validating that MCR remains active during these spectroscopic transformations with methyl-SCoM and CoB_6SH, as recently shown for the reaction with CoBSH (*33*).

To further study the conversion of methyl-SCoM to methane, we used the rapid chemical-quench method under conditions similar to those of the stopped-flow experiments. A solution containing MCR_{red1} (20 μM, after mixing) and [^{14}C]methyl-SCoM (20 μM, after mixing) was rapidly mixed with CoB_6SH (500 μM, after mixing), and incubated for various time points between 0.62 and 90 s. The reaction was quenched by mixing with 0.2 M perchloric acid and analyzed by liquid scintillation counting. The amount of remaining [^{14}C]methyl-SCoM was plotted versus time (Fig. 2B) and fit to a single-exponential curve, revealing a limiting rate constant of 0.31 ± 0.04 s^{-1}. The results demonstrated that conversion of the methyl group of methyl-SCoM to methane is limited by the same rate constant (~0.30 s^{-1}) as conversion of Ni(I) to Ni(II)/Ni(III) in the stopped-flow experiment. These results support mechanism II (or III), because the reactive methyl radical (and methyl anion) would have very transient existence and would immediately abstract a hydrogen atom or proton, respectively, from CoBSH to generate methane with the same rate constant as that of Ni(I) decay. However, because methyl-Ni(III) is relatively stable, methane formation by mechanism I requires another step and, thus, would occur more slowly than Ni(I) decay. For example, the alkyl-Ni(III) state of MCR reacts slowly with thiolates. The methyl-Ni(III) state of MCR reacts with CoMSH to generate methyl-SCoM and MCR_{red1} at a rate constant of 0.04 s^{-1} (at 25°C) (*23*). Similarly, the alkyl-Ni(III) state generated from bromopropanesulfonate or various brominated acids react slowly (k_{max} ~ 0.005 s^{-1}) with small thiolates and even slower with CoBSH (and analogs) to generate MCR_{red1} and the thioether (*22*, *24*, *36*).

To identify and characterize any EPR-detectable intermediates formed during the MCR reaction, we performed rapid freeze-quench (RFQ) EPR experiments under similar conditions as those for the stopped-flow and rapid chemical-quench experiments. We observed a dominant (~90%) decrease in intensity of the characteristic Ni(I) EPR spectrum of MCR_{red1} at g values of 2.249, 2.084, and 2.061 (Fig. 2C). Comparable results are observed in two similar FQ-EPR experiments (fig. S1). The decay curve fits to a single-exponential equation with a limiting rate constant of 0.53 ± 0.25 s^{-1} (circles in Fig. 2D). We did not observe any EPR-active species [e.g., methyl-Ni(III)] that accumulate to an amplitude similar to that of MCR_{red1} decay (~43 μM), suggesting that the Ni-based product of this reaction is a Ni(II) EPR-silent species. However, two other EPR signals formed at low levels during the time course of Ni(I) decay (insets, Fig. 2C and fig. S1). A rhombic signal (g = 2.212, 2.183, and 2.150) identical to that of MCR_{ox1} (*30*), present at very low levels in the initial sample, slightly increased in intensity to 3% of the initial MCR_{red1} according to a rate constant of 0.69 ± 0.24 s^{-1}. A radical-type species (g ~ 2.0), observed earlier (*35*), formed with a rate constant of 0.52 ± 0.32 s^{-1} to an amplitude that reached 6 to 7% of the initial Ni(I)-MCR_{red1} signal (fig. S1). The low level of accumulation of

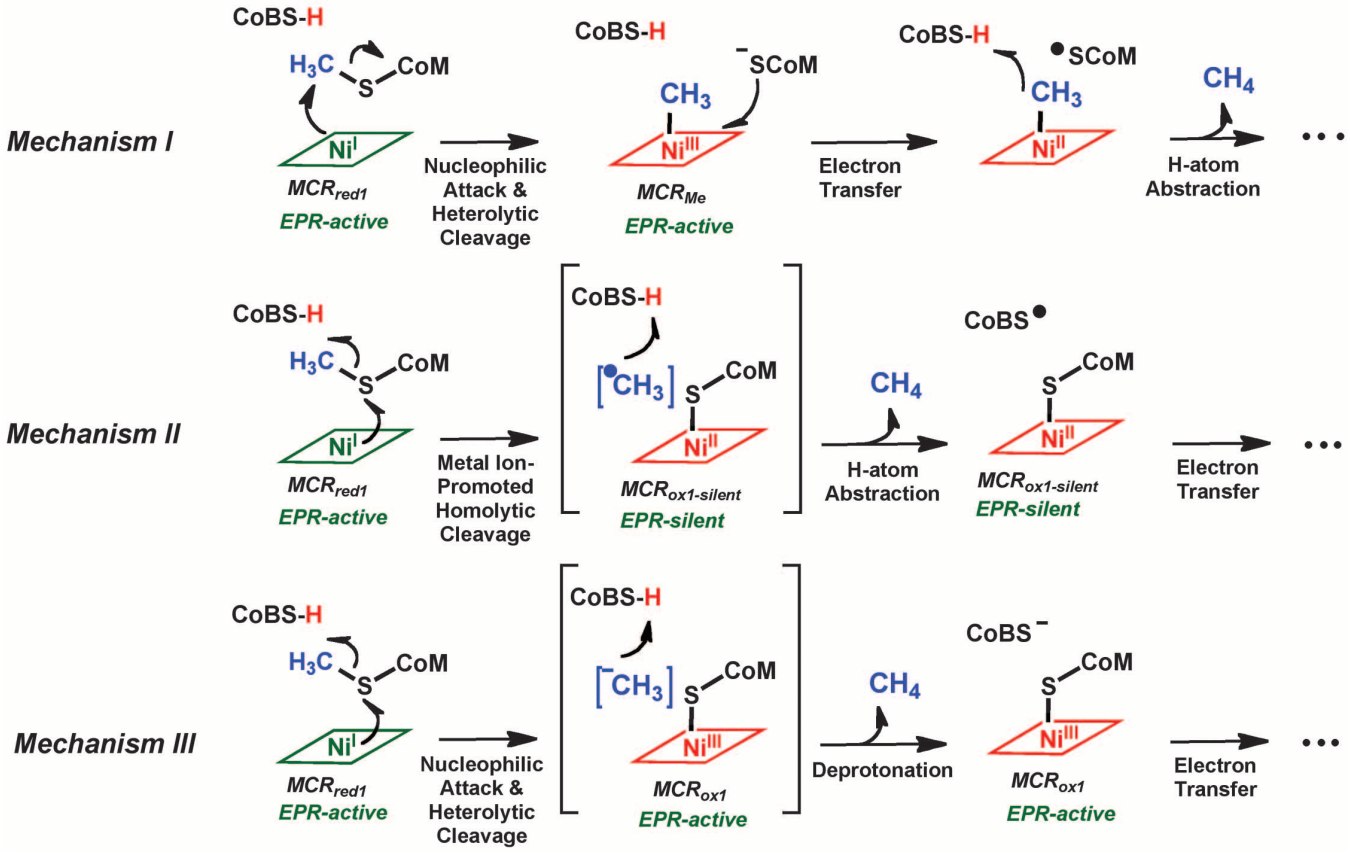


Fig. 1. Initial steps in three mechanisms of MCR catalysis. Mechanism I involves nucleophilic attack of Ni(I)-MCR_{red1} on the methyl group of methyl-SCoM to generate a methyl-Ni(III) intermediate (*34*). This mechanism is similar to that of B_{12}-dependent methyltransferases (*48*), which generate a methyl-cob(III)alamin intermediate. In mechanism II, Ni(I) attack on the sulfur atom of methyl-SCoM promotes the homolytic cleavage of the methyl-sulfur bond to produce a methyl radical (•CH_3) and a Ni(II)-thiolate. Mechanism III involves nucleophilic attack of Ni(I) on the sulfur of methyl-SCoM to form a highly reactive methyl anion and Ni(III)-SCoM (MCR_{ox1}).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

this radical species may be due to the difficulty of observing thiyl radicals due to their short lifetimes and large spin-orbit coupling with the sulfur atom (*37*). Regardless, a sulfur radical(s) is predicted to be an intermediate in all three mechanisms and, thus, is not diagnostic of which one is correct.

The above results suggest that, when active MCR_{red1} is incubated with methyl-SCoM and CoB_6SH, it converts nearly quantitatively to an EPR-silent Ni(II)-thiolate species, consistent with the predictions of mechanism II. Therefore, we turned our attention to a spectroscopic method that could positively identify that intermediate. The MCD spectra of MCR_{red1}, MCR_{ox1}, and $MCR_{ox1\text{-}silent}$ are distinct (*31*). Figure S2 shows the MCD data of MCR_{red1} and $MCR_{ox1\text{-}silent}$, prepared for comparison. The temperature dependence of the data showed that both species are paramagnetic, with MCR_{red1} and $MCR_{ox1\text{-}silent}$ having ground states of $S = ½$ and $S = 1$, respectively. MCR_{red1} has a characteristic negative feature (at 21,620 cm^{-1}) that can be used to monitor the reaction and decay of this species. The kinetic studies indicated that the first intermediate in methanogenesis forms at a rate constant of 0.35 s^{-1} ($t_{1/2}$ = 2 s) and remains for at least 70 s; thus, to directly observe this intermediate by MCD, we mixed a solution containing MCR_{red1} and methyl-SCoM with CoB_6SH, rapidly froze this mixture in liquid nitrogen within ~10 s, and performed MCD experiments of the samples prepared before and after reaction with substrates. The dominant changes in the MCD spectrum indicate an almost quantitative conversion of MCR_{red1} to a species nearly identical to $MCR_{ox1\text{-}silent}$ (Fig. 3). For example, the negative band at 21,620 cm^{-1} disappeared, direct evidence of the consumption of MCR_{red1}, as all of the characteristic bands associated with $MCR_{ox1\text{-}silent}$ appear, positively identifying that Ni(II) species as the reaction product. Although there is some difference in absolute intensity between the data sets (which is likely due to problems with depolarization of the circular polarized light by the frozen glass MCD sample), the spectra are almost identical in band shape—i.e., the number of features and their relative intensities—confirming formation of $MCR_{ox1\text{-}silent}$ as the major reaction product. The CD spectrometer also records a single-channel voltage curve that can be converted into a low-resolution UV-Vis absorption spectrum. These data demonstrate that the MCD samples were initially in the Ni(I) state and underwent quantitative conversion after reaction (fig. S2D). Altogether, our spectroscopic results provide direct evidence that the EPR-silent Ni(II)-thiolate intermediate proposed in mechanism II is the key intermediate in methanogenesis.

Mechanisms I, II, and III propose the formation of distinct intermediates. We found no evidence for a methyl-Ni(III) species in our RFQ EPR studies. This [and other alkyl-Ni(III)] species is relatively stable when bound to MCR (*23*, *38*) and should have been observed if it had formed. Moreover, there was no evidence for the MCR_{red2a} or MCR_{red2r} states (*g* values of 2.273, 2.077, 2.073 and 2.288, 2.231, 2.175, respectively) (*39*) during the transient kinetic reaction, strongly indicating that neither of these species, assigned as a Ni(III)-hydride or a Ni(I)-thiolate, respectively, serve as an intermediate in the MCR mechanism. A recent computational study suggested that the Ni(III)-H should be reassigned to a catalytically inactive species in which the proton of CoMSH binds near the Ni(I) of MCR_{red1} (*27*). Regardless, the data do not support a mechanism involving a methyl-Ni(III), Ni(III)-hydride or side-on C-S coordination to the Ni (*39*).

We further explored whether the minor MCR_{ox1} signal could have any catalytic relevance for methane formation, perhaps as a parallel pathway via mechanism III (fig. S3). If MCR_{ox1} forms with a rate constant of 0.5 s^{-1} and accumulates to 3% of the initial amount of MCR_{red1}, it must decay to another Ni-based intermediate with a rate constant (k_2) of ~ 15 s^{-1} (fig. S4). However, the concentration of MCR_{ox1} did not change for at least 70 s, strongly suggesting that the slight increase in the MCR_{ox1} EPR signal occurred by an off-pathway process that is unrelated to catalysis.

Computational studies rule out the Ni(III)-thiolate MCR_{ox1}-like intermediate

DFT calculations were undertaken to assess the relative energetics of the initial steps in mechanisms I, II and III involving methyl-Ni(III), Ni(II)-$MCR_{ox1\text{-}silent}$, or the Ni(III)-MCR_{ox1} species, respectively (fig. S5). We used truncated forms of the F_{430} cofactor, CoBSH, and CoM and performed these computations as previously described (*25*–*27*, *40*). The formation of a $MCR_{ox1\text{-}silent}$-like intermediate

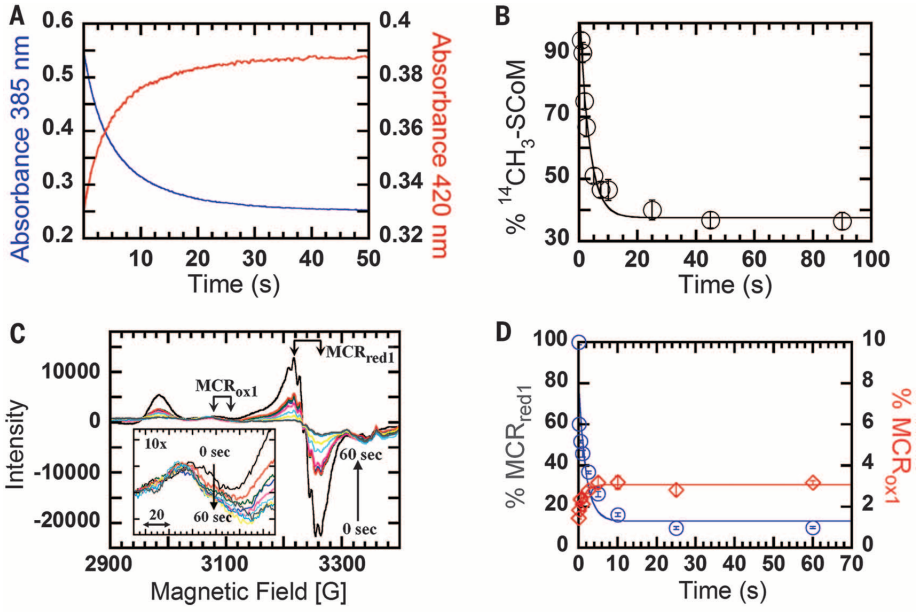


Fig. 2. Rapid kinetic studies of the reaction of the MCR:methyl-SCoM complex with CoB_6SH. (**A**) Stopped-flow. Kinetic traces of the reaction of a premixed solution containing MCR_{red1} (20 μM, after mixing) and methyl-SCoM (20 μM, after mixing) with CoB_6SH (500 μM, after mixing) in 50 mM Tris-HCl, pH 7.6. The reactions were performed under anaerobic conditions using the stopped-flow spectrophotometer at 18°C and monitored by following the decay of Ni(I) at 385 nm (blue line) and the formation of Ni(II)/Ni(III) at 420 nm (red line). The reaction showed monophasic kinetics with a rate constant of 0.35 ± 0.01 s^{-1}. (**B**) Rapid chemical-quench. Reactions of a premixed solution containing equimolar MCR_{red1} (20 μM, after mixing) and [^{14}C] methyl-SCoM (20 μM, after mixing) with CoB_6SH (500 μM, after mixing) were quenched with 0.2 M perchloric acid at various times using the rapid chemical-quench apparatus. Volatile methane product was lost from the solution, and the percentage conversion of [^{14}C]methyl-SCoM was determined by comparing the remaining concentration of [^{14}C] methyl-SCoM to the initial concentration. Plotting the percentage conversion versus time yielded a single-exponential curve with a rate constant of 0.31 ± 0.04 s^{-1}. The vertical brackets at each point indicate the standard deviation of the measurement. (**C** and **D**) RFQ EPR. A solution containing MCR_{red1} (48 μM) and methyl-SCoM (600 μM) was reacted with CoB_6SH (120 μM) and freeze-quenched at various times using an RFQ apparatus. Representative time-dependent EPR spectra are shown on (C). The inset shows the *g* ~ 2.2 region near the S-shaped feature of MCR_{ox1}. The percentage decay of MCR_{red1} (blue) and formation of MCR_{ox1} (red) were determined by comparing their doubly integrated signal intensities at various quenching times to their initial intensities. The data were plotted and fit to single-exponential curves in (D). The MCR_{red1} signal decayed by 90% during the first phase of the reaction with a rate constant of 0.53 ± 0.25 s^{-1}, whereas the MCR_{ox1}-like signal increased by ~3% (relative to the initial MCR_{red1}), with a rate constant of 0.69 ± 0.24 s^{-1}. A radical formed with a rate constant of 0.52 ± 0.32 s^{-1} and reached ~7% of the initial MCR_{red1} (see fig. S1). A vertical line at each point indicates the standard deviation of each measurement.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

via the homolytic process corresponding to mechanism II is thermodynamically favored (ΔG^0 = −3.5 kJ/mol) and has an activation energy ($\Delta G^{\ddagger}$) of 72 kJ/mol, whereas generation of the methyl-Ni(III) state (by mechanism I) is highly endoergic (ΔG^0 = 100 kJ/mol) with a $\Delta G^{\ddagger}$ of 102 kJ/mol.

We were unable to identify an MCR_{ox1}-like state, specifically a F_{430}-Ni(III)-SCoM/$CoBS^-$ intermediate, from direct DFT calculations; however, performing a wave function optimization (*41*) imposing a −1 *e* charge on the CoBS fragment, we found this MCR_{ox1}-like state residing at 136.0 kJ/mol above the $MCR_{ox1\text{-}silent}$ baseline. A similar computational strategy applied to identify an anionic transition state connecting MCR_{red1} to MCR_{ox1} failed. Nevertheless, starting from the homolytic transition state of mechanism II, we identified an excited state with a strong anionic CH_3^- character, residing at 134.6 kJ/mol above the MCR_{red1} baseline. The inability to optimize any transition state corresponding to MCR_{ox1} and the corresponding methyl anion could be due to the reduced size of the model; however, the large estimated energetic difference between the MCR_{ox1} and $MCR_{ox1\text{-}silent}$ states is unlikely to be substantially affected by increasing the model size.

Additional calculations suggested that the MCR_{ox1}-like state is inaccessible because of the very positive reduction potential of the Ni(III)-SCoM intermediate species, E_0' = 1.4 V versus the normal hydrogen electrode (NHE) (see the supplementary materials). Indeed, with a calculated redox potential for the CoBS•/$CoBS^-$ couple in the enzyme cavity of E_0' = 0.0 V versus NHE, Ni(III)-SCoM would promptly oxidize $CoBS^-$ during turnover. Taken together, these computational findings, along with the kinetic simulations described above, rule out the possibility of a MCR_{ox1}-like intermediate (and of mechanism III) in methane synthesis.

Temperature-dependence studies map the MCR reaction profile

One way to discriminate among different mechanisms is to compare experimental activation energy barriers with the transition-state barriers obtained by DFT computations. To obtain the experimental thermodynamic values, a premixed solution containing equimolar MCR_{red1} and methyl-SCoM was rapidly mixed with a saturating concentration of CoB_7SH or CoB_6SH at various temperatures between 10° and 50°C, and the reaction was monitored by stopped-flow spectrophotometry (Fig. 4 and fig. S6).

For the reactions with both CoB_7SH and CoB_6SH, k_{obs} increased as the temperature increased (Fig. 4A), giving a biphasic Arrhenius plot with a transition temperature at ~30°C (Fig. 4B). For CoB_7SH, the activation energies (E_a) above and below the transition temperature were calculated to be 51 and 84 kJ/mol, respectively (Table 1), indicating that there is a temperature-dependent structural transition that alters the MCR mechanism. The Eyring plot (Fig. 4C) also showed a similar nonlinear response with a transition temperature at the same temperature. The enthalpy of activation ($\Delta H^{\ddagger}$) and the entropy of activation ($\Delta S^{\ddagger}$) above the transition temperature are 48 kJ/mol and −56 J/(mol K), respectively (Table 1). Between 30° and 50°C, the Gibbs energy of activation ($\Delta G^{\ddagger}$) increased due to a more negative entropy of activation ($\Delta S^{\ddagger}$). The $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ values below the transition temperature were 82 kJ/mol and 54 J/(mol K), respectively.

The experimental values of E_a (51 kJ/mol), $\Delta H^{\ddagger}$ (48 kJ/mol), and calculated $\Delta G^{\ddagger}$ at 30°C (65 kJ/mol) were congruent with previous (*25*) and current DFT-calculated transition-state barriers for the formation of a Ni(II)-thiolate (about 63 kJ/mol and 71 kJ/mol, respectively) via mechanism II, given that only the enzyme's active site and truncated substrates were considered in the DFT model. In contrast, the DFT-calculated transition barrier ($\Delta G^{\ddagger}$) for formation of the methyl-Ni(III) intermediate was 102 kJ/mol, which is much higher than the experimental value. In addition to the high kinetic barrier, formation of methyl-Ni(III) from MCR_{red1} is computed to be highly unfavorable thermodynamically (ΔG^0 = 100 kJ/mol). Thus, our temperature-dependent studies strongly support the formation of the Ni(II)-$MCR_{ox1\text{-}silent}$-like species as the key intermediate in the MCR reaction, consistent with our spectroscopic data.

Similar temperature-dependence studies measured the thermodynamic values for the reaction with CoB_6SH. As with CoB_7SH, the Arrhenius and Eyring plots were biphasic with a transition temperature at ~29°C. The E_a values above and below the transition temperature were 47 and 127 kJ/mol, respectively (Fig. 4B). The Eyring analysis resulted in $\Delta H^{\ddagger}$ of 45 and 125 kJ/mol and $\Delta S^{\ddagger}$ of −90 and 176 J/mol.K, above and below the transition temperature, respectively (Fig. 4C).

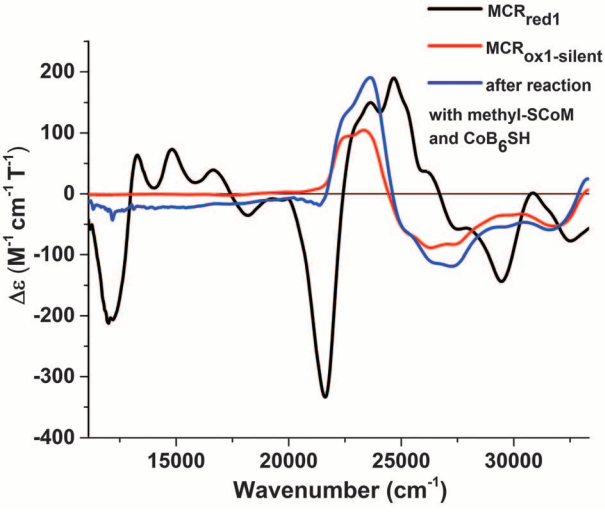


Fig. 3. MCD studies of the reaction of the MCR:methyl-SCoM complex with CoB_6SH. MCD spectra were taken at 2 K of MCR_{red1} before and after the reaction with methyl-SCoM and CoB_6SH. Samples were prepared in 50 mM GPT buffer [(50 mM glycine, 50 mM phosphate, and 50 mM Tris), pH 7.6, containing 0.05 mM Ti(III) citrate] with 50% glycerol for MCR_{red1} and 73% glycerol for the reaction with CoB_6SH.

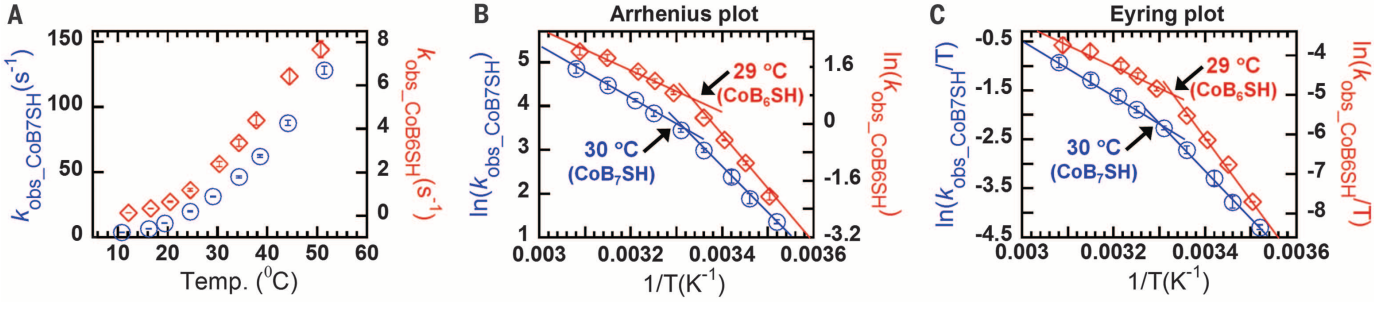


Fig. 4. Effect of temperature on the presteady-state reaction of the MCR:methyl-SCoM complex with CoB_6SH and CoB_7SH. Reactions of a premixed solution containing equimolar MCR_{red1} (20 μM, after mixing) and methyl-SCoM (20 μM, after mixing) with the saturating concentration of CoB_7SH (1 mM, after mixing) (circle blue) or CoB_6SH (500 μM, after mixing) (diamond red) at various temperature (10° to 50°C) were investigated using stopped-flow spectrophotometry (**A**). (**B**) is the Arrhenius plot and (**C**) is the Eyring plot for formation of a Ni(II)-thiolate showing a transition temperature (arrow). A vertical line at each point indicates a standard deviation of the measurement. The thermodynamic values (E_a, $\Delta H^{\ddagger}$, and $\Delta S^{\ddagger}$) and k_{obs} from rapid kinetics are summarized in Table 1.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Table 1. Activation barriers and observed rate constants for the first step in the MCR reaction. Activation barriers and rate constants were determined by presteady-state kinetic experiments (Fig. 4). These experiments were performed at the transition temperature, 30°C.

Substrate	Temp. range	E_a* (kJ/mol)	$\Delta H^{\dagger}$ (kJ/mol)	$\Delta S^{\dagger}$ (J/mol.K)	k_{obs} (s^{-1}), 30°C§
CoB_7SH	10–30°C	84	82	54	31
	30–50°C	51	48	−56	
CoB_6SH	10–29°C	127	125	176	2
	29–50°C	47	45	−90	

*The activation energy was calculated from the Arrhenius plot. †Enthalpy and entropy of the transition state were determined from the Eyring plot. §k_{obs} values at the temperature break.

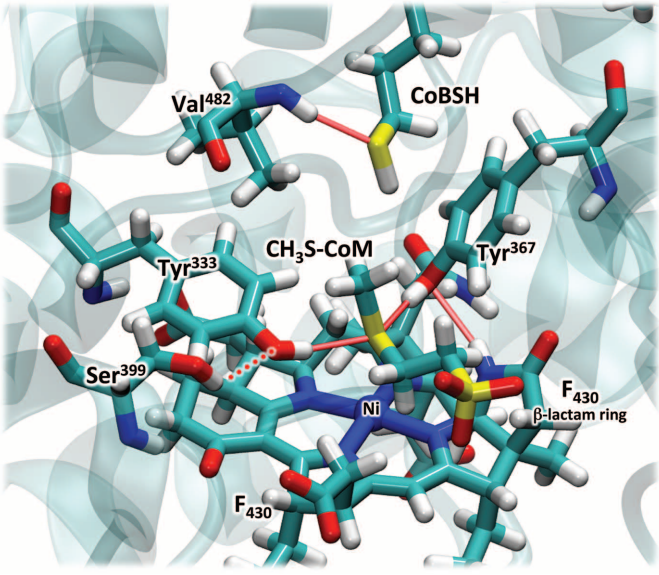


Fig. 5. Hydrogen bonding interactions among MCR active site residues. Red sticks indicate hydrogen bonds at 25°C. The dashed line indicates the weak hydrogen bond between Ser^{399} and Tyr^{333} above 30°C. Residues are numbered according to MCR from *Methanothermobacter marburgensis*. See also table S1.

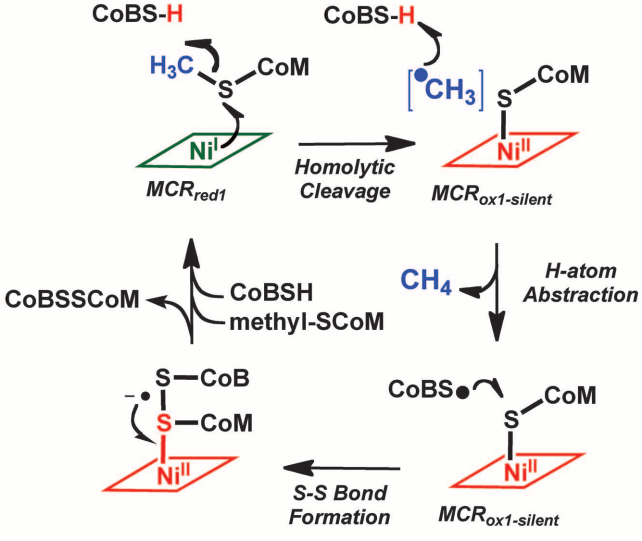


Fig. 6. Proposed steps of mechanism II. In the first step, Ni(I) attack on the sulfur of methyl-SCoM leads to homolytic cleavage of the C-S bond and generation of a methyl radical and a Ni(II)-thiolate ($MCR_{ox1\text{-}silent}$). Next, H-atom abstraction from CoBSH generates methane and the CoBS• radical, which in the third step combines with the Ni-bound thiolate of CoM to generate the Ni(II)-disulfide anion radical. Then, one-electron transfer to Ni(II) generates MCR_{red1} and the heterodisulfide (CoBSSCoM) product, which dissociates leading to ordered binding of methyl-SCoM and CoBSH and initiation of the next catalytic cycle.

Between 30° and 50°C, the E_a (47 kJ/mol) and $\Delta H^{\ddagger}$ (45 kJ/mol) for the CoB_6SH reaction are in the same range as those for CoB_7SH (E_a = 51 kJ/mol and $\Delta H^{\ddagger}$ = 48 kJ/mol). These results indicate that the same type of temperature-dependent transition occurs in the MCR reaction with CoB_6SH as with the natural CoB_7SH substrate. The $\Delta S^{\ddagger}$ value between 30° and 50°C for the reaction with CoB_6SH was 1.6 times lower than that for CoB_7SH, suggesting that the slow substrate has a higher degree of freedom in the MCR active site than the longer natural substrate. As expected, the $\Delta G^{\ddagger}$ at 30°C for the reaction of CoB_6SH is 7 kJ/mol higher than that for CoB_7SH (72 kJ/mol versus 65 kJ/mol), in agreement with the k_{obs} at 30°C for CoB_6SH being lower by a factor of 15 than that with CoB_7SH (2 s^{-1} versus 31 s^{-1}). These results strongly indicate that MCR employs the same rate-limiting chemical step in its reaction mechanism, whether the second substrate is CoB_6SH or CoB_7SH. Furthermore, it appears that the major reason for the factor of 1000 slower rate with CoB_6SH is the increased entropy of the slowly reacting substrate, CoB_6SH, at the active site and in the substrate channel relative to that of CoB_7SH.

Molecular dynamics simulations highlight the presence of temperature-dependent structural changes at the active site

Long time-scale molecular dynamics simulations over a wide range of temperatures (from room temperature to 50°C) were performed to identify residues at the active site that could be responsible for the observed biphasic temperature dependence of the kinetics (Table 1 and Fig. 4). The simulations suggested that above 25°C, Tyr^{333}, which is one of the two tyrosine residues that is hydrogen-bonded to the thioether sulfur of methyl-CoM, moves away from CoM, establishing a weak hydrogen bond with Ser^{399} and allowing methyl-SCoM to approach the Ni center of F_{430} (Fig. 5). This conformational change lengthens the Tyr-OH-(S)CoM hydrogen bond distance from 2.5 ± 0.4 Å (25°C) to 3.1 ± 0.5 Å at around 30°C as the CoM(S)-Ni average distance decreases from 3.7 ± 0.2 Å to 3.2 ± 0.3 Å (table S1 and Fig. 5). Shortening the CoM(S)-Ni distance is expected to facilitate cleavage of the CoM(S)-CH_3 bond. None of the other hydrogen bond interactions at the active site changed as the temperature increased (table S1 and fig. S7).

Implications

In the first step of mechanism II (Fig. 6), attack of Ni(I) on the sulfur of methyl-SCoM leads to homolytic cleavage of the C-S bond and generation of a methyl radical and a Ni(II)-thiolate, known as $MCR_{ox1\text{-}silent}$. Carbon (1.04) and secondary deuterium (1.19) isotope effect studies (*42*, *43*) indicate that the transition state for the rate-limiting C-S bond cleavage involves a trigonal planar carbon, consistent with formation of a methyl radical in the first step in methane synthesis. Because anaerobic methane oxidation occurs by direct reversal of its synthesis (*3*), the methyl radical will also be formed in the transition

Downloaded from http://science.sciencemag.org/ on May 19, 2016

state for the final step in reverse methanogenesis. The second step involves H-atom abstraction from CoBSH, generating methane and the CoBS• radical, which in the third step combines with the Ni-bound thiolate of CoM to generate a Ni(II)-disulfide anion radical, poised for one-electron transfer to Ni(II) to generate Ni(I)-MCR_{red1} and the heterodisulfide (CoBSSCoM) product. The final mechanistic step involves dissociation of the heterodisulfide to allow the ordered binding of methyl-SCoM and CoBSH and initiate the next catalytic cycle.

Understanding the mechanism of these reactions has large implications for developing technologies to catalytically generate and activate methane (*44*, *45*). The latter process is one of the most challenging chemical reactions because a catalyst must selectively break the first carbon-hydrogen bond of methane at a huge free-energy cost (438.9 kJ/mol) without cleaving any of the remaining weaker carbon-hydrogen bonds. Exploitation of methane for energy and chemistry is timely because natural gas reserves are predicted to increase by >40% in the United States over the next 30 years (*46*) and because anthropogenic sources of methane contribute a large proportion (20%) of the world's annual greenhouse gas warming potential (*48*).

REFERENCES AND NOTES

1. S. K. Atreya, P. R. Mahaffy, A. Wong, *Planet. Space Sci.* **55**, 358–369 (2007).
2. R. K. Thauer, *Microbiology* **144**, 2377–2406 (1998).
3. S. J. Hallam *et al.*, *Science* **305**, 1457–1462 (2004).
4. M. Krüger *et al.*, *Nature* **426**, 878–881 (2003).
5. K. Nauhaus, A. Boetius, M. Krüger, F. Widdel, *Environ. Microbiol.* **4**, 296–305 (2002).
6. W. Michaelis *et al.*, *Science* **297**, 1013–1015 (2002).
7. S. Shima *et al.*, *Nature* **481**, 98–101 (2011).
8. S. Scheller, M. Goenrich, R. Boecher, R. K. Thauer, B. Jaun, *Nature* **465**, 606–608 (2010).
9. S. Shima, R. K. Thauer, *Curr. Opin. Microbiol.* **8**, 643–648 (2005).
10. J. Ellermann, A. Kobelt, A. Pfaltz, R. K. Thauer, *FEBS Lett.* **220**, 358–362 (1987).
11. A. A. DiMarco, T. A. Bobik, R. S. Wolfe, *Annu. Rev. Biochem.* **59**, 355–394 (1990).
12. U. Ermler, W. Grabarse, S. Shima, M. Goubeaud, R. K. Thauer, *Science* **278**, 1457–1462 (1997).
13. L. Signor *et al.*, *Chemistry* **6**, 3508–3516 (2000).
14. W. Grabarse *et al.*, *J. Mol. Biol.* **309**, 315–330 (2001).
15. E. C. Duin, M. L. McKee, *J. Phys. Chem. B* **112**, 2466–2482 (2008).
16. V. Pelmenschikov, M. R. A. Blomberg, P. E. M. Siegbahn, R. H. Crabtree, *J. Am. Chem. Soc.* **124**, 4039–4049 (2002).
17. M. Goubeaud, G. Schreiner, R. K. Thauer, *Eur. J. Biochem.* **243**, 110–114 (1997).
18. D. F. Becker, S. W. Ragsdale, *Biochemistry* **37**, 2639–2647 (1998).
19. S.-K. Lin, B. Jaun, *Helv. Chim. Acta* **74**, 1725–1738 (1991).
20. G. K. Lahiri, A. M. Stolzenberg, *Inorg. Chem.* **32**, 4409–4413 (1993).
21. S. Scheller, M. Goenrich, S. Mayr, R. K. Thauer, B. Jaun, *Angew. Chem. Int. Ed. Engl.* **49**, 8112–8115 (2010).
22. M. Dey, R. C. Kunz, D. M. Lyons, S. W. Ragsdale, *Biochemistry* **46**, 11969–11978 (2007).
23. M. Dey *et al.*, *J. Am. Chem. Soc.* **129**, 11030–11032 (2007).
24. R. C. Kunz, M. Dey, S. W. Ragsdale, *Biochemistry* **47**, 2661–2667 (2008).
25. S.-L. Chen, M. R. A. Blomberg, P. E. M. Siegbahn, *Chemistry* **18**, 6309–6315 (2012).
26. S.-L. Chen, V. Pelmenschikov, M. R. A. Blomberg, P. E. M. Siegbahn, *J. Am. Chem. Soc.* **131**, 9912–9913 (2009).
27. S.-L. Chen, M. R. A. Blomberg, P. E. M. Siegbahn, *Phys. Chem. Chem. Phys.* **16**, 14029–14035 (2014).
28. S. P. J. Albracht *et al.*, *Biochim. Biophys. Acta* **955**, 86–102 (1988).
29. D. Hinderberger *et al.*, *Angew. Chem. Int. Ed. Engl.* **45**, 3602–3607 (2006).
30. J. Harmer *et al.*, *J. Am. Chem. Soc.* **127**, 17744–17755 (2005).
31. J. L. Craft, Y.-C. Horng, S. W. Ragsdale, T. C. Brunold, *J. Am. Chem. Soc.* **126**, 4068–4069 (2004).
32. E. C. Duin *et al.*, *J. Biol. Inorg. Chem.* **9**, 563–576 (2004).
33. T. Wongnate, S. W. Ragsdale, *J. Biol. Chem.* **290**, 9322–9334 (2015).
34. Y.-C. Horng, D. F. Becker, S. W. Ragsdale, *Biochemistry* **40**, 12875–12885 (2001).
35. M. Dey, X. Li, R. C. Kunz, S. W. Ragsdale, *Biochemistry* **49**, 10902–10911 (2010).
36. R. C. Kunz, Y.-C. Horng, S. W. Ragsdale, *J. Biol. Chem.* **281**, 34663–34676 (2006).
37. M. C. R. Symons, *J. Chem. Soc., Perkin Trans.* **2**, 1618–1620 (1974).
38. N. Yang, M. Reiher, M. Wang, J. Harmer, E. C. Duin, *J. Am. Chem. Soc.* **129**, 11028–11029 (2007).
39. J. Harmer *et al.*, *J. Am. Chem. Soc.* **130**, 10907–10920 (2008).
40. V. Pelmenschikov, P. E. M. Siegbahn, *J. Biol. Inorg. Chem.* **8**, 653–662 (2003).
41. P. H. Dederichs, S. Blügel, R. Zeller, H. Akai, *Phys. Rev. Lett.* **53**, 2512–2515 (1984).
42. S. Scheller, M. Goenrich, R. K. Thauer, B. Jaun, *J. Am. Chem. Soc.* **135**, 14975–14984 (2013).
43. S. Scheller, M. Goenrich, R. K. Thauer, B. Jaun, *J. Am. Chem. Soc.* **135**, 14985–14995 (2013).
44. C. A. Haynes, R. Gonzalez, *Nat. Chem. Biol.* **10**, 331–339 (2014).
45. T. J. Mueller *et al.*, *J. Ind. Microbiol. Biotechnol.* **42**, 391–401 (2015).
46. U. S. E. I. Administration, "Annual Energy Outlook 2015 With Projections to 2040" (Office of Integrated and International Energy Analysis, U.S. Department of Energy, Washington, DC, 2015).
47. "U.S. Environmental Protection Agency. Summary Report: Global Anthropogenic Non-CO2 Greenhouse Gases Emissions: 1990–2030" (U.S. Environmental Protection Agency, Washington, DC, 2012).
48. R. Banerjee, S. W. Ragsdale, *Annu. Rev. Biochem.* **72**, 209–247 (2003).

ACKNOWLEDGMENTS

Data presented here are available through the Ragsdale LabGuru portal (https://my.labguru.com/knowledge/projects/361/milestones/711). This work was supported by U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, under award DE-FG02-08ER15931 and by U.S. DOE, Advanced Research Projects Agency – Energy, under award number DE-AR0000426. Computer resources were provided by the W. R. Wiley Environmental Molecular Sciences Laboratory, a DOE Office of Science User Facility located at Pacific Northwest National Laboratory and sponsored by DOE's Office of Biological and Environmental Research. Computer resources were also provided by the National Energy Research Computing Center at the Lawrence Berkeley National Laboratory. Author contributions: T.W. generated the data, performed the data analysis, and prepared Figs. 2, A and B, and 4, and Table 1; and performed the data analysis and generated Fig. 2, C and D, and figs. S1, S4, and S6. D. Sliwa prepared CoBSH substrates and enzyme; generated the data and performed preliminary data analysis for Fig. 2, C and D, and fig. S1; and prepared the samples for the MCD measurements described in Fig. 3 and fig. S2. B.G. and S.R. performed and interpreted the computational studies described in Fig. 5; figs. S5, S7, S8, and S9; and tables S1 to S3. D. Smith performed initial computational experiments on the MCR mechanisms. M.W. and N.L. performed, analyzed, and interpreted the MCD spectroscopic experiments and prepared Fig. 3 and fig. S2. S.W.R. conceptualized and refined the research idea; prepared Figs. 1 and 6 and fig. S3; and coordinated the various collaborations and guided analysis and interpretation of the biochemical experiments. All authors were involved to varying degrees in writing and/or editing the manuscript.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/953/suppl/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S3
References (*49*–*79*)

11 December 2015; accepted 5 April 2016
10.1126/science.aaf0616

REPORTS

SUPERCONDUCTIVITY

Ubiquitous signatures of nematic quantum criticality in optimally doped Fe-based superconductors

Hsueh-Hui Kuo,[1,2]* Jiun-Haw Chu,[1,3]*† Johanna C. Palmstrom,[1,3] Steven A. Kivelson,[1,4] Ian R. Fisher[1,3]†

A key actor in the conventional theory of superconductivity is the induced interaction between electrons mediated by the exchange of virtual collective fluctuations (phonons in the case of conventional *s*-wave superconductors). Other collective modes that can play the same role, especially spin fluctuations, have been widely discussed in the context of high-temperature and heavy Fermion superconductors. The strength of such collective fluctuations is measured by the associated susceptibility. Here we use differential elastoresistance measurements from five optimally doped iron-based superconductors to show that divergent nematic susceptibility appears to be a generic feature in the optimal doping regime of these materials. This observation motivates consideration of the effects of nematic fluctuations on the superconducting pairing interaction in this family of compounds and possibly beyond.

A growing body of evidence suggests the possibility of an intimate connection between electronic nematic phases (*1*) and high-temperature superconductivity. However, it is currently unclear to what extent there is any causal relationship between nematic fluctuations and superconductivity. Strongly anisotropic electronic phases have been found in the underdoped regime of both cuprate (*2*–*6*) and Fe-based (*7*–*12*) high-temperature superconductors.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

In the case of underdoped Fe-based systems, recent measurements of the elastoresistance (*13*–*15*), Raman spectra (*16*–*18*), and elastic moduli (*19*, *20*) of the representative electron-doped system $Ba(Fe_{1-x}Co_x)_2As_2$ reveal a divergence of the electronic nematic susceptibility upon approaching the tetragonal-to-orthorhombic structural phase transition; this definitively establishes that the phase transition is driven by electronic correlation. In the case of the cuprates, recent x-ray diffraction (*21*–*24*) and nuclear magnetic resonance (*25*) measurements have shown evidence for short-range charge density wave order in underdoped crystals. Although details of the charge-ordered states are still being established, these initial observations have at least motivated discussion of a possible "vestigial" nematic order (*26*). Perhaps importantly, in the phase diagrams of both families of compounds, optimal doping is located close to putative quantum critical points (*27*–*29*) that potentially have a nematic character (*13*, *26*, *30*).

From a theoretical perspective, recent treatments indicate that nematic quantum criticality (i.e., quantum critical fluctuations caused by proximity to a nematic quantum critical point) can enhance an existing pairing interaction. In particular, a pure nematic phase does not break the translational symmetry of the original crystal lattice; consequently, the nematic fluctuations have a wave-vector $q = 0$ so that pairing through the exchange of nematic fluctuations enhances the critical temperature T_c in all symmetry channels (*31*–*33*). It is therefore of considerable interest to empirically establish whether nematic fluctuations are a characteristic feature of optimally doped high-temperature superconductors, as well as to determine the extent to which these nematic fluctuations show intrinsically quantum behavior. Here we show that this is the case for iron pnictide and chalcogenide superconductors by considering the representative 122-type pnictide materials $Ba(Fe_{1-x}Co_x)_2As_2$ and $Ba(Fe_{1-x}Ni_x)_2As_2$ (electron-doped), $Ba_{1-x}K_xFe_2As_2$ (hole-doped), and $BaFe_2(As_{1-x}P_x)_2$ (isovalent substitution) and the 11-type iron chalcogenide $FeTe_{1-x}Se_x$. Furthermore, we find that the nematic susceptibility obeys a simple Curie-Weiss power law for all five optimally doped Fe-based superconductors over a wide temperature range. For the electron- and hole-doped 122-type pnictides, a sub–Curie-Weiss deviation was observed at low temperatures, which we tentatively attribute to an enhanced sensitivity to disorder in a quantum critical regime.

Nematic order couples linearly to anisotropic strain of the same symmetry. Consequently, the nematic susceptibility of a material can be measured by considering the electronic anisotropy that is induced by anisotropic in-plane strain. In the regime of infinitesimal strains, all forms of electronic anisotropy are linearly proportional. Hence, the rate of change of resistivity anisotropy with respect to anisotropic strain, defined in the limit of vanishing strain, is linearly proportional to the nematic susceptibility (*34*). The proportionality constant depends on microscopic physics, but away from any quantum critical point, it does not exhibit any singular behavior. Thus, the induced resistivity anisotropy reveals the essential divergence of the nematic susceptibility upon approaching a thermally driven nematic phase transition (*14*, *15*). For a tetragonal material, nematic order has either d_{xy} symmetry (the B_{2g} irreducible representation of the D_{4h} point group, corresponding to nematic order oriented along a nearest-neighbor Fe-Fe bond, i.e., along the [110] or [−110] crystal axes) or $d_{x^2-y^2}$ symmetry (B_{1g}, corresponding to the [100] or [010] crystal axes). Anisotropic strains ε with appropriate symmetry are then $\varepsilon_6 = (\varepsilon_{xy} + \varepsilon_{yx})/2$ and $\varepsilon_1 - \varepsilon_2 = \varepsilon_{xx} - \varepsilon_{yy}$, respectively. The strain-induced changes in resistivity ρ can be described using the dimensionless elastoresistivity tensor, m_{ij} (*14*)

$$(\Delta\rho/\rho)_i = \sum_{j=1}^{6} m_{ij}\varepsilon_j \qquad (1)$$

where the indices are defined using standard Voigt notation (1 = *xx*, 2 = *yy*, and so forth). Consequently, the corresponding components of the nematic susceptibility tensor χ_N are given by

$$\chi_{N(B_{2g})} = c \times 2m_{66} \qquad (2)$$

$$\chi_{N(B_{1g})} = c' \times (m_{11} - m_{12}) \qquad (3)$$

where c and c' are proportionality constants that depend on microscopic physics (*34*).

We measured elastoresistivity by applying an in situ tunable anisotropic strain, using a piezoelectric lead-zirconate-titanate (PZT) stack.

[1]Stanford Institute of Energy and Materials Science, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park 94025, CA 94305, USA. [2]Department of Material Science and Engineering and Geballe Laboratory for Advanced Materials, Stanford University, Stanford, CA 94305, USA. [3]Department of Applied Physics and Geballe Laboratory for Advanced Materials, Stanford University, Stanford, CA 94305, USA. [4]Department of Physics and Geballe Laboratory for Advanced Materials, Stanford University, Stanford, CA 94305, USA.
*These authors contributed equally to this work. **†Corresponding author. Email: arrowchu@gmail.edu (J.-H.C.); irfisher@stanford.edu (I.R.F.)**

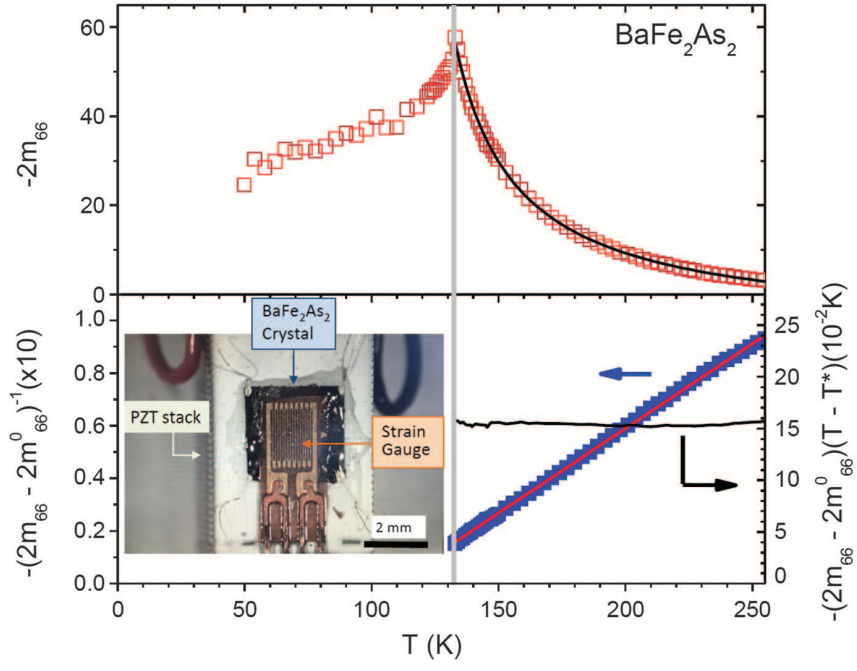


Fig. 1. Temperature dependence of the B_{2g} elastoresistance $2m_{66}$ of $BaFe_2As_2$. The data exhibit Curie-Weiss behavior, which is the anticipated mean-field temperature dependence of the nematic susceptibility of a material approaching a thermally driven nematic phase transition (*13*–*15*). The upper panel shows $-2m_{66}$, which is proportional to the nematic susceptibility $\chi_{N(B_{2g})}$, in the tetragonal phase. The black line shows the Curie-Weiss fit. The quality of fit can be better appreciated by considering the inverse susceptibility $-(2m_{66} - 2m_{66}^0)^{-1}$, which is perfectly linear (left axis of lower panel, as indicated by the blue arrow; the fit is shown by the red line), and the Curie constant $-(2m_{66} - 2m_{66}^0)^{-1}(T - T^*)$ (right axis of lower panel, as indicated by the black arrow), which is independent of temperature. The Weiss temperature T^* obtained from the Curie-Weiss fit, which gives the bare mean-field nematic critical temperature, is 109 ± 1.4 K. Coupling to the lattice renormalizes the critical temperature, leading to a coupled nematic-structural phase transition at T_s = 134 K (vertical gray line). The inset shows a photograph of a square crystal glued on a PZT piezoelectric stack for differential elastoresistance measurements, using Montgomery's geometry. Four electrical contacts were made at the corners, and a strain gauge was glued on the top surface (*36*).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

In this approach, a square plate sample (typical dimensions, 750 × 750 × 20 μm) is glued on the side wall of the PZT stack with a commercial two-part epoxy. The PZT stack deforms when a voltage is applied and hence strains the sample glued on top of it (*35*). The amount of strain can be measured by a strain gauge glued either on the backside of the PZT stack or on the top surface of a larger sample [the latter arrangement enables a full determination of the strain transmission (*36*, *37*)]. The in-plane resistivity tensor of the sample is measured via the Montgomery

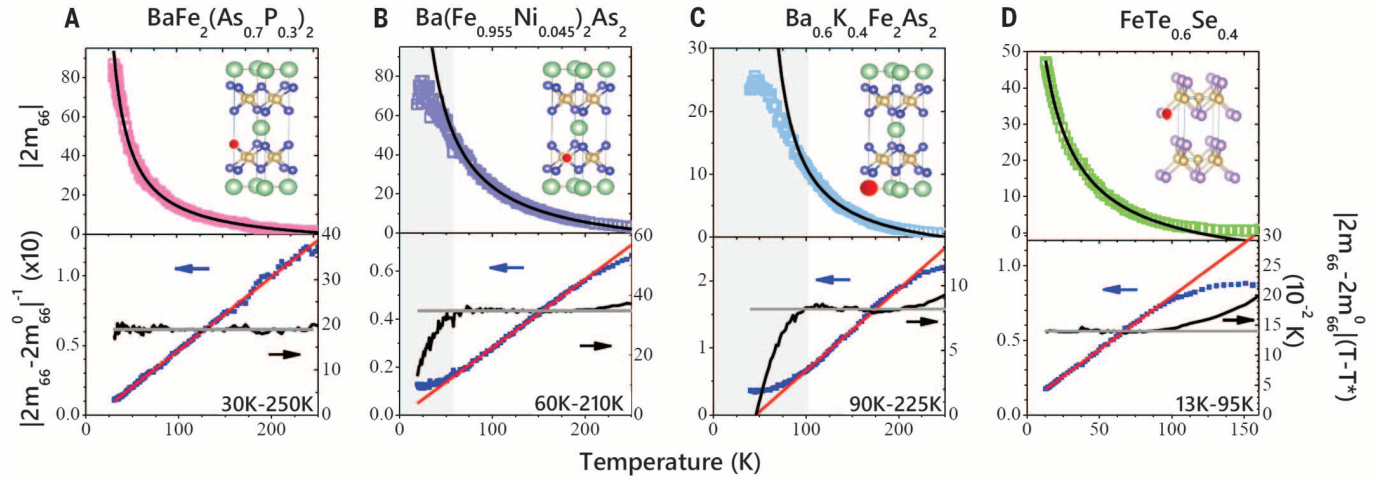


Fig. 2. Divergence of the B_{2g} elastoresistance $2m_{66}$ of several different families of optimally doped iron pnictide and chalcogenide superconductors. (**A**) Optimally doped $BaFe_2(As_{0.68}P_{0.32})_2$ (isovalent substitution), (**B**) optimally doped $Ba(Fe_{0.955}Ni_{0.045})_2As_2$ (electron-doped), (**C**) optimally doped $Ba_{0.58}K_{0.42}Fe_2As_2$ (hole-doped), and (**D**) optimally doped $FeTe_{0.58}Se_{0.42}$. Insets indicate the dopant site (red) in the respective unit cells of each material (Fe and Ni, yellow; As and P, purple; Ba and K; Te and Se, pink). Induced resistivity anisotropies of comparable magnitudes are observed for each case (*36*). Upper panels show $|2m_{66}|$; lower panels show $|2m_{66} - 2m_{66}^0|^{-1}$ (left axes, blue squares) and $|2m_{66} - 2m_{66}^0|(T - T^*)$ (right axes, black curves). Black (upper panels) and red (lower panels) lines show Curie-Weiss fits of $2m_{66}$ and $|2m_{66} - 2m_{66}^0|^{-1}$, respectively. Gray horizontal lines (lower panels) show the average values of $|2m_{66} - 2m_{66}^0|(T - T^*)$ in the fitting temperature range. Regions of deviation from Curie-Weiss behavior in (B) and (C) are indicated by gray shaded regions. For (A) and (B), $2m_{66}$ is negative. For (C) and (D), $2m_{66}$ is positive. Fit parameters, together with the temperature range over which a good fit is obtained, are listed in (*36*). The fits shown in the upper and lower panels are extrapolated beyond that range to emphasize the deviations from Curie-Weiss behavior.

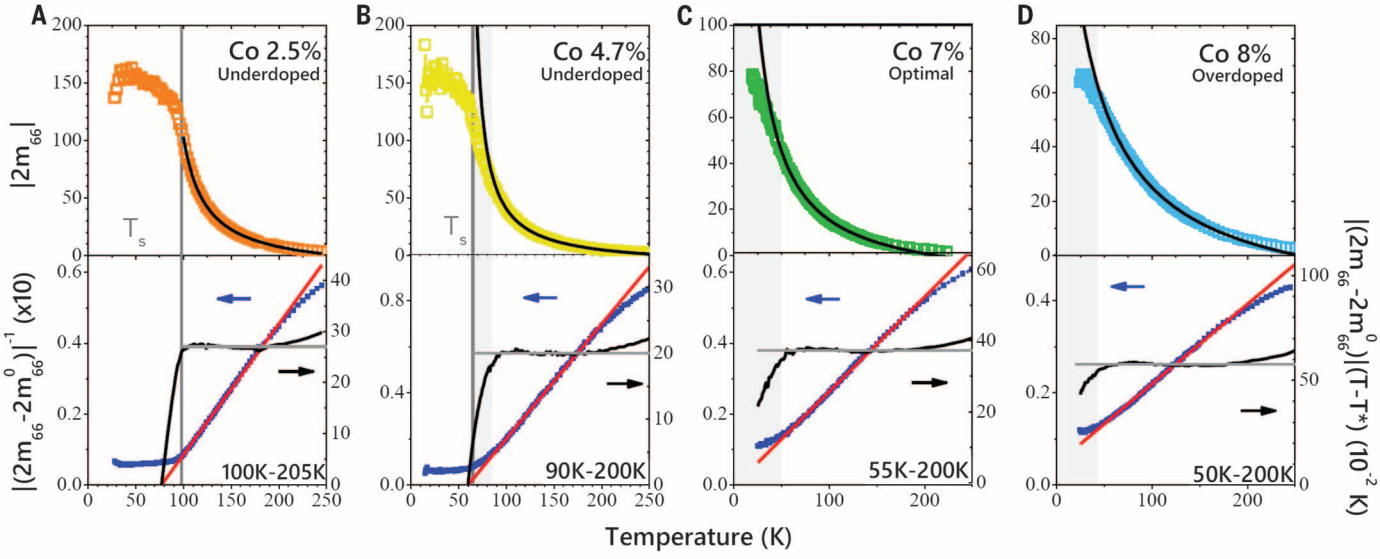


Fig. 3. Variation of the B_{2g} elastoresistance of $Ba(Fe_{1-x}Co_x)_2As_2$ for four representative compositions. (**A** and **B**) Underdoped compositions $Ba(Fe_{0.975}Co_{0.025})_2As_2$ and $Ba(Fe_{0.953}Co_{0.047})_2As_2$; (**C**) optimally doped $Ba(Fe_{0.93}Co_{0.07})_2As_2$; and (**D**) overdoped $Ba(Fe_{0.92}Co_{0.08})_2As_2$. The very underdoped composition (A) is well described by a Curie-Weiss temperature dependence over the entire temperature range (black lines in upper panels), with only subtle deviations at the highest temperatures that are sensitive to the value of m_{66}^0 in the fit. For the compositions near optimal doping, $2m_{66}$ can be well fit by a Curie-Weiss T dependence at intermediate temperatures. At temperatures below a characteristic value (different for each composition; indicated by the gray shaded regions), a strong downward deviation from Curie-Weiss behavior is observed; this deviation is also evident in the inverse susceptibility [upward deviation of the data (blue symbols) from an inverse Curie-Weiss fit (red line)] and in $|2m_{66} - 2m_{66}^0|(T - T^*) \propto \chi_N(T - T^*)$ (strong downturn, black curve), which are shown in the lower panels. Gray vertical lines indicate T_s; gray horizontal lines indicate average values of $|2m_{66} - 2m_{66}^0|(T - T^*)$. The deviation from Curie-Weiss behavior is the strongest at optimal doping and diminishes on either side of the phase diagram. Fit parameters, together with the temperature range over which a good fit is obtained, are given in (*36*). As in Fig. 2, the fits shown in the upper and lower panels are extrapolated beyond that range to emphasize the deviations from Curie-Weiss behavior.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

technique, with electrical contacts made at the four corners of the square sample (*36*). Representative data collected using this new technique are shown in Fig. 1 for the specific case of $BaFe_2As_2$. As has been previously demonstrated (*14*, *15*), the data can be fit very well by a Curie-Weiss temperature dependence

$$2m_{66} = 2m_{66}^0 + \frac{\lambda}{a(T - T^*)} \quad (4)$$

where λ/a is the Curie constant, and T^* is the Weiss temperature.

In Fig. 2, we show B_{2g} elastoresistance data for a range of optimally doped materials, including $BaFe_2(As_{1-x}P_x)_2$ (isovalently substituted), $Ba(Fe_{1-x}Ni_x)_2As_2$ (electron-doped), $Ba_{1-x}K_xFe_2As_2$ (hole-doped), and $FeTe_{1-x}Se_x$. In all cases, $2m_{66}$ rises pronouncedly with decreasing temperature, with comparably large values for each compound. The observation of this effect across such a wide variety of doping methods, including in the case of the iron chalcogenide $FeTe_{0.6}Se_{0.4}$, suggests strongly that a divergence of the nematic susceptibility in the B_{2g} channel is a generic feature of optimally doped Fe-based superconductors.

Regardless of microscopic models, from a purely empirical perspective, it is apparent that the optimally doped superconductor is born out of an electronic state that is characterized by strongly fluctuating orientational (nematic) order in this specific symmetry channel. This result is especially notable for $FeTe_{0.6}Se_{0.4}$, given that the orientation of both the magnetic ordering wave vector and the in-plane component of the structural distortion of undoped FeTe are 45° away from the orientation of those in the iron arsenide materials.

The m_{66} coefficient of optimally doped $BaFe_2(As_{0.68}P_{0.32})_2$ (Fig. 2A) follows a perfect Curie-Weiss temperature dependence from $T = 250$ K (the highest temperature at which strain can be effectively transmitted by the epoxy) down to T_c (below which the resistance is zero), with a T^* close to zero. For $FeTe_{0.6}Se_{0.4}$, the m_{66} coefficient can be perfectly fitted with the Curie-Weiss dependence down to T_c but substantially deviates at temperatures greater than 100 K, possibly related to the loss of quasiparticle coherence as a result of its extremely small Fermi energy, as has been observed in photoemission spectroscopy and transport (*38*, *39*). For the electron- and hole-doped 122-type pnictides, a downward deviation from Curie-Weiss behavior is apparent below a characteristic temperature that is different for each of the materials. The quality of fit to the Curie-Weiss functional form for $BaFe_2(As_{0.68}P_{0.32})_2$, in comparison with all of the other optimally doped compositions that we studied, can be readily appreciated when the data are plotted on log axes (fig. S11). Only the data for the P-substituted system can be fit by a single power law $[|2m_{66} - 2m_{66}^0| \sim (T - T^*)^{-\gamma}]$ over the entire temperature range, yielding the power law exponent $\gamma = 0.985 \pm 0.005$ and $T^* = 11.7 \pm 3.1$ K.

We also measured a slightly overdoped $BaFe_2(As_{0.64}P_{0.36})_2$ sample; the elastoresistivity coefficient m_{66} for this case can also be well fitted by a Curie-Weiss temperature dependence over the entire temperature range, with a negative $T^* = -11.5 \pm 2.3$ K (fig. S14). The small value of the Weiss temperature observed in $BaFe_2(As_{1-x}P_x)_2$ near optimal doping, and the fact that it crosses zero as doping increases, motivates consideration of the nematic susceptibility in the context of a quantum critical point. An Ising nematic phase transition in an insulator would generally be expected to have a dynamical exponent $z = 1$, so the effective dimensionality of the system would be $d + z = 3 + 1$. Because the materials in question are metallic, the situation is more complicated (*40*); the Hertz-Millis paradigm is expected to break down in the case of electronic nematic order in metallic systems, for which Landau damping associated with the gapless Fermions leads to a larger value of the dynamical exponent, $z = 3$. Although the analysis of this problem is not straightforward, both analytic (*41*) and numerical studies (*42*, *43*) indicate that the nematic susceptibility at criticality should diverge in proportion to $1/T$ (up to possible logarithmic corrections), consistent with the observed behavior of the B_{2g} elastoresistance of optimally doped $BaFe_2(As_{1-x}P_x)_2$.

To gain insight into the physical origin of the low-temperature downward deviation from Curie-Weiss behavior in the electron- and hole-doped pnictides, we consider the evolution of the B_{2g} elastoresistance as a function of composition for the specific case of Co-doped $BaFe_2As_2$.

The progression of $2m_{66}$ as a function of Co doping for several representative compositions [more were measured (*36*)], from underdoped through optimal doping to the overdoped regime, is shown in Fig. 3, A to D. For very underdoped compositions ($x = 0.025$), $2m_{66}$ can be well described by the mean-field Curie-Weiss T dependence down to the critical temperature for the tetragonal-to-orthorhombic structural phase transition, T_s. It has been previously established that this behavior is essentially independent of disorder, at least when comparing undoped $BaFe_2As_2$ materials with different residual resistance ratios and Co- and Ni-substituted crystals with the same Neel temperature T_N (*15*). The T^* extracted from high-temperature fits to Curie-Weiss behavior crosses zero close to optimal doping (Fig. 4). However, for slightly underdoped $Ba(Fe_{0.553}Co_{0.047})_2As_2$, a downward deviation from mean-field behavior at low temperatures is noticeable, and this pattern is more pronounced for optimally doped $Ba(Fe_{0.93}Co_{0.07})_2As_2$; similar deviations are apparent for optimally Ni- and K-doped $BaFe_2As_2$. A similar effect has been observed in recent measurements of the sheer modulus through three-point bending experiments (*20*). The deviation from Curie-Weiss behavior diminishes again as the doping is further increased: For the overdoped composition (Fig. 3D), the data can be fit to the Curie-Weiss functional form down to a lower temperature than for optimal doping, and the magnitude of the deviation below ~45 K is smaller than for optimal doping. The $2m_{66}$ of four additional Co dopings has also been measured, showing a similar nonmonotonic doping dependence of the deviation from Curie-Weiss behavior. The fact that the effect is maximal near optimal doping

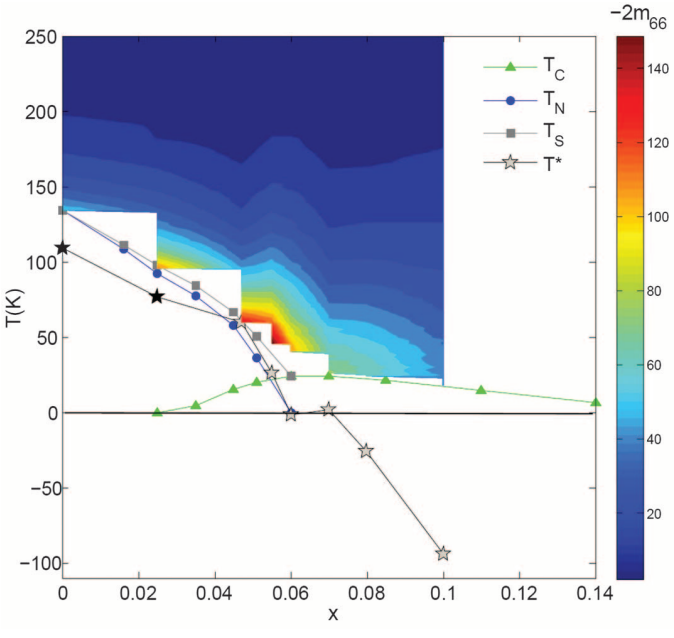


Fig. 4. Phase diagram of $Ba(Fe_{1-x}Co_x)_2As_2$, showing the variation of $2m_{66}$ in the x-T plane (color scale). Squares, circles, and triangles indicate T_s, T_N, and T_c respectively. Stars indicate the bare mean-field nematic critical temperatures (T^*) extracted from the Curie-Weiss fits of $2m_{66}$ above the temperature at which disorder effects suppress the nematic susceptibility (light gray stars are used for cases in which deviations from Curie-Weiss behavior are observed at low temperatures). As has been previously determined by longitudinal elastoresistance measurements (*13*), and as established in this study by the full B_{2g} differential elastoresistance, the Weiss temperature T^* goes through zero as a function of x close to optimal doping. The color scale shows the magnitude of $-2m_{66}$, which peaks between slightly underdoped to optimally doped compositions.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

when $T^* \sim 0$ suggests that it is associated with proximity to the putative nematic quantum phase transition.

P-substituted $BaFe_2As_2$ is the least disordered of all of the known 122-type iron pnictide families, as evidenced by the fact that quantum oscillations can be observed across the phase diagram (*44*, *45*). The deviation from Curie-Weiss behavior for optimally doped compositions of all other dopants in $BaFe_2As_2$ suggests that disorder plays an important role in the quantum critical regime. All forms of quenched disorder produce locally anisotropic effective strains, which thus couple to the orientation of the nematic order; this is known as random-field disorder (*26*). Analysis of the random-field Ising model yields several generically expected effects of random-field disorder, including suppression of the nematic susceptibility below mean-field expectations for a clean system, and, for the case of a quantum phase transition, the enhanced sensitivity of quantum critical phenomena to disorder (*36*).

The temperature dependence of the nematic susceptibility can also be extracted from measurements of the elastic moduli (*20*). The two measurements (elastoresistance and elastic moduli) are in broad agreement—for example, in terms of the Curie-Weiss T dependence of χ_N for underdoped compositions of $Ba(Fe_{1-x}Co_x)_2As_2$, and in terms of the deviation from Curie-Weiss behavior near optimal doping. However, there is an important distinction in the relative magnitude of the measured quantities as a function of doping. In particular, the normalized lattice softening $(c_{66}^0 - c_{66})/c_{66}^0$ extracted in (*20*) monotonically decreases in magnitude and extends over a smaller window of temperature as the Co concentration x is increased. In contrast, the quantity $|2m_{66} - 2m_{66}^0| = c\chi_N$ initially increases with x, peaking for slightly underdoped compositions where $x \sim 0.05$. The apparent enhancement of the elastoresistance $2m_{66}$ over the softening of the elastic modulus for compositions near optimal doping is potentially related to renormalization of the quasiparticle effective mass in the quantum critical regime, as has been observed in P-substituted $BaFe_2As_2$ (*29*). Such a mass renormalization is an expected consequence of nematic quantum critical fluctuations (*40*). It is precisely these low-energy quasiparticles, which are responsible for the large elastoresistivity, that are also involved in the eventual superconductivity.

REFERENCES AND NOTES

1. E. Fradkin, S. A. Kivelson, M. J. Lawler, J. P. Eisenstein, A. P. Mackenzie, *Annu. Rev. Condens. Matter Phys.* **1**, 153–178 (2010).
2. Y. Ando, K. Segawa, S. Komiya, A. N. Lavrov, *Phys. Rev. Lett.* **88**, 137005 (2002).
3. C. Howald, H. Eisaki, N. Kaneko, M. Greven, A. Kapitulnik, *Phys. Rev. B* **67**, 014533 (2003).
4. V. Hinkov *et al.*, *Science* **319**, 597–600 (2008).
5. R. Daou *et al.*, *Nature* **463**, 519–522 (2010).
6. S. A. Kivelson *et al.*, *Rev. Mod. Phys.* **75**, 1201–1241 (2003).
7. T. M. Chuang *et al.*, *Science* **327**, 181–184 (2010).
8. J. H. Chu *et al.*, *Science* **329**, 824–826 (2010).
9. M. A. Tanatar *et al.*, *Phys. Rev. B* **81**, 184508 (2010).
10. A. Dusza *et al.*, *Euro. Phys. Lett.* **93**, 37002 (2011).
11. M. Yi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 6878–6883 (2011).
12. X. Lu *et al.*, *Science* **345**, 657–660 (2014).
13. J.-H. Chu, H.-H. Kuo, J. G. Analytis, I. R. Fisher, *Science* **337**, 710–712 (2012).
14. H.-H. Kuo, M. C. Shapiro, S. C. Riggs, I. R. Fisher, *Phys. Rev. B* **88**, 085113 (2013).
15. H.-H. Kuo, I. R. Fisher, *Phys. Rev. Lett.* **112**, 227001 (2014).
16. Y. Gallais *et al.*, *Phys. Rev. Lett.* **111**, 267001 (2013).
17. V. K. Thorsmølle *et al.*, *Phys. Rev. B* **93**, 054515 (2016).
18. F. Kretzschmar *et al.*, *Nat. Phys.* (2016).
19. M. Yoshizawa *et al.*, *J. Phys. Soc. Jpn.* **81**, 024604 (2012).
20. A. E. Böhmer *et al.*, *Phys. Rev. Lett.* **112**, 047001 (2014).
21. J. Chang *et al.*, *Nat. Phys.* **8**, 871–876 (2012).
22. G. Ghiringhelli *et al.*, *Science* **337**, 821–825 (2012).
23. E. Blackburn *et al.*, *Phys. Rev. Lett.* **110**, 137004 (2013).
24. J. M. Tranquada, B. J. Sternlieb, J. D. Axe, Y. Nakamura, S. Uchida, *Nature* **375**, 561–563 (1995).
25. T. Wu *et al.*, *Nature* **477**, 191–194 (2011).
26. L. Nie, G. Tarjus, S. A. Kivelson, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7980–7985 (2014).
27. S. A. Kivelson, E. Fradkin, V. J. Emery, *Nature* **393**, 550–553 (1998).
28. B. J. Ramshaw *et al.*, *Science* **348**, 317 (2015).
29. J. G. Analytis *et al.*, *Nat. Phys.* **10**, 194–197 (2014).
30. K. Fujita *et al.*, *Science* **344**, 612–616 (2014).
31. M. A. Metlitski, D. F. Mross, S. Sachdev, T. Senthil, *Phys. Rev. B* **91**, 115111 (2015).
32. S. Lederer, Y. Schattner, E. Berg, S. A. Kivelson, *Phys. Rev. Lett.* **114**, 097001 (2015).
33. T. A. Maier, D. J. Scalapino, http://arxiv.org/abs/1405.5238 (2014).
34. There are several different ways to define a nematic order parameter in terms of equilibrium electronic correlation functions. Correspondingly, there is ambiguity in defining the nematic susceptibility. Thus, the constants c and c' depend not only on microscopic details but also on the precise microscopic definition given to the nematic order parameter.
35. Typical strains using this method are on the order of 10^{-3} to 10^{-4}.
36. Materials and methods are available as supplementary materials on *Science* Online.
37. As described in (*36*), samples that have in-plane dimensions that are larger than ~700 × 700 μm do not exhibit a size dependence of the strain transmission for typical thicknesses. Although most of the samples used in this study are in this size range, the P-substituted iron arsenide samples have in-plane dimensions of ~300 × 300 μm, resulting in a smaller strain transmission. This does not affect the measured Weiss temperature (*36*).
38. A. Tamai *et al.*, *Phys. Rev. Lett.* **104**, 097002 (2010).
39. A. Pourret *et al.*, *Phys. Rev. B* **83**, 020504(R) (2011).
40. A. L. Fitzpatrick, S. Kachru, J. Kaplan, S. Raghu, *Phys. Rev. B* **88**, 125116 (2013).
41. S. A. Hartnoll, R. Mahajan, M. Punk, S. Sachdev, *Phys. Rev. B* **89**, 155130 (2014).
42. Preliminary confirmation that the nematic susceptibility at criticality diverges as 1/T has been obtained from determinantal Monte Carlo calculations of a two-dimensional metallic nematic quantum critical system (*43*).
43. Y. Shattner, S. Lederer, S. A. Kivelson, E. Berg, http://arxiv.org/abs/1511.03282 (2015).
44. J. G. Analytis, J.-H. Chu, R. D. McDonald, S. C. Riggs, I. R. Fisher, *Phys. Rev. Lett.* **105**, 207004 (2010).
45. H. Shishido *et al.*, *Phys. Rev. Lett.* **104**, 057008 (2010).

ACKNOWLEDGMENTS

The authors thank S. Raghu, S. Lederer, and E. M. Spanton for helpful discussions. J.C.P. is supported by a Gabilan Stanford Graduate Fellowship and a NSF Graduate Research Fellowship (grant DGE-114747). This work was supported by the Department of Energy, Office of Basic Energy Sciences, under contract no. DE-AC02-76SF00515.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/958/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S16
Table S1
References (*46*–*63*)

27 February 2015; accepted 15 April 2016
10.1126/science.aab0103

ARTIFICIAL SPIN ICE

Rewritable artificial magnetic charge ice

Yong-Lei Wang,[1,2]* Zhi-Li Xiao,[1,3]* Alexey Snezhko,[1] Jing Xu,[1,3] Leonidas E. Ocola,[4] Ralu Divan,[4] John E. Pearson,[1] George W. Crabtree,[1,5,6,7] Wai-Kwong Kwok[1]

Artificial ices enable the study of geometrical frustration by design and through direct observation. However, it has proven difficult to achieve tailored long-range ordering of their diverse configurations, limiting both fundamental and applied research directions. We designed an artificial spin structure that produces a magnetic charge ice with tunable long-range ordering of eight different configurations. We also developed a technique to precisely manipulate the local magnetic charge states and demonstrate write-read-erase multifunctionality at room temperature. This globally reconfigurable and locally writable magnetic charge ice could provide a setting for designing magnetic monopole defects, tailoring magnonics, and controlling the properties of other two-dimensional materials.

Artificial ices are structures in which the constituents obey analogs of the "two-in, two-out" Pauling's ice rule that determines the proton positional ordering in water ice. These structures provide a material-by-design approach to physical properties and functionalities (*1*–*27*). Artificial ice can be created from ferromagnetic islands and connected wires (*3*, *5*), topological components such as superconducting vortices (*24*–*26*), and nonmagnetic colloidal particles (*27*). Among them, artificial spin ice is the most investigated system (*1*–*21*) that was first demonstrated in a square lattice of elongated interacting ferromagnetic nanoislands (*1*). In this case, the ice rule corresponds to two spins pointing inward and two pointing outward at the vertex of a square lattice. Extensive experiments have been conducted using various thermal (*6*–*8*) and magnetization approaches (*9*–*11*) to obtain ordered states of the spin ice. Long-range ordering was realized in the diagonally

Downloaded from http://science.sciencemag.org/ on May 19, 2016

polarized magnetic state through the magnetization method (*11*, *14*). For the nominal ground state, sizable domains and crystallites were obtained in as-grown samples (*11*), and larger domains were obtained in samples heated above their Curie temperatures (~500°C for permalloy island arrays) (*6*). At room temperature, long-range ordering of the ground state has been achieved only in ultrathin (3-nm-thick) samples via thermal relaxation (*8*). Other spin and charge configurations have only been observed locally and at crystallite boundaries (*5*, *6*). The difficulty in creating tailored multiple ordered states limits the experimental investigation of the spin and magnetic charge dynamics that can emerge from or between the ordered states (*6*, *10*–*13*), especially for a thermally stable (athermal) sample at room temperature. This also hinders the potential applications of artificial ice for data storage, memory, and logic devices (*3*, *4*) or as a medium for reconfigurable magnonics (*28*).

We designed an artificial spin structure in which we can conveniently obtain multiple long-range orderings of a magnetic charge ice lattice at room temperature. In a typical square spin ice (Fig. 1A), the single-domain magnetic islands are considered to be macro-Ising spins (*11*). Each macrospin can be replaced with a dumbbell of magnetic charges, one positive and one negative (*3*, *4*, *29*) (Fig. 1B). If we break the connections between the pairs of magnetic charges with opposite signs (Fig. 1C), we can design a different pattern of connections (Fig. 1D). Figure 1E shows the resulting artificial spin structure explicitly. The structure consists of a square lattice of plaquettes (labeled M and N) containing ferromagnetic nanoislands with three orientations (horizontal, vertical, and diagonal). The magnetic charge distribution is indicated by the calculated stray field distribution (Fig. 1F), which leads to a magnetic charge ice with the charge ice rule of two negative and two positive charges within each square plaquette.

The magnetic charge distribution in Fig. 1F is exactly the same as that of the ground state of a square spin ice structure (fig. S1). However, in contrast to the square spin ice, which has spins on four sides of the square plaquette and all oriented toward the plaquette center or vertex (Fig. 1A), the structure in Fig. 1E places one spin within the plaquette while removing two from the sides, breaking the fourfold symmetry of the square lattice. Moreover, the three spins associated with each plaquette do not meet at a vertex. Our design contains two types of plaquettes, one rotated by 180° from the other (denoted as M and N in Fig. 1E). Each plaquette (M or N) consists of three islands, each with two degrees of freedom coming from spin, resulting in a total multiplicity of eight configurations of magnetic charges. (See figs. S1 and S2 for detailed comparisons of the spin and charge configurations between the standard square spin ice and our design.) The eight ordered charge configurations (Fig. 1G) that we predicted are separated into three groups on the basis of their energies (see the calculated energies in fig. S3): a twofold degenerate type I ground state (I_1 and I_2), a twofold degenerate excited type II state (II_1 and II_2), and a fourfold degenerate excited type III state (III_1, III_2, III_3, and III_4) (Fig. 1G).

Because of the large energy barrier for the spins, the spin system in Fig. 1E is athermal at room temperature. An external applied magnetic

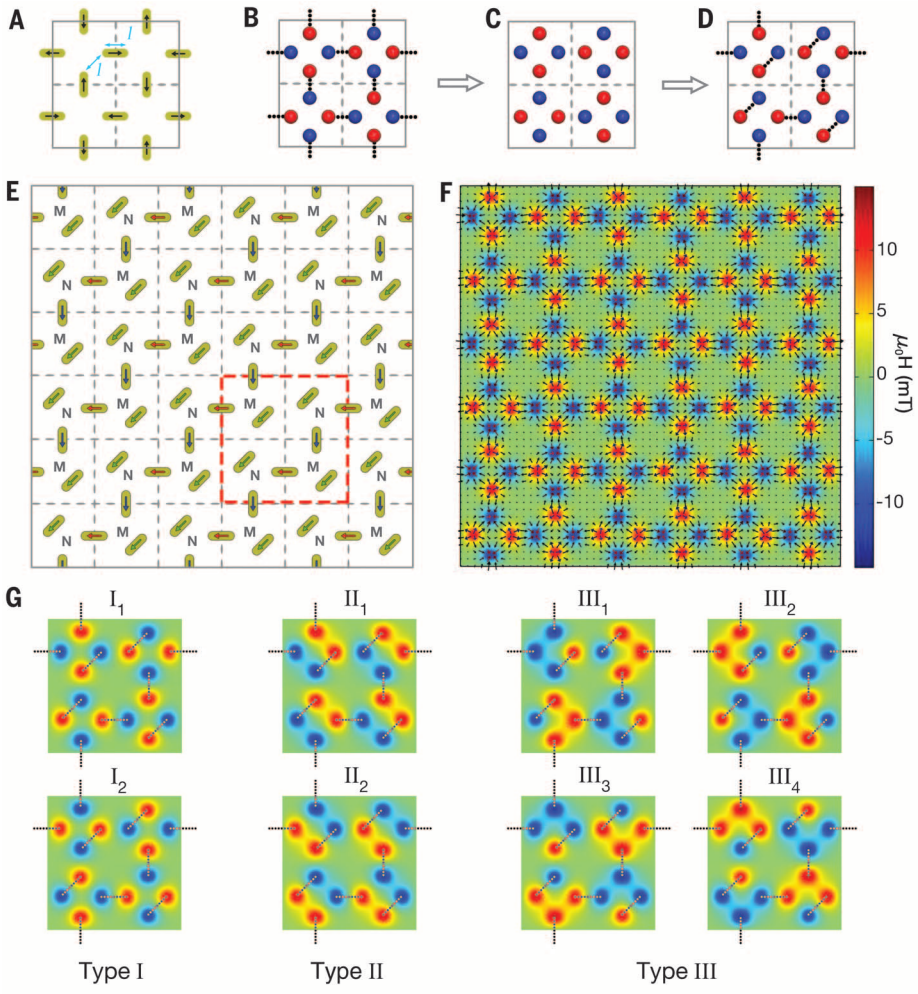


Fig. 1. Design of magnetic charge ices. (**A**) Typical artificial square spin ice in which the length of the magnetic nanoisland equals the separation of the ends of the nearest islands. Each island is an Ising spin (black arrows). The spin configuration of the lowest-energy ground state is shown. (**B**) Distribution of magnetic charges corresponding to the square spin ice shown in (A). Pairing of the positive (red) and negative (blue) charges is indicated by black dotted lines. (**C**) Distribution of magnetic charges with charge connections removed. (**D**) Redesign of the connections of paired charges with positive and negative polarities. (**E**) Design of the magnetic nanostructure based on (D). The arrows indicate the spin configuration of the ground state and are color coded for the three subsets of islands placed in three different orientations (horizontal, vertical, and diagonal). Two types of plaquettes are marked with M and N. (**F**) Calculated magnetic stray field associated with the spin configuration of the structure in (A) for islands with dimensions of 300 nm by 80 nm by 25 nm. Arrows indicate the local field directions. Color is encoded by the out-of-plane component of the field (red, out of the plane; blue, into the plane). The red and blue spots represent the positive and negative magnetic charges, respectively. μ_0, vacuum permeability; H, magnetic field. (**G**) Calculated magnetic stray fields of a 2-by-2 square plaquette, highlighted by the red frame in (E), for the eight possible spin- and charge-ordered configurations, separated into three charge types. Dotted lines denote the orientation of the islands containing the pair of magnetic charges (negative charge, blue; positive charge, red). To compare the magnetic simulations with the experimental magnetic force microscopy images, the stray field was calculated at a plane 100 nm above the surface of the sample.

[1]Materials Science Division, Argonne National Laboratory, Argonne, IL 60439, USA. [2]Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA. [3]Department of Physics, Northern Illinois University, DeKalb, IL 60115, USA. [4]Center for Nanoscale Materials, Argonne National Laboratory, Argonne, IL 60439, USA. [5]Department of Physics, University of Illinois at Chicago, Chicago, IL 60607, USA. [6]Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607, USA. [7]Department of Mechanical and Industrial Engineering, University of Illinois at Chicago, Chicago, IL 60607, USA.
***Corresponding author. Email: ylwang@anl.gov (Y.-L.W.); xiao@anl.gov (Z.-L.X.)**

Downloaded from http://science.sciencemag.org/ on May 19, 2016

field can be used to overcome the energy barrier. For a given island, the minimal magnetic field required to flip its spin moment varies with the angle between the applied field and the island (fig. S4). This enables separate control of the spin moments of each differently oriented island. In a square spin ice, there are two orientations of the islands, and the spins can only be aligned in diagonal directions by an applied magnetic field, enabling solely the type II phases with long-range ordering (*11*, *14*). Our design (Fig. 1E) provides an additional freedom for aligning the spins: There

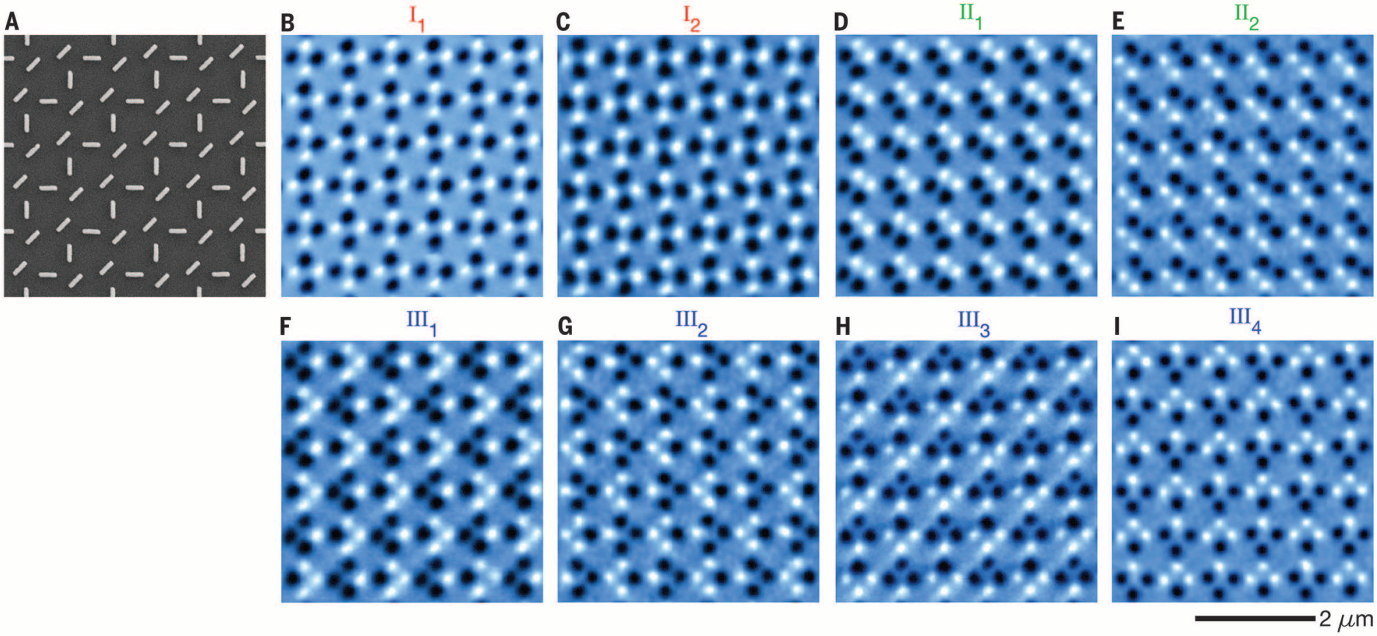


Fig. 2. Realization of magnetic charge ices. (**A**) Scanning electron microscopy image of permalloy ($Ni_{80}Fe_{20}$) magnetic islands (300 nm long, 80 nm wide, and 25 nm thick). (**B** to **I**) Magnetic force microscopy images of the various ordered states corresponding to all of the configurations in Fig. 1G. (B and C) Twofold degenerate type I ground states: I_1 (B) and I_2(C). (D and E) Twofold degenerate excited type II states: II_1 (D) and II_2 (E). (F to I) Fourfold degenerate excited type III states: III_1 (F), III_2 (G), III_3 (H), and III_4 (I). The lift height of the MFM scanning is 100 nm.

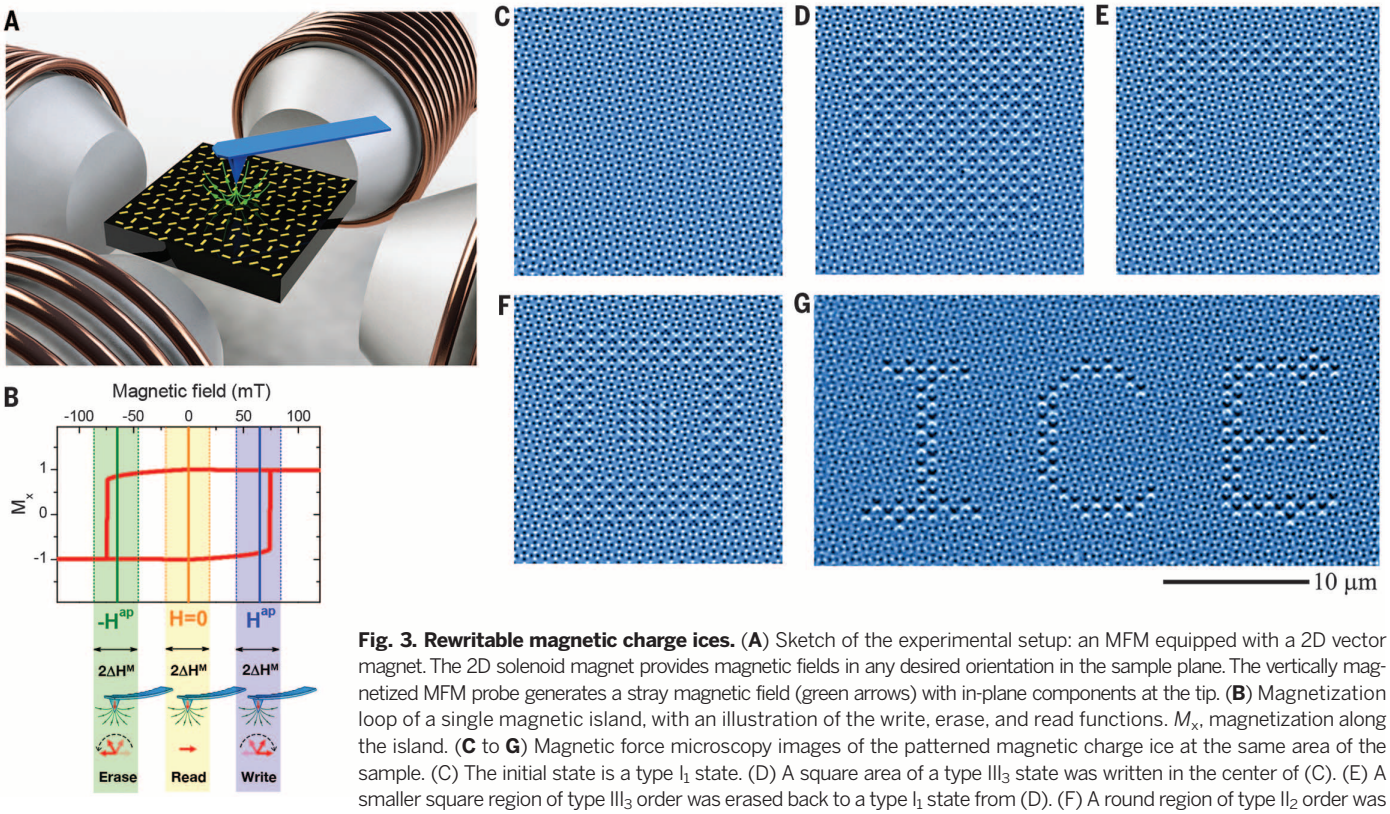


Fig. 3. Rewritable magnetic charge ices. (**A**) Sketch of the experimental setup: an MFM equipped with a 2D vector magnet. The 2D solenoid magnet provides magnetic fields in any desired orientation in the sample plane. The vertically magnetized MFM probe generates a stray magnetic field (green arrows) with in-plane components at the tip. (**B**) Magnetization loop of a single magnetic island, with an illustration of the write, erase, and read functions. M_x, magnetization along the island. (**C** to **G**) Magnetic force microscopy images of the patterned magnetic charge ice at the same area of the sample. (C) The initial state is a type I_1 state. (D) A square area of a type III_3 state was written in the center of (C). (E) A smaller square region of type III_3 order was erased back to a type I_1 state from (D). (F) A round region of type II_2 order was written onto the freshly erased area from (E). (G) "ICE" letters of type III_4 states were scribed on a type I background state.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

are three sets of islands oriented in the horizontal, vertical, and diagonal directions. More importantly, for each of the predicted ordered states shown in Fig. 1G, the spin moments of each of the three oriented islands (horizontal, vertical, and diagonal) are all magnetized in the same direction. This enables the creation of long-range ordering for all of the predicted charge configurations by tuning the in-plane external magnetic field angle and amplitude. We calculated the field-angle dependence of the moment-flipping curves for the three sets of islands (fig. S4C) and designed an effective magnetization protocol to realize the various ordered charge states (*30*).

To experimentally demonstrate the magnetic charge ordering, we fabricated arrays of permalloy ($Ni_{0.8}Fe_{0.2}$) nanoislands (300 nm long, 80 nm wide, and 25 nm thick) onto a Si/SiO_2 substrate (*30*), according to the design in Fig. 1E. A scanning electron microscopy image of the sample is presented in Fig. 2A. To control and visualize the charge configurations, we used a customized magnetic force microscope (MFM) equipped with a two-dimensional (2D) vector magnet. Figure 3A shows a schematic drawing of our experimental setup. The 2D electromagnetic solenoid magnet provides in-plane magnetic fields in any desired orientation, enabling us to accurately tune the field angle and amplitude.

The as-grown (fig. S5) and demagnetized (fig. S6) samples show mixed charge states of all eight charge configurations at both the M and N plaquettes. Results of statistical analysis for the demagnetized samples show that the charge-neutral configurations (type I) are strongly favored and the collective interaction can be enhanced by reducing the charge separation (fig. S6). Using the designed magnetization protocol (fig. S7), we successfully obtained all eight configurations of magnetic charge ordering, as shown by the magnetic force microscopy images in Fig. 2, B to I. Each of these ordered states possesses long-range ordering and can be reproduced over the entire patterned sample area (80 μm by 80 μm). For type I and II ordered states, there is no net charge in each plaquette, and the magnetic charge follows the "two-positive, two-negative" charge ice rule. In the type III ordered states, each plaquette with magnetic charges following the "three-positive (or negative), one-negative (or positive)" charge ice rule has an effective magnetic charge of two, with opposite signs on the M and N plaquettes. This distribution of magnetic charges in the square lattice of the type III state resembles the distribution of electrical charges in ionic compounds, such as $Mg^{2+}O^{2-}$. Thus, the type III states (Fig. 2, F to I) resemble an ionic crystal with magnetic charges (*31*), where we can associate $M^{2+}N^{2-}$ with the type III_2 and III_4 states and $M^{2-}N^{2+}$ with the type III_1 and III_3 states. In principle, all of these magnetic charge distributions could also exist in a square spin ice structure (figs. S1E and S2). In fact, our micromagnetic simulation result indicates that the square spin ice and our engineered spin structure not only produce the same magnetic charge arrangements but also have the same excitation energies for the type I, II, and III configurations (fig. S3). However, with the square spin ice arrangement, long-range ordering of type I states has only been realized in a thermally relaxed sample, and that of the type III states has not yet been experimentally realized because of the spin arrangement of the islands.

The reconfigurable ordered magnetic charges can be used, for example, as templates to form other artificial ices, such as superconducting vortex ices (*25*, *26*), by introducing icelike pinning potentials for superconducting vortices. These charges can also be applied to couple with other electronic materials, such as 2D electron gas (*32*, *33*) and graphene (*34*), by producing reconfigurable periodically distributed field potentials. For applications such as data storage, memory, and logic devices (*3*, *35*, *36*), however, local control of the magnetic charge states is desired. Toward this end, we developed a 2D magnetic field–assisted magnetic force microscopy patterning technique, which allows us to conveniently manipulate the local charge configurations.

As illustrated in Fig. 3A, the magnetic tip of an MFM generates an in-plane component of stray magnetic fields near the tip. The interaction of the MFM tip's stray field, ΔH^m, with a single ferromagnetic island can be tuned by adjusting the height of the MFM tip from the sample, which is 100 nm in this experiment. To locally switch the spin states of an island, we apply an in-plane magnetic field H^{ap} slightly below the ferromagnetic island's spin-moment–flipping field H^f (Fig. 3B). At this field value, the spin states of the entire sample will not be altered because $H^{ap} < H^f$. When the MFM probe scans over an island, the total magnetic field on that island will change in the range $H^{ap} \pm \Delta H^m$. We adjust H^{ap} and ΔH^m to satisfy the condition $H^{ap} < H^f < (H^{ap} + \Delta H^m)$ (light blue region in Fig. 3B). In this case, the spin of the underlying island flips when the MFM tip scans over it, providing a "write" function. A subsequent applied magnetic field in the opposite direction with $-(H^{ap} + \Delta H^m) < -H^f < -(H^{ap} - \Delta H^m)$ (green region in Fig. 3B) will switch the spin back, implementing the "erase" function (Fig. 3B). When the applied field is zero, the stray field provided by the MFM probe (yellow region in Fig. 3B) is too small to flip any islands, and this works as the "read" mode. Because the value of H^f depends on the angle between the applied field and the island, similar to the global control of the charge ordering, we can locally manipulate the charge states into any desired configuration.

We demonstrate the experimental realization of the write, read, and erase functions in Fig. 3, C to F. We first prepare the entire sample in the type I ground state (Fig. 3C) by applying and zeroing an in-plane magnetic field of 90 mT along the diagonal direction. We then write a square area of type III ordered state in the center (Fig. 3D) and subsequently erase a smaller square region in the central area by switching the type III state back into the type I state (Fig. 3E). Finally, we write a type II state into a small circular area inside the type III square (Fig. 3F). We can also write letters, as presented in Fig. 3G where the word "ICE" is scribed with type III order on a type I background. Such magnetic charge patterning can easily be realized by programming the 2D magnet to turn on or off and to switch the field directions during the magnetic force microscopy scanning, resembling the patterning process applied in electron beam lithography and in optical lithography using a laser pattern generator. See fig. S8 for more patterns of magnetic charge arrays. These rewritable magnetic charge patterns could be transferred to other materials, for example, through magnetolithography (*37*).

In addition to the aforementioned applications on coupling with superconducting vortices and other electronic systems, this reconfigurable magnetic charge ice can provide a platform to explore phenomena such as the ground state of a frustrated lattice (*6*). Combined with other control parameters such as the thickness of the islands (*11*), temperature (*6*), or oscillating magnetic field (*10*), our platform provides a versatile system to study and tailor phase transitions and defect formation (fig. S9). For example, the single spin control enabled by our method allows the creation of magnetic defects such as magnetic monopoles and Dirac strings (*16*, *17*, *20*) at any desired location. It also provides a direct technique to program magnetic logic circuits (*35*, *36*). Our strategy to decouple the arrangement of spins and magnetic charges should stimulate further creation and exploration of exotic phases of magnetic charges and their phase transitions and should also foster applications. We also note that Fig. 1E is not the only spin arrangement for achieving the same magnetic charge distribution. In fig. S10, we present several other possible designs of the spin and charge arrangements, including the type IV charge-ordered state in which all positive or negative charges are confined within a single plaquette. Furthermore, it is not necessary to keep the length of all ferromagnetic islands the same, as recently reported in Shakti spin ices (*18*, *19*). Our strategy could also be applied to other artificial spin ices to produce artificial structures with controllable magnetic charge orders, which would provide reconfigurable platforms for magnonic investigations, such as programming spin-wave band structures and designing spin-wave transmission channels (*5*, *28*).

REFERENCES AND NOTES

1. R. F. Wang *et al.*, *Nature* **439**, 303–306 (2006).
2. W. R. Branford, S. Ladak, D. E. Read, K. Zeissler, L. F. Cohen, *Science* **335**, 1597–1600 (2012).
3. L. J. Heyderman, R. L. Stamps, *J. Phys. Condens. Matter* **25**, 363201 (2013).
4. L. J. Heyderman, *Nat. Nanotechnol.* **8**, 705–706 (2013).
5. C. Nisoli, R. Moessner, P. Schiffer, *Rev. Mod. Phys.* **85**, 1473–1490 (2013).
6. S. Zhang *et al.*, *Nature* **500**, 553–557 (2013).
7. C. Nisoli *et al.*, *Phys. Rev. Lett.* **105**, 047205 (2010).
8. A. Farhan *et al.*, *Phys. Rev. Lett.* **111**, 057204 (2013).
9. C. Nisoli *et al.*, *Phys. Rev. Lett.* **98**, 217203 (2007).
10. X. Ke *et al.*, *Phys. Rev. Lett.* **101**, 037205 (2008).
11. Z. Budrikis *et al.*, *Phys. Rev. Lett.* **109**, 037203 (2012).
12. J. P. Morgan, A. Stein, S. Langridge, C. H. Marrows, *Nat. Phys.* **7**, 75–79 (2011).
13. V. Kapaklis *et al.*, *Nat. Nanotechnol.* **9**, 514–519 (2014).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

14. A. Remhof *et al*., *Phys. Rev. B* **77**, 134409 (2008).
15. Z. Budrikis, P. Politi, R. L. Stamps, *Phys. Rev. Lett.* **105**, 017201 (2010).
16. S. Ladak, D. E. Read, G. K. Perkins, L. F. Cohen, W. R. Branford, *Nat. Phys.* **6**, 359–363 (2010).
17. E. Mengotti *et al*., *Nat. Phys.* **7**, 68–74 (2011).
18. I. Gilbert *et al*., *Nat. Phys.* **10**, 670–675 (2014).
19. G. W. Chern, M. J. Morrison, C. Nisoli, *Phys. Rev. Lett.* **111**, 177201 (2013).
20. C. Phatak, A. K. Petford-Long, O. Heinonen, M. Tanase, M. De Graef, *Phys. Rev. B* **83**, 174431 (2011).
21. A. Schumann, B. Sothmann, P. Szary, H. Zabel, *Appl. Phys. Lett.* **97**, 022509 (2010).
22. L. Anghinolfi *et al*., *Nat. Commun.* **6**, 8278 (2015).
23. I. Gilbert *et al*., *Nat. Phys.* **12**, 162–165 (2015).
24. A. Libál, C. J. Reichhardt, C. Reichhardt, *Phys. Rev. Lett.* **102**, 237004 (2009).
25. M. L. Latimer, G. R. Berdiyorov, Z. L. Xiao, F. M. Peeters, W. K. Kwok, *Phys. Rev. Lett.* **111**, 067001 (2013).
26. J. Trastoy *et al*., *Nat. Nanotechnol.* **9**, 710–715 (2014).
27. A. Libál, C. Reichhardt, C. J. Reichhardt, *Phys. Rev. Lett.* **97**, 228302 (2006).
28. M. Krawczyk, D. Grundler, *J. Phys. Condens. Matter* **26**, 123202 (2014).
29. C. Castelnovo, R. Moessner, S. L. Sondhi, *Nature* **451**, 42–45 (2008).
30. Supplementary materials are available on *Science* Online.
31. G. Möller, R. Moessner, *Phys. Rev. B* **80**, 140409 (2009).
32. A. Singha *et al*., *Science* **332**, 1176–1179 (2011).
33. P. D. Ye *et al*., *Phys. Rev. Lett.* **74**, 3013–3016 (1995).
34. M. Taillefumier, V. K. Dugaev, B. Canals, C. Lacroix, P. Bruno, *Phys. Rev. B* **84**, 085427 (2011).
35. A. Imre *et al*., *Science* **311**, 205–208 (2006).
36. I. Amlani *et al*., *Science* **284**, 289–291 (1999).
37. A. Bardea, R. Naaman, *Small* **5**, 316–319 (2009).

ACKNOWLEDGMENTS

We thank W. J. Jiang, S. Zhang, W. Zhang, and M. P. Smylie for critical comments. This work was supported by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division. Z.-L.X. and J.X. were supported by NSF grant no. DMR-1407175. Use of the Center for Nanoscale Materials, an Office of Science user facility, was supported by the DOE, Office of Science, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/962/suppl/DC1
Materials and Methods
Figs. S1 to S10
Reference (*38*)

3 November 2015; accepted 13 April 2016
10.1126/science.aad8037

SUPERCONDUCTIVITY

Supercurrent in the quantum Hall regime

F. Amet,[1,2]*† C. T. Ke,[1]† I. V. Borzenets,[3] J. Wang,[1] K. Watanabe,[4] T. Taniguchi,[4] R. S. Deacon,[5] M. Yamamoto,[3,6] Y. Bomze,[1] S. Tarucha,[3,5] G. Finkelstein[1]*

A promising route for creating topological states and excitations is to combine superconductivity and the quantum Hall (QH) effect. Despite this potential, signatures of superconductivity in the QH regime remain scarce, and a superconducting current through a QH weak link has been challenging to observe. We demonstrate the existence of a distinct supercurrent mechanism in encapsulated graphene samples contacted by superconducting electrodes, in magnetic fields as high as 2 tesla. The observation of a supercurrent in the QH regime marks an important step in the quest for exotic topological excitations, such as Majorana fermions and parafermions, which may find applications in fault-tolerant quantum computing.

The interplay of the quantum Hall (QH) effect with superconductivity is expected to result in excitations with nontrivial braiding statistics such as Majorana fermions and non-Abelian Majorana anyons (*1*–*4*). When a QH region is contacted by two superconducting electrodes, the gapped QH bulk prevents the flow of a supercurrent (*5*–*11*). However, it has been predicted that the supercurrent may still be mediated by QH edge states (*12*). Because of its chiral nature, a single edge can conduct charge carriers in only one direction, so both edges must be involved in establishing a supercurrent between the two contacts. This situation is fundamentally different from that of the Josephson junctions made out of two-dimensional (2D) topological insulators, where each edge can support its own supercurrent (*13*–*16*). Contrary to the case of topological insulators, the magnetic field in the QH regime breaks time-reversal symmetry, which is essential for s-wave pairing of conventional superconductors. Working in this regime, we nonetheless observe a robust supercurrent, which we attribute to an unconventional form of Andreev bound states circulating along the perimeter of the QH region and involving electron and hole trajectories separated by several micrometers.

We performed transport measurements on four Josephson junctions ($J_{1\text{-}4}$) made of graphene encapsulated in boron nitride and contacted by electrodes made of a molybdenum-rhenium alloy (Fig. 1A) (*11*), a type II superconductor with a high upper critical field of H_{c2} = 8 T. The high quality of these heterostructures allowed us to observe Fabry-Perot oscillations of the junctions' resistance and critical current, indicating that the transmission of charge carriers between the contacts is ballistic (*17*). The supercurrent is uniformly distributed along the width of the contacts, as evidenced by the regular Fraunhofer pattern (*18*) measured at small magnetic fields (*17*). All junctions demonstrate supercurrent in the QH regime (figs. S6 to S12); for consistency, we choose to present data measured on sample J_1. It has a distance between contacts of length L = 0.3 μm and a width of contacts W = 2.4 μm (Fig. 1B).

Recently, the observation of a supercurrent through encapsulated graphene in moderate magnetic fields was reported (*19*); in that setup, the diameter of the cyclotron orbit was larger than but comparable to the length of the junction, $2r_C \geq L$. (Here, $r_C = \hbar k_F/eB$ is the cyclotron radius and k_F is the Fermi wavenumber.) This supercurrent was attributed to Andreev bound states made of closed trajectories connected by several elastic and Andreev reflections, which yield pockets of superconductivity at random values of density and field. We further explore this regime in (*17*). In the main text, we demonstrate that a very different regime emerges at even larger magnetic fields, when r_C is much smaller than the device dimensions and the mean free path. In this regime, the bulk of the junction is gapped by Landau quantization so that a current may only flow along the edges.

Figure 1, C and D, show the differential resistance of the sample, $R \equiv dV/dI$, plotted versus back-gate voltage, V_G, and magnetic field, B. The resistance is measured in a four-terminal configuration where four MoRe electrodes merge into two contacts on each side of the junction (Fig. 1B). The map in Fig. 1C is measured with an AC excitation current I_{AC} = 50 pA applied on top of a large DC current of I_{DC} = 6 nA, which suppresses supercurrent and highlights the QH features. As B increases, a fan diagram characteristic of the QH effect in graphene emerges: Resistance plateaus follow contours of constant filling factor $\nu \equiv nh/eB = \pm 2, 6, 10, \ldots$ (*20*). This quantization becomes visible as soon as B exceeds the red parabolic contour $2r_C = L$ because device dimensions prevent the development of the QH effect at lower fields. The dark region under the parabola ($2r_C > L$) indicates a vanishing differential resistance as a supercurrent of tens of nanoamperes may flow in this semiclassical regime (*19*).

Figure 1D shows $R(V_G, B)$ measured simultaneously with Fig. 1C using exactly the same AC excitation of 50 pA, but without applying a DC current. Notably, pockets of supercurrent extend far into the QH regime. They are visible as dark

[1]Department of Physics, Duke University, Durham, NC 27708, USA. [2]Department of Physics and Astronomy, Appalachian State University, Boone, NC 28607, USA. [3]Department of Applied Physics, University of Tokyo, Bunkyo-ku,Tokyo, 113-8656, Japan. [4]Advanced Materials Laboratory, National Institute for Materials Science, 1-1 Namiki, Tsukuba, 305-0044, Japan. [5]Center for Emergent Matter Science (CEMS), RIKEN, Wako-shi, Saitama 351-0198, Japan. [6]PRESTO, Japan Science and Technology Agency, Kawaguchi-shi, Saitama 332-0012, Japan.
***Corresponding author. Email: ametf@appstate.edu (F.A.); gleb@phy.duke.edu (G.F.)** †These authors contributed equally to this work.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

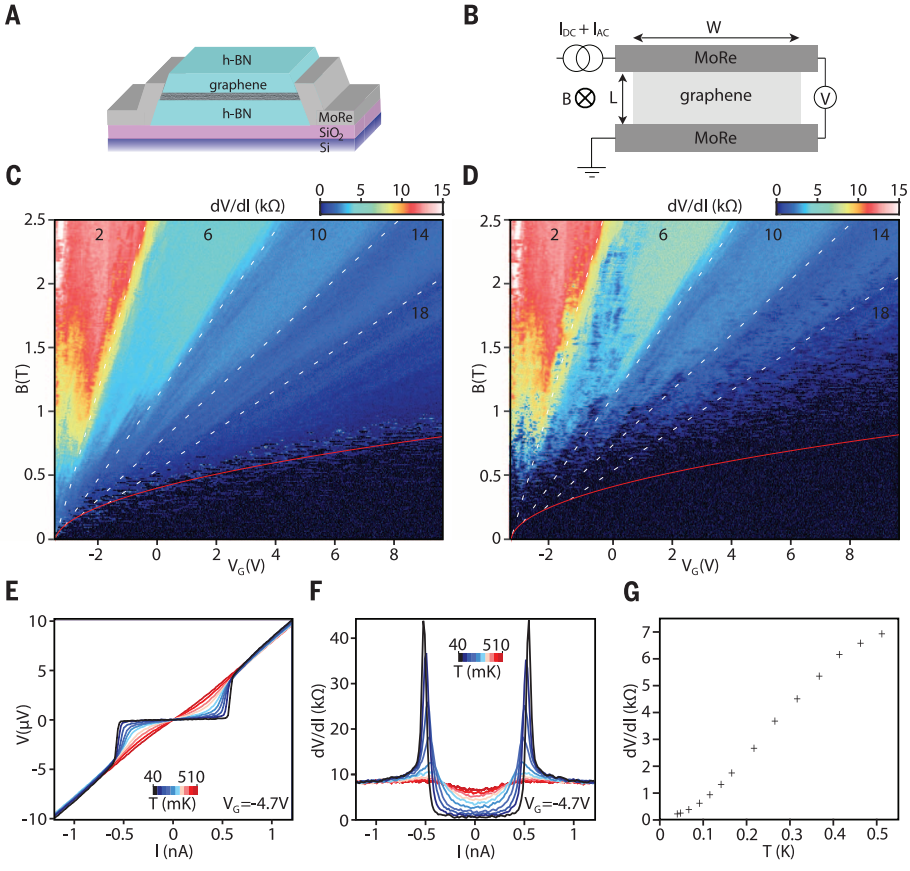


Fig. 1. Coexistence of the QH effect and superconductivity. (**A**) Illustration of the encapsulated graphene heterostructure with molybdenum rhenium contacts. (**B**) Schematic of the measurement setup. (**C** and **D**) Fan diagrams of the differential resistance *dV/dI* plotted versus gate voltage V_G and magnetic field *B*. In (C), *dV/dI* is measured at a finite current bias of I_{DC} = 6 nA, which suppresses superconductivity in the QH regime and reveals the quantized plateaus. In (D), *dV/dI* is measured at zero DC current and shows superconducting pockets extending beyond the semiclassical parabolic region of $2r_C \geq L$. (**E**) *I-V* curves measured in one of the superconducting pockets at *B* = 1 T and V_G = −4.7V. The supercurrent branch is clearly visible at the lowest temperature (40 mK) for *I* < 0.5 nA. *I-V* curves are measured at the following temperatures: 41, 47, 65, 91, 116, 141, 166, 217, 264, 315, 369, 413, 469, and 509 mK. (**F**) The temperature dependence of the corresponding differential resistance, *dV/dI*. The resistance reaches maximum at I_S ~ 0.5 nA, at which point the junction switches from the superconducting to the normal branch. (**G**) *dV/dI* measured as a function of temperature at *I* = 0, showing gradual suppression of superconductivity at elevated temperatures owing to the phase diffusion.

spots of vanishing resistance above the parabolic contour. These pockets occur at somewhat random values of V_G, but are highly reproducible as the gate voltage is swept back and forth. To check that these regions do indeed correspond to a supercurrent, in Fig. 1E we show the *I-V* curves measured in one of the superconducting pockets at *B* = 1 T, V_G = − 4.7 V. The curves demonstrate a clear supercurrent branch, which extends up to 0.5 nA at the lowest temperature of 40 mK. We stress that at that particular point, $r_C \approx 25$ nm << *L*/2 = 150 nm. The corresponding differential resistance (*dV/dI*) vanishes in the same range of currents (Fig. 1F). The maximum of resistance is reached at $I_S \approx 0.5$ nA, at which point the sample switches from the superconducting to the normal branch.

The curves in Fig. 1, E and F, are extremely sensitive to temperature, the supercurrent being washed away by T ≈ 500 mK. This energy scale is orders of magnitude below the critical temperature of MoRe (≈10 K) and the energy splitting of the lowest Landau levels in graphene (>100 K). It is, however, close to the Josephson energy $E_J = \hbar I_C/2e$, which is tens of millikelvin for critical currents of a few nanoamperes. For temperatures comparable to the Josephson energy, thermal fluctuations are expected to suppress the apparent switching current I_S with respect to the true critical current I_C. This explains the observed I_S of only 0.5 nA in Fig. 1E. The thermal fluctuations also result in phase diffusion (*18*), which yields a finite junction resistance even at zero DC current (Fig. 1G).

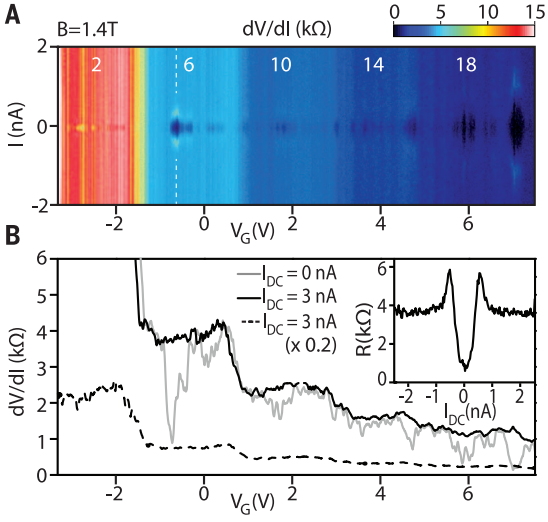


Fig. 2. Supercurrent on top of the QH plateaus. (**A**) Differential resistance *dV/dI* measured at 45 mK as a function of bias current *I* and gate voltage V_G at constant *B* = 1.4 T. Filling factors are indicated in bold white font on top of the map. QH plateaus are clearly visible as stripes of different colors around filling factors ν = 4(*n* + 1/2) with integer *n*. Pockets of superconductivity appear as dark regions close to zero current. (**B**) Line cuts of *dV/dI*, measured in the same range of V_G at zero DC current (gray) and at I_{DC} = 3 nA (black). The dashed curve is obtained by scaling the black curve by a factor 0.2 to demonstrate the ν = 2 plateau. Inset: *dV/dI* versus *I* for a prominent superconducting pocket, which is indicated in (A) by a vertical dashed line.

To further illustrate the coexistence of the QH and the superconducting pockets, we show the differential resistance of the same junction measured as a function of V_G and the current bias *I* at *B* = 1.4 T (Fig. 2A). The QH plateaus are visible in Fig. 2A as vertical stripes of different colors. Pockets of superconductivity appear around zero bias as dark minima of *dV/dI*. (At this field, the cyclotron radius $r_C = 15\nu^{0.5}$ nm is much smaller than device dimensions throughout the map.) The solid black line in Fig. 2B shows the cross-section of the *dV/dI* map taken with a finite current bias of I_{DC} = 3 nA, high enough to suppress any superconducting features. Plateaus are clearly visible close to quantized values of $R = h/(\nu e^2)$, with ν = 2, 6, 10, ... The deviations from perfect

Downloaded from http://science.sciencemag.org/ on May 19, 2016

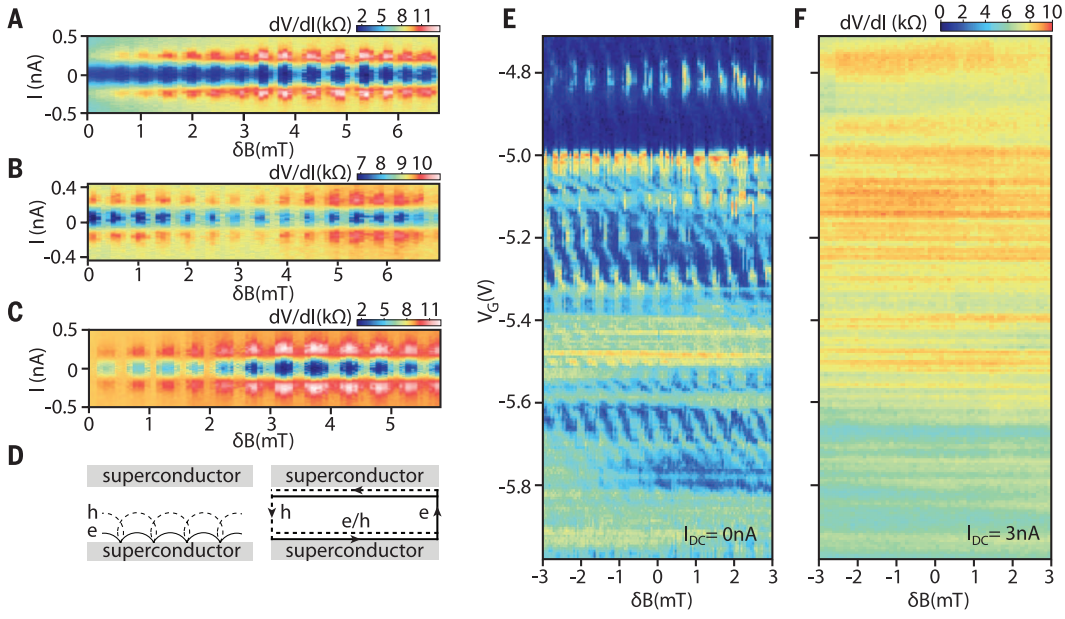


Fig. 3. Periodicity with magnetic field. (**A** to **C**) Differential resistance *dV/dI* measured at 40 mK as a function of bias current *I* and incremental magnetic field δ*B* around *B* = 1 T. The three panels correspond to the superconducting pockets that are located at gate voltages of (A) −5.1 V, (B) −2.2 V, and (C) −2.6 V (see fig. S5). In all three cases, the critical current oscillates with the same period of 0.5 mT, which is close to the period of the Fraunhofer pattern at small fields (fig. S4). (**D**) Right: schematics of the Andreev bound states, made of counterpropagating electron and hole states on the opposite sides of the sample, which couple through the hybrid electron-hole modes running along the superconductor-QH interfaces. Left: these chiral hybrid modes are made of electron and hole edge states mixed through the Andreev processes, as described in the main text. (**E** and **F**) Maps of *dV/dI* as a function of δ*B* and V_G measured at (E) zero I_{DC} and (F) I_{DC} = 3 nA when the supercurrent is suppressed. The superconducting features in (E) demonstrate the same magnetic field periodicity of 0.5 mT as found in (A) to (C), whereas the normal resistance in (F) is almost field-insensitive in this field range.

quantization are common in QH samples in which the current and voltage probes are combined (*21*). The gray line corresponds to the cross-section measured at zero DC current and clearly shows the regions of suppressed differential resistance formed on top of the plateaus thanks to superconductivity.

Figure 3, A to C, shows the differential resistance measured at three superconducting pockets as a function of the bias current and magnetic field, which is varied in steps of 0.1 mT around *B* = 1 T. As the magnetic field is varied, the critical current exhibits a robust interference pattern with a period of 0.5 mT. Notably, this value is close to the period of the Fraunhofer pattern measured on the same junction at very low fields (*B* < 10 mT), when the current distribution along the width of the junction is uniform (fig. S4). However, the current distribution becomes spatially inhomogeneous in the intermediate magnetic fields of tens of milliteslas and beyond, resulting in a very irregular pattern of the supercurrent versus magnetic field (*19*) (figs. S10 to S12). Therefore, the periodicity recovered at high field must be attributed to a very different mechanism.

Because at 1 T, the bulk of graphene is clearly gapped, the periodic supercurrent must be mediated by the edge states. However, the edge states with opposite momenta are located on the opposite sides of the sample, separated by 2.5 to 4.5 μm in our junctions. This scale greatly exceeds the coherence length of the MoRe electrodes (a few nanometers), which prevents the direct coupling of the edges through a simple Andreev reflection. Below, we discuss a mechanism that couples the edge states through hybrid electron-hole modes that are formed at the interfaces between the superconducting contacts and the QH region (*22*–*24*).

Because of the pairing gap, single particles cannot enter the superconducting electrodes and must form edge states along the superconductor-QH interfaces (Fig. 3D). The electron and hole states propagate in the same direction and are hybridized by the superconducting proximity, resulting in chiral hybrid modes. Quasiclassically, one can picture the hybrid electron-hole mode as a skipping orbit in which an electron and a hole are converted into one another on each bounce from the superconductor (*23*) (Fig. 3D). Depending on the transparency of the interface, such a mode could have various degrees of mixing between its electron and hole components. In particular, for perfectly transparent interfaces, it becomes a neutral mixture of the two carrier types, similar to the Majorana modes.

The hybrid modes provide a coherent reservoir of correlated electrons and holes, which is spread over the length of the superconducting electrode and thus could mediate the coupling between the edge states on the opposite sides of the sample (Fig. 3D). Specifically, an electron approaching the top contact along the right edge of graphene must be converted to the hybrid electron-hole mode, which then propagates along the graphene-superconductor interface to the left. Here, it has a finite amplitude of coupling to the left QH edge state as a hole, which then flows to the bottom contact. The loop is completed by the hole conversion into the hybrid mode at that contact and by its subsequent coupling to the original electronic state at the bottom right corner (*24*). This mechanism shuttles one Cooper pair between the contacts, coupling in the process the single-particle edge states separated by micrometers.

To substantiate this scenario, we study the dependence of the superconducting features on the magnetic field and the gate voltage (Fig. 3, E and F). Here, the data in Fig. 3F were measured at I_{DC} = 3 nA, which exceeds the critical current, whereas the data in Fig. 3E were taken at zero DC current and show suppressed resistance when supercurrent flows between the contacts. The features in Fig. 3F are almost field-insensitive, whereas the superconducting features in Fig. 3E exhibit the same magnetic field periodicity as in Fig. 3, A to C. Remarkably, the phase of these features depends on V_G.

The behavior in Fig. 3E is very different from that observed in Josephson junctions based on 2D topological insulators. There, the magnetic interference patterns demonstrate a phase that is independent of the gate voltage (*14*). (This would be equivalent to contours of minimal and maximal current forming vertical stripes in Fig. 3E.) Instead, we observe inclined contours superposed on top of the superconducting pockets, which appear here as wide horizontal stripes.

A general framework for understanding this behavior is provided in (*24*). In the limit relevant to our measurement, it shows that the critical current is proportional to the product of two terms, responsible for the pockets and the inclined contours, respectively. The origin of the first term is the vector potential difference between the superconductor and the edge state: $\delta A = Bd$, where d is the distance of the edge state from the superconductor. The associated wavevector mismatch $\delta Ae/\hbar$ accumulates along the width of the contact, resulting in a phase $e\delta AW/\hbar = Wd/l_0^2$. In our case, this phase factor is ~100 and should strongly depend on the gate voltage: Changing d by just 1 nm should change the phase by 2π. The mesoscopic variations of d should therefore be responsible for the irregular appearance of the superconducting pockets versus V_G.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

The second factor depends on the flux through the Andreev bound state, $\pi BW(L - 2d)/\Phi_0 = \pi\Phi/\Phi_0 - 2Wd/l_0^2$. This phase explicitly combines the magnetic flux through the geometrical area of the junction, $\Phi = BWL$, and the gate-dependent contribution $2Wd/l_0^2$. The maximal current is reached when $\pi\Phi/\Phi_0 - 2Wd/l_0^2 = 2\pi \times$ integer, resulting in the inclined contours in the B-V_G plain of Fig. 3E.

The magnetic field periodicity in Fig. 3 is close to the one observed in the Fraunhofer pattern measured at very small fields (fig. S4). This is different from the result of (*24*), which predicts twice the period in the QH regime. Doubling of frequency has been observed in magnetic interference patterns measured in 2D topological insulators. It has been attributed to charge poisoning (*25*), and a similar mechanism may be at play in our case. Alternatively, it is possible that the results of (*24*) may not be entirely applicable in our case; we hope that our results will stimulate further theoretical work.

The Andreev bound states that we observed in the QH regime are noninvariant under time reversal. This is an important step toward the realization of QH-superconducting hybrids in the quest for Majorana fermions and other exotic topological excitations proposed in (*1*–*3*). Fractional QH states in graphene (*26*, *27*), which emerge in fields as small as 5 T (*28*), should bring these proposals into the realm of possibility. We also anticipate that control of these excitations will be greatly facilitated in 2D graphene nanostructures, where edge channels can be easily manipulated, split, and combined by the application of gate voltages.

REFERENCES AND NOTES

1. N. H. Lindner, E. Berg, G. Refael, A. Stern, *Phys. Rev. X* **2**, 041002 (2012).
2. D. J. Clarke, J. Alicea, K. Shtengel, *Nat. Commun.* **4**, 1348 (2013).
3. R. S. K. Mong *et al.*, *Phys. Rev. X* **4**, 011036 (2014).
4. P. San-Jose, J. L. Lado, R. Aguado, F. Guinea, J. Fernandez-Rossier, *Phys. Rev. X* **5**, 041042 (2015).
5. H. Takayanagi, T. Akazaki, *Physica B* **249-251**, 462–466 (1998).
6. T. D. Moore, D. A. Williams, *Phys. Rev. B* **59**, 7308–7311 (1999).
7. K. Komatsu, C. Li, S. Autier-Laurent, H. Bouchiat, S. Gueron, *Phys. Rev. B* **86**, 115412 (2012).
8. P. Rickhaus, M. Weiss, L. Marot, C. Schönenberger, *Nano Lett.* **12**, 1942–1945 (2012).
9. P. Kumaravadivel, X. Du, Signatures of pseudo diffusive transport in ballistic suspended graphene superconductor junctions. Preprint at: http://arxiv.org/abs/1504.06338.
10. Z. Wan *et al.*, *Nat. Commun.* **6**, 7426 (2015).
11. V. E. Calado *et al.*, *Nature Nanotechnol.* **10**, 761–764 (2015).
12. M. Ma, A. Y. Zyuzin, *Europhys. Lett.* **21**, 941–945 (1993).
13. I. Knez, R. R. Du, G. Sullivan, *Phys. Rev. Lett.* **109**, 186603 (2012).
14. S. Hart *et al.*, *Nat. Phys.* **10**, 638–643 (2014).
15. V. S. Pribiag *et al.*, *Nature Nanotechnol.* **10**, 593–597 (2015).
16. J. Wiedenmann *et al.*, *Nat. Commun.* **7**, 10303 (2016).
17. Supplementary materials are available on *Science* Online.
18. M. Tinkham, *Introduction to Superconductivity* (McGraw-Hill, New York, 1996).
19. M. Ben Shalom *et al.*, *Nat. Phys.* **12**, 318–322 (2015).
20. M. I. Katsnelson, *Graphene: Carbon in Two Dimensions* (Cambridge Univ. Press, 2012).
21. J. R. Williams, D. A. Abanin, L. DiCarlo, L. S. Levitov, C. M. Marcus, *Phys. Rev. B* **80**, 045408 (2009).
22. Y. Takagaki, *Phys. Rev. B* **57**, 4009–4016 (1998).
23. H. Hoppe, U. Zülicke, G. Schön, *Phys. Rev. Lett.* **84**, 1804–1807 (2000).
24. J. A. M. van Ostaay, A. R. Akhmerov, C. W. J. Beenakker, *Phys. Rev. B* **83**, 195441 (2011).
25. S. P. Lee, K. Michaeli, J. Alicea, A. Yacoby, *Phys. Rev. Lett.* **113**, 197001 (2014).
26. X. Du, I. Skachko, F. Duerr, A. Luican, E. Y. Andrei, *Nature* **462**, 192–195 (2009).
27. K. I. Bolotin, F. Ghahari, M. D. Shulman, H. L. Stormer, P. Kim, *Nature* **462**, 196–199 (2009).
28. F. Amet *et al.*, *Nat. Commun.* **6**, 5838 (2015).

ACKNOWLEDGMENTS

We thank H. Baranger, P. Gallagher, K. Matveev, K. Michaeli, A. Smirnov, J. Williams, and S. Yarmolenko for fruitful discussions and technical help. The work done in the United States was supported by the Division of Materials Sciences and Engineering, Office of Basic Energy Sciences, U.S. Department of Energy, under Award no. DE- SC0002765. F.A. acknowledges support from the Fritz London postdoctoral fellowship. I.V.B. and M.Y. are funded by the Canon foundation and Grants-in-Aid for Scientific Research on Innovative Areas, Science of Atomic Layers. R.S.D. is funded by Grants-in-Aid for Young Scientists B (no. 26790008). S.T. is funded by Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research S (no. 26220710).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/966/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Table S1
References (*29*–*33*)

12 October 2015; accepted 12 April 2016
10.1126/science.aad6203

CATALYSIS

Quantifying the promotion of Cu catalysts by ZnO for methanol synthesis

Sebastian Kuld,[1] Max Thorhauge,[1] Hanne Falsig,[1] Christian F. Elkjær,[1,2] Stig Helveg,[1] Ib Chorkendorff,[2] Jens Sehested[1]*

Promoter elements enhance the activity and selectivity of heterogeneous catalysts. Here, we show how methanol synthesis from synthesis gas over copper (Cu) nanoparticles is boosted by zinc oxide (ZnO) nanoparticles. By combining surface area titration, electron microscopy, activity measurement, density functional theory calculations, and modeling, we show that the promotion is related to Zn atoms migrating in the Cu surface. The Zn coverage is quantitatively described as a function of the methanol synthesis conditions and of the size-dependent thermodynamic activities of the Cu and ZnO nanoparticles. Moreover, experimental data reveal a strong interdependency of the methanol synthesis activity and the Zn coverage. These results demonstrate the size-dependent activities of nanoparticles as a general means to design synergetic functionality in binary nanoparticle systems.

Nanoparticles (NPs) used to catalyze chemical reactions generally expose a variety of surface sites (e.g., facets, steps, and corners), each with their distinct reactivity (*1*), and in some cases the function of catalysts based on metal NPs can be predicted with first-principles methods (*2*). The addition of minor amounts of suitably chosen promoter elements can further enhance or suppress surface reactivity and selectivity by orders of magnitude. In general, promotion is thought to originate from either an electronic modification of the catalytic active site or a morphological change of the NPs that enhances the abundance of the active sites, but a detailed understanding is often difficult to establish because of the lower abundance of the promoter atoms. Adding to the difficulty, numerous experimental observations have shown that the catalytic active state dynamically adjusts to the working conditions (*1*). That must be included to understand catalyst promotion.

[1]Haldor Topsøe A/S, Haldor Topsøes Allé 1, DK-2800 Kgs. Lyngby, Denmark. [2]Center for Individual Nanoparticle Functionality (CINF), Department of Physics, Technical University of Denmark, Building 312, DK-2800 Kgs. Lyngby, Denmark.
***Corresponding author. Email: jss@topsoe.dk**

Here, we focus on the catalyzed transformation of synthesis gas (mixture of H_2, CO, and CO_2) into methanol. Methanol is consumed at the scale of 65 million tons/year (2013) and considered as a future energy carrier (*3*). The typical catalyst consists of Cu NPs mixed with NPs of ZnO and Al_2O_3. Although Cu can function alone as a methanol synthesis catalyst, its activity is substantially boosted by the interaction with ZnO, which on its own has only negligible catalytic activity (*4*). The Cu-ZnO system has emerged as a prototype for studying complex promotional interactions in catalysis (*1*, *5*–*22*), and several possible explanations have been suggested, including (i) gas-dependent morphological changes of Cu on ZnO (*6*, *7*, *9*); (ii) support-induced strain in Cu (*22*); and (iii) ZnO_x species/layers covering part of the Cu NPs (*5*, *6*, *9*, *10*, *14*, *19*–*21*), often denoted as a strong metal-support interaction (SMSI). More recently, the active site was associated with step sites in the Cu surface and the Cu-ZnO synergy proposed to reflect the incorporation of Zn atoms into the Cu steps (*5*, *18*). This picture is attractive because of its consistency with observations of fully or partly reduced Zn in the Cu surface after methanol synthesis over Zn-decorated Cu foils and

Downloaded from http://science.sciencemag.org/ on May 19, 2016

single crystals and after reduction of industrial-type $Cu/ZnO/Al_2O_3$ catalysts (*11*, *13*, *15*, *19*). As this dynamic phenomenon may be typical for metal-promoter interactions in heterogeneous catalysis, it is important to reinforce such a picture with a quantitative extension that enables a full description of the Cu-ZnO interaction over the examined reactions' conditions.

We initially studied the $Cu/ZnO/Al_2O_3$ catalyst in the reduced state with high-resolution transmission electron microscopy (TEM). The reduced catalyst consists of an agglomeration of Cu, ZnO, and Al_2O_3 NPs in intimate contact (Fig. 1A). Moreover, Fig. 1B shows that the Cu lattice fringes extend to the projected surfaces of the NPs without changes in their spacing and structure, indicating that the Cu surfaces are in direct contact with the gas environment. The Cu surfaces can thus only be covered with a submonolayer of Zn, consistent with x-ray photoelectron spectroscopy (XPS) measurements (*11*). For that reason, the present work focused on describing the coverage of $Zn(O_x)$ in the Cu surface during methanol synthesis.

First, the Zn coverage, θ_{Zn}, at steady state was determined in the Cu surface of a $Cu/ZnO/Al_2O_3$ catalyst under the exposure to synthesis gas. The concentration in bulk Cu of Zn atoms from ZnO was obtained from the thermodynamics of the ZnO reduction reaction, and the energetic driving force for Zn to migrate to different Cu facets was determined from density functional theory (DFT) calculations, which were used instead of a simple broken-bond model (*23*) to improve the precision. By these calculations, θ_{Zn} is estimated at different gas pressures and temperatures, representative for the methanol synthesis, for a series of Cu surfaces, including steps, edges, and corners. Different Cu surface sites are considered because the methanol synthesis reaction is structure sensitive (*5*, *18*, *24*). Formation of Zn-Zn bonds is included in the calculations, as these are important at high θ_{Zn}. Moreover, the calculations are also done for Cu and ZnO NPs because their reduced size affects the thermodynamic calculations and the abundance of different Cu surface sites.

The formation of a Cu-Zn bulk alloy by incorporation of Zn atoms into Cu NPs can be described by reaction 1 or 2.

$$ZnO(s) + CO = Zn + CO_2 \quad (1)$$
$$ZnO(s) + H_2 = Zn + H_2O \quad (2)$$

Under conditions where the water-gas shift (WGS) is not in equilibrium, the kinetics of the reduction/oxidation reactions determine whether reaction 1 or 2 defines the reduction potential of the gas. Here, the CO/CO_2 ratio was chosen because high CO and CO_2 concentrations were generally applied in the experimental measurements to increase the kinetic importance of this reaction pair. The equilibrium constant K_1 for reaction 1 is given by

$$K_1 = a_{Zn}a_{CO2}/(a_{ZnO}a_{CO}) = \exp(-dG_{red}/RT_K) \quad (3)$$

where dG_{red} is the free energy, a is the activity of the species, R is the gas constant, and T_K is the temperature in Kelvin. The free energy change of reaction 1 for Zn = Zn(s) is 60.95 kJ/mol at 220°C (*25*). Thus, pure Zn(s) cannot form in synthesis gas at 200° to 300°C, but Zn can thermodynamically form a disordered bulk α-alloy phase with Cu (*23*, *26*, *27*). The thermodynamic activity of Zn in Cu is given by $a_{Zn} = \gamma_{Zn}X_{Zn}$, where X_{Zn} is the fraction of Zn in the alloy, and the activity coefficient of Zn in Cu, γ_{Zn}, is calculated from (*26*) as $\ln(\gamma_{Zn}) = -0.736 - 2977/T_K$. The fraction of Zn in the alloy with Cu is thus given by

$$\mathrm{Ln}(X_{Zn}) = -dG_{red}/RT_K - \ln(\gamma_{Zn}) - \ln(a_{CO2}/a_{CO}) \quad (4)$$

Equation 4 is fulfilled when steady state is reached. The diffusion coefficient of Zn in Cu indicates that establishing equilibrium concentrations of Zn in Cu NPs takes years at 200°C, days at 250°C, and hours at 300°C (*27*). In contrast, high θ_{Zn} is observed within a few hours at 220°C (*11*).

To describe the situation where θ_{Zn} reaches steady state at the Cu surface, it is convenient to use Eq. 4, although bulk equilibrium of Zn is not established in the experiment. The bulk contribution in Eq. 4, however, has to be corrected for the effect of the lower atom coordination in NPs. This change in chemical potential of ZnO, $\Delta\mu_{ZnO}$, depends on the ZnO crystallite diameter, d_{ZnO}, and is given by (*28*)

$$\Delta\mu_{ZnO} = 4\gamma_{ZnO}M_{ZnO}/(d_{ZnO}\rho_{ZnO}) \quad (5)$$

where the surface energy, γ_{ZnO}, is 0.74 J/m^2 (*29*), and M_{ZnO} and ρ_{ZnO} are the molar weight and density of ZnO, respectively. The additional chemical potential of Cu NPs relative to bulk Cu is described by a similar term, except that the increased chemical potential gives rise to a lower thermodynamic driving force for Cu-Zn alloy formation. Thus, combining these considerations yields

$$\mathrm{Ln}(X_{Zn}) = -dG_{red}/RT_K - \ln(\gamma_{Zn}) - \ln(a_{CO2}/a_{CO}) + 4\gamma_{ZnO}M_{ZnO}/(d_{ZnO}\rho_{ZnO}RT_K) - 4\gamma_{Cu}M_{Cu}/(d_{Cu}\rho_{Cu}RT_K) \quad (6)$$

where 1.95 J/m^2 (*30*) may be used for γ_{Cu}.

Finally, to calculate θ_{Zn} in Cu surfaces, the equilibrium constant for Zn segregation in bulk Cu, K_{seg}, is calculated (see fig. S1 and tables S1 to S5). Zn will tend to segregate to the surface of Cu because atoms in surfaces have fewer nearest neighbors than in the bulk, and the bond energy is lower for Cu-Zn than for Cu-Cu bonds. By means of DFT, the segregation energies of Zn from bulk Cu to Cu(111), Cu(100), Cu(211)(edge), and Cu(532)(kink) surfaces were calculated to 23.7, 26.1, 27.0, and 29.8 kJ/mol at 0 K, respectively, and the entropy contributions to the segregation were estimated for Cu(111) and Cu(211) to be 5.8 to 7.1 J/K/mol. Because Eq. 6 provides the calculation of the bulk concentration of Zn in Cu, and the segregation energies afford the relative abundances of Zn in the bulk and at the surface of Cu, θ_{Zn} can be calculated for the most abundant Cu facets, steps, and corners from $K_{seg} = \frac{\theta_{Zn}(1-X_{Zn})}{X_{Zn}(1-\theta_{Zn})}$.

Figure 2A shows the calculated θ_{Zn} at Cu(111), Cu(100), Cu(211)(edge), Cu(532)(corner), and X_{Zn} at 220°C as a function of the CO/CO_2 ratio. A random distribution of Zn in the Cu surface is assumed as observed experimentally for a Cu(111) surface by Sano *et al.* (*31*) up to θ_{Zn} = 0.18. At higher θ_{Zn}, Zn-Zn bonds are established in the Cu surface and reduce the segregation energies (fig. S2). This Zn-Zn interaction energy at Cu(111), Cu(100), and Cu(211) is estimated by DFT, and the calculations of θ_{Zn} include this correction. Moreover, θ_{Zn} is also estimated for a catalyst with both Cu and ZnO NPs (Fig. 2A). The calculations use NPs of ZnO, with d_{ZnO} = 87 Å determined by x-ray diffraction (XRD) and a regular cubo-octahedral shape for Cu NPs, having a size of d_{Cu} = 88 Å (which is closest to the crystal size of d_{Cu} = 85 Å, obtained by XRD). The fractions of the different types of surface sites were then obtained according to the model of Van Hardeveld and Hartog (*32*) (table S6). At the terraces, edges, and corners of the cubo-octahedron, θ_{Zn} was estimated as the coverage at the Cu(111), Cu(100), Cu(211), and Cu(532) surface sites, respectively. Taking these coverages together results in a θ_{Zn} for the Cu NP that closely follows that of the more abundant (111) facet site as a function

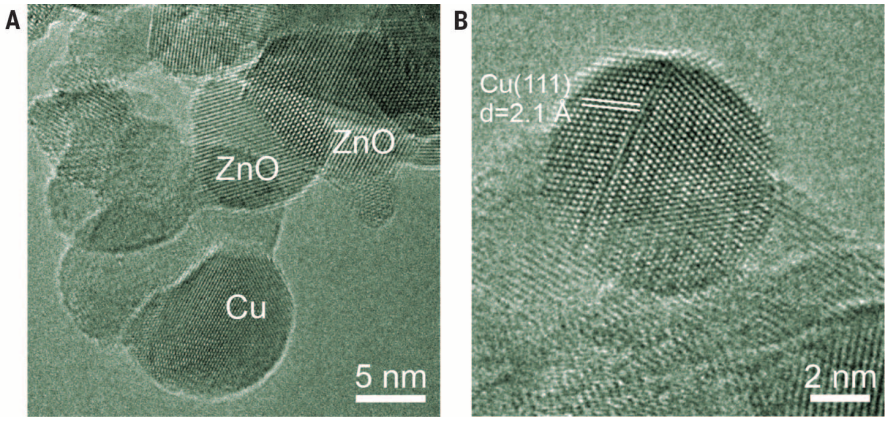


Fig. 1. Electron micrographs of an industrial-type $Cu/ZnO/Al_2O_3$ methanol catalyst acquired in situ during exposure to 1 mbar H_2 at 300°C. (**A**) The image shows the arrangement of nanoparticles in the catalyst. (**B**) The image reveals a Cu NP with surfaces having direct gas accessibility.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

of the CO/CO_2 ratio (Fig. 2A). That is, θ_{Zn} increases steeply for lower CO/CO_2 ratios and more slowly at higher CO/CO_2 ratios, because of the decrease in the Zn segregation energies as a result of Zn-Zn interactions. Creation of additional undercoordinated sites because of bulk defects (*5*) is not considered. Nevertheless, the difference between the coverages of Zn at steps and terraces is not large, as seen from Fig. 2A, and thus, the number of added Zn atoms caused by a small number of additional step sites is not expected to result in substantial changes of the total θ_{Zn}.

Several studies have proposed the presence of ZnO_x layers or Zn attached to oxygenated species partly covering the Cu particle surface in $Cu/ZnO/Al_2O_3$ catalysts (*1*, *4*, *5*, *8*, *12*, *13*, *15*, *16*, *18*, *21*, *22*, *24*, *29*, *33*–*38*). Previously (*11*), we provided strong evidence that the identity of these species is a Cu-Zn surface alloy using XPS, N_2O-RFC (reactive frontal chromatography), H_2-TPD (temperature programmed desorption), and H_2-TA (transient adsorption). Furthermore, Nakamura *et al.* (*13*) found a good correlation between the number of Zn atoms in the Cu surface attached to formate at 90°C and the methanol synthesis activity, strongly indicating that single Zn atoms, possibly with adsorbed species like formate, and not ZnO_x overlayers are the active phase in methanol synthesis, in agreement with our present TEM images.

Then, we note that Zn in the Cu surface must be partly oxidized during the methanol synthesis reaction due to the attachment of formate or other oxygenated reaction intermediates to Zn in the Cu surface (*5*, *18*). Formate is reported theoretically and experimentally to be the most abundant adsorbate in synthesis gas (*5*, *13*, *18*, *35*, *37*). Our experimental work shows that the overall adsorbate coverage at the Cu surface is so low that adsorbates do not affect the Zn stability, and hence coverage, substantially at 220°C. This finding is supported by the theoretical work (*5*, *18*, *35*) (figs. S3 and S4 and tables S7 and S8) suggesting only high coverage of formate at the few step sites and low coverage at the abundant terrace sites of Cu NPs under methanol synthesis conditions. Thus, our experimental work does not rule out that Zn is partly oxidized during the reaction, but the key point for our model is that the experimental results demonstrate that any such effect is sufficiently weak that it does not affect the Zn coverage. Therefore, we find it most probable that the main part of the Zn in Cu is in the reduced state at 220°C and above, in agreement with the theoretical studies.

However, we find high coverages of adsorbates, most probably formate (*13*), at the Cu surface after methanol synthesis at 1 bar and 90°C. Hence, our Zn model does not cover these conditions, but the temperatures in industrial methanol synthesis are generally higher—220° to 280°C—resulting in lower adsorbate coverages. DFT calculations also show that CO will not compete effectively with formate at Cu(211) steps with and without Zn in CO_2-containing synthesis gases (*18*).

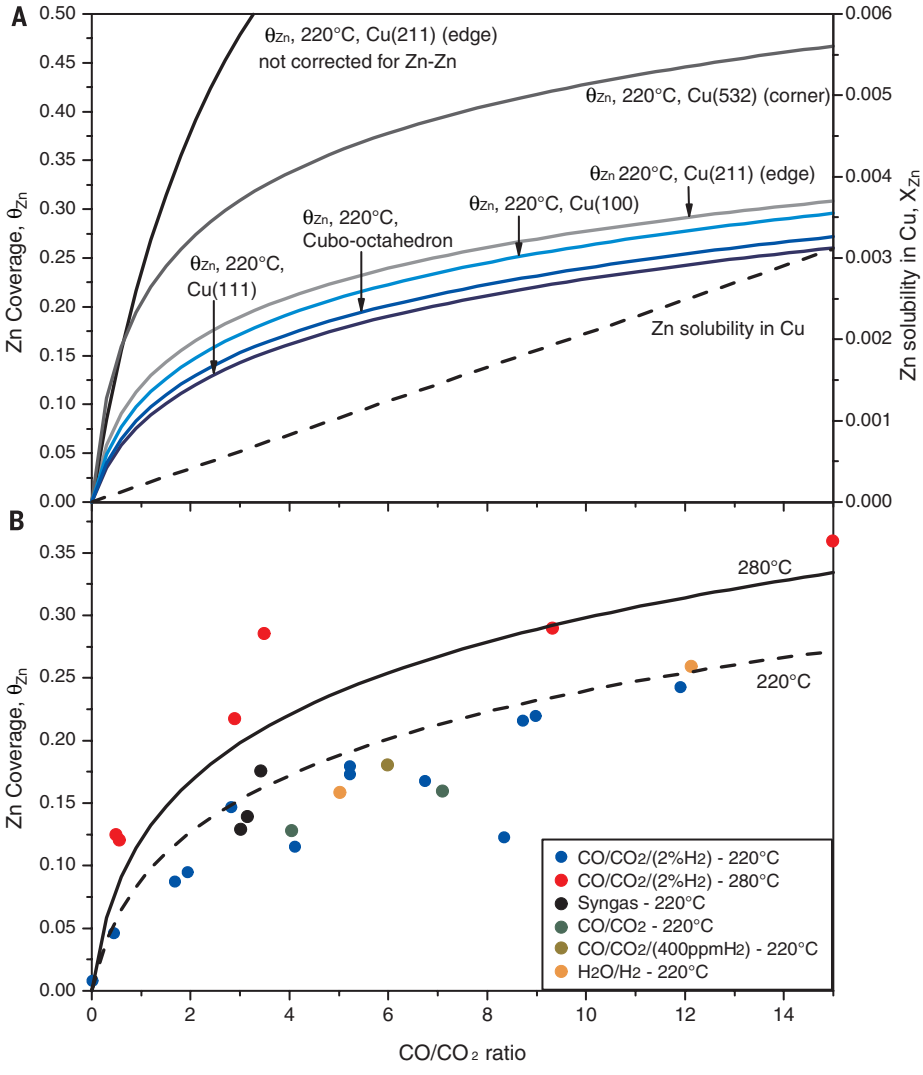


Fig. 2. Modeling and measurements of the Zn coverages at Cu nanoparticles in a $Cu/ZnO/Al_2O_3$ methanol catalyst. θ_{Zn} at specific sites of a Cu surface and on the surface of a Cu cubo-octahedron with 88 Å in diameter and d_{ZnO} = 87 Å (closest to a catalyst with d_{Cu} = 85 Å and d_{ZnO} = 87 Å, as determined by XRD) are calculated using a model based on thermodynamics and DFT. (**A**) Zn solubility (X_{Zn}) in Cu and Zn coverages, including Zn-Zn interactions [apart from one of the curves for edge sites], plotted as a function of the CO/CO_2 ratio at 220°C. (**B**) Zn coverages determined by H_2-TPD compared with those predicted theoretically at 220°C (dashed line) and 280°C (solid line).

In our recent work (*11*), a method for measuring θ_{Zn} in industrial-type methanol catalysts was established by means of N_2O-RFC, H_2-TPD, H_2-TA, and XPS. The results showed that values of θ_{Zn} determined by H_2-TPD are proportional to those determined by the complementary methods, as seen from fig. S5 (*11*). In the present work, the gas-induced Zn decoration of the Cu NPs was only addressed using H_2-TPD. This method was applied because the probe interacts only mildly with the surface (*8*) and provides reliable and reproducible values of θ_{Zn} with high precision (*11*). A series of catalysts were treated in different CO/CO_2 gas mixtures with small amounts of H_2 added and then analyzed by the H_2-TPD method. H_2 was added to mimic a real synthesis gas, but in low concentrations (H_2/CO_x = 0.02) to suppress possible surface-formate and methanol formation, to ensure that reduction of ZnO proceeds via CO rather than via H_2 and to diminish the reverse WGS. With 2% H_2 added, the change in the CO/CO_2 ratio at WGS equilibrium is <0.004 at 220°C and <0.006 at 280°C. In three experiments, H_2/CO_x ratios up to 3.17 were used. For high H_2/CO_x ratios and non-WGS equilibrium, H_2 will at some point dominate the reduction of ZnO(s), and a more complicated analysis than that applied here will be needed. In all experiments, the CO/CO_2 ratios were determined by direct measurements as the average between the inlet and the outlet values.

Experimental series were initiated and terminated by a (standard) H_2-TPD measurement consisting of a reduction in 1% H_2 at 220°C for 5 hours, flushing in He for 1 hour, exposure of the sample to 16 bar of H_2 at –33°C, cooling to –196°C,

Downloaded from http://science.sciencemag.org/ on May 19, 2016

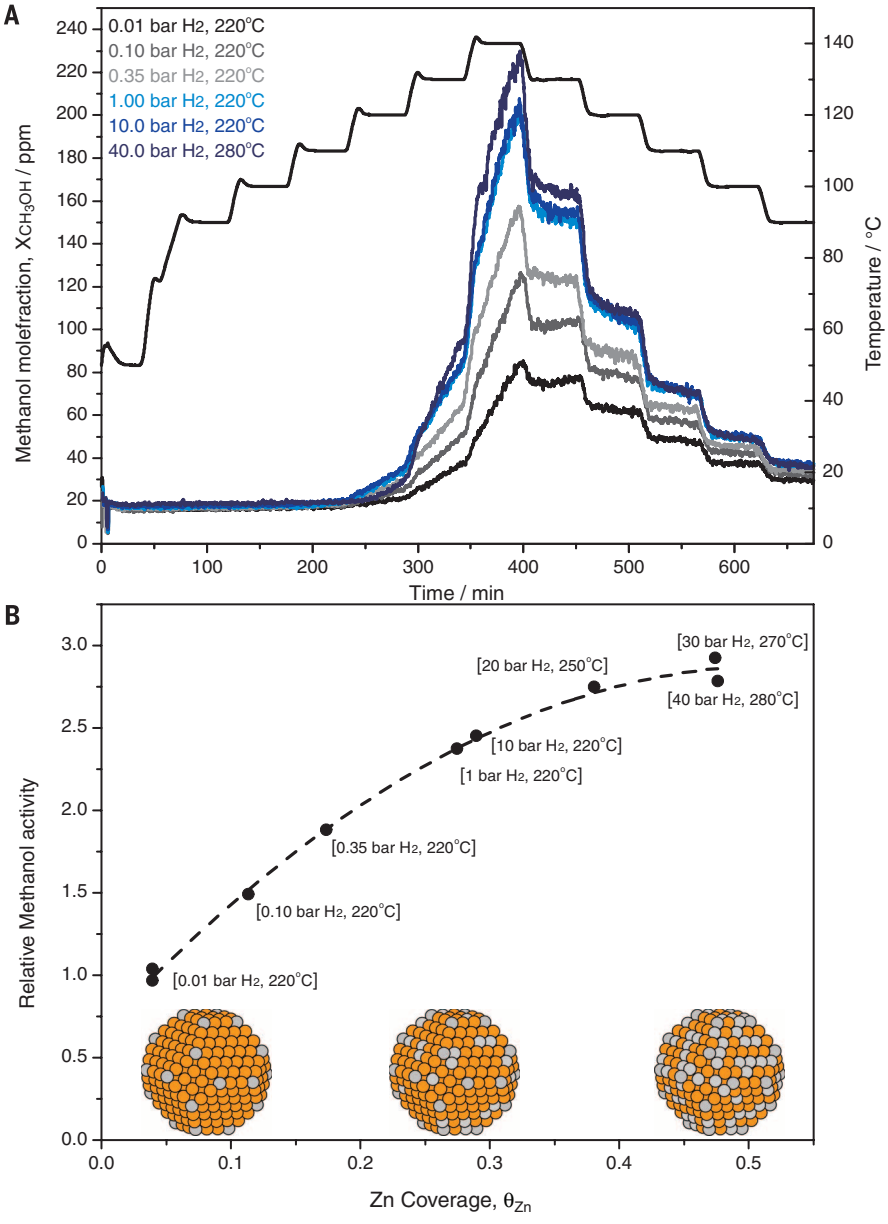


Fig. 3. Methanol activities as a function of the Zn coverages at the Cu nanoparticles of a Cu/ZnO/Al_2O_3 methanol catalyst. (**A**) Temperature ramp and exit methanol concentrations after pretreatment as a function of time and temperature in a $CO/CO_2/H_2$ = 18/18/64 gas mixture at ambient pressure and SV = 30 to 33 Nl/g/h. (**B**) Relative measured methanol exit concentrations at 130°C (temperature ramp-down) as a function of postreaction values of θ_{Zn}. The varying values of θ_{Zn} are obtained by pretreatments in H_2 at different pressures and temperatures prior to activity tests. The dashed line is a second-order polynomial fit to the data.

and heating in He to 210°C. In between these two standard measurements, 5 to 20 hours of pretreatments in different gases followed by H_2-TPD measurements were performed. Hereby, θ_{Zn} was determined using Eq. 7 (fig. S6).

$$\theta_{Zn} = \frac{n_{H_2\text{-}std}/0.96 - n_{H_2\text{-}TPD}}{n_{H_2\text{-}std}/0.96} \quad (7)$$

$n_{H_2\text{-}TPD}$ is the hydrogen desorption capacity after a pretreatment, and $n_{H_2\text{-}std}$ is the weighted average of the hydrogen desorption capacity obtained from the initial and the final standard H_2-TPD measurements, assuming a similar change of surface area in each treatment cycle. θ_{Zn} is corrected for the value of θ_{Zn} = 0.04 that was found after a standard reduction in 1% H_2 at 220°C by XPS (*11*).

Most of the Zn measurements were performed after treatments in $CO/CO_2/H_2$(2%) gas mixtures. The applicability of the CO/CO_2 ratio as descriptor and the establishment of the proposed steady state of the catalyst were addressed by several additional experiments. First, θ_{Zn} was measured after pretreatments in pure CO/CO_2 and H_2O/H_2 gases, in $CO/CO_2/H_2$(400 ppm) (parts per million) gas, and in a synthesis gas ($CO/CO_2/H_2$ = 18/6/76) at 1, 10, and 40 bar. The reduction potential of the H_2O/H_2 gas in terms of a CO/CO_2 ratio was calculated by converting the H_2O/H_2 ratio to a CO/CO_2 ratio using the equilibrium constant for the WGS reaction: $\frac{p_{CO}}{p_{CO_2}} K_{shift} = \frac{p_{H_2}}{p_{H_2O}}$. Any formate present at the Cu/Zn surface would decompose in the He purge flow at 220°C (1 hour) before the H_2-TPD and ZnO_x or Zn-OH species in Cu would be measured as a Zn atom (*11*). The data from these experiments match those obtained with $CO/CO_2/H_2$(2%) mixtures (Fig. 2B), which shows the robustness of the model and the method for determination of θ_{Zn}. Adsorbates at the surfaces of Cu and the oxide support during methanol synthesis could potentially modify the stability of Zn in the surface of Cu, and hence θ_{Zn}. However, the fact that treatments in synthesis gas and in CO/CO_2, H_2O/H_2, $CO/CO_2/H_2$(400 ppm) and $CO/CO_2/H_2$(2%) gas mixtures resulted in similar deviations of θ_{Zn} from the model verifies that our model is also valid at synthesis conditions, even though the model is derived for reduced Zn. Second, variation of space velocity (3 to 30 Nl g^{-1} h^{-1}) did not change θ_{Zn} (<5.4%). Third, experiments showed that the standard 5-hour pretreatment time provides values of θ_{Zn} that are negligibly (<2%) lower than those obtained using pretreatment times of 10 or 20 hours.

The good correspondence between measured and modeled data (SD = 0.037) in Fig. 2B, obtained without any fitting or optimization, and the shapes of the θ_{Zn} versus CO/CO_2 model curves and the experimental data being similar, indicate that the physical principles in the model are sufficient to describe the dynamic dependency of θ_{Zn} on the reduction potential of the methanol synthesis gas for an industrial-type methanol catalyst. The model describes quantitatively the interaction between Cu and ZnO over a wide range of relevant methanol synthesis gas compositions and temperatures and is a function of the nanostructural architecture of the Cu/ZnO/Al_2O_3 catalyst as well.

Next, to determine the promoting effect of Zn in an industrial-type methanol synthesis catalyst, the methanol activity was related experimentally to θ_{Zn}. Specifically, pretreatments in H_2 at different pressures were used as a convenient method to induce different θ_{Zn}, as described in detail previously (*11*). The resulting θ_{Zn} after these treatments was determined by kinetics rather than thermodynamics. The methanol activities were then measured upon heating from room temperature to 140°C (in steps of 10°C in the range 90° to 140° and back to 90°C) at ambient pressure in a $CO/CO_2/H_2$ = 18/18/64 gas mixture and at a space velocity (SV) of 30 to 33 Nl/g/h. Figure 3A shows the measured methanol exit concentrations as a function of temperature. Upon heating from 90°C, activity was absent, and, after passing 110°C, activity increased substantially with time at the different temperature plateaus up to 140°C. A similar induction time was observed by Yang *et al.* (*39*) at 140°C, and

Downloaded from http://science.sciencemag.org/ on May 19, 2016

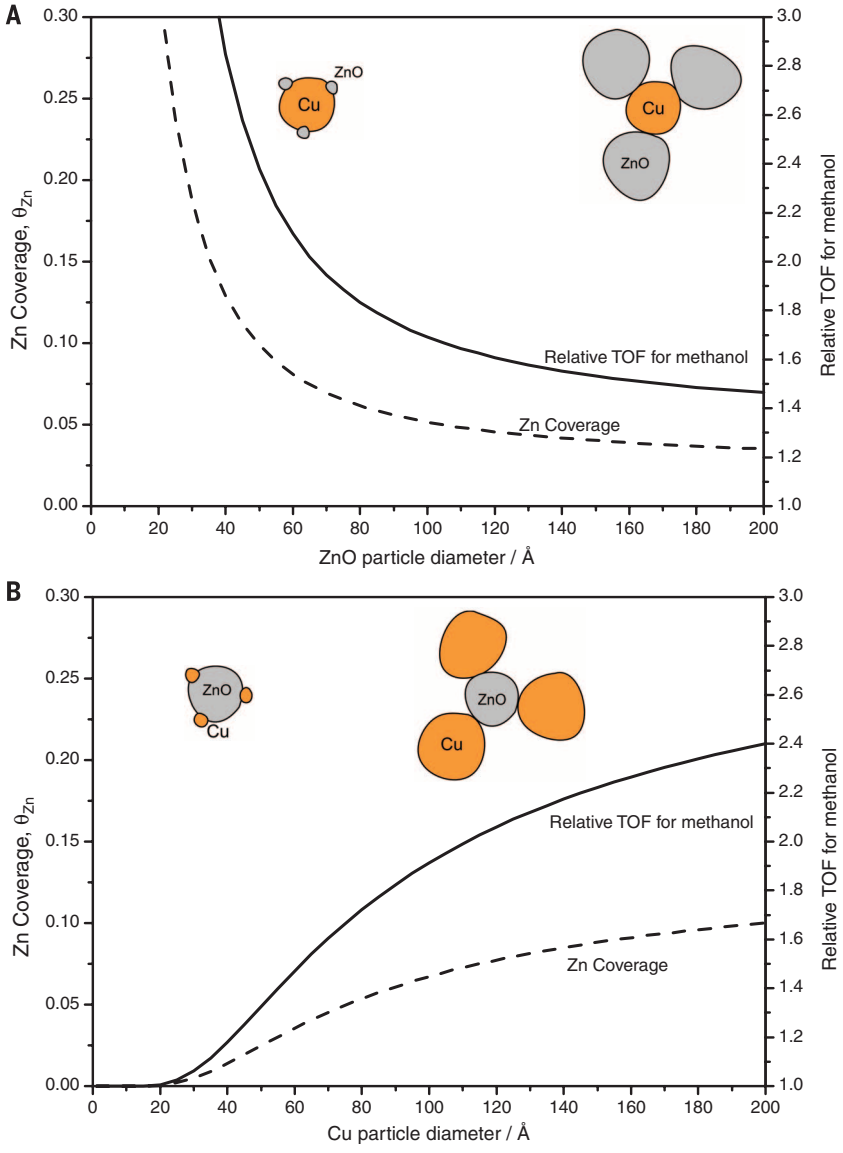


Fig. 4. Modeling of Zn coverages at Cu nanoparticles and relative methanol turnover numbers of a Cu/ZnO/Al_2O_3 catalyst as a function of ZnO and Cu particle sizes. (**A**) θ_{Zn} (dashed line) and $TOF_{MeOH,rel}$ (solid line) relative to that at θ_{Zn} = 0, plotted as a function of the ZnO particle diameter. (**B**) θ_{Zn} (dashed line) and $TOF_{MeOH,rel}$ relative to that at d_{Cu} = 0 (θ_{Zn} = 0) as a function of Cu particle diameter (solid line). The parameters in the model are d_{ZnO} = 87 Å, d_{Cu} = 88 Å, T = 220°C, and CO/CO_2 = 0.5. A fit to the activity data in Fig. 3B was used to estimate $TOF_{MeOH,rel}$ for a given value of θ_{Zn}. The uncertainties in θ_{Zn} and $TOF_{MeOH,rel}$ were estimated to 0.037 and 0.06–0.36 for θ_{Zn} = 0–0.3, based on the deviation of the data and the model in Fig. 2B and on the fit to the data in Fig. 3B.

they ascribed this phenomenon to a lack of water vapor. The most likely explanation seems to be some kind of delay in the accumulation af active species at the surface of the catalyst, which possibly could involve formate (*37*). Nevertheless, the activity stabilized upon the following stepwise cooling of the catalyst. After the activity measurements, θ_{Zn} was measured as described in more detail in the supplementary material (fig. S7). To rank the performance of the catalyst at different θ_{Zn}, the constant activities at 130°C during the cooling sequence, relative to that obtained after reduction in 1% H_2, are plotted as a function of θ_{Zn} in Fig. 3B.

Figure 3B shows that the methanol synthesis activity is strongly dependent on θ_{Zn}. The activity increases by a factor of 3 for θ_{Zn} in the range 0.04 to 0.47 and is fitted well by a second-order polynomial equation. The data indicate that the methanol activity may reach a maximum near $\theta_{Zn} \approx 0.47$. Previously, Nakamura and co-workers related the methanol synthesis to θ_{Zn} physically deposited onto polycrystalline Cu particles and single crystals (*13*, *15*), and they also reported a maximum in the methanol turnover frequency at $\theta_{Zn} \approx 0.20$. The reason for this difference is not fully understood, but as suggested by Nakamura *et al.* (*13*), the loss of activity at higher values of θ_{Zn} resulted from large coverage of ZnO, in agreement with our model suggesting that ZnO is the stable phase of Zn under their conditions. In Cu/ZnO/Al_2O_3 catalysts, part of the Cu area is already covered by ZnO, as seen from Fig. 1A. Thus, θ_{Zn} is a key parameter for understanding the activity of the industrial-type Cu/ZnO/Al_2O_3 methanol catalyst, and Zn-covered Cu sites at Cu facets or steps or both constitute the active site for methanol synthesis. The model enables modification of the Cu-ZnO interaction, hence optimizing the activity through control of the reaction conditions as well as tuning the sizes of Cu and ZnO. The Cu and ZnO size dependencies of θ_{Zn} and the relative turnover frequency of methanol, $TOF_{MeOH,rel}$, were determined from the model and using the second-order polynomial fit to the data in Fig. 3B. The results are depicted in Fig. 4A. Interestingly, the present model predicts that catalysts prepared with small ZnO particles will be very active because of an enhanced spillover of Zn to the Cu surfaces (Fig. 4A); hence, it is the thermodynamic activity of the promoter phase, ZnO, that determines promotion of the methanol synthesis activity. These effects readily explain the findings by Kurtz *et al.* (*16*) and Liao *et al.* (*17*), suggesting that higher methanol synthesis activities are obtained for Cu/ZnO catalysts with nanosized ZnO or high-energy ZnO facets relative to those of catalysts containing larger ZnO particles or low-energy ZnO facets. Similarly, Fichtl *et al.* (*40*) report that deactivation of Cu/ZnO/Al_2O_3 methanol synthesis catalysts cannot be related solely to the loss of Cu surface area, and they observe growth in ZnO crystallite sizes during sintering. Another, and perhaps even more surprising, implication of the model is that θ_{Zn} varies with Cu particle sizes, d_{Cu}. Normally, a catalyst's activity increases as the constituent particle size decreases because of the enlargement of the gas-accessible surface area. In contrast, the model predicts a decrease in θ_{Zn} with decreasing d_{Cu} and hence lower TOF for smaller Cu NPs (Fig. 4B). This effect will tend to reduce the benefits of decreasing the Cu particle sizes in methanol catalysts. To our knowledge, this effect was not observed experimentally so far, presumably because the changes in the methanol activity with Cu particle size for Cu/ZnO catalysts were only studied for d_{Cu} > ~100 Å (*38*, *41*).

An alternative approach to optimizing the performance of Cu-ZnO–based methanol catalysts is to enhance θ_{Zn} in the Cu surface by increasing the reduction potential of the synthesis gas. Interestingly, a peak in the activity is observed for CO/CO_2 ≈ 10, when the reduction potential of a synthesis gas is varied by substituting CO with CO_2 (*4*, *33*). This maximum in activity for high CO/CO_2 ratios is surprising, because isotope labeling shows that carbon atoms in methanol formed over Cu/ZnO catalysts mainly originate from CO_2 (*18*). The present work demonstrates that high CO/CO_2 ratios results in high θ_{Zn} and, thus, high activities. We suggest that a CO/CO_2 ratio of ~10 combines a high θ_{Zn} (~0.25 at 220°C) (see Fig. 2B) and a CO_2 partial pressure of 0.4 to 0.5 bar that ensures an adequate coverage of

Downloaded from http://science.sciencemag.org/ on May 19, 2016

formate (*5*, *18*, *35*) or another key precursor (*37*) at Zn-decorated Cu sites.

Moreover, the present model also consistently explains other dynamical effects reported on methanol catalysts. For example, shape changes observed for Cu NPs on a ZnO support were observed by in situ extended x-ray absorption fine structure and in situ TEM (*6*, *7*, *9*). In the present model, the Zn atoms incorporated into the Cu surface under reducing conditions result in a decrease in the surface energy of Cu, and this will, according to (*7*), contribute to the increased wetting of Cu NPs on ZnO and provides an explanation of transient activity profiles.

REFERENCES AND NOTES

1. R. Schlögl, *Angew. Chem. Int. Ed.* **54**, 3465–3520 (2015).
2. K. Honkala *et al.*, *Science* **307**, 555–558 (2005).
3. G. A. Olah, *Angew. Chem. Int. Ed.* **52**, 104–107 (2013).
4. J. B. Hansen, P. E. Højlund Nielsen, in *Handbook of Heterogeneous Catalysis*, G. Ertl, H. Knözinger, F. Schüth, J. Weitkamp, Eds. (Wiley-VCH, Weinheim, Germany, ed. 2, 2008), pp. 2920–2949.
5. M. Behrens *et al.*, *Science* **336**, 893–897 (2012).
6. P. L. Hansen *et al.*, *Science* **295**, 2053–2055 (2002).
7. P. C. K. Vesborg *et al.*, *J. Catal.* **262**, 65–72 (2009).
8. M. Muhler, E. Törnqvist, L. P. Nielsen, B. S. Clausen, H. Topsøe, *Catal. Lett.* **25**, 1–10 (1994).
9. J.-D. Grunwaldt, A. M. Molenbroek, N.-Y. Topsøe, H. Topsøe, B. S. Clausen, *J. Catal.* **194**, 452–460 (2000).
10. R. Naumann d'Alnoncourt *et al.*, *Phys. Chem. Chem. Phys.* **8**, 1525–1538 (2006).
11. S. Kuld, C. Conradsen, P. G. Moses, I. Chorkendorff, J. Sehested, *Angew. Chem. Int. Ed.* **53**, 5941–5945 (2014).
12. J. Graciani *et al.*, *Science* **345**, 546–550 (2014).
13. J. Nakamura, I. Nakamura, T. Uchijima, T. Watanabe, T. Fujitani, *Stud. Surf. Sci. Catal.* **101B**, 1389–1399 (1996).
14. J. Schumann *et al.*, *ACS Catal.* **5**, 3260–3270 (2015).
15. J. Nakamura, Y. Choi, T. Fujitani, *Top. Catal.* **22**, 277–285 (2003).
16. M. Kurtz *et al.*, *Catal. Lett.* **92**, 49–52 (2004).
17. F. Liao *et al.*, *Angew. Chem. Int. Ed.* **50**, 2162–2165 (2011).
18. F. Studt *et al.*, *Chem. Cat. Chem.* **7**, 1105–1111 (2015).
19. E. D. Batyrev, J. C. van den Heuvel, J. Beckers, W. P. A. Jansen, H. L. Castricum, *J. Catal.* **229**, 136–143 (2005).
20. W. P. A. Jansen *et al.*, *J. Catal.* **210**, 229–236 (2002).
21. T. Lunkenbein, J. Schumann, M. Behrens, R. Schlögl, M. G. Willinger, *Angew. Chem. Int. Ed.* **54**, 4544–4548 (2015).
22. I. Kasatkin, P. Kurr, B. Kniep, A. Trunschke, R. Schlögl, *Angew. Chem. Int. Ed.* **119**, 7465–7468 (2007).
23. M. S. Spencer, *Surf. Sci.* **192**, 336–343 (1987).
24. J. Yoshihara, C. T. Campbell, *J. Catal.* **161**, 776–782 (1996).
25. I. Barin, *Thermodynamic Data of Pure Substances* (Weinheim, Basel, Switzerland; Cambridge, New York, VCH, 1989).
26. E. J. Rapperport, J. P. Pemsler, *Metall. Trans.* **3**, 827–831 (1972).
27. M. S. Spencer, *Surf. Sci.* **192**, 329–335 (1987).
28. R. Ouyang, J.-X. Liu, W.-X. Li, *J. Am. Chem. Soc.* **135**, 1760–1771 (2013).
29. D. B. Rasmussen *et al.*, *J. Catal.* **293**, 205–214 (2012).
30. L. Vitos, A. V. Ruban, H. L. Skriver, J. Kollár, *Surf. Sci.* **411**, 186–202 (1998).
31. M. Sano, T. Adaniya, T. Fujitani, J. Nakamura, *J. Phys. Chem.* **106**, 7627–7633 (2002).
32. R. Van Hardeveld, F. Hartog, *Surf. Sci.* **15**, 189–230 (1969).
33. O. Martin, J. Pérez-Ramírez, *Catal. Sci. Technol.* **3**, 3343–3352 (2013).
34. I. A. Fisher, A. T. Bell, *J. Catal.* **184**, 357–376 (1999).
35. W. J. van Rensburg, M. A. Petersen, M. S. Datt, J.-A. van den Berg, P. van Helden, *Catal. Lett.* **145**, 559–568 (2015).
36. L. C. Grabow, M. Mavrikakis, *ACS Catal.* **1**, 365–384 (2011).
37. Y. Yang, D. Mei, C. H. F. Peden, C. T. Campbell, C. A. Mims, *ACS Catal.* **5**, 7328–7337 (2015).
38. G. C. Chinchen, K. C. Waugh, D. A. Whan, *Appl. Catal.* **25**, 101–107 (1986).
39. Y. Yang, C. A. Mims, D. H. Mei, C. H. F. Peden, C. T. Campbell, *J. Catal.* **298**, 10–17 (2013).
40. M. B. Fichtl *et al.*, *Appl. Catal. A.* **502**, 262–270 (2015).
41. M. Kurtz, H. Wilmer, T. Genger, O. Hinrichsen, M. Muhler, *Catal. Lett.* **86**, 77–80 (2003).

ACKNOWLEDGMENTS

We thank P. G. Moses for fruitful discussions and commenting on the work at an early stage. The Center for Individual Nanoparticle Funtionality is sponsored by the Danish National Research Foundation (DNRF54).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/969/suppl/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 to S8
References (*42*–*49*)

15 December 2015; accepted 15 April 2016
10.1126/science.aaf0718

CATALYSIS

Self-assembly of noble metal monolayers on transition metal carbide nanoparticle catalysts

Sean T. Hunt,[1] Maria Milina,[1] Ana C. Alba-Rubio,[2]* Christopher H. Hendon,[1] James A. Dumesic,[2] Yuriy Román-Leshkov[1]†

We demonstrated the self-assembly of transition metal carbide nanoparticles coated with atomically thin noble metal monolayers by carburizing mixtures of noble metal salts and transition metal oxides encapsulated in removable silica templates. This approach allows for control of the final core-shell architecture, including particle size, monolayer coverage, and heterometallic composition. Carbon-supported $Ti_{0.1}W_{0.9}C$ nanoparticles coated with Pt or bimetallic PtRu monolayers exhibited enhanced resistance to sintering and CO poisoning, achieving an order of magnitude increase in specific activity over commercial catalysts for methanol electrooxidation after 10,000 cycles. These core-shell materials provide a new direction to reduce the loading, enhance the activity, and increase the stability of noble metal catalysts.

Noble metal (NM) catalysts critically enable many existing and emerging technologies, such as catalytic converters (*1*), reforming (*2*), and fuel cells (*3*). However, their scarcity and high cost necessitate the development of catalytic systems with reduced NM loadings, increased activity, and improved durability. In this respect, various nanostructured architectures have been investigated, including atomically dispersed NM catalysts (*4*), hollow nanocages (*5*, *6*), alloyed nanoparticles (NPs) (*7*), and core-shell structures (*8*, *9*). In particular, core-shell NPs composed of an earth-abundant core coated with an atomically thin NM shell are a promising platform that offers both design flexibility and reduced precious-metal loadings. However, achieving independent control over the particle size, core composition, shell composition, and shell thickness poses a substantial challenge (*8*, *9*). State-of-the-art synthetic methods are predominantly limited to a few earth-abundant metallic cores (e.g., Fe, Co, Ni, and Cu) that allow for more precise synthetic control; however, these metal cores form intrinsically metastable core-shell particles that restructure during heating (*10*–*12*) or electrochemical cycling (*13*, *14*).

Early transition metal carbides (TMCs) are earth-abundant ceramics with ideal topochemical properties for supporting precious-metal shells (*15*–*17*). First, TMCs exhibit metallic electrical conductivity, corrosion resistance, and high melting points (*18*). Second, precious metals tend to bind strongly to metal-terminated early TMC surfaces (fig. S1) but cannot readily form stable carbides (*19*). Thus, NMs should coat TMC surfaces but should not alloy with the underlying core. In particular, tungsten carbide (WC) is inexpensive (fig. S2), exhibits a "platinum-like" density of electronic states (*20*, *21*), and its metal-terminated surface forms interfacial Pt-WC bonds that are ~90 kJ mol^{-1} stronger than interfacial Pt-Pt bonds (fig. S1). Although experimental studies on model thin-film systems have corroborated these attractive properties (*15*, *22*, *23*), synthetic efforts to date have not achieved TMC NPs coated with NM monolayers. Until recently, TMC NPs alone were difficult to engineer with controlled properties because their synthesis requires high temperatures (above ~700°C), followed by passivation with dilute oxygen (*24*). Without proper synthetic control, these conditions result in sintered TMC particles covered in both graphitic coke and an oxide surface layer. Such surface impurities preclude core-shell formation because of the unfavorable binding energies between precious metals and contaminated TMC surfaces (fig. S1).

We present here a high-temperature self-assembly method to synthesize size-tunable TMC NPs (<10 nm) coated with monometallic or heterometallic NM surface shells of controlled thicknesses ranging from submonolayer to multilayer coverages. These core-shell materials achieve

[1]Department of Chemical Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. [2]Department of Chemical and Biological Engineering, University of Wisconsin–Madison, Madison, WI 53706, USA. *Present address: Department of Chemical and Environmental Engineering, University of Toledo, Toledo, OH 43606, USA. †Corresponding author. Email: yroman@mit.edu

Downloaded from http://science.sciencemag.org/ on May 19, 2016

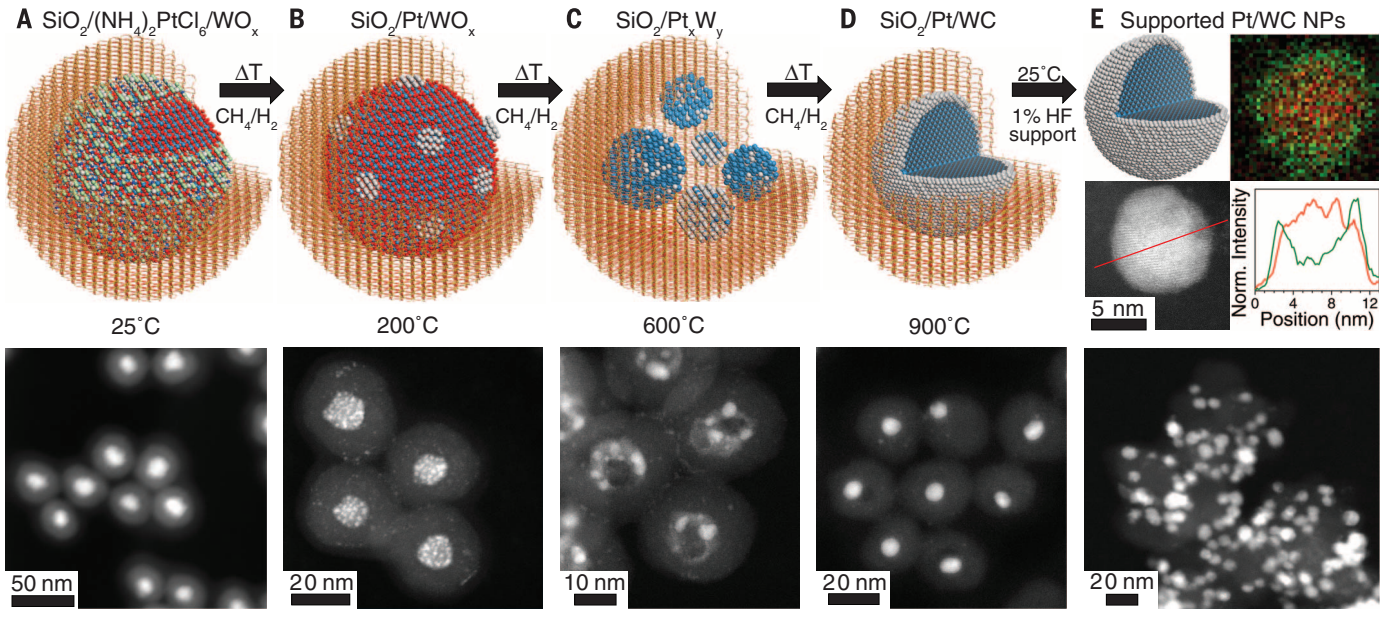


Fig. 1. High-temperature self-assembly of NM monolayers on TMC cores. (**A** to **D**) Schematic representations and corresponding STEM images of (A) silica-encapsulated $(NH_4)_2PtCl_6/WO_x$ NPs synthesized in a one-pot reactor at room temperature and subsequently heated to (B) 200°C, (C) 600°C, and (D) 900°C in a CH_4/H_2 atmosphere. ΔT, temperature change. (**E**) STEM image, EDX map, and line scan (Pt signal in green, W signal in red) of a resulting core-shell Pt/WC NP and a STEM image of Pt/WC formulated on a carbon black support after silica removal.

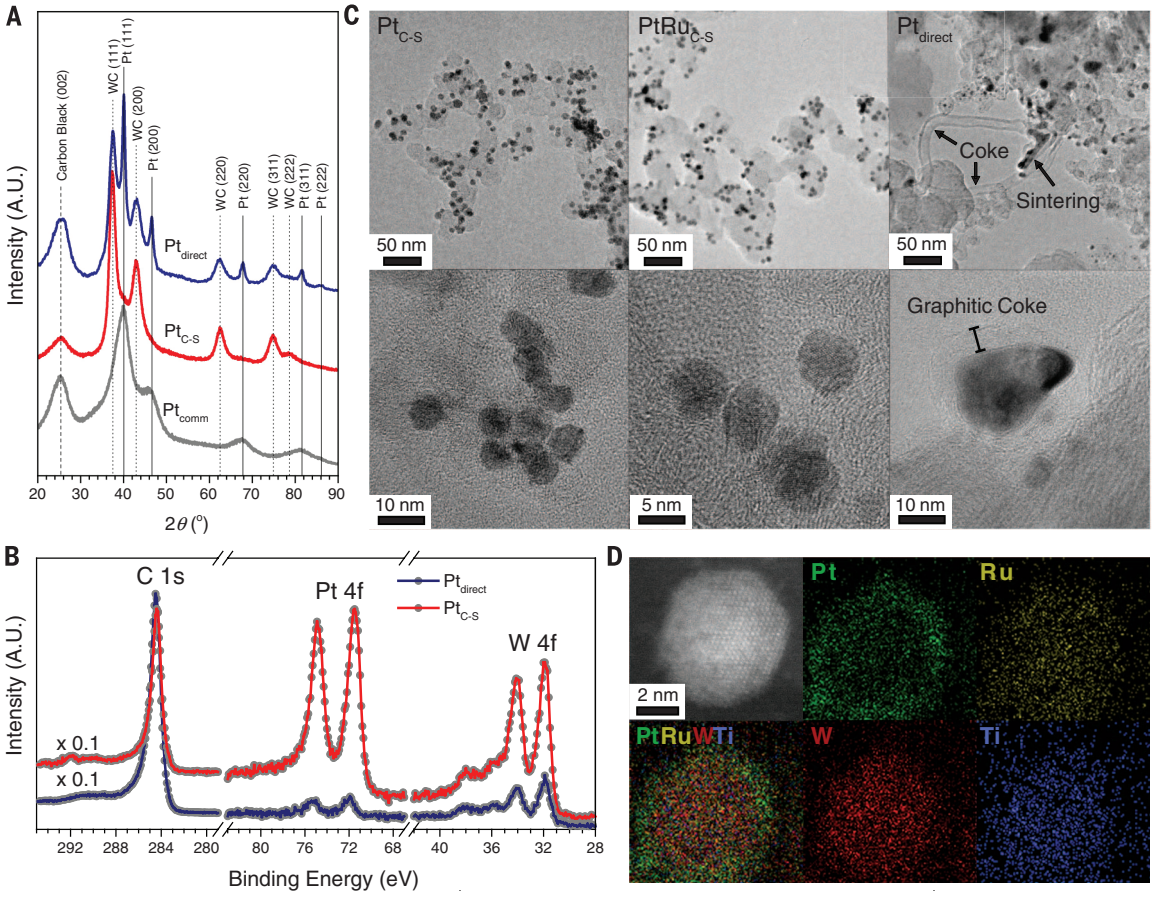


Fig. 2. Experimental corroboration of core-shell structure. (**A**) PXRD diffractograms of Pt_{C-S} and Pt_{direct} compared to Pt_{comm}. A.U., arbitrary units. (**B**) XPS comparison of the C 1s, Pt 4f, and W 4f signals of Pt_{C-S} and Pt_{direct}. (**C**) TEM images of Pt_{C-S}, $PtRu_{C-S}$, and Pt_{direct}. (**D**) STEM and EDX maps of $PtRu_{C-S}$.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

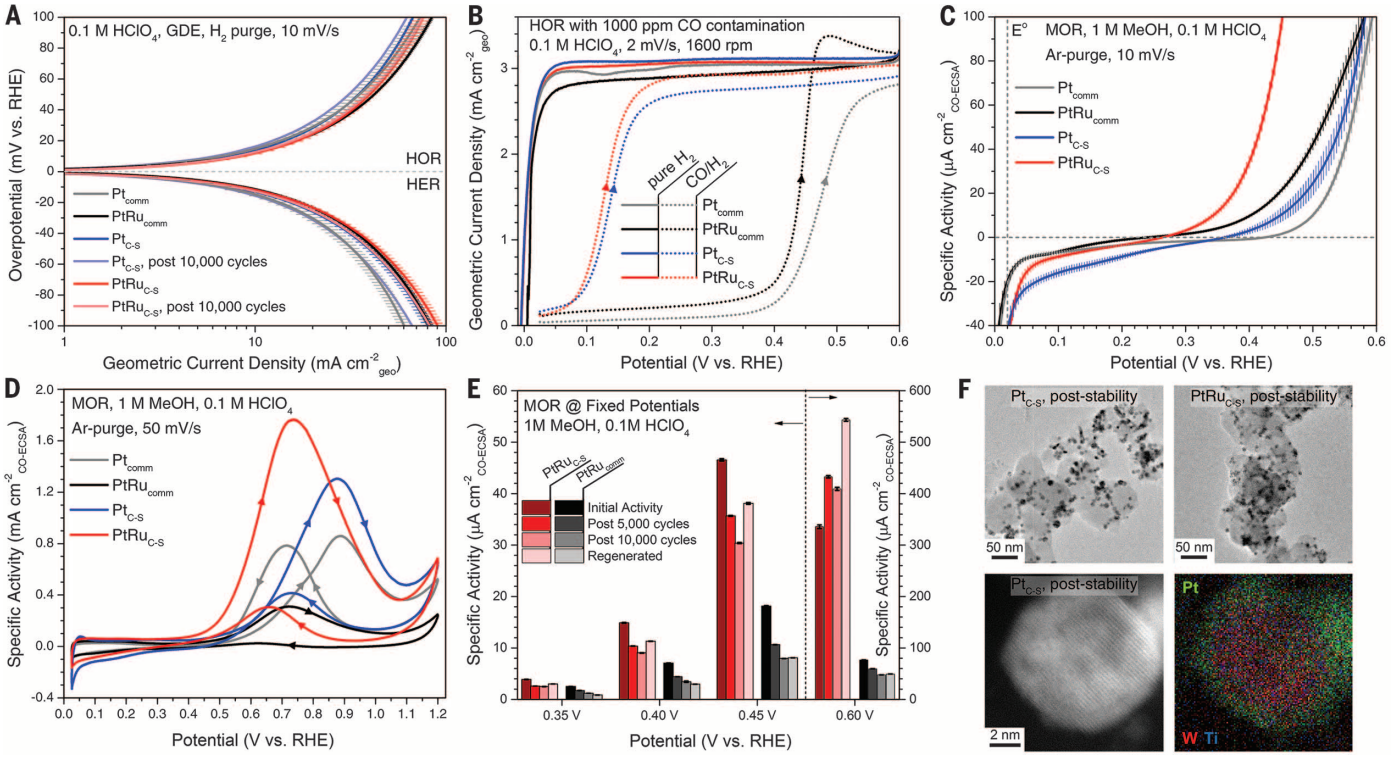


Fig. 3. Electrochemical activity and stability of $Pt_{C\text{-}S}$ and $PtRu_{C\text{-}S}$ compared to commercial catalysts. (**A**) HOR/HER Tafel plots. RHE, reversible hydrogen electrode. (**B**) HOR LSVs with and without CO contamination. (**C**) LSVs and (**D**) CVs for MOR normalized by CO-ECSA roughness factors. (**E**) Steady-state specific activity at fixed potentials after stability cycling and regeneration in alkaline media. (**F**) TEM images of $Pt_{C\text{-}S}$ and $PtRu_{C\text{-}S}$ and a STEM image with EDX map of a $Pt_{C\text{-}S}$ NP after stability cycling.

superior catalytic activity, improved stability, and reduced NM loadings as compared to state-of-the-art commercial catalysts for electrochemical applications. They are also sinter-resistant, and the core-shell structure remains stable at high temperatures under various atmospheres.

Aberration-corrected scanning transmission electron microscopy (STEM) images (Fig. 1) depict the stages of the NM/TMC (shell/core) self-assembly process as a function of temperature, with Pt/WC NP formation as the representative example. In the first step, WO_x NPs were impregnated uniformly with a $(NH_4)_2PtCl_6$ salt and encapsulated in silica nanospheres via a room-temperature reverse microemulsion (RME) (Fig. 1A). This method generated $SiO_2/(NH_4)_2PtCl_6/WO_x$ composites with controlled NP size and Pt loading [see the supporting materials for comprehensive synthetic details (*25*)].

The $SiO_2/(NH_4)_2PtCl_6/WO_x$ composites were then subjected to a temperature ramp under a 15% CH_4/85% H_2 flow (Fig. 1, B to D). At temperatures lower than 200°C, H_2 permeated through the silica nanospheres (*24*), reducing the encapsulated Pt salt into Pt nanoclusters over the WO_x domains (Fig. 1B and fig. S3). By 600°C, the central WO_x NPs underwent reduction, forming metallic mixtures of Pt and W that were trapped within the silica nanospheres (Fig. 1C and figs. S3 and S4). Near 900°C, these small metallic clusters sintered to form single central NPs, while C from methane decomposition intercalated into the W-rich domains, forming WC (Fig. 1D). Because NMs are insoluble in TMC lattices, Pt phase-segregates from the WC domains and wets the central carbide core as an atomically thin shell, resulting in the self-assembly of uniform core-shell NPs as determined by an energy-dispersive x-ray spectroscopy (EDX) map and line scan (Fig. 1E). The silica template was dissolved, and the resulting NPs could then be dispersed in solution with or without a capping agent (fig. S5) or onto a high-surface-area matrix (Fig. 1E). The final Pt/WC particle size and Pt shell thickness were controlled by engineering the WO_x NP size and $(NH_4)_2PtCl_6$ loading in the RME before silica encapsulation and carburization.

Silica encapsulation is critical for controlling core-shell NP formation. In Fig. 2, A to C, we compare two carbon-supported NM/TMC materials, one with silica encapsulation [denoted as $Pt_{C\text{-}S}$, 28%Pt/72%$Ti_{0.1}W_{0.9}C$ supported at 28 weight % (wt%)], and the other without silica encapsulation (denoted as Pt_{direct}, 20%Pt/80% $Ti_{0.1}W_{0.9}C$ supported at 20 wt%). Because TiC is the most electrochemically stable carbide (*26*), a bimetallic TiWC core was used to enhance stability without affecting the WC lattice parameter by more than 1%. The powder x-ray diffraction (PXRD) pattern for $Pt_{C\text{-}S}$ shows reflections consistent with phase-pure face-centered cubic (fcc) WC [powder diffraction file (PDF) no. 00-020-1316] without additional fcc Pt reflections (PDF no. 00-004-0802), whereas the pattern for Pt_{direct} exhibits distinct, sintered fcc Pt crystallites (Fig. 2A). These data are consistent with core-shell formation for $Pt_{C\text{-}S}$ but Pt phase-segregation for Pt_{direct}. In addition, $Pt_{C\text{-}S}$ showed a difference between the bulk and surface Pt:TiW ratios [28 versus 49%, as determined by inductively coupled plasma mass spectrometry (ICP MS) and x-ray photoelectron spectroscopy (XPS), respectively]. This surface ratio enhancement is indicative of Pt monolayers screening a TiW-rich core. In contrast, such surface screening was not observed for Pt_{direct}, where the bulk and surface Pt:TiW ratios were 20 and 18%, respectively.

Silica encapsulation prevented undesirable coking during carburization, as verified by a sixfold lower C-to-metal surface ratio for $Pt_{C\text{-}S}$ as compared to Pt_{direct} (Fig. 2B). Characteristic graphitic coke fibrils and sintered particles encapsulated in 4 to 5 nm of graphitic coke are visible in the TEM images of Pt_{direct} (Fig. 2C). In contrast, $Pt_{C\text{-}S}$ shows well-dispersed crystalline NPs with a uniform particle size distribution (PSD) of 6 to 8 nm and the absence of detectable coke layers. A heterometallic 27% $Pt_{0.67}Ru_{0.33}$/73% $Ti_{0.1}W_{0.9}C$ material (denoted $PtRu_{C\text{-}S}$) was synthesized analogously to $Pt_{C\text{-}S}$ (Fig. 2C and figs. S6 to S8). Although Ru could partition into the carbide core because it is less noble than Pt, STEM-EDX mapping (Fig. 2D) and XPS (fig. S7) indicate a core-shell structure with a Ru-enriched surface.

We compared the electrocatalytic properties of $Pt_{C\text{-}S}$ and $PtRu_{C\text{-}S}$ (8 wt% NM) to 20 wt% carbon-supported commercial (Premetek) electrocatalysts, denoted as Pt_{comm} and $PtRu_{comm}$, using gas diffusion electrodes (GDEs) (Fig. 3A) and rotating disk electrodes (RDEs) (Fig. 3, B to E). Carbon monoxide stripping voltammetry was used to determine the electrochemically active

Downloaded from http://science.sciencemag.org/ on May 19, 2016

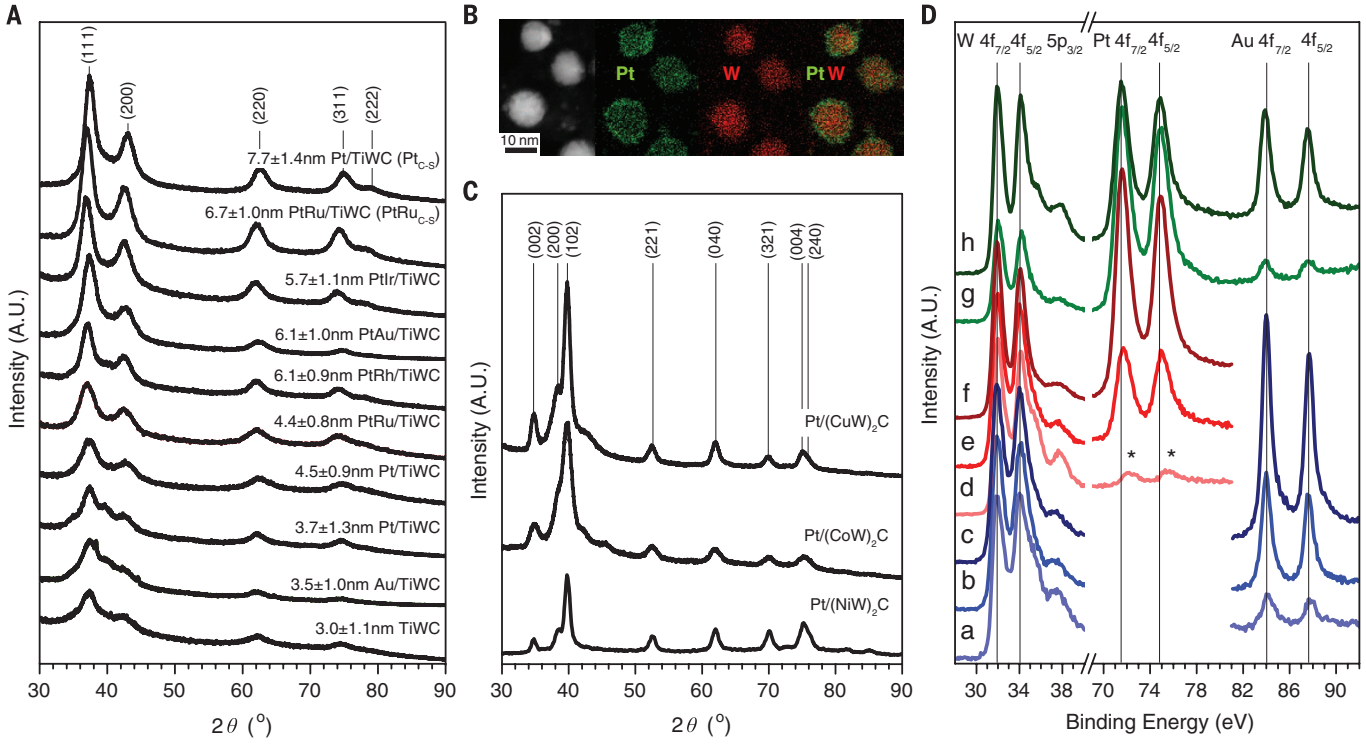


Fig. 4. Experimental exploration of various core-shell NM/TMC architectures. (**A**) PXRD diffractograms of NM/TiWC NPs of various sizes, compositions, and NM coverages. (**B**) STEM image and EDX maps of carbon-supported Pt/(CuW)$_2$C NPs. (**C**) PXRD diffractograms of Pt monolayers on various bimetallic semicarbide core NPs. (**D**) XPS spectra of (a to c) Au/TiWC NPs, (d to f) Pt/TiWC NPs, and (g and h) PtAu/TiWC NPs with submonolayer, monolayer, and multilayer NM shell thicknesses. The asterisks in (D) mark the Pt^{2+} signal.

surface area (CO-ECSA) and roughness factors of all materials (table S1 and fig. S9). Both Pt_{comm} and $PtRu_{comm}$ consist of 1- to 3-nm NPs (fig. S10) and have high CO-ECSAs of 68 ± 6 $m^2\ g^{-1}_{NM}$ and 99 ± 7 $m^2\ g^{-1}_{NM}$, respectively. Despite having larger PSDs (6 to 8 nm) than the commercial samples, both Pt_{C-S} and $PtRu_{C-S}$ achieved comparable CO-ECSAs of 50 ± 2 $m^2\ g^{-1}_{NM}$ and 73 ± 2 $m^2\ g^{-1}_{NM}$, respectively. Pt_{direct} did not exhibit a measureable CO-ECSA (fig. S9) and behaved analogously to carbon during cycling from 0.025 to 1 V (fig. S11).

The reactivity and stability of Pt_{C-S} and $PtRu_{C-S}$ were characterized by using density functional theory (DFT) and various probe reactions, including hydrogen evolution (HER), hydrogen oxidation (HOR), HOR under CO contamination, and methanol electrooxidation (MOR). Both Pt_{C-S} and $PtRu_{C-S}$ exhibited improved specific activity for HER and HOR, indicating that TiWC cores are excellent supports for NM monolayers and can favorably alter catalytic activity. Despite a 60% reduction in NM loading, Pt_{C-S} and $PtRu_{C-S}$ exhibited a symmetric activity profile during HER and HOR linear sweep voltammetry (LSV), similar to that for the commercial catalysts (Fig. 3A). Both core-shell materials exhibited HER and HOR Tafel slopes of ~30 mV dec^{-1} even after 10,000 cycles between −50 and 600 mV (table S2 and figs. S12 and S13). At an HER overpotential of 50 mV, both core-shell materials showed a fourfold improvement in specific activity and a threefold improvement in mass activity over the commercial catalysts (table S3) and maintained this enhancement after cycling.

Enhanced catalytic activity was corroborated by DFT calculations for thermally equilibrated Pt/TiWC slabs. Specifically, our calculations indicate Fermi level matching with Pt, which translates to minimal alterations to the work function of surface Pt by subsurface TiWC (*27*) (figs. S14 and S15), but show that the d-band center for a two-monolayer Pt shell downshifts from −2.7 to −2.8 eV (fig. S16). This downshift provides access to potentially favorable surface properties, such as a substantial weakening of the CO binding energy by ~40 kJ mol^{-1} (*28*). In turn, this effect can increase the tolerance of Pt_{C-S} and $PtRu_{C-S}$ toward CO poisoning, thereby overcoming a major challenge afflicting the performance of industrial Pt-based catalysts in many electrochemical applications (*11*, *28*).

Indeed, HOR experiments performed in the presence of high CO concentrations confirm that the TiWC cores mitigate poisoning on Pt and PtRu monolayers. Specifically, although 1000 parts per million of CO contamination markedly increased the HOR onset potential by ~400 mV for $PtRu_{comm}$ (Fig. 3B) and by ~200 mV for a state-of-the-art Pt/PtSn catalyst (*29*), both Pt_{C-S} and $PtRu_{C-S}$ catalyzed HOR with an overpotential as low as 50 mV. Under pure CO, both Pt_{C-S} and $PtRu_{C-S}$ achieved a ~200 mV lower onset potential and a 30-fold enhancement in specific activity for CO electrooxidation at 400 mV when compared to the commercial catalysts (fig. S17).

By decreasing the CO binding strength, the TiWC core is responsible for enhancing the MOR kinetics observed for Pt_{C-S} as compared to Pt_{comm}. The former features a higher specific activity over a wide potential window and a ~100 mV lower onset potential than Pt_{comm} (Fig. 3, C and D, and fig. S18). Ru is known to improve the MOR performance of Pt at low overpotentials via a bifunctional mechanism (*30*). Despite extensive efforts in the literature, little improvement in the activity and durability of MOR catalysts has been demonstrated over commercial PtRu/C in acidic media (*31*). Whereas both $PtRu_{C-S}$ and $PtRu_{comm}$ achieve a low onset potential of ~250 mV, $PtRu_{C-S}$ displays a higher steady-state turnover frequency (TOF) of 15.9 min^{-1} at 0.6 V as compared to 3.6 min^{-1} using $PtRu_{comm}$ (table S4), further evidencing the favorable activity modulation achievable using a TiWC core.

$PtRu_{C-S}$ also demonstrates enhanced stability as compared to $PtRu_{comm}$ (Fig. 3E and figs. S19 and S20). After 10,000 cycles, $PtRu_{comm}$ lost more than 50% of its steady-state activity at 0.35, 0.40, and 0.45 V, whereas the $PtRu_{C-S}$ activity decreased by only 35% at these potentials and actually improved at 0.6 V. A simple 2-min alkaline dip partially regenerated the activity of $PtRu_{C-S}$ at all potentials but had no appreciable benefit for $PtRu_{comm}$. Specifically, after regeneration, the overall loss in activity at 0.45 V was 18 and 55% for $PtRu_{C-S}$ and $PtRu_{comm}$, respectively. The final TOF at 0.6 V after 10,000 cycles and regeneration was 25.7 min^{-1} for $PtRu_{C-S}$, 2.4 min^{-1} for $PtRu_{comm}$, 1.8 min^{-1} for Pt_{C-S}, and 1.1 min^{-1} for Pt_{comm} (table S4), which represents an order of magnitude improvement of our core-shell material over the commercial catalysts.

No appreciable deactivation via particle sintering was observed for the core-shell materials after

Downloaded from http://science.sciencemag.org/ on May 19, 2016

10,000 cycles (Fig. 3F). HR-STEM and EDX mapping of $Pt_{C\text{-}S}$ after stability cycling show a highly crystalline composite NP with an intact Pt shell and a well-alloyed TiWC core (Fig. 3F and fig. S21). We attribute the improved stability of the core-shell materials to the strong NM-TMC interfacial bonds and to the lower surface free energies of large NPs relative to the surface free energies of ultrasmall NPs. These results demonstrate the ability of our method to reformulate a classic bimetallic NM catalyst, such as PtRu, into an architecture that preserves the complexity of the original bimetallic surface chemistry while favorably modulating catalytic performance through subsurface strain and ligand effects. As a class of materials that breaks the traditional metal-adsorbate scaling relations for transition metals (*32*), TMCs not only serve as promising core candidates to reduce NM loadings but could favorably affect catalytic activity for industrially relevant reactions, such as HOR, MOR, water-gas shift, or methanol synthesis, in which CO coverage effects deeply influence catalyst performance.

The high-temperature self-assembly process used here permits comprehensive control of the entire core-shell architecture for a variety of early and late transition metals (Fig. 4). Using TiWC cores, we synthesized NPs of controlled sizes (3 to 10 nm) with mono- and bimetallic shell compositions using mixtures of Ru, Rh, Ir, Pt, and Au (Fig. 4A and figs. S22 to S24). All materials crystallized into fcc WC lattices and displayed enhanced surface NM:TiW ratios, consistent with core-shell structures (figs. S25 to S27 and table S5). In addition, NM shells lattice-matched and self-assembled onto bimetallic semicarbide cores (PDF no. 00-020-1315) such as $(Cu_{0.2}W_{0.8})_2C$, $(Co_{0.2}W_{0.8})_2C$, and $(Ni_{0.3}W_{0.7})_2C$ (Fig. 4, B and C, and figs. S28 and S29). Although NM interfacial bond formation and stability were not explored on all possible TMC core formulations, our data suggest broader applicability of the method for synthesizing a variety of NM/TMC combinations.

For Au, Pt, and PtAu monolayers self-assembled on TiWC cores, the RME method also allowed control of the NM shell thickness from submonolayer [~0.5 monolayer (ML)] to multilayer (~3 ML) coverages (Fig. 4D). For each material, the extent of the XPS-determined surface NM:TiW ratio enhancement over the ICP-determined bulk NM:TiW ratio correlated with the monolayer coverage, ranging from 1 to 3% at submonolayer coverages to 10 to 20% screening at multilayer coverages (table S5). Unlike Au surfaces, the surface of Pt passivates under ambient conditions with a PtO layer, which is detectable as Pt^{2+} with XPS. As the monolayer coverage decreased for the Pt/TiWC system (Fig. 4D, spectra f to d), the Pt 4f signals shifted to higher binding energies, reaching 72.3 and 75.7 eV for the submonolayer sample (denoted as $Pt_{sub\text{-}ML}$). The presence of only PtO is consistent with submonolayer coverage, and such materials could have important applications in thermal catalysis, where both WC and Pt surface functionalities are accessible for catalytic transformations. When $Pt_{sub\text{-}ML}$ was supported on carbon and heated to 400° and 600°C in different atmospheres (H_2, dry N_2, or H_2O-saturated N_2), neither sintering nor discrete fcc Pt crystallites were detectable when using TEM and PXRD, and an enriched Pt:W ratio showing only Pt^{2+} surface species was observed with XPS (figs. S30 to S33). Collectively, TMC NPs coated with NM monolayers offer new, highly tunable pathways for decreasing NM loading requirements while increasing activity and stability in thermo- and electrocatalysis.

REFERENCES AND NOTES

1. A. T. Bell, *Science* **299**, 1688–1691 (2003).
2. G. A. Deluga, J. R. Salge, L. D. Schmidt, X. E. Verykios, *Science* **303**, 993–997 (2004).
3. J. Zhang, K. Sasaki, E. Sutter, R. R. Adzic, *Science* **315**, 220–222 (2007).
4. Y. Zhai *et al.*, *Science* **329**, 1633–1636 (2010).
5. L. Zhang *et al.*, *Science* **349**, 412–416 (2015).
6. C. Chen *et al.*, *Science* **343**, 1339–1343 (2014).
7. J. Greeley *et al.*, *Nat. Chem.* **1**, 552–556 (2009).
8. P. Strasser *et al.*, *Nat. Chem.* **2**, 454–460 (2010).
9. H. Yang, *Angew. Chem. Int. Ed.* **50**, 2674–2676 (2011).
10. F. Tao *et al.*, *Science* **322**, 932–934 (2008).
11. S. Alayoglu, A. U. Nilekar, M. Mavrikakis, B. Eichhorn, *Nat. Mater.* **7**, 333–338 (2008).
12. C. S. Bonifacio *et al.*, *Chem. Mater.* **27**, 6960–6968 (2015).
13. C. Cui, L. Gan, M. Heggen, S. Rudi, P. Strasser, *Nat. Mater.* **12**, 765–771 (2013).
14. C. Wang *et al.*, *Adv. Funct. Mater.* **21**, 147–152 (2011).
15. D. V. Esposito, J. G. Chen, *Energy Environ. Sci.* **4**, 3900–3912 (2011).
16. I. E. Stephens, A. S. Bondarenko, U. Grønbjerg, J. Rossmeisl, I. Chorkendorff, *Energy Environ. Sci.* **5**, 6744–6762 (2012).
17. J. A. Schaidle, N. M. Schweitzer, O. T. Ajenifujah, L. T. Thompson, *J. Catal.* **289**, 210–217 (2012).
18. S. T. Oyama, *The Chemistry of Transition Metal Carbides and Nitrides* (Blackie, Glasgow, 1996).
19. S. Ono, T. Kikegawa, Y. Ohishi, *Solid State Commun.* **133**, 55–59 (2005).
20. R. B. Levy, M. Boudart, *Science* **181**, 547–549 (1973).
21. L. H. Bennett, J. R. Cuthill, A. J. McAlister, N. E. Erickson, R. E. Watson, *Science* **184**, 563–565 (1974).
22. D. V. Esposito, S. T. Hunt, Y. C. Kimmel, J. G. Chen, *J. Am. Chem. Soc.* **134**, 3025–3033 (2012).
23. T. G. Kelly, A. L. Stottlemyer, H. Ren, J. G. Chen, *J. Phys. Chem. C* **115**, 6644–6650 (2011).
24. S. T. Hunt, T. Nimmanwudipong, Y. Román-Leshkov, *Angew. Chem. Int. Ed.* **53**, 5131–5136 (2014).
25. See the supporting materials on *Science* Online.
26. Y. C. Kimmel, X. Xu, W. Yu, X. Yang, J. G. Chen, *ACS Catal.* **4**, 1558–1562 (2014).
27. W. Schottky, *Annal. Physik* **362**, 541–567 (1918).
28. B. Hammer, Y. Morikawa, J. K. Nørskov, *Phys. Rev. Lett.* **76**, 2141–2144 (1996).
29. Z. Liu, G. S. Jackson, B. W. Eichhorn, *Angew. Chem. Int. Ed.* **49**, 3173–3176 (2010).
30. T. J. Schmidt, H. A. Gasteiger, R. J. Behm, *Electrochem. Commun.* **1**, 1–4 (1999).
31. A. K. Singh, Q. Xu, *ChemCatChem* **5**, 652–676 (2013).
32. R. Michalsky, Y. J. Zhang, A. J. Medford, A. A. Peterson, *J. Phys. Chem. C* **118**, 13026–13034 (2014).

ACKNOWLEDGMENTS

Work at MIT was supported by the U.S. Department of Energy, Office of Basic Energy Sciences (grant DE-FG02-12ER16352). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation (NSF) (grant ACI-1053575). The authors acknowledge the use of facilities and instrumentation supported by the University of Wisconsin Materials Research Science and Engineering Center (grant DMR-1121288). S.T.H. thanks NSF for financial support through the NSF Graduate Research Fellowship under grant no. 1122374. M.M. thanks the Swiss National Science Foundation (project number P2EZP2_159124) for financial support. J.A.D. acknowledges funding by the U.S. Department of Energy, Office of Basic Energy Sciences (grant DE-SC0014058).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/974/suppl/DC1
Materials and Methods
Figs. S1 to S33
Tables S1 to S5
References (*33*–*61*)

10 November 2015; accepted 15 April 2016
10.1126/science.aad8471

ROBOTICS

Perching and takeoff of a robotic insect on overhangs using switchable electrostatic adhesion

M. A. Graule,[1,2]* P. Chirarattananon,[2,3] S. B. Fuller,[2,4] N. T. Jafferis,[2] K. Y. Ma,[2] M. Spenko,[5] R. Kornbluh,[6] R. J. Wood[2]*

For aerial robots, maintaining a high vantage point for an extended time is crucial in many applications. However, available on-board power and mechanical fatigue constrain their flight time, especially for smaller, battery-powered aircraft. Perching on elevated structures is a biologically inspired approach to overcome these limitations. Previous perching robots have required specific material properties for the landing sites, such as surface asperities for spines, or ferromagnetism. We describe a switchable electroadhesive that enables controlled perching and detachment on nearly any material while requiring approximately three orders of magnitude less power than required to sustain flight. These electroadhesives are designed, characterized, and used to demonstrate a flying robotic insect able to robustly perch on a wide range of materials, including glass, wood, and a natural leaf.

Micro aerial vehicles (MAVs) with the capability to stay aloft for a prolonged time would be invaluable in many applications: providing a bird's-eye view of a disaster area, detecting hazardous chemical or biological agents, or enabling secure signal transmission in ad hoc communication networks. However, the flight time of aerial robots is restricted by the weight of their on-board power supplies and the lifetime of their mechanical components. Moreover, the endurance of current aerial robots decreases substantially as vehicle scale

Downloaded from http://science.sciencemag.org/ on May 19, 2016

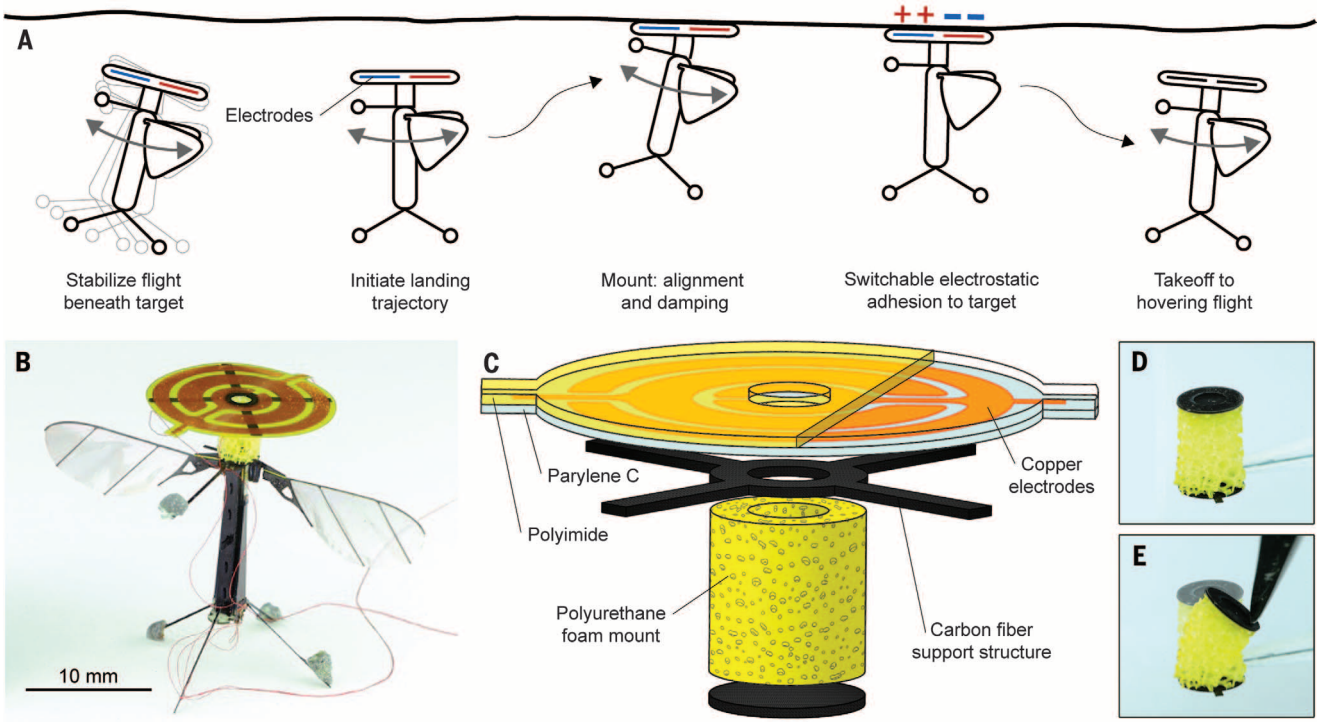


Fig. 1. Robot design and principle of operation. (**A**) Before initiating a perching maneuver, the robot attains stable hovering underneath the target surface. A compliant mount assures successful alignment between the adhesive patch and the target surface. Upon contact, the electrodes in the patch induce surface charges on the substrate, leading to an electrostatic attraction between the surface and the patch. These surface charges recombine when the voltage between the electrodes is switched off, allowing for a smooth detachment. (**B**) Depiction of the flapping wing MAV capable of landing on and relaunching from the underside of nearly any material. (**C**) The adhesion mechanism relies on compliant circular copper electrodes on a polyimide film that are embedded in Parylene C to generate electrostatic adhesion. This is supported by a carbon fiber cross and integrated with the robot through a polyurethane foam mount. The final version of the mechanism weighs 13.4 mg (robot without payload, 84mg). (**D** and **E**) The polyurethane foam mount provides damping and passive alignment to facilitate perching over a wide envelope of trajectories and allows us to integrate features to assist with characterizing the flight performance before perching experiments (fig. S11).

diminishes (*1*). Perching—defined in this case to mean alighting onto a surface or object and remaining attached—on commonly available overhangs such as trees, buildings, or powerlines would allow MAVs to continue their mission while conserving energy, thus expanding their mission time. Hanging underneath these structures would provide a clear path to the ground for vision or signal transmission and protection from extreme weather conditions.

Whereas birds, bats, and insects are capable of perching on compliant and wind-buffeted surfaces such as tree branches, leaves, and flowers (*2*–*4*), it has proven challenging to reproduce this aerial prowess to dynamically land MAVs on and relaunch from a broad range of natural and artificial targets. Demonstrations to date predominantly focus on bird-sized vehicles. Proposed approaches include a passive biomimetic gripper (*5*), directional dry adhesives in a spring-lever system (*6*), magnets with servo-actuated release (*7*), an articulated nondirectional dry adhesive that is repositioned when perching is desired (*8*), and microspines (*9*) or needles (*10*) driven into a soft target by a preloaded snapping mechanism. Perching and relaunch with a high success rate was achieved on rough walls with microspines that engage in surface asperities, which can be released through resistively heating a shape-memory alloy actuator (*11*). Additionally, advanced control strategies were developed for fixed-wing MAVs to land on specific targets via a high angle-of-attack stalling maneuver (*12*) or transitioning into hovering flight (*13*).

Insect-like aerial robots may exceed the agility of larger systems at lower costs, but their deployment comes with a number of additional challenges. For example, perching with articulated and actuated mechanisms becomes impractical at insect scales because of challenges in manufacturing and the destabilizing effect of asymmetrically moving parts on vehicle dynamics. Chemical adhesives may not require complex mechanisms for attachment but are either irreversible, require pressure for engagement, or cause destabilizing torques during detachment (*8*, *14*). An alternative is the use of electrostatic adhesives, which have been used in mechanically simple, low-power attachment mechanisms across a broad range of substrates (*15*, *16*) and were successfully implemented in decimeter-scale climbing robots for vertical walls (*17*–*20*). Although recent advances have combined this form of adhesion with gecko-inspired dry adhesives for improved performance over a variety of substrates (*21*), electroadhesives are not typically as strong as, for example, pressure-sensitive or thermoplastic adhesives (*22*). However, because a robot's surface area–to-volume ratio increases with decreasing size, electroadhesion becomes increasingly promising for smaller devices, and its electrical and mechanical simplicity make it particularly attractive for insect-sized vehicles.

We leveraged these prior insights regarding electrostatic adhesion and propose a controllable attachment principle for small-scale MAV perching. Our method enables repeatable transitions from flight to perching, as well as transitions from attachment to stable hovering flight on overhanging surfaces consisting of wood, glass, or a natural leaf (Fig. 1A) with a tethered insect-scale flapping wing robot (Fig. 1B) (*23*). The method relies on the electrostatic force between interdigitated circular electrodes (voltage difference, 1000 V) (Fig. 1C) and the opposing surface charges they induce on the target substrate. The electrodes are integrated with the robot through a foam mount, which provides damping and passive alignment so as to ensure reliable attachment for a broad range of landing

[1]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]John A. Paulson School of Engineering and Applied Sciences and the Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA 02138, USA. [3]Department of Mechanical and Biomedical Engineering, City University of Hong Kong, Kowloon, Hong Kong. [4]Department of Mechanical Engineering, University of Washington, Seattle, WA 98195, USA. [5]Mechanical, Materials, and Aerospace Engineering Department, Illinois Institute of Technology, Chicago, IL 60616, USA. [6]SRI International, Menlo Park, CA 94025, USA.
***Corresponding author. Email: graule@mit.edu (M.A.G.); rjwood@seas.harvard.edu (R.J.W.)**

Downloaded from http://science.sciencemag.org/ on May 19, 2016

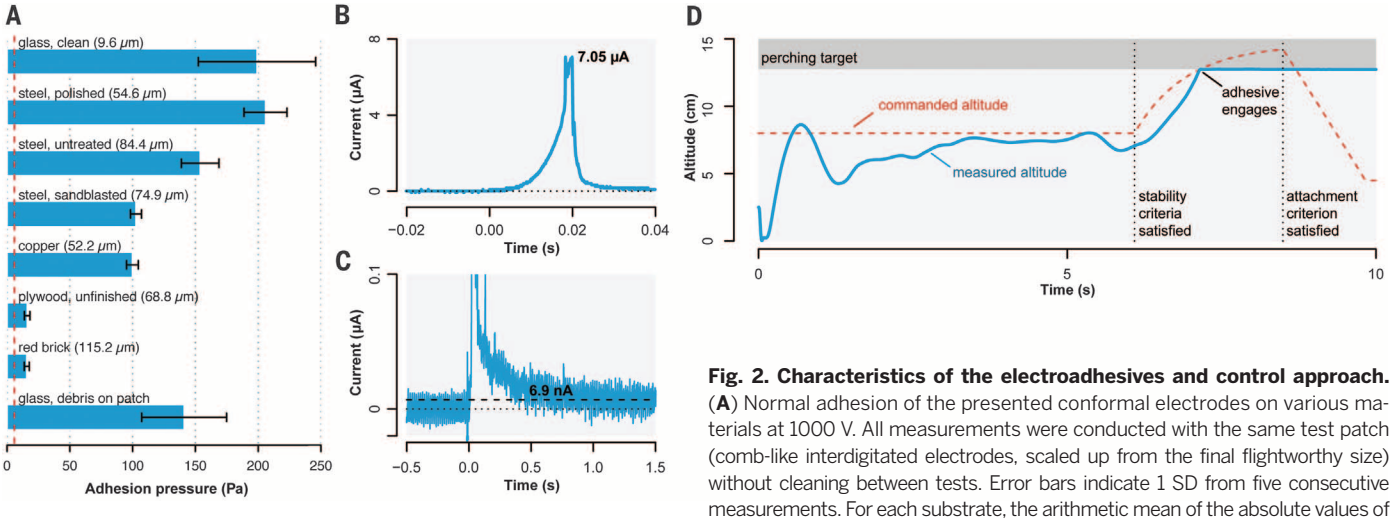


Fig. 2. Characteristics of the electroadhesives and control approach. (**A**) Normal adhesion of the presented conformal electrodes on various materials at 1000 V. All measurements were conducted with the same test patch (comb-like interdigitated electrodes, scaled up from the final flightworthy size) without cleaning between tests. Error bars indicate 1 SD from five consecutive measurements. For each substrate, the arithmetic mean of the absolute values of the surface asperities is stated in parentheses. The dashed red line indicates the pressure required to support the vehicle weight (including the adhesion mechanism), assuming a patch size as used in our demonstrations. (**B**) Charging current for the circular patch on glass at 1000 V [switched on at time (t) = 0 s]. Charging requires 37.7 μJ on glass, 37.4 μJ on wood, and 46.6 μJ on copper. (**C**) Current in the charging phase and during steady state on glass (1000 V). A leakage current of 6.9 nA occurs on glass (1.4 nA on wood; 1.0 nA on copper). The mechanism requires 6.9 μW to remain perched on glass (1.4 μW on wood; 1.0 μW on copper). This is substantially lower than the flight power of 19 mW (*23*). (**D**) Commanded and measured altitude during an exemplary perching flight. The logic module initiates a bio-inspired landing trajectory once the robot has achieved stable hovering in the target region underneath the substrate. The flapping amplitude is ramped down (corresponding to a decrease in commanded altitude) once the logic module detects a successful attachment.

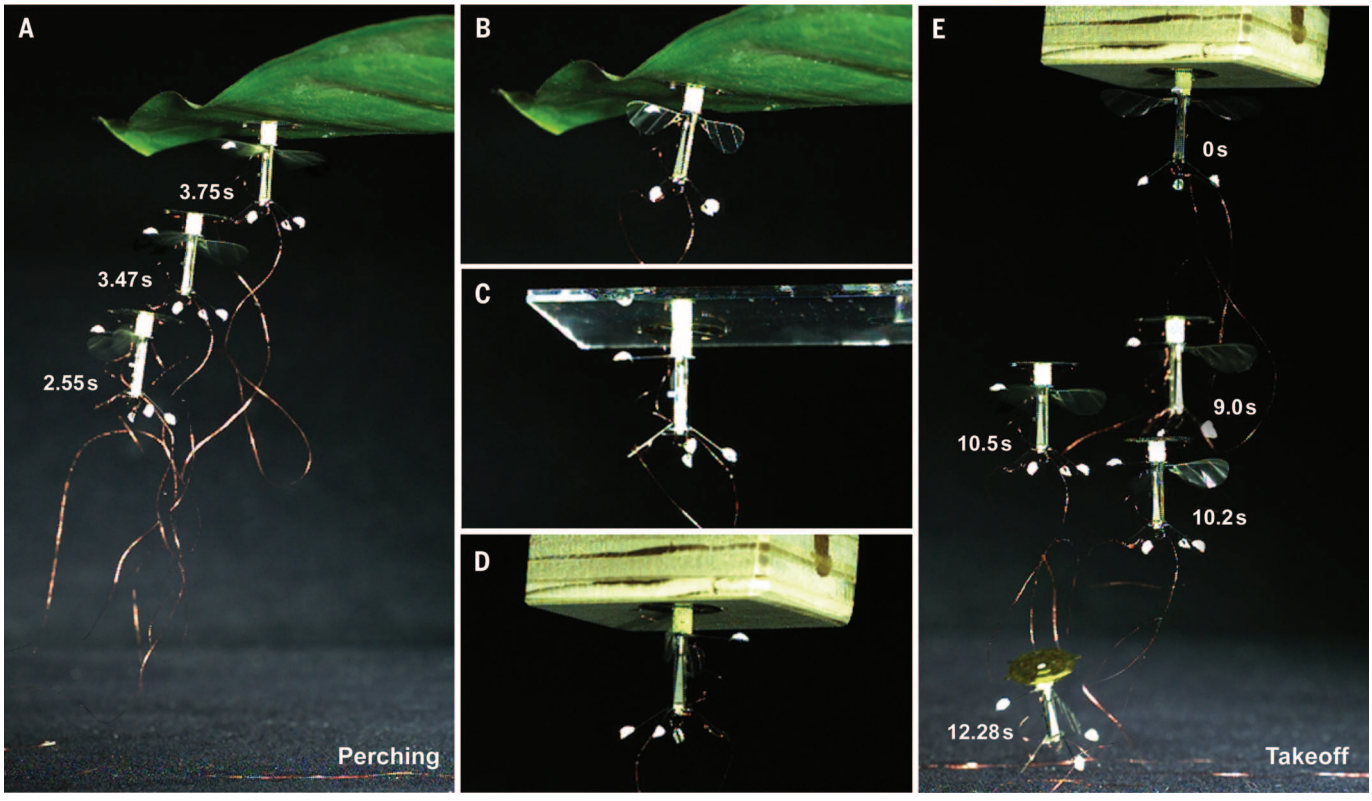


Fig. 3. Perching and relaunch demonstrations on a leaf, glass, and unfinished plywood. (**A**) Frame overlay from a high-speed video taken of a successful landing maneuver on a natural leaf. (**B** to **D**) The micro aerial vehicle after successfully landing on a leaf, glass, and unfinished plywood (wings turned off). (**E**) Frame overlay from a high-speed video taken of a successful relaunch from unfinished wood, followed by stable hovering flight (10.5 s) and a smooth landing on the ground (12.28 s).

trajectories. We further demonstrate torque-free detachment from unfinished wood through de-energizing the patch, followed by stable hovering flights (Fig. 1A). The electroadhesive patch and the compliant mount provide a lightweight (13.4 mg, less than 15% of the total body mass) and physically simple mechanism for effective perching.

Electrostatic adhesives were fabricated by embedding ~200-nm-thick interdigitated copper electrodes (fig. S1) between thin polymer layers so as to produce compliant adhesive patches with a high adhesion pressure per weight [fabrication details are provided in fig. S2 and (*24*)]. When placed on a target substrate with a voltage across the electrodes, the patch induces areas of net

Downloaded from http://science.sciencemag.org/ on May 19, 2016

charge on the surface of the substrate. These charges mirror the charges accumulated in the electroadhesive patch, leading to an attractive force between the patch and the substrate. This force increases with decreasing thicknesses of the insulating layer and the air gap at the interface (and thus is inversely related to the surface roughness), increases with increasing applied voltage (*19*, *25*), and is affected by humidity [detailed discussion is available in (*24*), section 2.4.2], material properties, leakage currents, and nonuniform charge distribution on the interface (*19*, *25*). The strong dependence on environmental parameters for this type of adhesive makes it challenging to accurately predict the attractive forces analytically or numerically (*24*).

We therefore pursued an experimental approach to characterize the normal adhesion pressures and guide the final design of our electrode patch [exact geometries are provided in fig. S3; experimental details are provided in (*24*), section 2.4, and fig. S4]. Electroadhesive patches with comb-like interdigitated electrodes (fig. S3) were created and tested on glass, steel (with varying surface texture), copper, unfinished plywood, and red brick. To appropriately mimic multiple attach/detach cycles as expected under normal use, the test patch was not cleaned between measurements. It therefore attracted both debris from the substrates and airborne particles over time. Tests on glass were performed at the start and end of the substrate test series in order to illustrate the effect of accumulated debris on the adhesion performance.

When activated at 1000 V, the test patch provided a normal adhesion pressure of at least 15.6 Pa on all the investigated substrates at 59 to 70% relative humidity. The repeated measurements on glass showed that the achievable adhesion pressure drops by <30% after 35 attachment/release iterations across different materials and 6 hours of exposure to airborne debris without cleaning (Fig. 2A). The normal pressure shows a high variance, which can be explained by its dependence on environmental parameters, such as the alignment of the patch with the target surface, the presence of debris or surface irregularities, and relative humidity [an analysis of surface roughness is available in (*24*), section 2.4, and figs. S5 and S6, and a discussion on humidity is available in fig. S7].

These experiments show that sufficient adhesion can be achieved with 1000 V, and this voltage was chosen for the final demonstrations. Although a higher voltage could provide greater adhesion, practical issues—such as the breakdown strength of the power wires used in flight experiments, or the feasibility of onboard high-voltage electronics on a future prototype—favor choosing lower voltages (*24*). At a minimal adhesion pressure of 15.6 Pa, a patch area of 0.63 cm^2 would be required to carry the weight of the microrobot (total weight, ~100 mg; robot, 84 mg; patch and glue, ~16mg). The size of the final patch was chosen to be 1.7 cm^2 in order to accommodate for the uncertainty in actual adhesion pressure caused by relative humidity, naturally rough or dirty surfaces, and degradation that will occur upon repeated attach and detach cycles. This patch was fabricated in the same manner as the larger test patch but relies on circular interdigitated electrodes to promote symmetric normal and shear forces, thus reducing the risk of destabilizing torques from residual charges during takeoff maneuvers (from a perched state). Simulated adhesion pressures are comparable for the two geometries [(*24*), section 2.2, and table S1]. We determined the charging and leakage current of the circular patch at 1000 V on glass, wood, and copper, finding that the power it requires to maintain the robot attached (<7 μW) is at least three orders of magnitude lower than the flight power of 19 mW required by our vehicle (Fig. 2, B and C, and fig. S8) (*24*).

Our control strategy for perching is inspired by the finding that honeybees hold the apparent rate of image expansion constant during a perching maneuver, which translates to an approach speed that is linear in the distance to the target (*26*). To realize a linearly decreasing approach speed, we implemented a controller that sets the reference height so that the distance to the target decays exponentially. In order to enable high reliability in the attachment maneuvers, the landing sequence is initiated only after the robot achieves a predefined stability metric (position and orientation errors and linear and angular velocities below specified thresholds) while hovering at a desired position underneath the perching target. The controller architecture relies on a logic module that detects relevant events such as stable hovering or a successful attachment. This module adjusts the reference position to initiate the perching and detachment maneuvers and ramps down the flapping amplitude once the landing is completed. The reference position is then fed into two adaptive controller blocks, one controlling altitude, the other controlling attitude and latitude (*23*). The system relies on position and orientation feedback from a motion-tracking arena to close the control loop (fig. S9). The functionality of the logic module is illustrated with data from an exemplary landing maneuver in Fig. 2D.

The fast dynamics of the underactuated microrobot, its inherent instability, and its susceptibility to disturbances (such as wind gusts), particularly in close proximity to the attachment point, render the precise tracking of a reference trajectory challenging. For example, we found that the lift of our flapping-wing vehicle increases by >40% in close proximity to ceilings (fig. S10). To address this, a tube-shaped, laser-cut polyurethane foam damper (chosen for its low density and low coefficient of restitution) was used to mount the patch onto the robot. This passive mechanism reduces the chance of rebound during high-velocity collisions and provides passive alignment between the patch and the perching target. This promotes successful attachments over a broad envelope of landing trajectories.

This combination of lightweight conformal electrodes with a flexible, energy-absorbing mount enabled reliable perching on a wide range of natural and artificial overhangs. We demonstrated this capability by executing subsequent transitions from free flight to stable attachment on a leaf (two perching attempts), glass (one perching attempt), and unfinished plywood (three perching attempts), as well as two takeoffs into hovering flight from wood (Fig. 3 and movie S1). One of the perching attempts on the leaf failed because the leaf occluded the robot from the motion capture system, causing an emergency shut-off of the actuation before alignment with the target was achieved [(*24*), section 2.5]. The other five attempts were successful.

Given the limited payload capacities and the energetically expensive nature of flight at small scales, perching presents an opportunity to increase mission durations, and hence utility, of MAVs. Deploying insect-like MAVs presents substantial challenges in manufacturing, actuation, sensing, power, and control. However, a decrease in scale does not only come with challenges, but also offers new capabilities. Motivated by the increased dominance of area-dependent forces at reduced sizes, we have demonstrated that electrostatic adhesives are an attractive option to achieve robust and dynamic perching behavior in an insect-like robot. The techniques for developing and integrating compliant electroadhesives may also find utility in other small-scale robotics tasks such as delicate micromanipulation and adhesion for inclined and inverted terrestrial locomotion.

REFERENCES AND NOTES

1. D. Floreano, R. J. Wood, *Nature* **521**, 460–466 (2015).
2. M. Tschapka, *Biol. J. Linn. Soc. Lond.* **80**, 281–288 (2003).
3. B. Anderson, W. W. Cole, S. C. H. Barrett, *Nature* **435**, 41–42 (2005).
4. M. Sazima, I. Sazima, *Can. J. Zool.* **77**, 47–51 (1999).
5. C. E. Doyle *et al.*, *IEEE/ASME Trans. Mechatron.* **18**, 506–517 (2013).
6. E. W. Hawkes *et al.*, "Dynamic surface grasping with directional adhesion," 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5487–5493 (2013).
7. J. F. Roberts, J. C. Zufferey, D. Floreano, "Energy management for indoor hovering robots," 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), vols 1–3, Conference Proceedings, pp. 1242–1247 (2008).
8. L. Daler, A. Klaptocz, A. Briod, M. Sitti, D. Floreano, "A perching mechanism for flying robots using a fibre-based adhesive," 2013 IEEE International Conference on Robotics and Automation (ICRA), pp. 4433-4438 (2013).
9. D. Mellinger, M. Shomin, V. Kumar, Control of quadrotors for robust perching and landing. *Proc. Int. Power. Lift Conf.* **2010**, 205–225 (2010).
10. M. Kovač, J. Germann, C. Hürzeler, R. Y. Siegwart, D. Floreano, *J. Micro-Nano Mech.* **5**, 77–91 (2009).
11. A. Lussier Desbiens, A. Asbeck, M. Cutkosky, *Int. J. Robot. Res.* **30**, 355–370 (2011).
12. R. Cory, R. Tedrake, Experiments in fixed-wing UAV perching. *Proc. AIAA Guid. Nav. Contr. Conf.* **2008**, 1–12 (2008).
13. A. Frank, J. McGrew, M. Valenti, D. Levine, J. P. How, *Hover, Transition, and Level Flight Control Design for a Single-Propeller Indoor Airplane* (Defense Technical Information Center, 2007).
14. M. L. Anderson *et al.*, "The sticky-pad plane and other innovative concepts for perching uavs," 47th AIAA Aerospace Sciences Meeting (2009).
15. G. J. Monkman, *Int. J. Robot. Res.* **16**, 1–10 (1997).
16. J. P. D. Téllez, J. Krahn, C. Menon, "Characterization of electroadhesives for robotic applications," 2011 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 1867–1872 (2011).
17. H. Prahlad, R. Pelrine, S. Stanford, J. Marlow, R. Kornbluh, "Electroadhesive robots—Wall climbing robots enabled by a novel, robust, and electrically controllable adhesion technology," 2008 IEEE International Conference on Robotics and Automation (ICRA), pp. 3028–3033 (2008).
18. R. Liu, R. Chen, H. Shen, R. Zhang, *Int. J. Adv. Robot. Syst.* **10**, 36 (2013).
19. R. Chen, R. Liu, H. Shen, "Modeling and analysis of electric field and electrostatic adhesion force generated by interdigital electrodes for wall climbing robots," 2013 IEEE/RSJ

Downloaded from http://science.sciencemag.org/ on May 19, 2016

International Conference on Intelligent Robots and Systems (IROS), pp. 2327–2332 (2013).
20. H. Wang, A. Yamamoto, T. Higuchi, "Electrostatic-motor-driven electroadhesive robot," 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 914–919 (2012).
21. D. Ruffatto 3rd, A. Parness, M. Spenko, *J. R. Soc. Interface* **11**, 20131089 (2014).
22. L. Wang, L. Graber, F. Iida, *IEEE Trans. Robot.* **29**, 863–874 (2013).
23. K. Y. Ma, P. Chirarattananon, S. B. Fuller, R. J. Wood, *Science* **340**, 603–607 (2013).
24. Materials and methods are avialable as supplementary materials on *Science* Online.
25. M. R. Sogard, A. R. Mikkelson, M. Nataraju, K. T. Turner, R. L. Engelstad, *J. Vac. Sci. Technol. B* **25**, 2155–2161 (2007).
26. E. Baird, N. Boeddeker, M. R. Ibbotson, M. V. Srinivasan, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 18686–18691 (2013).

ACKNOWLEDGMENTS

This material is based on work supported by the National Science Foundation (award CMMI-1251729), the Wyss Institute for Biologically Inspired Research, and the Swiss Study Foundation.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/978/suppl/DC1
Materials and Methods
Figs. S1 to S11
Table S1
References (*27*–*34*)
Movie S1

18 December 2015; accepted 23 March 2016
10.1126/science.aaf1092

SYNAPSE FORMATION

Control of neuronal synapse specification by a highly dedicated alternative splicing program

Lisa Traunmüller,* Andrea M. Gomez,* Thi-Minh Nguyen, Peter Scheiffele†

Alternative RNA splicing represents a central mechanism for expanding the coding power of genomes. Individual RNA-binding proteins can control alternative splicing choices in hundreds of RNA transcripts, thereby tuning amounts and functions of large numbers of cellular proteins. We found that the RNA-binding protein SLM2 is essential for functional specification of glutamatergic synapses in the mouse hippocampus. Genome-wide mapping revealed a markedly selective SLM2-dependent splicing program primarily consisting of only a few target messenger RNAs that encode synaptic proteins. Genetic correction of a single SLM2-dependent target exon in the synaptic recognition molecule neurexin-1 was sufficient to rescue synaptic plasticity and behavioral defects in *Slm2* knockout mice. These findings uncover a highly selective alternative splicing program that specifies synaptic properties in the central nervous system.

Alternative splicing provides a key mechanism for neuron-specific gene expression (*1*–*3*). An array of RNA-binding proteins broadly expressed in neuronal cells has been implicated in controlling developmentally regulated and neuron-specific alternative splicing programs, with single proteins regulating hundreds of target transcripts (*4*–*6*). However, some RNA-binding proteins are selectively expressed in neuronal populations, raising the possibility that they may control cell type– and synapse-specific functions (*7*, *8*).

The KH-domain RNA-binding protein SLM2 is highly expressed in glutamatergic pyramidal cells of the mouse hippocampus and in a specific subset of γ-aminobutyric acid (GABA)–releasing interneurons (*9*, *10*). In $Slm2^{KO}$ knockout (KO) mice, glutamatergic spine synapses formed at normal numbers on the primary apical dendrites of hippocampal CA1 neurons (Fig. 1, A and B). Western blot analysis of synaptosome fractions from wild-type (WT) and $Slm2^{KO}$ hippocampi revealed overall normal concentrations of glutamatergic synapse proteins. However, there was an increase in the AMPA-type glutamate receptor (AMPAR) subunit GluA1, in particular in detergent-soluble fractions from adult $Slm2^{KO}$ mouse synaptosomes (Fig. 1, C and D). In acute slices from adolescent mice (postnatal day 25), GluA1 amounts were elevated in total cell lysates and cell surface fractions, whereas *N*-methyl-D-aspartate (NMDA)–receptor GluN1 subunit expression was unaltered (Fig. 1, E and F). Whole-cell voltage-clamp recordings from CA1 neurons in $Slm2^{KO}$ hippocampal slices showed no difference in the miniature excitatory postsynaptic current (mEPSC) amplitude and frequency between WT and $Slm2^{KO}$ mice (Fig. 2, A and B). However, mEPSC events showed a modest increase in the speed of rise and decay times in $Slm2^{KO}$ CA1 neurons (Fig. 2, C and D). AMPAR/NMDAR ratios were significantly increased in $Slm2^{KO}$ mice (Fig. 2, E and F). In $Slm2^{KO}$ CA1 neurons, stimulation of Schaffer collaterals elicited larger postsynaptic responses as compared to WT (Fig. 2G; see also fig. S1A for data from field EPSP recordings). Paired-pulse facilitation was normal in $Slm2^{KO}$ (Fig. 2H). In sum, these experiments demonstrate an elevation in postsynaptic AMPAR surface expression and function in CA1 neurons of $Slm2^{KO}$ mice. Finally, long-term potentiation (LTP) induced by theta-burst stimulation of Schaffer collaterals was significantly reduced in acute slices from $Slm2^{KO}$ mice (Fig. 2I).

Candidate gene approaches have identified some transcripts that are altered in $Slm2^{KO}$ brains (*9*–*12*). However, a comprehensive global analysis of SLM2 targets is lacking. Using Illumina paired-end sequencing we mapped SLM2-dependent alternative splicing events at a genome-wide level. This analysis revealed highly correlated expression of transcripts between genotypes, indicating that SLM2 does not play a major role in tuning overall transcript levels (Fig. 3A). Moreover, transcripts encoding ionotropic and metabotropic glutamate receptors were not significantly changed (fig. S1, B to E). Most significantly altered was the transcript encoding the SLM2 paralogue SLM1, which has previously been shown to be upregulated in $Slm2^{KO}$ hippocampus (*11*).

Genome-wide splicing patterns were extracted on the basis of annotations from FAST DB, and splicing indices calculated. The vast majority of exons remained essentially unchanged between WT and $Slm2^{KO}$ hippocampi (Fig. 3B and fig. S2A; analysis includes 4965 microexons). Notably, alternative exons in four genes showed disproportionately strong deregulation in the $Slm2^{KO}$ hippocampus. There was a >2-fold increase in the incorporation of exons at the alternatively spliced segment four (AS4) of neurexins (*Nrxn1*, *2*, and *3*), three genes that encode synaptic cell surface receptors. Moreover, exon 24 incorporation was 1.58-fold elevated in tomosyn-2 (*Stxbp5l*), a component of the vesicle fusion machinery. Another seven exons were identified that exhibited significant alterations ($P < 0.01$) although with modest fold-change (Fig. 3C and table S1). Independent experimental validation confirmed that deregulation of additional candidate target exons was small or in some cases not detectable (Fig. 3, D and E, and fig. S2B). The paralogue SLM1 may compensate for the loss of SLM2 (*11*). However, comparison of candidate exons in $Slm1^{KO}$, $Slm2^{KO}$, and $Slm1{:}Slm2^{DKO}$ (double knockout) mice revealed that deregulation was not significantly more severe in the DKO mice (fig. S2C).

Alternative splicing at *Nrxn* AS4 regulates selective trans-synaptic interactions of neurexins in the presynaptic terminal with several synaptic receptors (*13*–*16*). For an unbiased identification of interaction partners regulated by this alternative splicing event, we performed affinity purifications on recombinant NRX1β4(+) and NRX1β4(-) isoforms, followed by shot-gun mass spectrometry (Fig. 4A and fig. S3, A and B). We identified 21 candidate binding partners, including known neurexin-binding proteins (Cbln2, the leucine-rich repeat proteins LRRTM1,2,4, neuroligin-1,2, 3, and the sortilin-related VPS10-domain containing receptor

Biozentrum, University of Basel Klingelbergstrasse 50-70, 4056 Basel, Switzerland
*These authors contributed equally to this work. †Corresponding author. Email: peter.scheiffele@unibas.ch

Downloaded from http://science.sciencemag.org/ on May 19, 2016

SORCS) and a number of previously unknown candidate ligands (complement C3, chondroadherin-like protein, neuronal pentraxin and neuronal pentraxin receptor 1, and astrotactin). For six of these proteins, we observed significant differences in the interaction with AS4(+) and AS4(-) NRX1β isoforms, including neuroligins, Chadl, LRRTMs, and complement C3 (Fig. 4B and table S2; $P < 0.05$). Thus, there is an array of synaptic interactions that can be modified by deregulation of neurexin alternative splicing in *Slm2*^{KO} mice.

We hypothesized that the modification of endogenous neurexin alternative splicing in *Slm2*^{KO} hippocampus may disrupt interactions with these splice insertion-sensitive ligands. In coimmunoprecipitation experiments with an antibody against neurexin-1 (anti-neurexin-1), we observed that neuroligin-1 (NL1) and neuroligin-3 (NL3) are abundant in immunoprecipitates from WT mice. However, both NLs were reduced in precipitates prepared from *Slm2*^{KO} mice (Fig. 4C). Conversely, coprecipitation of C3, a component of the complement system that has been implicated in synapse elimination (*17*), was slightly elevated in the same precipitates from *Slm2*^{KO} mice (fig. S3C). Thus, the loss of SLM2 indeed switches synaptic receptor-ligand interactions. NL1 recruits synaptic NMDA receptors (*18*, *19*), whereas NL3 and LRRTMs mediate the synaptic recruitment and stabilization of AMPA receptors (*20*). Loss of the trans-synaptic interactions with these neurexin ligands might result in the postsynaptic glutamate receptor deficits in *Slm2*^{KO} mice. To test this hypothesis, we generated *Nrxn1*$^{\Delta ex21}$ mice in which *Nrxn1* exon 21 (the alternative cassette exon at AS4) is specifically deleted but total *Nrxn* transcript levels are unchanged (fig. S4, A to C). Heterozygous removal of one *Nrxn1*$^{\Delta ex21}$ allele restored normal *Nrxn1* AS4(-) transcript levels in

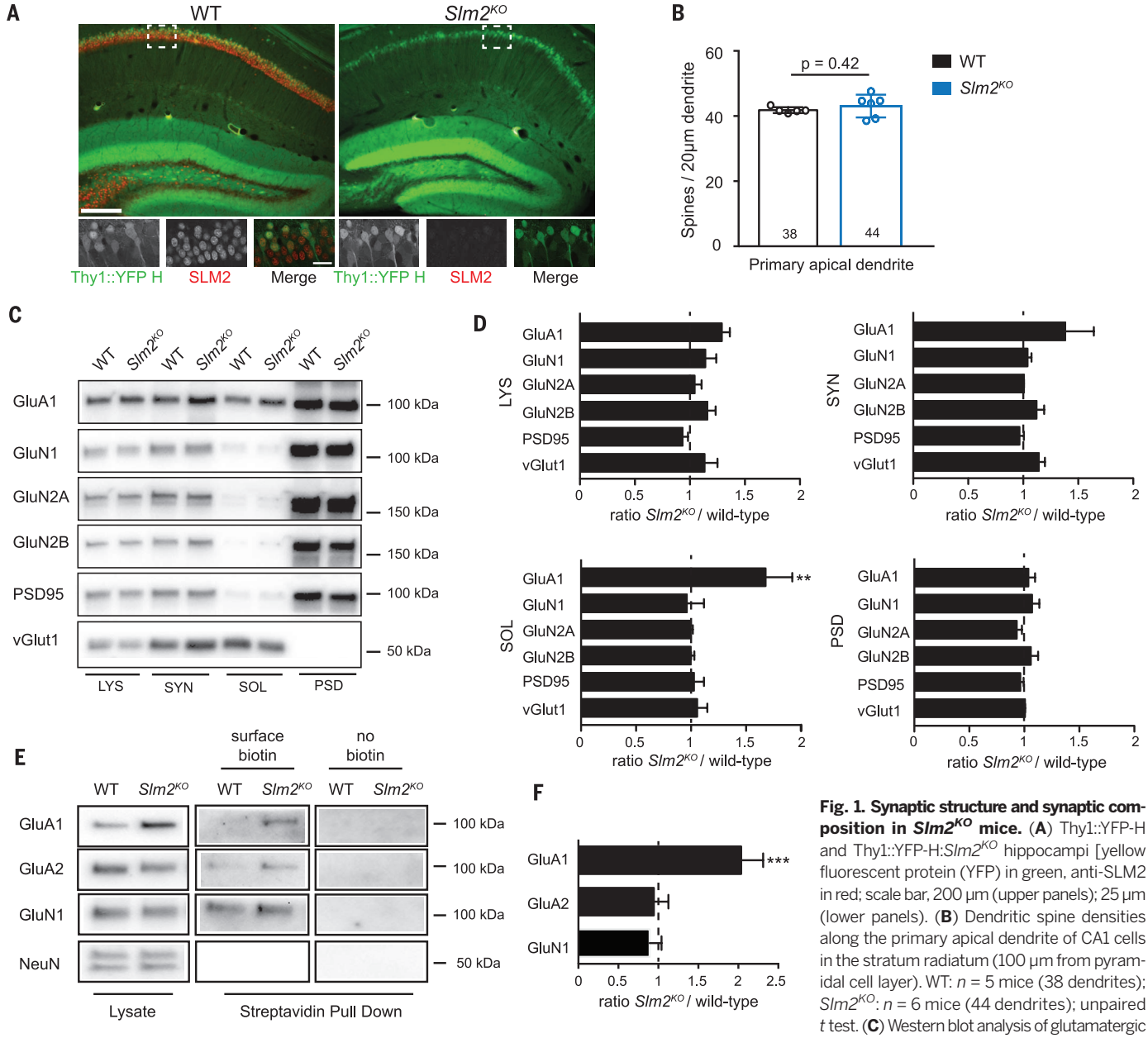


Fig. 1. Synaptic structure and synaptic composition in *Slm2*^{KO} mice. (**A**) Thy1::YFP-H and Thy1::YFP-H:*Slm2*^{KO} hippocampi [yellow fluorescent protein (YFP) in green, anti-SLM2 in red; scale bar, 200 µm (upper panels); 25 µm (lower panels). (**B**) Dendritic spine densities along the primary apical dendrite of CA1 cells in the stratum radiatum (100 µm from pyramidal cell layer). WT: $n = 5$ mice (38 dendrites); *Slm2*^{KO}: $n = 6$ mice (44 dendrites); unpaired *t* test. (**C**) Western blot analysis of glutamatergic proteins in total lysate (LYS), synaptosomes (SYN), Triton X100–soluble fraction at pH = 6 (SOL), and Triton X100–insoluble postsynaptic density (PSD) in hippocampal tissue from adult WT and *Slm2*^{KO} mice. (**D**) Mean intensities of expression levels in WT and *Slm2*^{KO} hippocampi ($n \geq 3$ mice per genotype, One-way analysis of variance (ANOVA) with Dunnett's multiple comparison test, $^{**}P < 0.01$). (**E**) Total and surface glutamate receptor pools were assessed in acute slices from P25 mice by surface biotinylation and streptavidin pull-down assay; nonbiotinylated samples ("no biotin") served as negative control. (**F**) Quantification for Fig. 1E (one-way ANOVA with Tukey's post-hoc test, $^{***}P < 0.001$, $n \geq 4$ for WT and *Slm2*^{KO}).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

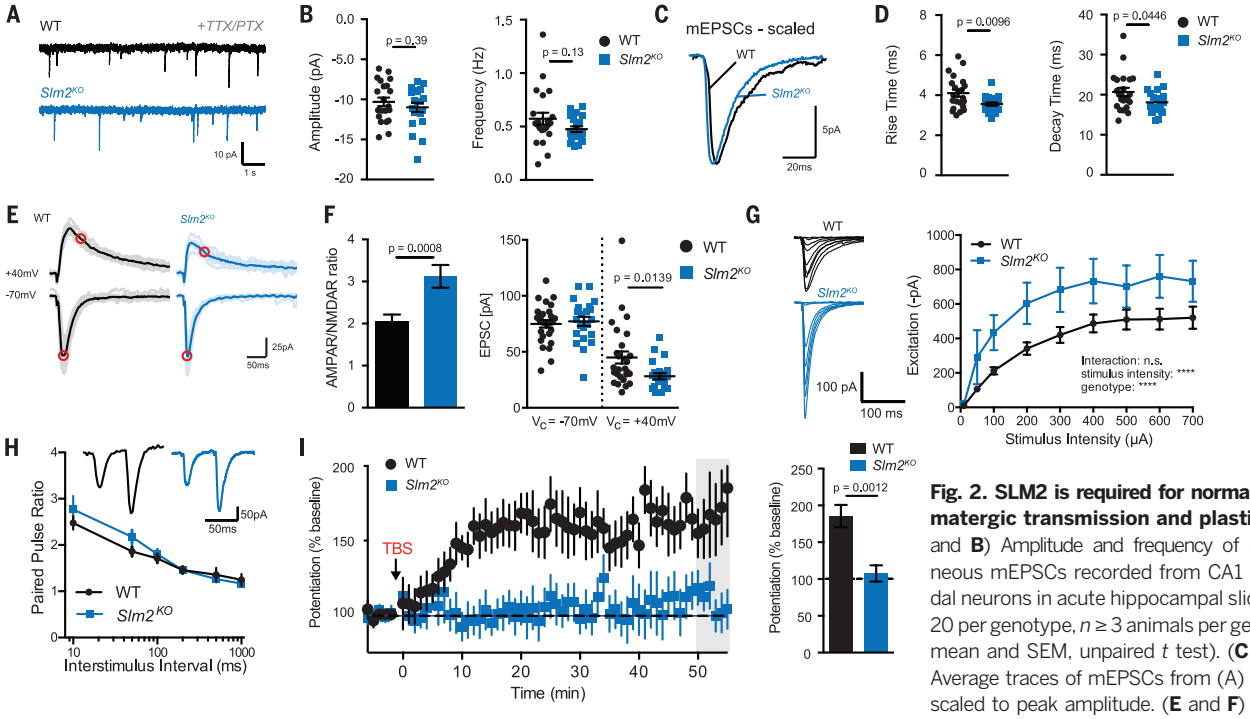
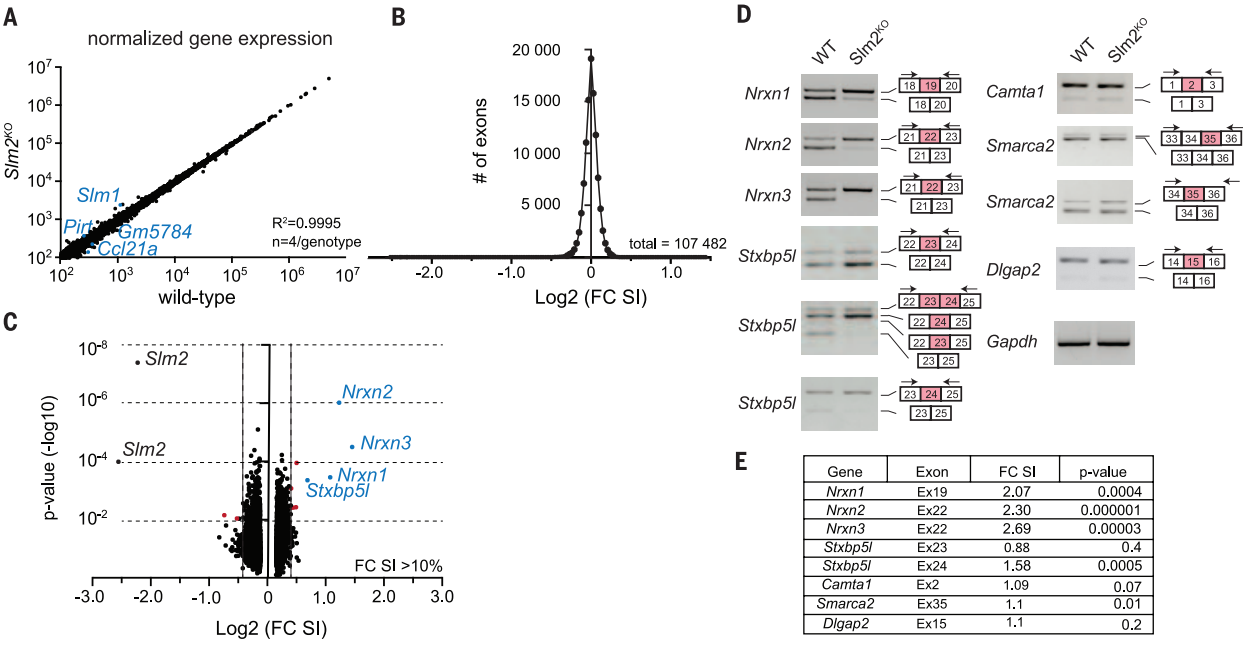


Fig. 2. SLM2 is required for normal glutamatergic transmission and plasticity. (**A** and **B**) Amplitude and frequency of spontaneous mEPSCs recorded from CA1 pyramidal neurons in acute hippocampal slices ($n \geq 20$ per genotype, $n \geq 3$ animals per genotype, mean and SEM, unpaired t test). (**C** and **D**) Average traces of mEPSCs from (A) and (B) scaled to peak amplitude. (**E** and **F**) Average traces and summary data of evoked EPSCs at –70 mV (lower trace) and +40 mV (upper trace), respectively. Peak EPSC amplitudes were measured at –70 mV (peak marked by red circle). NMDAR currents were measured at +40 mV (red circle, $n \geq 22$ per genotype, $n \geq 4$ animals; mean and SEM, unpaired t test). (**G**) Representative traces of EPSCs evoked with varied stimulation intensities in WT or *Slm2*^{KO} CA1 neurons ($n \geq 8$ per genotype, $n \geq 3$ animals; mean and SEM, two-way ANOVA). (**H**) Paired-pulse ratio in CA1 neurons from acute hippocampal slices ($n \geq 11$ per genotype, $n \geq 4$ animals, mean and SEM, two-way ANOVA). (**I**) Averaged responses from *Slm2*^{KO} CA1 neurons following a TBS (theta-burst stimulus) delivered at V_c= –70 mV. The gray bar indicates the interval quantified in the histogram ($n \geq 7$ per genotype, $n \geq 6$ animals, mean and SEM, unpaired t test).



E

Gene	Exon	FC SI	p-value
Nrxn1	Ex19	2.07	0.0004
Nrxn2	Ex22	2.30	0.000001
Nrxn3	Ex22	2.69	0.00003
Stxbp5l	Ex23	0.88	0.4
Stxbp5l	Ex24	1.58	0.0005
Camta1	Ex2	1.09	0.07
Smarca2	Ex35	1.1	0.01
Dlgap2	Ex15	1.1	0.2

Fig. 3. Genome-wide mapping of SLM2-dependent alternative splicing program. (**A**) Correlation analysis of $\log_{10}$-transformed, normalized total transcript counts (for all genes with count ≥ 100 in *Slm2*^{KO}). Transcripts with significant alteration between genotypes ($P < 0.01$, $\Delta FC > 50\%$) are highlighted in blue ($n = 4$ mice per genotype). (**B**) Frequency distribution of exon incorporation rates for 107,482 exons expressed as fold change in splicing index (FC SI) plotted on a $\log_2$ scale. (**C**) Fold change SI and P-values for all exons with an FC SI > 10% (total of 9110 exons). Exons exhibiting changes of FC SI > 30% and $P < 0.01$ are marked in red (*Pecam1*, *Epha5*, *Dgkb*, *Gm1673*, *Cpne5*, *Hnrnpul1*, and *Cdk16*; threshold indicated by dashed lines). Highly deregulated exons (FC SI > 50% and $P < 0.01$) are marked in blue (*Nrxn1*, *Nrxn2*, *Nrxn3*, *Stxbp5l*). *Slm2* exon 2, deleted in *Slm2*^{KO}, is not displayed. (**D**) Experimental assessment of candidate SLM2-dependent exons by reverse transcription–polymerase chain reaction (RT-PCR) in WT and *Slm2*^{KO} hippocampi; alternative exons (red), primer sites (arrows). (**E**) Table for fold change SI and P-values for validated exons shown in (D).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

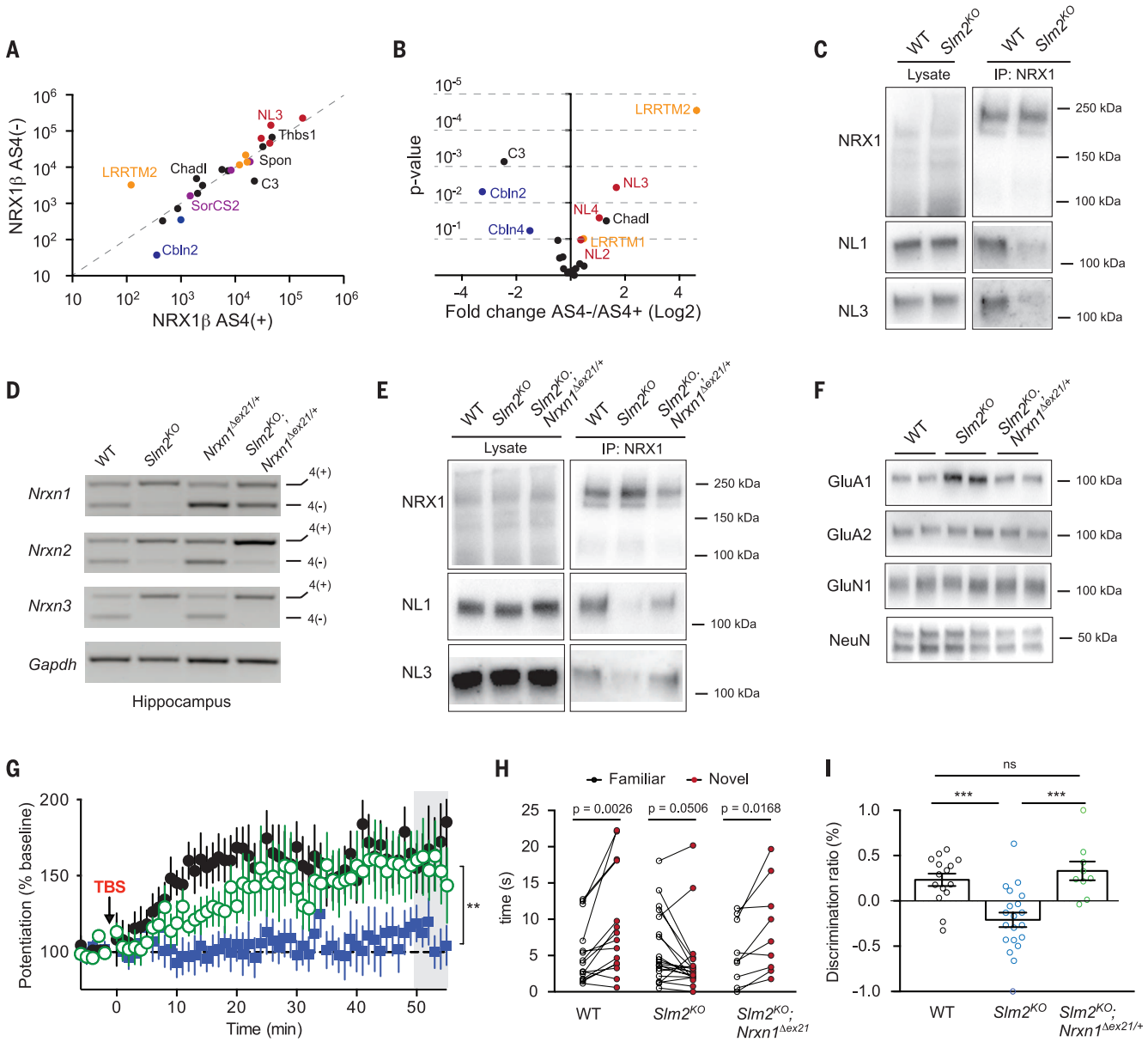


Fig. 4. Heterozygous deletion of *Nrxn1* exon 21 restores synaptic phenotypes in *Slm2*^{KO} hippocampus. (**A**) Affinity purification with recombinant NRX1β AS4(+) and NRX1β AS4(-). Correlation of $\log_{10}$-transformed median spectra counts for all extracellular proteins identified with ≥2 peptides recovered on at least one of the two neurexin isoforms. Paralogues highlighted: NLs (red), Cblns (blue), LRRTMs (orange), SORCs (purple). (**B**) Fold change against *P*-value for all bound extracellular proteins when comparing NRX1β AS4(+) to NRX1β AS4(-). (**C**) Anti-NRX1 immunoprecipitation from forebrain lysates of WT and *Slm2*^{KO} animals probed with anti-NRX1, anti-NL1, and anti-NL3. (**D**) Representative RT-PCR for analysis of *Nrxn1,2,3* AS4 in the hippocampus of WT, *Slm2*^{KO}, *Nrxn*$^{\Delta ex21/+}$, and *Slm2*^{KO};*Nrxn1*$^{\Delta ex21/+}$ animals. (**E**) NRX1 immunoprecipitation from forebrain lysates of WT, *Slm2*^{KO}, and *Slm2*^{KO};*Nrxn1*$^{\Delta ex21/+}$ animals probed with anti-NRX1, anti-NL1 and anti-NL3. (**F**) Glutamate receptor expression in acute slices from P25 WT and *Slm2*^{KO} and *Slm2*^{KO};*Nrxn1*$^{\Delta ex21/+}$ hippocampi. Protein concentrations were probed by Western blotting of total lysates (average FC *Slm2*^{KO}/WT = 1.94, *n* = 5, average FC *Slm2*^{KO};*Nrxn1*$^{\Delta ex21/+}$/WT = 1.13, *n* = 3). (**G**) Averaged responses to a theta-burst stimulation (TBS) in *Slm2*^{KO};*Nrxn1*$^{\Delta ex21/+}$ CA1 neurons (green) is overlaid on WT (gray) and *Slm2*^{KO} responses (blue) from Fig. 2I (*n* = 8 cells, *n* ≥ 5 animals per genotype). Comparison of the responses at 40 to 60 mins after induction between *Slm2*^{KO} and *Slm2*^{KO};*Nrxn1*$^{\Delta ex21+/-}$ by ANOVA with Tukey's multiple comparison test (**$P < 0.01$). (**H** and **I**) Behavioral alterations of WT, *Slm2*^{KO}, and *Slm2*^{KO};*Nrxn1*$^{\Delta ex21}$ mice in object recognition task. Time spent investigating a novel and a familiar object during a 5-min trial 1 hour after acquisition and discrimination ratio are plotted [(H), paired *t* test; (I), unpaired *t* test; WT, *n* = 15; *Slm2*^{KO}, *n* = 20; *Slm2*^{KO};*Nrxn1*$^{\Delta ex21}$, *n* = 9, ns $P > 0.05$; ***$P < 0.001$].

the *Slm2*^{KO} background (Fig. 4D). Restoration of *Nrxn1* AS4(-) transcripts rescued trans-synaptic cell surface interactions of endogenous neurexin with NL1 and NL3 in the coimmunoprecipitation assays (Fig. 4E), normalized GluA1 amounts in acute hippocampal slices (Fig. 4F), and partially recovered theta burst–induced Schaffer collateral LTP (Fig. 4G). In an object recognition test, a behavioral task that involves the hippocampus (*21*), *Slm2*^{KO} mice differed significantly from WT mice in that they did not preferentially explore novel objects (Fig. 4, H and I, and fig. S4D). Also, this phenotype and other behavioral alterations were rescued in *Slm2*^{KO};*Nrxn1*$^{\Delta ex21/+}$ mice (Fig. 4, H and I, and fig. S4, E to G). Thus, the control of a single alternative exon by SLM2 has a major contribution to the specification of glutamatergic synapse function, plasticity, and mouse behavior.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Forced expression of *Nrxn3* AS4(+) isoforms was previously shown to reduce postsynaptic AMPAR localization and to impair LTP in the subiculum (*22*). By contrast, in *Slm2*^{KO} CA1 neurons, total surface AMPAR amounts and AMPAR function are elevated. These differences are most likely due to the simultaneous disruption of alternative splicing in all three neurexin genes in the *Slm2*^{KO}, as well as the different cell type–specific context. Compared to other RNA-binding proteins, SLM2 exhibits highly selective neuronal cell type–specific expression, and only a small set of alternative exons is particularly reliant on SLM2 function. Our genetic rescue experiments demonstrate that restoration of a single alternative exon has a major impact on the *SLM2*^{KO} phenotype. We hypothesize that targeted, cell type–specific splicing regulation of surface receptor recognition systems as reported here for the SLM2-neurexin system represents a general mechanism for the control of synapse specification in neuronal circuits.

REFERENCES AND NOTES

1. S. Zheng, D. L. Black, *Trends Genet.* **29**, 442–448 (2013).
2. B. Raj, B. J. Blencowe, *Neuron* **87**, 14–27 (2015).
3. D. Schreiner *et al.*, *Neuron* **84**, 386–398 (2014).
4. J. Ule *et al.*, *Nat. Genet.* **37**, 844–852 (2005).
5. M. Quesnel-Vallières, M. Irimia, S. P. Cordes, B. J. Blencowe, *Genes Dev.* **29**, 746–759 (2015).
6. Q. Li *et al.*, *eLife* **3**, e01201 (2014).
7. D. Schreiner, T. M. Nguyen, P. Scheiffele, *Curr. Opin. Neurobiol.* **27**, 25–30 (2014).
8. A. D. Norris *et al.*, *Mol. Cell* **54**, 946–959 (2014).
9. I. Ehrmann *et al.*, *PLOS Genet.* **9**, e1003474 (2013).
10. T. Iijima, Y. Iijima, H. Witte, P. Scheiffele, *JCB* **204**, 331–342 (2014).
11. L. Traunmüller, C. Bornmann, P. Scheiffele, *J. Neurosci.* **34**, 16755–16761 (2014).
12. T. Iijima *et al.*, *Cell* **147**, 1601–1614 (2011).
13. A. A. Boucard, A. A. Chubykin, D. Comoletti, P. Taylor, T. C. Südhof, *Neuron* **48**, 229–236 (2005).
14. B. Chih, L. Gollan, P. Scheiffele, *Neuron* **51**, 171–178 (2006).
15. T. Uemura *et al.*, *Cell* **141**, 1068–1079 (2010).
16. K. Matsuda, M. Yuzaki, *Eur. J. Neurosci.* **33**, 1447–1461 (2011).
17. B. Stevens *et al.*, *Cell* **131**, 1164–1178 (2007).
18. B. Chih, H. Engelman, P. Scheiffele, *Science* **307**, 1324–1328 (2005).
19. E. C. Budreck *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 725–730 (2013).
20. S. L. Shipman *et al.*, *Nat. Neurosci.* **14**, 718–726 (2011).
21. S. J. Cohen, R. W. Stackman Jr.., *Behav. Brain Res.* **285**, 105–117 (2015).
22. J. Aoto, D. C. Martinelli, R. C. Malenka, K. Tabuchi, T. C. Südhof, *Cell* **154**, 75–88 (2013).

ACKNOWLEDGMENTS

We thank O. Mauger, D. Schreiner, and L. Xiao for insightful comments on the manuscript and Y. Hayashi and J. Bischofberger for expert advice. We thank C. Bornmann and P. Demougin, F. Lemoine, P. de la Grange, A. Schmidt, and M. Rüegg for suggestions and experimental support. L.T. was financially supported by the Boehringer Ingelheim Fonds. A.M.G. was supported by a European Molecular Biology Organization long-term fellowship. T.-M.N. was supported by the International PhD Program University of Basel. This work was supported by funds to P.S. from the Swiss National Science Foundation; European Autism Interventions EU-AIMS, which receives support from the Innovative Medicines Initiative; and the Kanton Basel-Stadt. Sequencing data were deposited at the functional genomics data repository Gene Expression Omnibus (accession number GSE79902).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/982/suppl/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References (*23*–*32*)

12 January 2016; accepted 15 April 2016
10.1126/science.aaf2397

PLANT SCIENCE

Control of eukaryotic phosphate homeostasis by inositol polyphosphate sensor domains

Rebekka Wild,[1]* Ruta Gerasimaite,[2]* Ji-Yul Jung,[3]* Vincent Truffault,[4] Igor Pavlovic,[5] Andrea Schmidt,[2] Adolfo Saiardi,[6] Henning Jacob Jessen,[5,7] Yves Poirier,[3]† Michael Hothorn,[1]† Andreas Mayer[2]†

Phosphorus is a macronutrient taken up by cells as inorganic phosphate (P_i). How cells sense cellular P_i levels is poorly characterized. Here, we report that SPX domains—which are found in eukaryotic phosphate transporters, signaling proteins, and inorganic polyphosphate polymerases—provide a basic binding surface for inositol polyphosphate signaling molecules (InsPs), the concentrations of which change in response to P_i availability. Substitutions of critical binding surface residues impair InsP binding in vitro, inorganic polyphosphate synthesis in yeast, and P_i transport in *Arabidopsis*. In plants, InsPs trigger the association of SPX proteins with transcription factors to regulate P_i starvation responses. We propose that InsPs communicate cytosolic P_i levels to SPX domains and enable them to interact with a multitude of proteins to regulate P_i uptake, transport, and storage in fungi, plants, and animals.

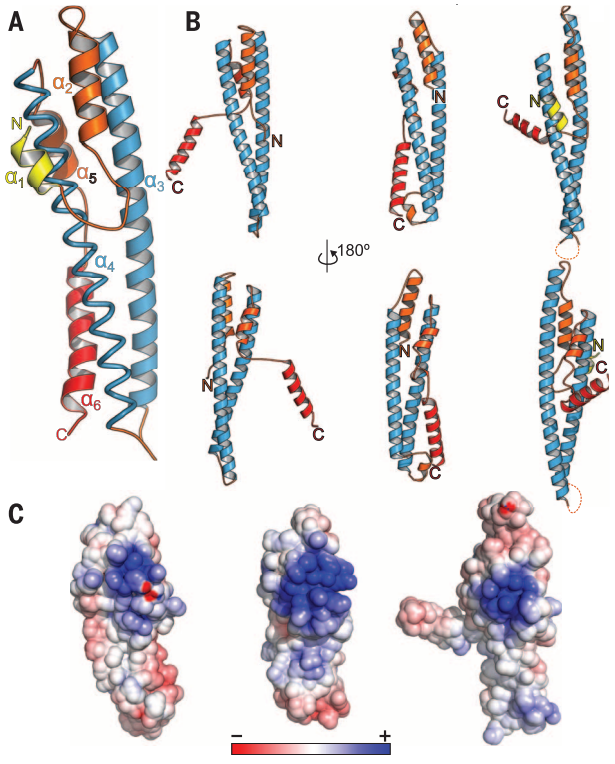
Inorganic phosphate (P_i) uptake, transport, storage, and signaling systems in eukaryotes (*1*, *2*) are supported by proteins containing SPX domains of unknown function (*3*). These small domains (135 to 380 residues) are located at the N-termini of P_i transporters (*4*–*7*), the inorganic polyphosphate (polyP) polymerase vacuolar transporter chaperone (VTC) (*8*), and P_i signaling proteins (*9*, *10*) or occur as independent, single-domain proteins (fig. S1) (*11*). Mutations in the SPX domain in plant and human P_i exporters impair their transport capacity and affect P_i signaling (*12*). The P_i transporters Pho87 and Pho90 interact via their SPX domains with the yeast Spl2 protein, and this interaction downregulates P_i uptake (fig. S1) (*4*, *13*). In plants, SPX domains bind phosphate starvation response (PHR) transcription factors, which mediate P_i starvation–induced gene expression under P_i-limiting conditions (*11*, *14*). Formation of a SPX domain–PHR complex prevents the transcription factor from binding its target promoters, thus reducing the expression of starvation-induced genes under P_i-sufficient growth conditions (*15*, *16*). P_i itself may promote the association of SPX proteins with PHRs, and thus, SPX domains may sense cellular P_i levels (*15*, *16*).

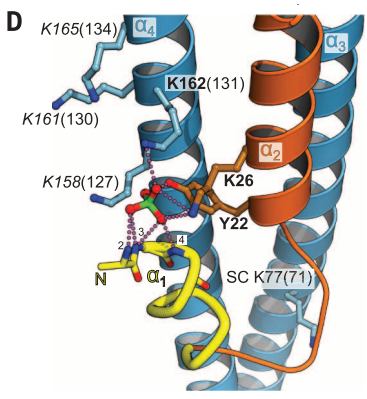
[1]Structural Plant Biology Laboratory, Department of Botany and Plant Biology, University of Geneva, Geneva, Switzerland. [2]Department of Biochemistry, University of Lausanne, Lausanne, Switzerland. [3]Department of Plant Molecular Biology, University of Lausanne, Lausanne, Switzerland. [4]Department of Biochemistry, Max Planck Institute for Developmental Biology, Tübingen, Germany. [5]Department of Chemistry and Pharmacy, University of Zürich, Zürich, Switzerland. [6]Medical Research Council Laboratory for Molecular Cell Biology, University College London, London, UK. [7]Institute of Organic Chemistry, Albert-Ludwigs-University Freiburg, Freiburg, Germany.
*These authors contributed equally to this work. **†Corresponding author. Email: yves.poirier@unil.ch (Y.P.); michael.hothorn@unige.ch (M.H.); andreas.mayer@unil.ch (A.M.)**

To investigate whether SPX domains are eukaryotic sensors for P_i, we mapped their domain boundaries and expressed and purified fungal, plant, and human SPX domains (fig. S2). We determined three independent crystal structures of SPX^{ScVtc4} (residues 1 to 178) from the yeast VTC complex (*8*) at 2.1 to 3.0 Å resolution (table S1). Structures of the SPX domain of the *Chaetomium thermophilum* glycerophosphodiesterase (residues 1 to 184) and of the human phosphate transporter XPR1 (xenotropic and polytropic retrovirus receptor 1; residues 1 to 207) were solved at 1.95 and 2.43 Å resolution, respectively. These structures highlight key features of the SPX domain: Its core consists of two long (~80 Å) helices, α3 and α4, connected by linkers of variable length (Fig. 1, A and B). The core helices and two smaller C-terminal helices, α5 and α6, contribute to the formation of a three-helix bundle (Fig. 1, A and B). Helix α6 adopts various orientations in our different structures (Fig. 1B and fig. S3). The N terminus of the domain folds into a helical hairpin formed by α1 and α2, but in several "apo" structures, helix α1 appears disordered (Fig. 1, A and B). SPX domains share no major structural homology with proteins of known function (*17*).

We located a conserved basic surface cluster at the N terminus of the SPX helical bundle (Fig. 1C). Several invariant Lys residues and Tyr22^{HsXPR1} from helices α2 and α4 represent SPX sequence fingerprints and contribute to the formation of the surface cluster (Fig. 1, C and D, and fig. S2) (*3*). A sulfate ion is bound to a pocket in the surface cluster of SPX^{HsXPR1} (Fig. 1D). SO_4^{2-}, which in crystal structures is often bound in place of P_i, is coordinated by the backbone amides of Lys^2, Phe^3, and Ala^4 from helix α1; by Tyr^{22} and Lys^{26} from α2; and by Lys^{162} from α4 [phosphate binding cluster (PBC)] in SPX^{HsXPR1} (Fig. 1D).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

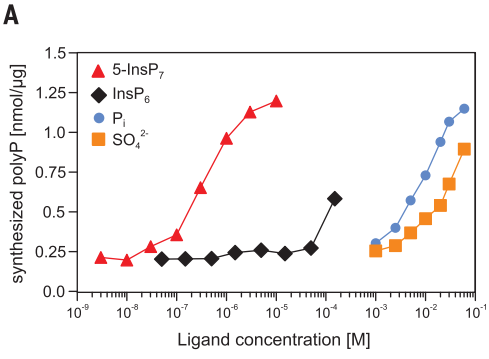




E

Protein	Ligand	K_d [mM]
$SPX^{ScVtc2\ wt}$	P_i	4.8 ± 0.6
$SPX^{ScVtc2\ \mathbf{Y22F}}$	P_i	47.1 ± 22.5
$SPX^{ScVtc2\ \mathbf{K26A}}$	P_i	19.5 ± 1.7
$SPX^{ScVtc2\ \mathbf{K131A}}$	P_i	9.1 ± 1.0
$SPX^{ScVtc2\ \mathbf{PBC}}$	P_i	93.1 ± 38.4
$SPX^{ScVtc2\ KSC}$	P_i	6.1 ± 2.4
$SPX^{ScVtc2\ SC}$	P_i	7.5 ± 1.6
$SPX^{ScVtc2\ \Delta N14}$	P_i	45.3 ± 3.1
$SPX^{ScVtc2\ meth}$	P_i	60.5 ± 18.4
$SPX^{ScVtc2\ wt}$	SO_4^{2-}	7.1 ± 0.3
$SPX^{ScVtc2\ wt}$	PP_i	0.53 ± 0.13
$SPX^{HsXPR1\ wt}$	P_i	22.8 ± 0.3

Fig. 1. SPX helical bundles provide a positively charged ligand binding surface. (**A**) Ribbon diagram of apo SPX^{ScVtc4} with core helices in blue and surrounding helices colored from N to C terminus in yellow to red. (**B**) Ribbon diagrams of the reductively methylated SPX^{ScVtc4} (left), SPX^{CtGde1} (middle), and SPX^{HsXPR1} (right) shown in two orientations [colors as in (A)]. (**C**) Electrostatic surface potential of the SPX domains from (B) reveals the SPX basic surface cluster. (**D**) A sulfate ion is bound to the SPX^{HsXPR1} basic surface cluster. Interacting residues are shown in bonds representation (PBC, in bold; KSC, in italics); a distant Lys residue (SC) is shown alongside. Residue numbers correspond to SPX^{HsXPR1} (SPX^{ScVtc2} in parentheses). (**E**) Table summaries for NMR titrations of wild-type and mutant SPX^{ScVtc2} and SPX^{HsXPR1} with different candidate ligands.



B

Protein	Ligand	K_d [µM]
$SPX^{ScVtc2\ *}$	$5\text{-}InsP_7$	0.05 ± 0.002
$SPX^{ScVtc2\ *}$	$InsP_6$	0.04 ± 0.001
SPX^{ScVtc2}	$InsP_6$	0.44 ± 0.19
SPX^{CtVtc4}	$InsP_6$	0.07 ± 0.02
SPX^{CtGde1}	$InsP_6$	0.28 ± 0.02

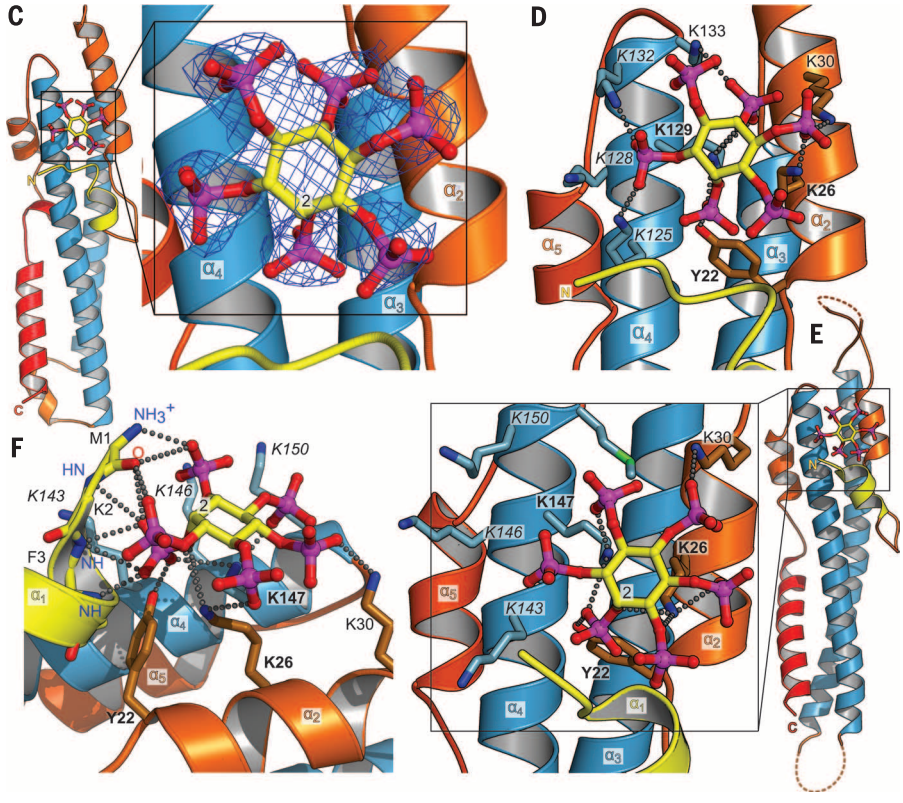


Fig. 2. SPX domains sense InsP signaling molecules. (**A**) VTC-dependent polyP synthesis by isolated yeast vacuoles in response to increasing concentrations of externally supplied P_i, SO_4^{2-}, $InsP_6$, and $5\text{-}InsP_7$. (**B**) Binding affinities of different SPX domains versus $5\text{-}InsP_7$ and $InsP_6$ by means of isothermal titration calorimetry and microscale thermophoresis (the latter indicated by an asterisk) (± fitting errors). (**C**) Ribbon diagram of SPX^{CtGde1} in complex with $InsP_6$ (in bonds representation, colors as in Fig. 1A) and including a $2F_o–F_c$ omit electron density map contoured at 1.5σ. (**D**) Close-up view of the SPX^{CtGde1} with $InsP_6$–protein interactions depicted as dotted lines. (**E**) Ribbon diagram and close-up view of the SPX^{CtVtc4}–$InsP_6$ complex. (**F**) Rotated view of the SPX^{CtVtc4} complex highlighting the interactions of $InsP_6$ with main-chain atoms of helix α1.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

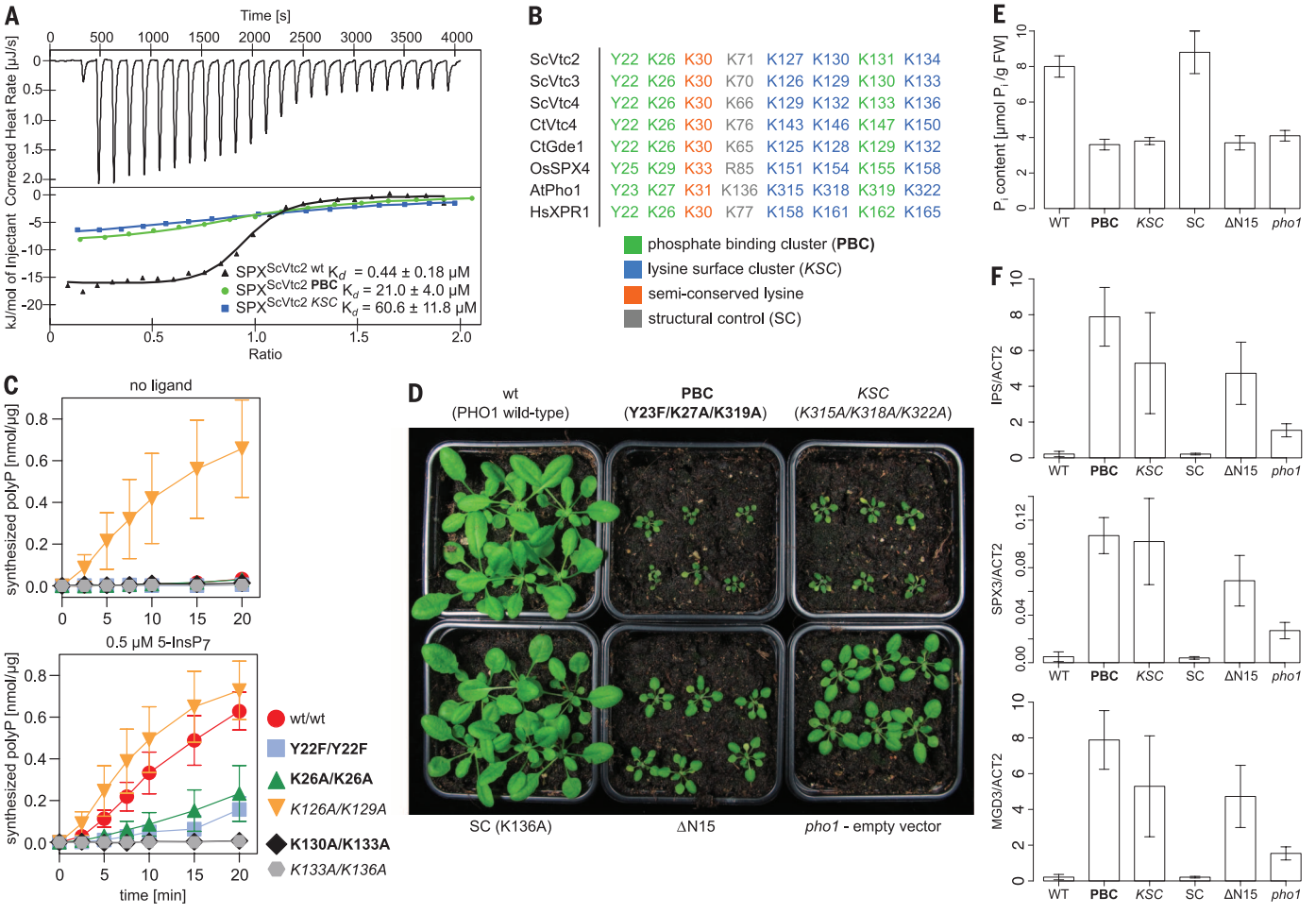


Fig. 3. Mutations in the SPX InsP-binding site affect polyP synthesis in yeast and plant P_i transport and P_i starvation signaling. (**A**) Isothermal titration calorimetry of wild type and of PBC and KSC mutant SPX^{ScVtc2} versus $InsP_6$ (± fitting errors). (**B**) Table summaries of the conserved surface cluster residues in the different SPX domain–containing proteins used in this study. (**C**) PolyP synthesis by isolated vacuoles carrying VTC complexes, with corresponding substitutions in the SPX basic surface clusters of subunits Vtc3 and Vtc4 and in the presence or absence of 0.5 µM 5-$InsP_7$. (**D**) Shoot phenotypes of 28-day-old T2 *pho1-2* transgenic lines, carrying promPHO1::PHO1:GFP (labeled wild type) or its mutant variants. GFP, green fluorescent protein. (**E**) Shoot P_i content of the transgenic lines from (D). (**F**) Normalized relative expression of selected P_i starvation–induced genes in the different *pho1-2* complementation lines.

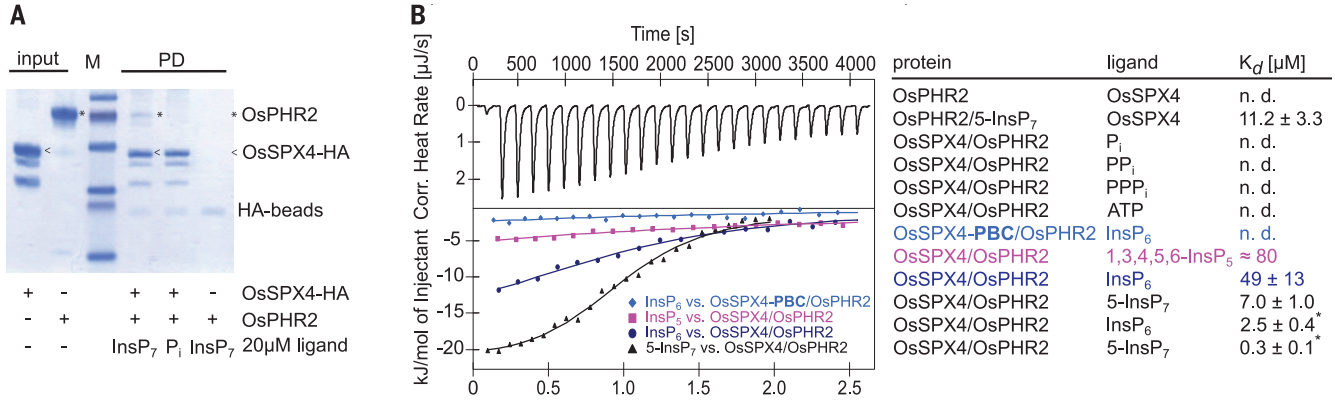
Residues in contact with SO_4^{2-} are invariant among SPX domains from different eukaryotes (fig. S2), and thus, the sulfate ion could mark the previously suggested P_i binding site (*15*, *16*).

Nuclear magnetic resonance (NMR) titrations of wild-type SPX^{ScVtc2} and SPX^{HsXPR1} revealed a dissociation constant (K_d) for P_i of ~5 and 20 mM, respectively (Fig. 1E and fig. S4). Mutation of PBC residues in SPX^{ScVtc2} reduces P_i binding, as does deletion of the N-terminal α1 helix (ΔN14) or reductive methylation of surface lysines (Fig. 1E). The reduced binding affinity of the ΔN14 mutant, the observed structural changes in our SPX domain crystal structures, and the number of peak shifts in the NMR titration experiments indicate that SPX domains undergo conformational changes upon ligand binding, possibly involving the α1 helix (Fig. 1, A to E, and figs. S3 and S4). In contrast to earlier reports (*15*, *16*), we found that SPX^{ScVtc2} cannot discriminate between P_i and sulfate (Fig. 1E and fig. S4). Also, the SPX basic surface cluster contains many more conserved lysine residues than required for P_i coordination [lysine surface cluster (KSC); $Lys^{ScVtc2}127$, $Lys^{ScVtc2}130$, and $Lys^{ScVtc2}134$], and mutation of all three KSC lysines to Ala has no effect on P_i binding (Fig. 1, D and E, and fig. S4). Larger ligands, such as pyrophosphate (PP_i), interact more tightly with SPX^{ScVtc2} (Fig. 1E and fig. S4). Thus, SPX domains harbor a large, positively charged surface able to interact with a phosphate-containing ligand but show little specificity and selectivity for P_i itself.

In order to define other potential ligands for eukaryotic SPX domains, we analyzed polyP synthesis in isolated yeast vacuoles, which is catalyzed by the SPX-containing VTC complex and regulated by P_i levels in vivo. Deletion of the $InsP_6$ kinase Kcs1, which produces inositol pyrophosphate (PP-InsP) signaling molecules, eliminates VTC-generated polyP stores (fig. S5) (*18*–*20*). We synthesized 5-$InsP_7$, a Kcs1 reaction product, and found that it stimulated VTC-catalyzed polyP synthesis at concentrations above 100 nM (Fig. 2A and fig. S5). In contrast, P_i and SO_4^{2-} activated VTC only at millimolar concentrations (Fig. 2A). SPX^{ScVtc2} and other SPX domains bind 5-$InsP_7$ and $InsP_6$ with ~50 to 500 nM affinity and with 1:1 binding stoichiometry in different in vitro assays (Fig. 2B, figs. S4 and S6, and table S2). SPX^{ScVtc2} discriminates between 5-$InsP_7$/$InsP_6$ and $InsP_5$, $InsP_4$, or $InsP_3$, and consequently, these molecules cannot stimulate VTC-catalyzed polyP synthesis (figs. S6 and S7 and table S2).

Isolated SPX domains recognize 5-$InsP_7$ and $InsP_6$ with similar binding affinities, and $InsP_6$ stimulates VTC activity above 100 µM. We thus used $InsP_6$ in our subsequent crystallization experiments as a commercially available substitute for 5-$InsP_7$, of which we could only synthesize low-milligram quantities. Complex structures of SPX^{CtGde1} and SPX^{CtVtc4} reveal that $InsP_6$ interacts with the basic surface cluster via variable hydrogen-bond interactions (Fig. 2, C to F; fig. S8; and table S1). This results in a high structural plasticity of the binding site, as seen in pleckstrin homology domains (*21*). The C2-attached axial phosphate group of $InsP_6$ is

Downloaded from http://science.sciencemag.org/ on May 19, 2016



protein	ligand	K_d [μM]
OsPHR2	OsSPX4	n. d.
OsPHR2/5-InsP$_7$	OsSPX4	11.2 ± 3.3
OsSPX4/OsPHR2	P$_i$	n. d.
OsSPX4/OsPHR2	PP$_i$	n. d.
OsSPX4/OsPHR2	PPP$_i$	n. d.
OsSPX4/OsPHR2	ATP	n. d.
OsSPX4-**PBC**/OsPHR2	InsP$_6$	n. d.
OsSPX4/OsPHR2	1,3,4,5,6-InsP$_5$	≈ 80
OsSPX4/OsPHR2	InsP$_6$	49 ± 13
OsSPX4/OsPHR2	5-InsP$_7$	7.0 ± 1.0
OsSPX4/OsPHR2	InsP$_6$	2.5 ± 0.4*
OsSPX4/OsPHR2	5-InsP$_7$	0.3 ± 0.1*

Fig. 4. InsPs trigger the association of plant SPX proteins with PHR transcription factors. (**A**) In vitro pull-down of full-length OsSPX4 containing a C-terminal hemagglutinin (HA)–tag. Interaction with full-length untagged OsPHR2 in the presence of either 20 μM 5-InsP$_7$ or P$_i$, respectively. (**B**) Isothermal titration calorimetry of OsSPX4/OsPHR2 versus different ligands (5-InsP$_7$, black; InsP$_6$, dark blue; 1,3,4,5,6-InsP$_5$, magenta; OsSPX4-PBC mutant versus InsP$_6$, light blue) and including table summaries for different titrations (± fitting errors; n.d., no detectable binding). The asterisks indicate experiments performed by means of microscale thermophoresis.

anchored in the PBC pocket occupied by SO_4^{2-} in the SPX^{HsXPR1} structure (Figs. 1D and 2, C to F, and fig. S8). PBC and KSC residues together form the SPX InsP binding site (Fig. 2, C to F). Additional hydrogen bonds with InsP$_6$ are established by the N-terminal amino group and by main-chain atoms from residues 1 to 4 in helix α1 (Fig. 2F). Binding of InsP ligands may thus stabilize the SPX α-helical hairpin motif.

Substitution of conserved surface cluster residues reduced InsP$_6$ and 5-InsP$_7$ binding to SPX^{ScVtc2} in vitro (Fig. 3, A and B; fig. S9; and table S3). Corresponding mutations in SPX^{ScVtc3} and SPX^{ScVtc4} in yeast cells abolished stimulation of polyP synthesis by 5-InsP$_7$, whereas mutation of Lys126^{ScVtc3}/Lys129^{ScVtc4} to Ala constitutively activated polyP production (Fig. 3C). Thus, VTC polyP synthesis is controlled by its SPX domains and InsPs, rationalizing previous genetic findings (*18*, *22*).

The *Arabidopsis* PHOSPHATE 1 (PHO1) protein, an ortholog of the human XPR1 P$_i$ exporter, controls P$_i$ transport from the root to the shoot and P$_i$-starvation responses (*5*, *12*). We substituted InsP-binding residues in the SPX^{AtPHO1} basic surface cluster and transformed the *pho1* knockout mutant (*5*) with these mutant AtPHO1 constructs. K136 → Ala control plants resembled plants transformed with a wild-type AtPHO1 construct, whereas the PBC, KSC, and ΔN15 mutants showed reduced growth (Fig. 3D and fig. S10). PBC and KSC plants had very similar phenotypes, suggesting that the entire surface cluster and not only the PBC residues are required for normal SPX domain function of AtPHO1 in vivo (Fig. 3D). This favors a role for InsPs rather than for P$_i$ in AtPHO1-mediated plant P$_i$ homeostasis. The ΔN15 mutant phenotype underlines the importance of helix α1 in SPX domain function, further supporting the idea of an InsP-induced conformational change (Fig. 3D). Shoot P$_i$ content in PBC, KSC, and ΔN15 plants was reduced compared with that of the wild type, and similar to *pho1* plants, suggesting that P$_i$ export into the xylem is compromised in these mutants (Fig. 3E). Expression of P$_i$ starvation–induced genes was up-regulated in shoots of PBC, KSC, and ΔN15 mutant lines, to an even larger extent when compared with the *pho1* knockout mutant (Fig. 3F). Targeting the ligand-binding function of SPX^{AtPHO1} thus impairs P$_i$ export and enhances P$_i$ starvation signaling in the P$_i$-deficient *pho1* mutant. P$_i$ efflux was also found impaired in patients with brain calcification–carrying mutations in SPX^{HsXPR1}, indicating that InsP control of P$_i$ transport is conserved across evolution (fig. S11) (*23*).

In our SPX-InsP$_6$ complexes, one ligand face remains accessible for additional coordination by target proteins (Fig. 2, C to F), which may serve as coreceptors for InsP signaling molecules, similarly as observed for the plant jasmonate receptor (*24*) or for the cyclin-dependent kinase complex Pho80–Pho85 and its inhibitor Pho81 (*25*). SPX proteins from rice interact with the OsPHR2 transcription factor in the presence of 15 mM P$_i$ (*16*, *26*). We found that OsSPX4 and OsPHR2 interact at 5-InsP$_7$ concentrations as low as 20 μM (Fig. 4A). No significant binding was detected in calorimetry experiments upon titrating P$_i$, PP$_i$, PPP$_i$, or adenosine 5´-triphosphate (ATP) into a 1:1 OsSPX4/OsPHR2 protein solution (Fig. 4B and fig. S12). InsP$_6$ bound with a K_d of ~50 μM, whereas 5-InsP$_7$ bound with ~7 μM affinity. In contrast to SPX^{ScVtc2}, the OsSPX4–OsPHR2 complex can thus sense the presence of the pyrophosphate group, indicating that interacting proteins may assist SPX domains in selective InsP recognition (Figs. 2B and 4B and fig. S12) (*4*, *13*, *15*, *16*). InsP$_6$ binding was disrupted in the OsSPX4 PBC triple mutant, again suggesting that the basic surface cluster contributes to InsP sensing. Last, we titrated OsSPX4 into a solution containing OsPHR2 preincubated with a twofold molar access of 5-InsP$_7$. We observed formation of a 1:1 SPX domain–transcription factor complex with a K_d of ~15 μM, whereas in the absence of InsPs, no binding occurred (Fig. 4B). Thus, InsPs can promote the specific interaction of plant SPX proteins with their transcription factor targets. Given that the P$_i$ concentration in the plant cytosol may not exceed 100 μM, InsPs but not P$_i$ itself should represent physiological ligands for plant SPX proteins (*27*, *28*). This would explain why *Arabidopsis* InsP$_5$ kinase mutants, which contain less than normal InsP$_6$ and PP-InsPs (*19*) but have higher cellular P$_i$, nonetheless show constitutive P$_i$ starvation responses (*29*).

Our crystallographic, biochemical, and genetic analysis of fungal, plant, and human SPX domain–containing proteins suggests that these domains function as cellular receptors for InsP signaling molecules (*19*). We report that different SPX domains bind InsP$_6$ and 5-InsP$_7$ with similar affinities in vitro, but several observations suggest that PP-InsPs may be the relevant signaling molecules in vivo: First, the PP-InsP 5-InsP$_7$—but not InsP$_6$—activates the yeast VTC complex in vitro (Fig. 2A). Second, polyP accumulation persists upon deletion of the InsP$_6$-producing kinase Ipk1, but no polyP can be detected in yeast strains lacking Kcs1, which generates PP-InsPs (*18*, *20*). Third, in the presence of their transcription factor target proteins, plant SPX domains show a binding preference for PP-InsPs (Fig. 4). Last, InsP$_7$ levels decrease under P$_i$ starvation in yeast, whereas InsP$_6$ levels remain constant (table S4) (*18*). PP-InsPs may thus signal the cellular P$_i$ status by binding to SPX domains in P$_i$-sufficient conditions, enabling them to interact with a multitude of proteins that regulate P$_i$ homeostasis in eukaryotic cells (*4*, *13*, *15*, *16*, *26*). We speculate that the highly plastic basic binding surface may allow SPX domains to sense different InsP/PP-InsP isomers, which could convey signaling inputs from various physiological processes (*19*). Further analyses will be required to elucidate this aspect in molecular detail. Our work now opens venues for specifically manipulating InsP/PP-InsP signaling in eukaryotic cells in order to uncover additional interaction partners for SPX domains. The P$_i$ signaling pathways we propose here may prove useful for engineering phosphate use efficiency in crop plants.

REFERENCES AND NOTES

1. M. E. Lenburg, E. K. O'Shea, *Trends Biochem. Sci.* **21**, 383–387 (1996).
2. T.-J. Chiou, S.-I. Lin, *Annu. Rev. Plant Biol.* **62**, 185–206 (2011).
3. D. Secco *et al.*, *New Phytol.* **193**, 842–851 (2012).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

4. H. C. Hürlimann, B. Pinson, M. Stadler-Waibel, S. C. Zeeman, F. M. Freimoser, *EMBO Rep.* **10**, 1003–1008 (2009).
5. D. Hamburger, E. Rezzonico, J. MacDonald-Comber Petétot, C. Somerville, Y. Poirier, *Plant Cell* **14**, 889–902 (2002).
6. J. Liu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6571–E6578 (2015).
7. D. Giovannini, J. Touhami, P. Charnet, M. Sitbon, J.-L. Battini, *Cell Rep.* **3**, 1866–1873 (2013).
8. M. Hothorn *et al.*, *Science* **324**, 513–516 (2009).
9. N. Ogawa *et al.*, *Mol. Cell. Biol.* **15**, 997–1004 (1995).
10. B. S. Park, J. S. Seo, N.-H. Chua, *Plant Cell* **26**, 454–464 (2014).
11. K. Duan *et al.*, *Plant J.* **54**, 965–975 (2008).
12. S. Wege *et al.*, *Plant Physiol.* **170**, 385–400 (2016).
13. D. D. Wykoff, A. H. Rizvi, J. M. Raser, B. Margolin, E. K. O'Shea, *Mol. Cell* **27**, 1005–1013 (2007).
14. V. Rubio *et al.*, *Genes Dev.* **15**, 2122–2133 (2001).
15. M. I. Puga *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14947–14952 (2014).
16. Z. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14953–14958 (2014).
17. L. Holm, S. Kääriäinen, P. Rosenström, A. Schenkel, *Bioinformatics* **24**, 2780–2781 (2008).
18. A. Lonetti *et al.*, *J. Biol. Chem.* **286**, 31966–31974 (2011).
19. S. B. Shears, *Adv. Biol. Regul.* **57**, 203–216 (2015).
20. C. Auesukaree, H. Tochio, M. Shirakawa, Y. Kaneko, S. Harashima, *J. Biol. Chem.* **280**, 25127–25133 (2005).
21. M. A. Lemmon, K. M. Ferguson, C. S. Abrams, *FEBS Lett.* **513**, 71–76 (2002).
22. N. Ogawa, J. DeRisi, P. O. Brown, *Mol. Biol. Cell* **11**, 4309–4321 (2000).
23. A. Legati *et al.*, *Nat. Genet.* **47**, 579–581 (2015).
24. L. B. Sheard *et al.*, *Nature* **468**, 400–405 (2010).
25. Y.-S. Lee, S. Mulugu, J. D. York, E. K. O'Shea, *Science* **316**, 109–112 (2007).
26. Q. Lv *et al.*, *Plant Cell* **26**, 1586–1597 (2014).
27. J. Pratt *et al.*, *Plant Physiol.* **151**, 1646–1657 (2009).
28. M. Desai *et al.*, *Plant J.* **80**, 642–653 (2014).
29. J. Stevenson-Paulik, R. J. Bastidas, S.-T. Chiou, R. A. Frye, J. D. York, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12612–12617 (2005).

ACKNOWLEDGMENTS

We thank G. Bräunlich for help with data analysis and J. Piehler for a sample of tris-NTA-Oregon Green488. Atomic coordinates and structure factors have been deposited with the Protein Data Bank with accession nos. 5IIG (SPX-TTM^{ScVtc4} form A), 5IIQ (SPX-TTM^{ScVtc4} form B), 5IIT (SPX^{ScVtc4}-macro), 5IJH (SPX^{HsXPR1} SO_4^{2-}), 5IJJ (SPX^{CtGde1} $InsP_6$), and 5IJP (SPX^{CtVtc4} $InsP_6$). We thank S. Wege for assistance with confocal microscopy and P_i export assays. This project was funded by grants from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC grant agreement 310856, the European Molecular Biology Organisation (EMBO) Young Investigator Programme (to M.H.), the Swiss National Fund (SNF) grants 31003A-138339 and 31003A-159998 (to Y.P.), an SNF Professorship PP00P2 157607 (to H.J.J.), and from ERC grant 233458 and SNF grant 144258 (to A.M.). Diffraction experiments were performed at beam lines PXII and PXIII of the Swiss Light Source, Paul Scherrer Institute, Villigen, and at beam-line ID29 of the European Synchrotron Radiation Facility, Grenoble, France. The supplementary materials contain additional data.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/986/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 to S4
References (*30*–*53*)

2 December 2015; accepted 31 March 2016
Published online 14 April 2016
10.1126/science.aad9858

SIGNAL TRANSDUCTION

Thresholds and ultrasensitivity from negative cooperativity

S. H. Ha[1] and J. E. Ferrell Jr.[1,2]*

Negative cooperativity is a phenomenon in which the binding of one or more molecules of a ligand to a multimeric receptor makes it more difficult for subsequent ligand molecules to bind. Negative cooperativity can make a multimeric receptor's response more graded than it would otherwise be. However, through theory and experimental results, we show that if the ligand binds the receptor with high affinity and can be appreciably depleted by receptor binding, then negative cooperativity produces a qualitatively different type of response: a highly ultrasensitive response with a pronounced threshold. Because ultrasensitivity and thresholds are important for generating various complex systems-level behaviors, including bistability and oscillations, negative cooperativity may be an important ingredient in many types of biological responses.

Receptors, signal transducers, and transcription factors are often present as oligomers, so understanding the interactions of multimeric complexes with their regulators is fundamental for understanding cellular regulation. Two simple schemes for the sequential interaction (*1*, *2*) of a stably dimeric receptor with a monomeric ligand are shown in Fig. 1, A and B. If the two equilibrium constants are equal ($K_1 = K_2$), the binding events can be viewed as independent; if the first binding is weaker than the second ($K_1 > K_2$), there is positive cooperativity; and if the first binding is stronger than the second ($K_1 < K_2$), there is negative cooperativity. The parameter $c = K_1/K_2$ can be taken as a measure of the cooperativity, with $c < 1$ corresponding to negative cooperativity and $c > 1$ to positive cooperativity.

If it is assumed that each receptor subunit is activated independently by ligand binding—i.e., that the singly-bound receptor is half as active as the doubly-bound receptor (model 1, Fig. 1A)—then the stimulus-response relations for various assumed degrees of cooperativity are as shown in Fig. 1C. If instead only the doubly-bound receptor is active (model 2, Fig. 1B), the relations are as shown in Fig. 1D. In either case, the greater the positive cooperativity, the more switchlike or ultrasensitive the response is (Fig. 1, C and D, blue curves), and as c approaches infinity, the effective Hill exponent (n_H) for the response approaches 2. Conversely, the stronger the negative cooperativity, the more graded the response is (Fig. 1, C and D, red curves). In model 1, high negative cooperativity makes the response subsensitive (*3*), with an effective Hill exponent less than 1. The system becomes less decisive but is better able to discriminate between various high levels of stimulus than it otherwise could. In model 2, the Hill exponent never falls below 1.

Implicit in this approach is the assumption that the number of ligand molecules is much larger than the number of receptors—either the ligand concentration is higher, or the ligand is distributed through a larger volume, or both—so that receptor binding has a negligible effect on the concentration of free ligand. This is a reasonable assumption for the binding of oxygen to hemoglobin, metabolites to metabolic enzymes, or circulating drugs to receptor proteins, for instance. But in some cases, particularly in intracellular signaling, the upstream regulator may be comparable to or even lower in concentration than the protein (or other target) that it is regulating. We therefore examined how the stimulus-response curves would be affected if the ligand were not assumed to be in infinite supply.

We derived expressions for the relation between the total ligand concentration L_{tot} (rather than the free ligand concentration L) and the equilibrium fraction of receptor in the unbound, singly-bound, and doubly-bound states (supplementary materials). Because ligand depletion is most pronounced when the affinity of the ligand is high, we initially examined the response in the limit where the K values approach zero.

If it is assumed that the receptor subunits are activated independently (Fig. 1A), the stimulus-response curves are unaffected by the cooperativity (model 3, Fig. 1E). On a linear plot, the response is a simple linear increase in receptor activity with total ligand concentration, until full activation is attained (not shown). Thus, the classical connection between positive cooperativity and ultrasensitivity is broken.

If it is assumed that two binding events are required to activate the receptor (Fig. 1B), and there is negative cooperativity in the binding, then the stimulus-response curve acquires a sharp threshold (model 4, Fig. 1F). This can perhaps be best appreciated on a linear plot, as shown in Fig. 1G. The smaller the value of c,

[1]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA 94305-5174, USA. [2]Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305-5307, USA.
*Corresponding author. Email: james.ferrell@stanford.edu

Downloaded from http://science.sciencemag.org/ on May 19, 2016

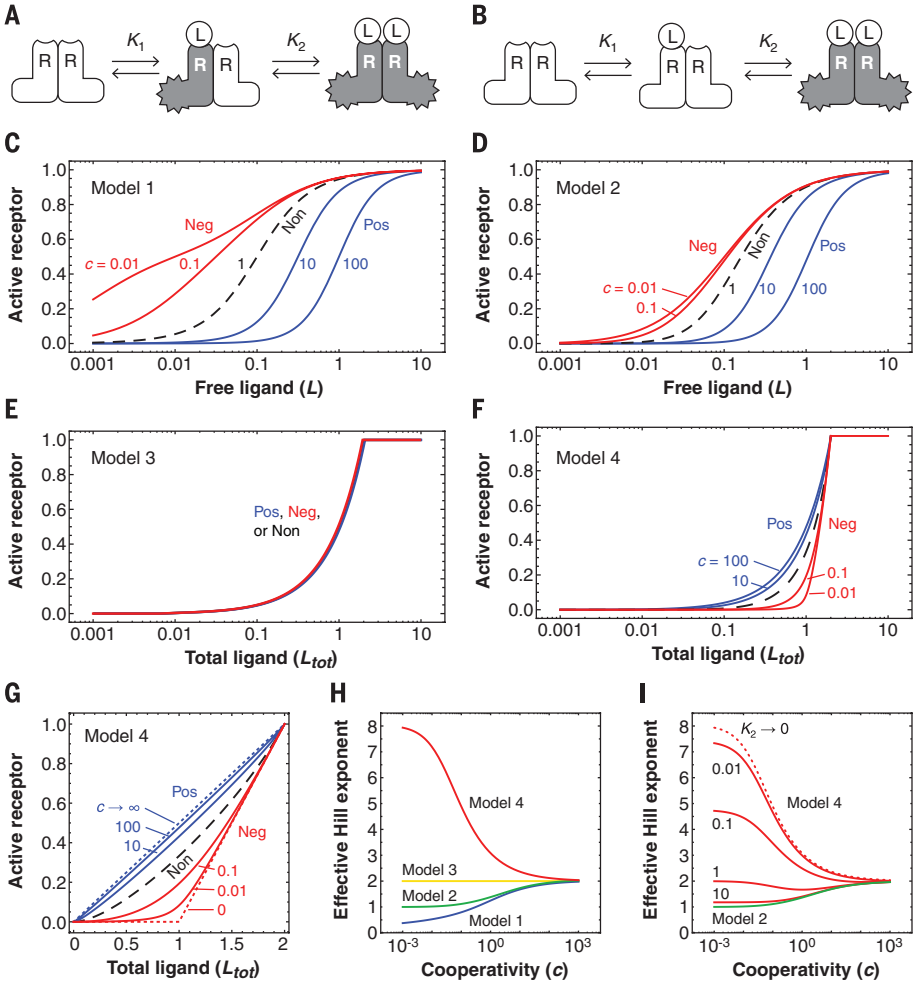


Fig. 1. Negative cooperativity can produce a sharp threshold in the response of a dimeric receptor to a high-affinity ligand. (**A** and **B**) Schematic of the sequential binding of a ligand (labeled L) to a stable dimeric receptor (labeled R), assuming (A) the receptor subunits are activated independently by ligand binding, or (B) the receptor is activated only when two ligands are bound. (The white receptors are inactive, whereas the gray receptors are active.) (**C** to **F**) Receptor activity as a function of [(C) and (D)] free ligand concentration or [(E) and (F)] total ligand concentration for various assumed degrees of cooperativity. In each panel, the blue curves correspond to positive cooperativity (Pos); the dashed black curve corresponds to independent noncooperative binding with c = 1 (Non); and the red curves correspond to negative cooperativity (Neg). In (C) and (D), it is assumed that K_2 = 0.1; higher or lower values would shift the curves to the right or left but would not alter the shapes of the curves. In (E) and (F), it is assumed that the binding affinities are very high and the equilibrium constants approach zero. (**G**) Linear plots of receptor activation as a function of total ligand concentration for model 4. In (C) to (G), x and y axis values are in arbitrary and fractional units, respectively. (**H**) The effective Hill exponents as a function of the cooperativity c for the four models. The Hill exponents are calculated as described in (*3*) and (*24*): $n_H = \log_{10}(81)/\log_{10}(EC_{90}/EC_{10})$, where EC_{90} is the concentration of ligand that yields a 90%-maximal response, and EC_{10} is the concentration that yields a 10%-maximal response. (**I**) The effect of decreasing the ligand-receptor affinity on the effective Hill exponents for model 4. Details of the theory can be found in the supplementary materials. Log-log plots of the stimulus-response curves and plots of the local response sensitivities are shown in fig. S1.

the sharper the threshold is. The effective Hill exponent for model 4's response is smallest when there is a high degree of positive cooperativity in the ligand binding, and it increases as c decreases, approaching a maximum value of ~8.04 as c approaches zero. Comparing all four models (Fig. 1H), it is negative cooperativity, rather than positive cooperativity, that produces the most highly ultrasensitive responses. Similar conclusions can be drawn using a local definition of response sensitivity, $d\ln(\text{activity})/d\ln(L)$ or $d\ln(\text{activity})/d\ln(L_{tot})$ (fig. S1) (*4*).

So far, we have assumed that the equilibrium constants for ligand binding are vanishingly small. If lower affinities are assumed, so that the receptor is less effective at depleting low concentrations of ligand, the ultrasensitivity of the response is lessened (Fig. 1I). Ultimately, the binding curves and effective Hill exponents obtained with model 4 approach those obtained with model 2 (Fig. 1I), as expected.

An intuitive explanation of these findings is illustrated in Fig. 2, A and B. If there is strong negative cooperativity, then the first site acts as a stoichiometric buffer, soaking up the first increments of the depletable ligand without producing a response (*5*–*7*). Only when the concentration of ligand exceeds the capacity of this buffer can the second binding event, and the consequent receptor activation, occur.

To experimentally test these theoretical findings, we engineered the high-affinity binding of two ligand molecules to a receptor under conditions of independent binding, positively cooperative binding, and negatively cooperative binding, and we quantitatively assessed the shapes of the binding curves. To accomplish this, we used DNA annealing, which made it easy to obtain high affinities and to manipulate the cooperativity. The basic idea was to use one strand of DNA as the equivalent of a dimeric receptor and use complementary strands as ligands. To avoid the possible formation of a stable hairpin structure between a ligand molecule and both binding sites on the receptor, we used a receptor DNA strand that bound two different ligands with similar affinities at two adjacent binding sites (*8*). We made the binding positively cooperative by using ligands that would abut each other when bound, allowing for favorable base-stacking interactions (*9*, *10*); noncooperative by engineering a one- or two-nucleotide gap between the ligand binding sites; or negatively cooperative, or at least less positively cooperative, by having the two ligand binding sites overlap by one to six nucleotides (Fig. 2C). We then measured, by means of native polyacrylamide gel electrophoresis, staining, and densitometry, how the equilibrium concentration of doubly-bound receptors varied with the total ligand concentration.

For the experiments shown in Fig. 2, the total concentration of the receptor oligonucleotide was held constant at 1000 nM, and the concentrations of each of the complementary ligand oligonucleotides were varied together. Pilot experiments showed that maximal complex formation was achieved at a ligand concentration of 1000 nM (yielding a stoichiometry of the two ligands and the receptor of 1:1:1), as expected from theory, so we focused on a detailed dose-response relation for ligand concentrations between 0 and 1000 nM. The theoretical binding equation (eq. S12) was then fitted to the data, with the cooperativity c as the only adjustable parameter. Examples of the primary data for pairs of ligands that were expected to exhibit positive cooperativity (ligands 14 and 28), independent binding (ligands 14 and 26), or negative cooperativity (ligands 17 and 28) are shown in Fig. 2, D to F. Figure 2G shows the cumulative data for all seven ligand pairs, from five (all of the ligand

Downloaded from http://science.sciencemag.org/ on May 19, 2016

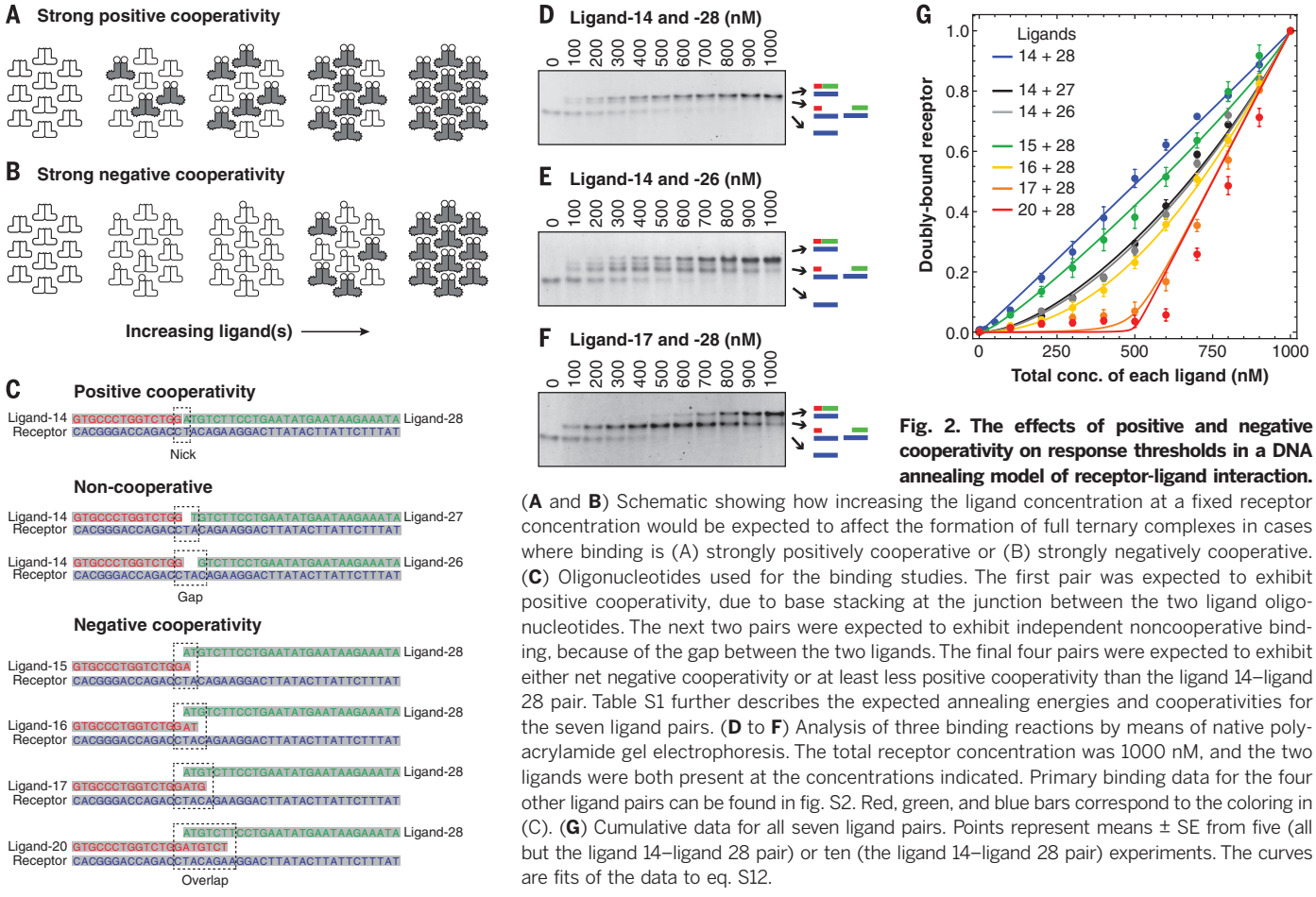


Fig. 2. The effects of positive and negative cooperativity on response thresholds in a DNA annealing model of receptor-ligand interaction. (**A** and **B**) Schematic showing how increasing the ligand concentration at a fixed receptor concentration would be expected to affect the formation of full ternary complexes in cases where binding is (A) strongly positively cooperative or (B) strongly negatively cooperative. (**C**) Oligonucleotides used for the binding studies. The first pair was expected to exhibit positive cooperativity, due to base stacking at the junction between the two ligand oligonucleotides. The next two pairs were expected to exhibit independent noncooperative binding, because of the gap between the two ligands. The final four pairs were expected to exhibit either net negative cooperativity or at least less positive cooperativity than the ligand 14–ligand 28 pair. Table S1 further describes the expected annealing energies and cooperativities for the seven ligand pairs. (**D** to **F**) Analysis of three binding reactions by means of native polyacrylamide gel electrophoresis. The total receptor concentration was 1000 nM, and the two ligands were both present at the concentrations indicated. Primary binding data for the four other ligand pairs can be found in fig. S2. Red, green, and blue bars correspond to the coloring in (C). (**G**) Cumulative data for all seven ligand pairs. Points represent means ± SE from five (all but the ligand 14–ligand 28 pair) or ten (the ligand 14–ligand 28 pair) experiments. The curves are fits of the data to eq. S12.

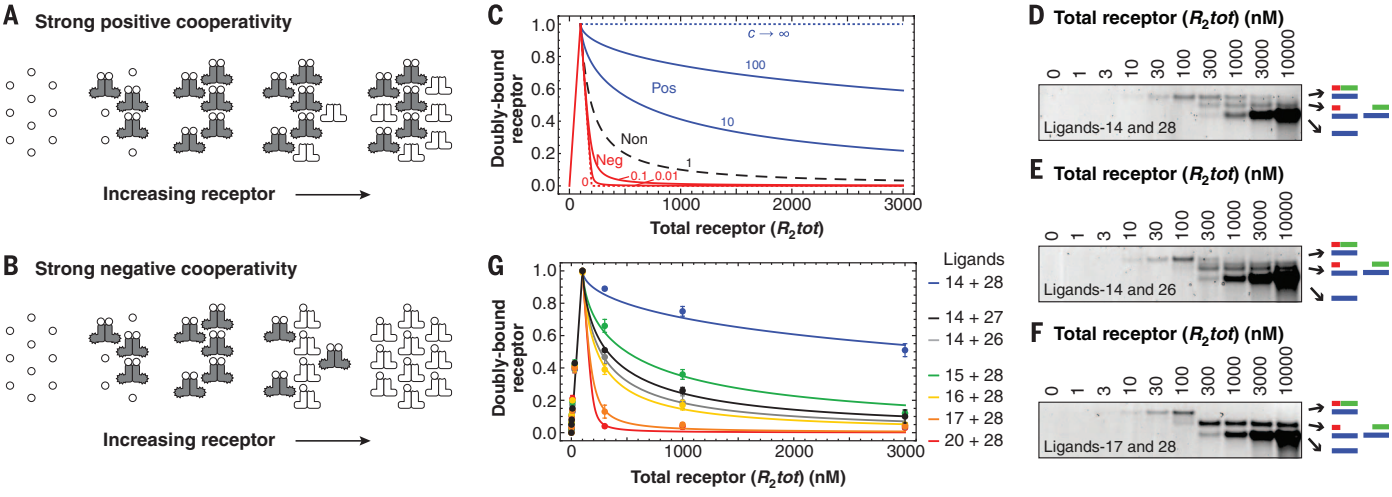


Fig. 3. Biphasic responses as the total receptor concentration is varied. (**A** and **B**) Schematic of the formation of ternary complexes as a function of receptor concentration when binding is (A) strongly positively cooperative or (B) strongly negatively cooperative. (**C**) Theoretical binding curves based on model 4 and eq. S12. The total ligand concentration was assumed to be 200 arbitrary units. (**D** to **F**) Analysis of three binding reactions by means of native polyacrylamide gel electrophoresis. In each case, the concentration of each ligand was 100 nM, and the total receptor concentration was varied as indicated. Red, green, and blue bars correspond to the coloring in Fig. 2C. Primary binding data for the four other ligand pairs can be found in fig. S3. (**G**) Cumulative binding data for all seven ligand pairs. Points represent means ± SE from three to eight experiments. The curves are fits of the data to eq. S12.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

pairs except ligand 14–ligand 28) or ten (ligand 14–ligand 28 pair) experiments.

The ligands that were expected to exhibit positive cooperativity (14 and 28) had a linear stimulus-response relation (Fig. 2G, blue curve); those that were expected to bind independently (14 paired with 26 or 27) yielded nonlinear stimulus-response curves with c values closer to 1 (Fig. 2G, black and gray curves, and table S1); and those expected to exhibit negative (or less positive) cooperativity (15, 16, 17, or 20 paired with 28) yielded a series of curves with increasingly marked thresholds and decreasing inferred values of c (Fig. 2G, green, yellow, orange, and red curves). Some of the inferred c values were higher than expected from simple energetic considerations (table S1) (*11*, *12*). Nevertheless, the data demonstrate that negative cooperativity can, as predicted, imbue a marked threshold and a high degree of ultrasensitivity on the formation of a ternary complex between a receptor and two high-affinity ligands.

Another feature of the classical sequential models (models 1 and 2) is that receptor activation increases linearly with the total concentration of receptor (eqs. S7 and S8). This is not true, however, for model 4, which considers ligand depletion. In the high-affinity limit, the response is predicted to be biphasic. As the receptor concentration increases, the output initially increases, peaks when the stoichiometric ratio of receptor to ligand is 1:2, and decreases thereafter (Fig. 3, A to C) (*13*). The narrowness of the response peak depends upon the cooperativity of the binding; positive cooperativity makes the fall-off in output at high receptor concentrations occur gradually, and negative cooperativity makes the fall-off more abrupt (Fig. 3, A to C). This phenomenon is conceptually related to the prozone effect, which was discovered in early studies of antigen-antibody interactions (*14*, *15*); to the phenomenon of transcriptional squelching (*16*–*18*); and to the inhibition of protein kinase signaling by overexpression of scaffold proteins (*19*–*22*). Essentially, the receptor is behaving like a bivalent adaptor, and high concentrations of the receptor drive the formation of binary ligand-receptor complexes at the expense of the full ternary complex.

We therefore determined whether complex formation was a biphasic function of receptor concentration, again using the oligonucleotide ligands shown in Fig. 2C. The formation of the full ternary complex was biphasic, in good agreement with theory (Fig. 3, D to G). The ligands that were expected to exhibit the highest degrees of negative cooperativity produced the sharpest biphasic peaks, and, as shown in table S1, the inferred cooperativity values agreed reasonably well with those obtained from the threshold data in Fig. 2.

Our results show that for the high-affinity interaction of a multimeric receptor with a depletable ligand, marked thresholds and high degrees of ultrasensitivity can arise as a result of negative cooperativity in the binding reactions. Such thresholds can allow a system to filter out small stimuli and yet respond decisively to suprathreshold stimuli. Moreover, the thresholds and ultrasensitivity that arise through this mechanism can be critical elements for the production of more complex systems-level behaviors, such as bistability and oscillations (*23*). In addition, multimeric receptors can exhibit a biphasic response to changes in the concentration of the receptor, and this biphasic response is sharpest when strong negative cooperativity is in effect. These basic aspects of the regulation of multimeric proteins may be important both for understanding the behavior of natural regulatory systems and for designing synthetic ones.

REFERENCES AND NOTES

1. D. E. Koshland Jr., G. Némethy, D. Filmer, *Biochemistry* **5**, 365–385 (1966).
2. A. Levitzki, D. E. Koshland Jr., *Proc. Natl. Acad. Sci. U.S.A.* **62**, 1121–1128 (1969).
3. D. E. Koshland Jr., A. Goldbeter, J. B. Stock, *Science* **217**, 220–225 (1982).
4. M. A. Savageau, *Curr. Top. Cell. Regul.* **6**, 63–130 (1972).
5. J. E. Ferrell Jr., *Trends Biochem. Sci.* **21**, 460–466 (1996).
6. N. E. Buchler, M. Louis, *J. Mol. Biol.* **384**, 1106–1119 (2008).
7. J. E. Ferrell Jr., S. H. Ha, *Trends Biochem. Sci.* **39**, 556–569 (2014).
8. N. Jullien, J. P. Herman, *Biotechniques* **51**, 267–269 (2011).
9. E. T. Kool, *Annu. Rev. Biophys. Biomol. Struct.* **30**, 1–22 (2001).
10. P. Yakovchuk, E. Protozanova, M. D. Frank-Kamenetskii, *Nucleic Acids Res.* **34**, 564–574 (2006).
11. R. Owczarzy, Y. You, C. L. Groth, A. V. Tataurov, *Biochemistry* **50**, 9352–9367 (2011).
12. IDT-Biophysics, "DNA thermodynamics and hybridization," 2015; http://biophysics.idtdna.com/.
13. S. H. Ha, S. Y. Kim, J. E. Ferrell Jr., *Cell Rep.* **14**, 1408–1421 (2016).
14. D. Bray, S. Lay, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13493–13498 (1997).
15. M. Heidelberger, F. E. Kendall, *J. Exp. Med.* **50**, 809–823 (1929).
16. M. A. Cahill, W. H. Ernst, R. Janknecht, A. Nordheim, *FEBS Lett.* **344**, 105–108 (1994).
17. S. Natesan, V. M. Rivera, E. Molinari, M. Gilman, *Nature* **390**, 349–350 (1997).
18. R. Prywes, H. Zhu, *Nucleic Acids Res.* **20**, 513–520 (1992).
19. A. Levchenko, J. Bruck, P. W. Sternberg, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5818–5823 (2000).
20. M. Dickens *et al.*, *Science* **277**, 693–696 (1997).
21. A. M. Cacace *et al.*, *Mol. Cell. Biol.* **19**, 229–240 (1999).
22. N. R. Michaud *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12792–12796 (1997).
23. J. E. Ferrell Jr., S. H. Ha, *Trends Biochem. Sci.* **39**, 612–618 (2014).
24. A. Goldbeter, D. E. Koshland Jr., *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6840–6844 (1981).

ACKNOWLEDGMENTS

We thank D. Herschlag and R. Owczarzy for helpful discussions and G. Anderson, B. Beltran, X. Cheng, and T. H. Lee for comments on the paper. This work was supported in part by grants from the U.S. NIH (R01 GM046383 and P50 GM107615) and from the National Research Foundation of Korea (NRF-2011-357-C00151 and 2012M3A9B9036680). S.H.H. and J.E.F. jointly designed the studies, made the figures, and wrote the paper. S.H.H. carried out the experiments and J.E.F. did the theoretical work.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/990/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S3
Table S1
Reference (*25*)

6 October 2015; accepted 7 April 2016
Published online 12 May 2016
10.1126/science.aad5937

DNA REPAIR

ppGpp couples transcription to DNA repair in *E. coli*

Venu Kamarthapu,[1,2] Vitaly Epshtein,[1] Bradley Benjamin,[1] Sergey Proshkin,[3] Alexander Mironov,[3] Michael Cashel,[4] Evgeny Nudler[1,2]*

The small molecule alarmone (p)ppGpp mediates bacterial adaptation to nutrient deprivation by altering the initiation properties of RNA polymerase (RNAP). ppGpp is generated in *Escherichia coli* by two related enzymes, RelA and SpoT. We show that ppGpp is robustly, but transiently, induced in response to DNA damage and is required for efficient nucleotide excision DNA repair (NER). This explains why *relA-spoT*-deficient cells are sensitive to diverse genotoxic agents and ultraviolet radiation, whereas ppGpp induction renders them more resistant to such challenges. The mechanism of DNA protection by ppGpp involves promotion of UvrD-mediated RNAP backtracking. By rendering RNAP backtracking-prone, ppGpp couples transcription to DNA repair and prompts transitions between repair and recovery states.

Bacteria respond to starvation by rapidly accumulating guanosine-3′,5′-(bis)pyrophosphate (ppGpp). This results in the inhibition of ribosomal and transfer RNA transcription and metabolic transition, known as the stringent response, which allows bacteria to conserve energy and adapt to adverse conditions (*1*, *2*). The global reprogramming of gene expression is a consequence of ppGpp binding to RNA polymerase (RNAP), which causes destabilization of open promoter complexes (*3*, *4*). ppGpp acts synergistically with the transcription factor DksA, which also interacts with RNAP (*5*, *6*). Evidence suggests ppGpp also functions independently in preserving genomic integrity (*7*–*10*). Here, we demonstrate that ppGpp is crucial for the process of nucleotide excision DNA repair (NER).

The majority of bulky DNA lesions are eliminated by NER (*11*). *Escherichia coli* resistance to helix-distorting mutagens, such as ultraviolet (UV) radiation, 4-nitroquinoline-1-oxide (4NQO), and nitrofurazone (NFZ), depends on the intracellular level of ppGpp (*8*, *10*) (Fig. 1A and fig. S1). Cells lacking *relA* and *spoT* (ppGpp^0) were sensitive to these stressors, whereas *relA256 spoT203*

Downloaded from http://science.sciencemag.org/ on May 19, 2016

cells, which produced excess ppGpp under normal exponential growth conditions, were more resistant than the parent control (Fig. 1A and fig. S1). ppGpp was transiently induced in response to these challenges (Fig. 1B). Its level increased more than 20 times within the first 15 min of genotoxic stress and then returned to basal levels within 45 min. Because DNA lesions caused by UV, 4NQO, and NFZ are eliminated predominantly by NER (*12*), we examined the potential role of ppGpp in NER.

RNAP facilitates the recognition of DNA damage by NER enzymes through transcription-coupled DNA repair (TCR) (*13*). ppGpp is a modulator of RNAP activity and, therefore, could support NER by functioning in TCR. To test this, we compared the ability of wild-type and ppGpp-deficient cells to remove the UV-induced cyclobutane pyrimidine dimers (CPDs) from the individual DNA strands of the *lac* operon (Fig. 1C and fig. S2). Wild-type isopropyl-β-D-thiogalactopyranoside (IPTG)–induced cells repaired CPDs in the transcribed strand of *lac* operon more rapidly than in the nontranscribed strand (*14*) (Fig. 1C and fig. S2). In contrast, ppGpp^0 cells displayed the same slow rate of repair on both strands, which was similar to the rate of repair of the nontranscribed strand in wild-type cells (Fig. 1C and fig. S2). Thus, ppGpp is important for TCR.

In the "forward translocation" TCR pathway, the translocase Mfd binds to the upstream edge of stalled RNAP (*15*) and pushes it forward against the lesion, which terminates transcription (*16*, *17*). Mfd then recruits UvrA to the site of damage (*16*) to initiate NER. The "backtracking" pathway relies on the 3′-5′ helicase UvrD, which binds RNAP and pulls it backward, away from the lesion site, which exposes the lesions to NER enzymes (*18*). To determine which of the two TCR pathways requires ppGpp, we combined the *relA spoT* deletions with either *uvrD* or *mfd* (Fig. 2A). The sensitivity of Δ*mfd* Δ*relAspoT* to 4NQO, NZF, or UV was approximately equal to that of the sum of the sensitivity of individual Δ*mfd* and ppGpp^0 mutant strains. An additive effect of Mfd and ppGpp indicates that the two factors act in different repair pathways. In contrast, the Δ*uvrD* Δ*relAspoT* mutant was as sensitive to genotoxic agents and UV as Δ*uvrD* alone (Fig. 2A and fig. S3). The epistatic relation between UvrD and ppGpp suggests that they act in the same, backtracking-mediated TCR pathway.

Transcript cleavage factors GreA and GreB interfere with UvrD-mediated TCR, as they compete with UvrD-mediated backtracking (*18*). Inactivation of these antibacktracking factors greatly suppressed the sensitivity of ppGpp^0 cells to being killed by genotoxic agents and UV (Fig. 2B and fig. S4); from this, we argue that ppGpp acts as a probacktracking factor in UvrD-mediated TCR. In bacteria, transcription and translation are coupled. The leading ribosome directly controls the rate of transcription elongation by preventing RNAP backtracking (*19*) and, thereby, also interferes with UvrD-mediated TCR (*18*). Analogous to the situation with Gre-deficiency, we show that slowing ribosome translocation with a sublethal dose of chloramphenicol (Cm) rendered ppGpp^0 cells more resistant to genotoxic chemicals and UV (Fig. 2C and fig. S5).

DksA has been shown to promote ppGpp activities in vivo and in vitro (*5*, *6*). Accordingly, DksA-deficient cells were sensitive to 4NQO, NFZ, and mitomycin C, whereas inactivation of GreA fully suppressed this sensitivity (*8*) (Fig. 2D and fig. S6). Moreover, DksA overexpression suppressed the sensitivity to genotoxic stress in the absence of ppGpp (fig. S7). Collectively, these in vivo results argue that ppGpp, in conjunction with DksA, promotes UvrD-mediated TCR by facilitating RNAP backtracking.

To test whether ppGpp assists UvrD in promoting RNAP backtracking, we examined the effect of ppGpp and UvrD on transcription elongation in a reconstituted single-round runoff assay (Fig. 3A and fig. S8). UvrD forced RNAP to stall at many sites so that only a small fraction of transcription complexes reached the end of the template to form a full-length runoff product (*18*). The majority of observed paused or arrested complexes were the result of UvrD-induced backtracking, hence, their sensitivity to GreB (Fig. 3B and table S1). Addition of 100 μM ppGpp, a concentration corresponding to the cellular level of ppGpp during the stringent response, greatly stimulated UvrD-mediated backtracking: Less than 20% of RNAP molecules traversed the middle of the template in the presence of ppGpp together with UvrD (Fig. 3A, lane 6), whereas more than 50% of RNAPs did so in the presence of UvrD alone (Fig. 3A, lane 3). DksA + ppGpp further promoted the probacktracking activity of UvrD (Fig. 3A, lanes 7 to 9). DksA alone potentiated UvrD-mediated backtracking only slightly (Fig. 3A, lanes 10 to 12).

The structural model of the ppGpp-RNAP complex suggests that ppGpp facilitates partial opening of the RNAP clamp (*4*), which may render RNAP prone to backtracking. To test this prediction, we monitored the front edge of a stable elongation complex (EC) stalled at position +32 (EC32) by probing it with exonuclease III (Fig. 3C and fig. S8). The footprinting of EC32 showed that RNAP was almost evenly distributed between post and pretranslocated states (lane 3). ppGpp shifts the equilibrium toward the pretranslocated state (lane 4). Longer incubation with exonuclease III resulted in progressive backtracking, which was more prominent in the presence of ppGpp than in its absence (lanes 5 and 6).

To confirm that ppGpp also facilitates RNAP backtracking in vivo, we used a plasmid (p1EC) in which the Lac repressor bound to its operator site blocks an isolated EC ~70 nucleotides

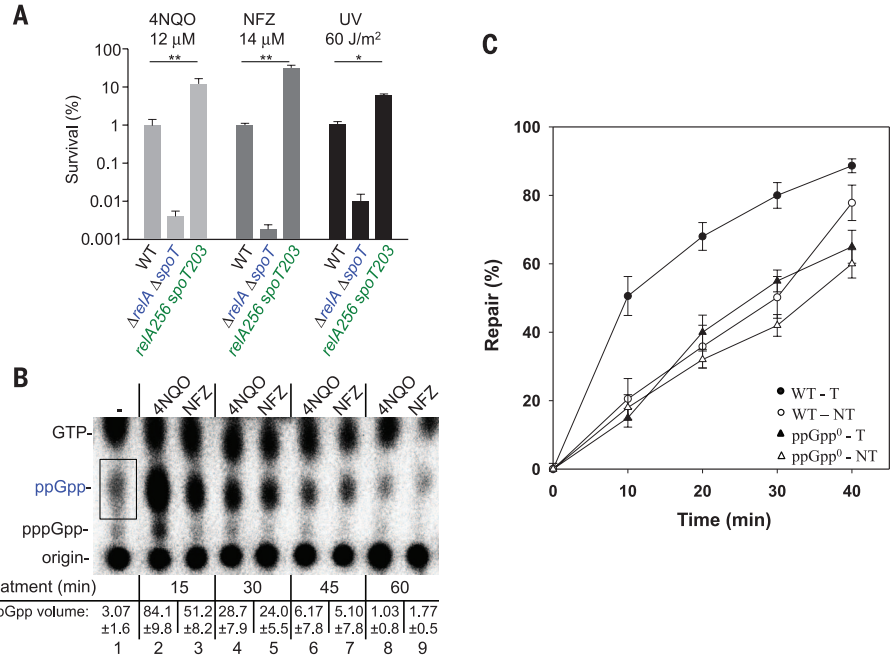


Fig. 1. ppGpp is required for genotoxic stress survival and TCR. (**A**) ppGpp^0 cells (Δ*relA* Δ*spoT*) or cells producing excessive amounts of ppGpp (*relA256spoT203*) (*29*) were treated with 4NQO, NFZ, or UV. Data from three independent experiments are presented as the means ± SEM; $^{**}P < 0.01$; $^{*}P < 0.05$. Representative efficiencies of colony formation are shown in fig. S1. (**B**) MG1655 cells were treated with 4NQO or NFZ for the indicated time. ppGpp was detected by thin-layer chromatography. The same area (rectangle) in each lane was used for ppGpp quantification. Values are the means (±SEM) of three independent experiments. (**C**) Strand-specific repair of the *lacZ* operon (6.6 kilobases of Apa I–Sst II fragment) was measured at 0, 10, 20, 30, and 40 min after UV irradiation (50 J/m^2). Data from three independent experiments (fig. S2) are presented as the means ± SEM. T, transcribed strand; NT, nontranscribed strand.

[1]Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY 10016, USA. [2]Howard Hughes Medical Institute, New York University School of Medicine, New York, NY 10016, USA. [3]Engelhardt Institute of Molecular Biology, Russian Academy of Science, Moscow 119991, Russia. [4]Division of Developmental Biology, Intramural Research Program, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA.
***Corresponding author. Email: evgeny.nudler@nyumc.org**

Downloaded from http://science.sciencemag.org/ on May 19, 2016

downstream of the transcription start site (Fig. 3D) (*20*). To monitor the effect of ppGpp and UvrD on the positioning of the halted EC, we performed in situ DNA footprinting using the single strand–specific probe, chloroacetaldehyde (CAA). Cells transformed with the plasmid pEC1 alone or pEC1 together with the UvrD overexpression plasmid (pUvrD) were treated with serine hydroxamate (SHX) to induce the stringent response (fig. S8). ppGpp induction caused the roadblocked EC to backtrack over a longer distance (Fig. 3D, compare lanes 3 and 4): CAA reactive sites were more prominent upstream of the footprint, and the reactivity of the downstream margin of the bubble was diminished. Backtracking was substantially more extensive in the SHX-treated cells that overexpressed UvrD (lane 5). SHX failed to promote backtracking in UvrD-deficient cells (fig. S9). Thus, as shown in vitro, ppGpp also facilitates UvrD-mediated RNAP backtracking in vivo. Consistently, SHX-pretreated cells are more resistant to genotoxic stress (fig. S10).

A stringent RNAP allele, *rpoB*35*, carries a point mutation [in which glutamine replaces His^{1244} (H1244Q)] in the beta subunit that mimics the presence of ppGpp (*7*). *rpoB*35* suppressed the sensitivity of $ppGpp^0$ cells to genotoxic agents and UV (*8*) (Fig. 2E and fig. S11). In the absence of UvrD, however, cells harboring *rpoB*35* were unable to effectively suppress their sensitivity to DNA-damaging agents and were almost as sensitive as $uvrD^-$ cells (fig. S12), which suggested that *rpoB*35* acted via UvrD-mediated TCR. To test this directly, we examined the effect of UvrD and ppGpp on *rpoB*35* RNAP in vitro. RpoB*35 was purified and used in a single-round assay (Fig. 3A). In the absence of UvrD, the RpoB*35 enzyme elongated slightly faster than wild-type enzyme (*21*) and was less responsive to certain strong pauses (compare Fig. 3A, lanes 1 and 13) (*21*). However, RpoB*35 became much more prone to backtracking in the presence of UvrD (Fig. 3A, lanes 14 and 15). The extent of UvrD-mediated backtracking by RpoB*35 was comparable to that of wild-type enzyme in the presence of ppGpp and DksA (compare Fig. 3A, lanes 8 and 9, as well as 14 and 15). Moreover, addition of ppGpp (with or without DksA) to the RpoB*35 reaction had no further effect on UvrD-mediated backtracking (lanes 17 and 18). We thus conclude that RpoB*35 mimics the effect of ppGpp on wild-type RNAP. These results provide direct support for the UvrD- and ppGpp-mediated TCR and explain how ppGpp couples transcription to DNA repair.

Here, we implicate ppGpp as an important component of UvrD-mediated TCR (fig. S13).

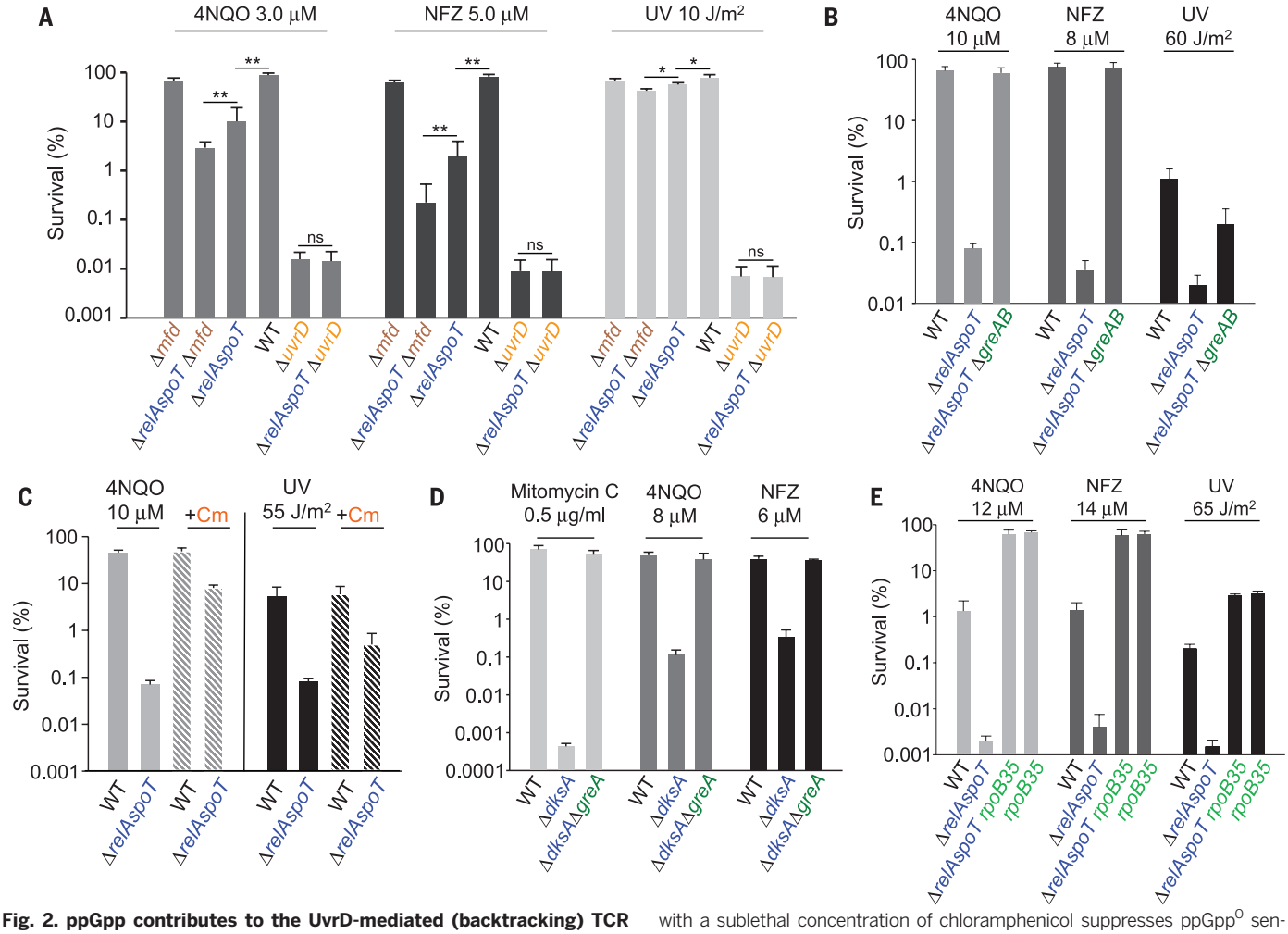


Fig. 2. ppGpp contributes to the UvrD-mediated (backtracking) TCR pathway. (**A**) MG1655 cells lacking Mfd, UvrD, or RelA and SpoT or combinations of Mfd-RelA-SpoT and UvrD-RelA-SpoT were exposed to the indicated amounts of 4NQO and NFZ. Data from three independent experiments are presented as the means ± SEM; $^{**}P < 0.01$; $^{*}P < 0.05$; ns, nonsignificant. Representative efficiencies of colony formation are shown in fig. S3. (**B**) Inactivating antibacktracking factors GreA and GreB suppress $ppGpp^0$ sensitivity to 4NQO, NFZ, and UV. Data from three independent experiments are presented as the means ± SEM ($P < 0.01$). Representative efficiencies of colony formation are shown in fig. S4. (**C**) Slowing ribosomal translocation with a sublethal concentration of chloramphenicol suppresses $ppGpp^0$ sensitivity to 4NQO or UV. Data from three independent experiments are presented as the means ± SEM ($P < 0.01$). Representative efficiencies of colony formation are shown in fig. S5. (**D**) Inactivating GreA suppresses the sensitivity of DksA-deficient cells to genotoxic stress caused by mitomycin C, 4NQO, or NFZ. Data from three independent experiments are presented as the means ± SEM ($P < 0.01$). Representative efficiencies of colony formation are shown in fig. S6. (**E**) The *rpoB*35* allele suppresses $ppGpp^0$ sensitivity to 4NQO, NFZ, and UV. Data from three independent experiments are presented as the means ± SEM ($P < 0.01$). Representative efficiencies of colony formation are shown in fig. S11.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

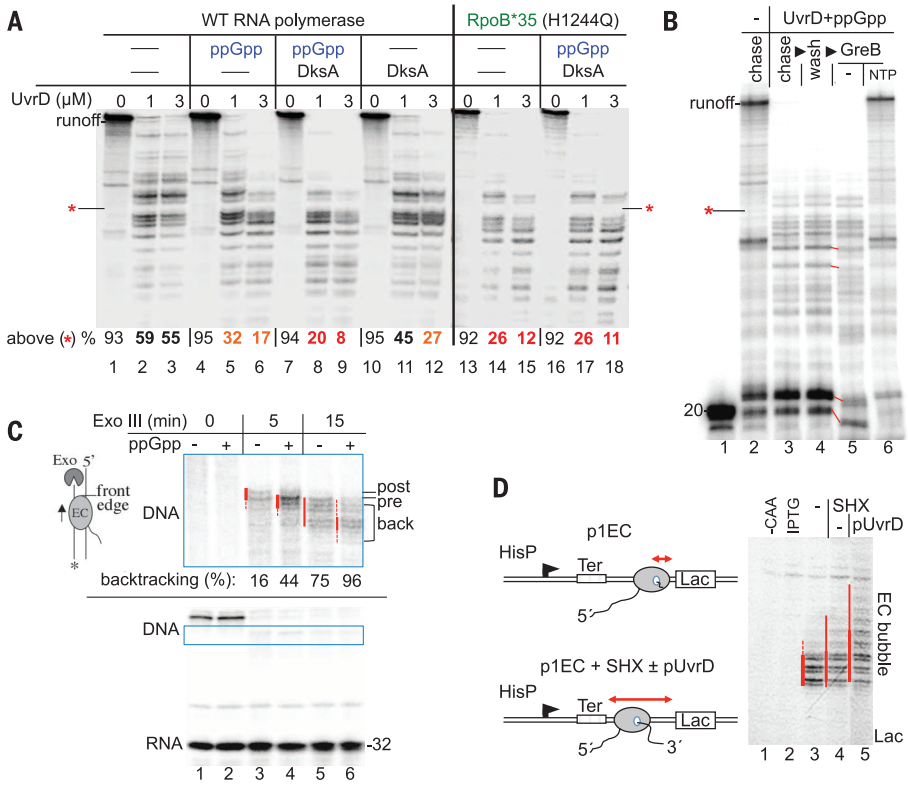


Fig. 3. ppGpp promotes UvrD-mediated RNAP backtracking in vitro and in vivo. (**A**) EC20 was formed by wild-type RNAP or RpoB*35 (lanes 13 to 18) at the T7A1 DNA template and then chased in the presence of specified amounts of UvrD. ppGpp and/or DksA were added to the chase reaction as indicated. The probacktracking activity of UvrD was assessed as a ratio (%) between the total amounts of RNA products located above and below the arbitrary midsection line indicated by red asterisks. Mean values ± SEM ($P < 0.05$) from three independent experiments are shown in table S1. (**B**) EC20 (lane 1) immobilized on Co^{++}-beads was chased with (lanes 3 to 6) or without UvrD+ppGpp (lane 2), followed by washing (lane 4). GreB without (lanes 5 and 6) or with nucleoside triphosphates (lane 6; second chase) was added. Red lines connect exemplary corresponding RNA from arrested ECs before and after GreB treatment (lanes 4 and 5) to illustrate transcript cleavage. The red asterisk denotes the same position as in (A). (**C**) The top and bottom panels resolve the same probes to show protection of the ^{32}P–end-labeled nontemplate DNA strand of EC32 from Exo III digestion in the presence or absence of ppGpp. Blue boxes correspond to the areas of Exo III footprinting at the front edge of EC32. The weight of the red lines illustrates the intensity of the footprint signal. Bands corresponding to the pre- and posttranslocated states, as well as backtracked species, are indicated. The extent of backtracking (%) was measured as the ratio between all backtracked signals and the total RNA signal in the corresponding lane. (**D**) The p1EC constructs (left) (*19*) and primer extension analyses (right). CAA modifications on the nontemplate strand of p1EC (lanes 2 and 3), p1EC in the presence of the ppGpp inducer, SHX (lane 4), or p1EC1+SHX+pUvrD (lane 5). The lac operator (Lac) and transcription bubble are indicated. Red lines show the position of blocked EC. Experimental workflows for A, B, C, and D are shown in fig. S8.

Our in vivo and in vitro data indicate that ppGpp and UvrD act together to promote backtracking and that this probacktracking activity accounts for most of their phenotypes with respect to DNA damage. Modeling suggests that RNAP undergoes a conformational change upon binding ppGpp, which widens its "claw-pincers" (*4*) and renders it prone to backtracking (*22*). In the absence of stress, the *E. coli* cell contains ~3×10^3 molecules of UvrD (*23*), at an approximately 1:1 ratio with RNAP. Considering the high affinity of UvrD for RNAP (K_d = 35 nM), most of the UvrD should be sequestered by RNAP under normal growth conditions. However, for UvrD to act as a helicase capable of backtracking, it must form a dimer (*18*, *24*). Such dimers are not stable ($K_d \approx 470$ nM) (*25*). As the induction of UvrD during the SOS response is only two- to threefold (*26*), UvrD dimerization is expected to occur only intermittently during stress. Thus, by lowering the energy barrier necessary to cause backtracking, ppGpp widens the critical window of opportunity for UvrD to act in TCR. DksA contributes substantially to UvrD- and/or ppGpp-mediated TCR (figs. S6 and S7), because it stabilizes RNAP in the backtracking-prone state imposed by ppGpp (Fig. 3A) (*4*) and competes with antibacktracking Gre factors for the same binding site on RNAP (*6*).

Codirectional collisions between the replication fork and backtracked RNAP often lead to DNA double-strand breaks (*21*). However, because ppGpp accumulates only during the critical phase of genotoxic stress (Fig. 1B), when it also inhibits replication (*9*, *27*), such detrimental collisions are effectively avoided. The timely decline of ppGpp triggers the recovery process (fig. S13), which relies on the concerted action of antibacktracking factors—GreAB, active ribosomes, and Mfd—that compete with probacktracking UvrD (Fig. 2 and figs. S4, S5, S14, and S15) (*18*, *19*, *28*). Thus, ppGpp not only helps to activate UvrD-TCR but also ensures that the process is deactivated promptly so that transcription may recover and that conflicts with replication are avoided.

REFERENCES AND NOTES

1. L. U. Magnusson, A. Farewell, T. Nyström, *Trends Microbiol.* **13**, 236–242 (2005).
2. K. Potrykus, H. Murphy, N. Philippe, M. Cashel, *Environ. Microbiol.* **13**, 563–575 (2011).
3. M. M. Barker, T. Gaal, C. A. Josaitis, R. L. Gourse, *J. Mol. Biol.* **305**, 673–688 (2001).
4. Y. Zuo, Y. Wang, T. A. Steitz, *Mol. Cell* **50**, 430–436 (2013).
5. B. J. Paul *et al.*, *Cell* **118**, 311–322 (2004).
6. A. Parshin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6862–E6871 (2015).
7. P. McGlynn, R. G. Lloyd, *Cell* **101**, 35–45 (2000).
8. B. W. Trautinger, R. P. Jaktaji, E. Rusakova, R. G. Lloyd, *Mol. Cell* **19**, 247–258 (2005).
9. J. DeNapoli, A. K. Tehranchi, J. D. Wang, *Mol. Microbiol.* **88**, 93–104 (2013).
10. K. E. Madison *et al.*, *Mol. Microbiol.* **92**, 28–46 (2014).
11. J. T. Reardon, A. Sancar, *Prog. Nucleic Acid Res. Mol. Biol.* **79**, 183–235 (2005).
12. D. P. Batty, R. D. Wood, *Gene* **241**, 193–204 (2000).
13. V. Kamarthapu, E. Nudler, *Curr. Opin. Microbiol.* **24**, 15–20 (2015).
14. I. Mellon, P. C. Hanawalt, *Nature* **342**, 95–98 (1989).
15. L. F. Westblade *et al.*, *Nucleic Acids Res.* **38**, 8357–8369 (2010).
16. C. P. Selby, A. Sancar, *Science* **260**, 53–58 (1993).
17. J. S. Park, M. T. Marr, J. W. Roberts, *Cell* **109**, 757–767 (2002).
18. V. Epshtein *et al.*, *Nature* **505**, 372–377 (2014).
19. S. Proshkin *et al.*, *Science* **328**, 504–508 (2010).
20. V. Epshtein *et al.*, *EMBO J.* **22**, 4719–4727 (2003).
21. D. Dutta *et al.*, *Cell* **146**, 533–543 (2011).
22. S. Sekine, Y. Murayama, V. Svetlov, E. Nudler, S. Yokoyama, *Mol. Cell* **57**, 408–421 (2015).
23. H. M. Arthur, P. B. Eastlake, *Gene* **25**, 309–316 (1983).
24. N. K. Maluf *et al.*, *J. Mol. Biol.* **325**, 913–935 (2003).
25. N. K. Maluf, T. M. Lohman, *J. Mol. Biol.* **325**, 889–912 (2003).
26. E. C. Siegel, *Mol. Gen. Genet.* **191**, 397–400 (1983).
27. G. Schreiber, E. Z. Ron, G. Glaser, *Curr. Microbiol.* **30**, 27–32 (1995).
28. B. J. Schalow, C. T. Courcelle, J. Courcelle, *J. Bacteriol.* **194**, 2637–2645 (2012).
29. E. Sarubbi, K. E. Rudd, M. Cashel, *Mol. Gen. Genet.* **213**, 214–222 (1988).

ACKNOWLEDGMENTS

We thank R. Lloyd for N4849 and N4235 strains and T. Artemyev for his contribution. This work was supported by the Russian Science Foundation grant 14-50-00060 and the Ministry of Education and Science of the Russian Federation grant 14.Z50.31.0014 (S.P. and A.M.), the National Institute of Child Health and Human Development, NIH Intramural Program (M.C.), NIH grant R01 GM107329, and by the Howard Hughes Medical Institute (E.N.).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/993/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 and S2
References (*30*–*41*)

21 October 2015; accepted 7 April 2016
10.1126/science.aad6945

Downloaded from http://science.sciencemag.org/ on May 19, 2016

HIV-1 ANTIBODIES

HIV-1 therapy with monoclonal antibody 3BNC117 elicits host immune responses against HIV-1

Till Schoofs,[1]* Florian Klein,[1,2,3]* Malte Braunschweig,[1,4]* Edward F. Kreider,[5] Anna Feldmann,[6] Lilian Nogueira,[1] Thiago Oliveira,[1] Julio C. C. Lorenzi,[1] Erica H. Parrish,[5] Gerald H. Learn,[5] Anthony P. West Jr.,[7] Pamela J. Bjorkman,[7] Sarah J. Schlesinger,[1] Michael S. Seaman,[8] Julie Czartoski,[9] M. Juliana McElrath,[9] Nico Pfeifer,[6] Beatrice H. Hahn,[5] Marina Caskey,[1] Michel C. Nussenzweig[1,10]†

3BNC117 is a broad and potent neutralizing antibody to HIV-1 that targets the CD4 binding site on the viral envelope spike. When administered passively, this antibody can prevent infection in animal models and suppress viremia in HIV-1–infected individuals. Here we report that HIV-1 immunotherapy with a single injection of 3BNC117 affects host antibody responses in viremic individuals. In comparison to untreated controls that showed little change in their neutralizing activity over a 6-month period, 3BNC117 infusion significantly improved neutralizing responses to heterologous tier 2 viruses in nearly all study participants. We conclude that 3BNC117-mediated immunotherapy enhances host humoral immunity to HIV-1.

Development of serum neutralization breadth during HIV-1 infection typically occurs several years after infection and is seen as a continuum, with ~50% of HIV-1–infected individuals developing some level of broad neutralization and a small fraction of individuals acquiring serum neutralizing activity of extraordinary breadth and potency (*1*–*4*). Antibody cloning experiments have revealed that this activity is due to one or more potent broadly neutralizing antibodies (bNAbs) that target one or more epitopes on the viral spike protein gp160 (*1*, *5*–*10*).

bNAbs show exceptional breadth and potency in vitro and can protect against or suppress active infection in humanized mice (*11*–*13*) and macaques (*14*, *15*). Moreover, in a phase I clinical trial, a single injection of 3BNC117, a CD4 binding-site–specific bNAb (*6*), was safe and effective in suppressing HIV-1 viremia by an average of 1.48 logs (*16*).

In addition to direct effects on target cells and pathogens, antibody-mediated immunotherapies have the potential to engage the host immune system and induce both innate and adaptive immune responses (*17*). In particular, the Fc domains of antibodies interact with receptors on innate cells such as natural killer (NK) cells and phagocytes to enhance the clearance of viral particles and the killing of infected cells (*18*). To test the hypothesis that bNAb-mediated immunotherapy can enhance immunity to HIV-1 in humans, we examined the serologic responses to the virus in individuals who received 3BNC117.

A single 3BNC117 infusion was administered to HIV-1–infected individuals at doses of 1, 3, 10, or 30 mg per kilogram of body weight (mg/kg) (Fig. 1A and table S1A) (*16*, *19*). To determine whether 3BNC117 therapy is associated with changes in viral sensitivity and serologic responses to autologous viruses, we cultured HIV-1 from peripheral blood mononuclear cells (PBMCs) of nine viremic individuals before [day 0 (d0)] and 4 weeks after 3BNC117 infusion (*16*). On d0, all but one of the cultured viruses were

[1]Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA. [2]Laboratory of Experimental Immunology, Center for Molecular Medicine Cologne, University of Cologne, 50931 Cologne, Germany. [3]Department of Internal Medicine, Center of Integrated Oncology Cologne-Bonn, University Hospital Cologne, 50937 Cologne, Germany. [4]Albert Ludwigs University of Freiburg, Freiburg, Germany. [5]Departments of Medicine and Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [6]Department of Computational Biology and Applied Algorithmics, Max Planck Institute for Informatics, Campus E14, 66123 Saarbrücken, Germany. [7]Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA. [8]Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA. [9]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [10]Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065, USA.
*These authors contributed equally to this work. †Corresponding author. Email: nussen@rockefeller.edu

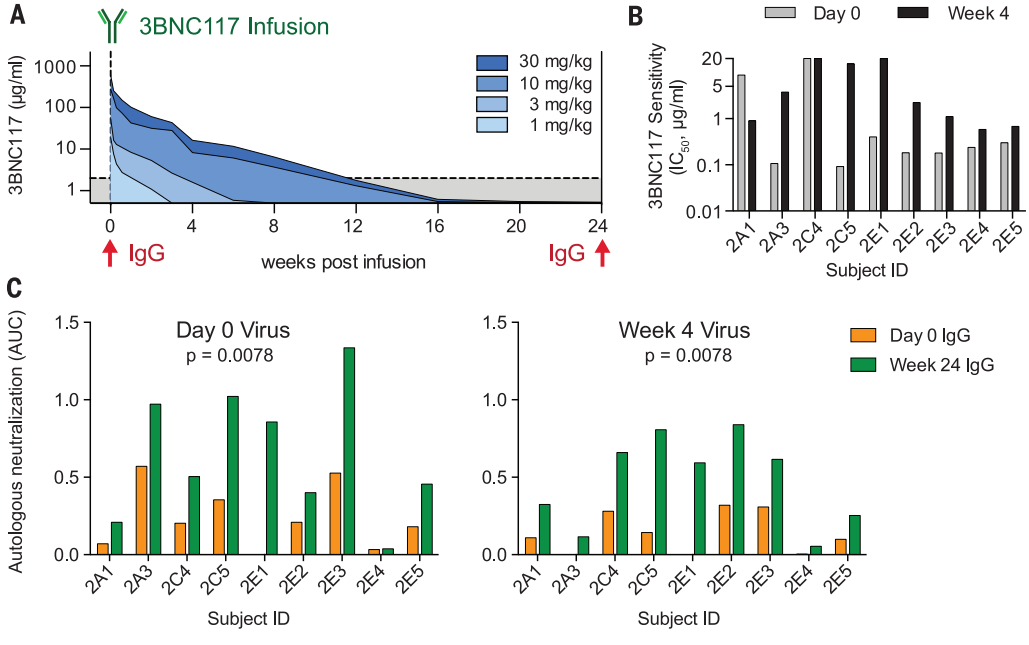


Fig. 1. Virus sensitivity to 3BNC117 and autologous antibody responses. (**A**) Kinetics of 3BNC117 antibody decay in HIV-1–infected individuals as determined by a validated anti-idiotype enzyme-linked immunosorbent assay (ELISA) (*16*). Shown are mean values of patients infused in each respective dose group. Each patient sample was measured in duplicates. The gray-shaded area indicates the lower level of accuracy of the assay (2 µg/ml). Red arrows indicate the time points of IgG purification. (**B**) Autologous virus sensitivity to 3BNC117 before (d0, gray) and 4 weeks after (black) 3BNC117 infusion. The *y* axis shows IC_{50}'s for 3BNC117 on viral culture supernatants from PBMCs determined by TZM.bl assay. Neutralization assays performed in duplicates. (**C**) The AUC of the neutralization curves of purified IgGs obtained from sera on d0 (orange) or week 24 (green) against d0 (left) or week 4 (right) autologous viruses. Neutralization assays were performed in duplicates. *P* values were determined by Wilcoxon signed-rank test.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

sensitive to 3BNC117, with median inhibitory concentration (IC_{50}) values ranging from 0.09 to 8.8 μg/ml [Fig. 1B and (*16*)]. At week 4, we found increased resistance to 3BNC117 in most individuals, indicating selection for viral escape variants [Fig. 1B and (*16*)].

When the same viral isolates were tested for sensitivity to the matched individual's immunoglobulins (IgGs) obtained before (d0) and 24 weeks after 3BNC117 infusion (Fig. 1A), we found increased neutralizing activity in the week 24 IgG against both d0 and week 4 autologous

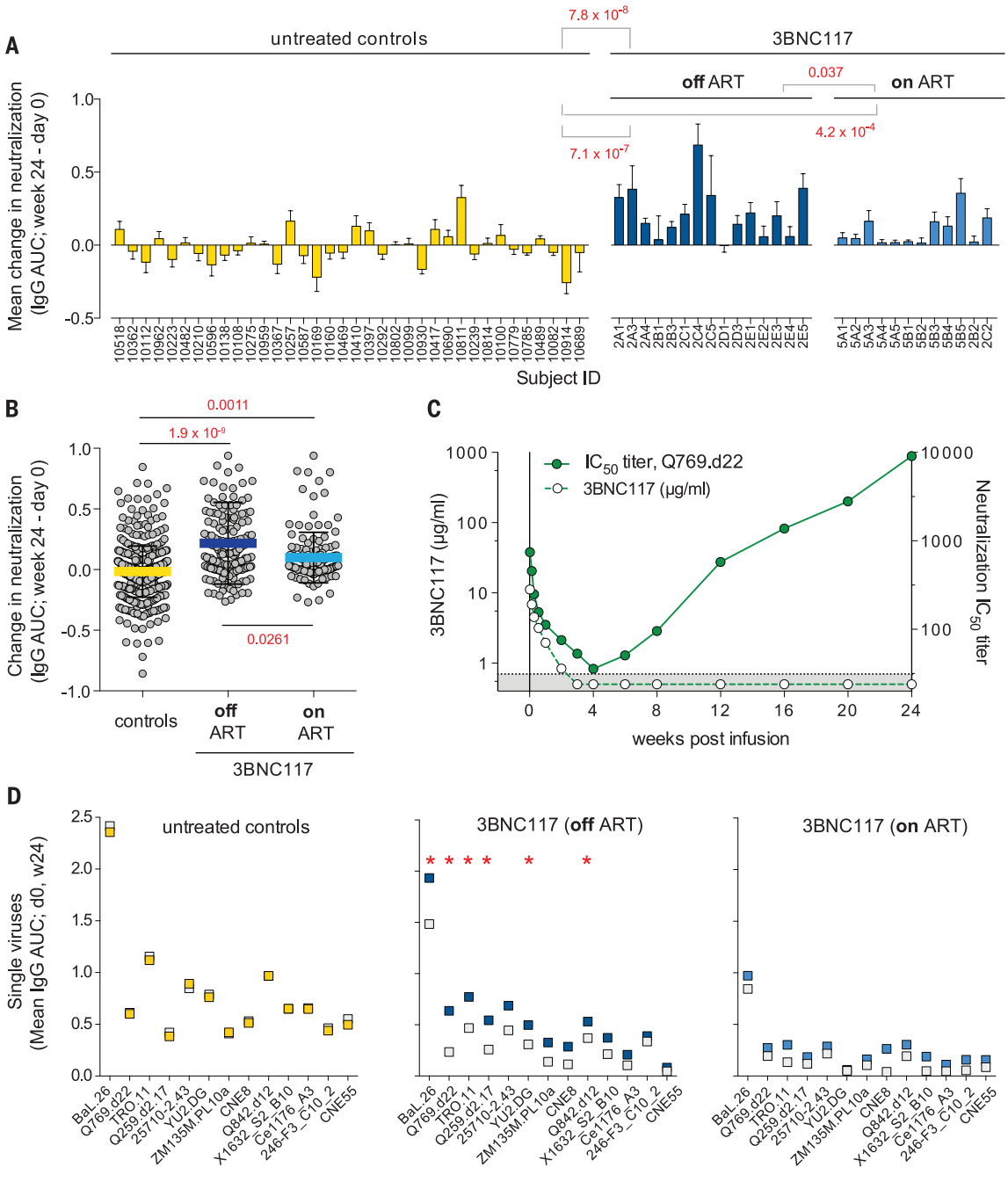


Fig. 2. Heterologous antibody responses. (**A**) The difference in overall AUC (mean AUC change) per individual in TZM.bl assays against 13 heterologous viruses (Fig. 2D) for d0 versus week 24 IgG obtained from 36 untreated viremic controls (mean sampling interval 26.8 weeks), 15 viremic individuals infused with 3BNC117 (mean sampling interval 24.1 weeks), and 12 ART-treated individuals receiving 3BNC117 infusion (mean sampling interval 24.0 weeks) (*16*). Neutralization assays performed in duplicates. *P* values were determined by unpaired Wilcoxon test (rank-sum test). (**B**) The aggregated differences in AUC between d0 and week 24 IgG assayed by TZM.bl for all viruses and all individuals. Each dot represents a single AUC difference for a single virus from one individual displayed in (A). Colored bars represent the mean of all AUCs. Whiskers show standard deviation. *P* values were determined using generalized estimating equations (*38*). (**C**) 3BNC117 antibody levels (ELISA, white) and TZM.bl neutralization titer against tier 2 strain Q769.d22 (green) in patient 2A3. (**D**) Mean AUCs of IgGs of all individuals at d0 (gray) and week 24 (color of respective group) for each HIV-1 pseudovirus tested. Changes in neutralization of viremic control individuals without 3BNC117 infusion are shown in yellow (left). Change in neutralization of 3BNC117-treated individuals are shown in dark blue (off ART, middle) and light blue (on ART, right). *P* values were determined by unpaired Wilcoxon test (rank-sum test). Red stars indicate significant *P* values after Bonferroni correction (threshold $P < 0.0038$).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

viruses (P = 0.0078, Fig. 1C and table S2). Thus, although 3BNC117 infusion selected for 3BNC117-resistant HIV-1 variants, neutralizing antibody responses continued to develop against autologous viruses (*20*).

To test for changes in heterologous neutralizing activity after 3BNC117 treatment, we assayed individuals' d0 and week 24 IgG against a panel of tier 1 (n = 1) and tier 2 (n = 12) HIV-1 pseudoviruses that included globally circulating HIV-1 strains (*21*) (Fig. 2 and tables S1, S3, and S4). Neutralizing activity was compared between the two time points by measuring the area under the neutralization curve (AUC) for patients' isolated IgG against each virus (table S4B). 15 participants that received 3BNC117 were not on antiretroviral therapy (ART) and had starting viral loads from 640 to 53,470 copies/ml (table S1A). Control IgGs were obtained from 36 viremic individuals who did not receive 3BNC117 and had starting viral loads ranging from 150 to 303,200 copies/ml (Fig. 2 and table S1B).

During a 6-month observation period, control individuals' neutralizing activity showed no consistent improvement in either breadth or potency (Fig. 2, A and B, figs. S1A and S2, and tables S4 and S5A) (*4*, *22*). In contrast, all but one of the 15 viremic individuals infused with 3BNC117 showed increased breadth and/or potency against the pseudovirus panel at week 24 (P = 7.1 × 10^{-7}, Fig. 2A, figs. S1B and S2, and tables S4 to S6). The absolute change in neutralizing activity varied between viruses and individuals, ranging from small effects to dramatic increases as observed in patient 2A3 for viral strain Q769.d22 (Fig. 2C and tables S4 to S6). Significant differences were also evident between 3BNC117-treated and control groups regardless of whether sera from all individuals were considered in aggregate or examined against individual viruses (P = 1.9 × 10^{-9}, Fig. 2, B and D).

In addition to viremic patients, we examined 12 individuals that received 3BNC117 while on ART, with no detectable or low-level viremia (<20 to 100 copies/ml). In comparison to viremic patients, the increase in heterologous neutralizing activity was significantly less pronounced in ART-treated individuals (P = 0.037; Fig. 2, A, B, and D; figs. S1B and S2; and tables S3 to S5).

The observed improvement in neutralizing activity could not be explained by confounding factors such as differences in initial viral load or $CD4^+$ T cell levels (fig. S3 and tables S1 and S7). Moreover, we found no correlation between d0 neutralizing activity and neutralization improvement (fig. S4). A comparison of the pattern of neutralization increase with 3BNC117's neutralization profile ruled out that remaining antibody was responsible for the effect (fig. S5 and table S8). We conclude that 3BNC117 enhances host immunity to heterologous tier 2 HIV-1 viruses irrespective of initial neutralization breadth and potency.

To examine the effects of 3BNC117 immunotherapy on the plasma viral population of treated individuals, we performed single-genome sequencing (SGS) of over 1000 plasma-derived gp160 *env* genes (gp160) before (d0) and 4 (*6*), 12, or 24 weeks after infusion (Fig. 3, A and B, figs. S6 to S10, and table S9). With the exception of two individuals who were sexual partners, all other volunteers had epidemiologically unrelated infections (Fig. 3A). On d0, *env* sequences from patients 2A1, 2A3, and 2C4 comprised multiple lineages, which was reflected in a multimodal distribution of pairwise diversity measurements from these individuals (Fig. 3B and fig. S6). Analysis of *env* sequences from subsequent time points revealed significant shifts in both nucleotide (six out of nine individuals, Fig. 3B) and amino acid sequence (seven out of nine individuals, fig. S6) diversity. Consistent with the observation that *env* diversity is associated with neutralization breadth (*23*–*25*), there was a strong correlation between the initial level of neutralizing activity and the initial diversity of the circulating viral swarm (correlation coefficient = 0.92, Fig. 3C).

We next evaluated viral sequence evolution in each of the 3BNC117-treated patients over time. Shifts in the viral quasispecies were evident

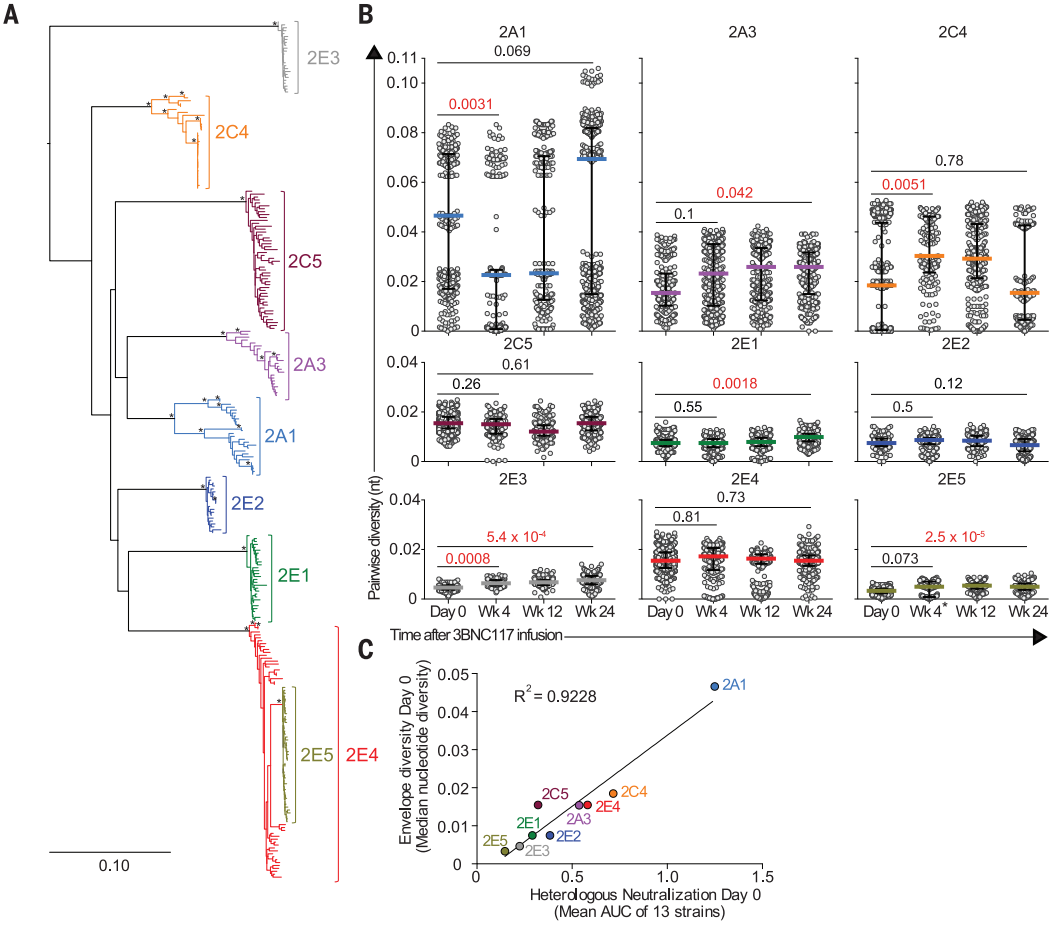


Fig. 3. HIV-1 quasispecies diversity before and after 3BNC117 infusion. (**A**) Maximum-likelihood phylogenetic tree of single-genome–derived *env* gene sequences from d0 plasma, before therapy with 3BNC117 (table S9). Asterisks indicate bootstrap values of 100%. Individual viral sequences are color-coded as indicated. (**B**) Scatter plots depicting pairwise nucleotide sequence diversity of plasma *env* sequences on d0 and weeks 4 (2E5, week 6), 12, and 24 after infusion. Each dot represents the pairwise genetic difference between two sequences at a given time point. Colored bars indicate median diversity, whereas black bars indicate the interquartile range. P values were determined using a two-sample U statistic–based Z test (*39*–*41*). (**C**) Graph shows the relationship between d0 mean heterologous neutralizing AUC against a panel of tier 1 (n = 1) and tier 2 (n = 12) viruses (x axis) and the median pairwise nucleotide diversity for each patient (y axis). R^2, correlation coefficient.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

regardless of initial 3BNC117 neutralization sensitivity and bNAb dose (Fig. 4 and fig. S7). However, the nature of these shifts differed depending on the individual (Fig. 4 and figs. S7 to S9). For example, in patient 2A1, 15 out of 27 (15/27) d0 sequences fell into a single clade marked "group A" (Fig. 4A and fig. S8). Four weeks after 3BNC117 infusion, group A viruses contracted (2/25 sequences) and group C viruses expanded (16/25). At week 24, the viral quasispecies was primarily composed of group B and D viruses (Fig. 4 and fig. S8). This pattern of "clade shifting" was also seen in patients 2A3 and 2C4 (fig. S7). Patients with lower initial *env* diversities, such as 2E1, did not harbor distinct viral sublineages at d0 (Figs. 3 and 4A) but continued to accrue mutations, some of which became fixed during the 24-week follow-up (for example, changes in V1/V2 in 2E1, fig. S9).

To assess viral sequence changes after 3BNC117 infusion, we generated longitudinal logo plots depicting 3BNC117 contact residues (*26*, *27*) for each patient (Fig. 4B and figs. S7 and S10). Although viruses from all nine patients exhibited mutations within 3BNC117 contact residues relative to the d0 consensus sequence, their number and position varied considerably, as exemplified by patients 2A1 and 2E1 (Fig. 4B and figs. S7 and S10). Using LASSIE (Longitudinal Antigenic Sequences and Sites from Intrahost Evolution) (*28*), we scanned the entire *env* protein sequence for sites selected within the 24-week time frame (selection cutoff ≥ 80%) (table S10). Although selected sites were identified in all patients, no consistent mutational pattern was observed (table S10). These data suggest that 3BNC117 immunotherapy is associated with shifts in circulating quasispecies and a number of different *env* mutations, some of which persist even after the infused antibody levels drop below detection.

To better understand the virus-host interactions that led to the development of enhanced heterologous neutralizing breadth, we performed neutralization assays on 63 pseudoviruses expressing the gp160s found in the circulation on d0 and weeks 4, 12, and 24 from five individuals (Fig. 4, fig. S7, and table S11). The pseudoviruses were tested for sensitivity to the corresponding individual's IgG obtained on d0 and week 24. In all cases, we were able to identify d0 or week 4 viruses that exhibited greater neutralization sensitivity to week 24 IgG as compared to d0 IgG (Fig. 4, fig. S7, and table S11). For example, all tested 2A1 and 2E1 viruses were 3BNC117-sensitive and exhibited a week 24/d0 fold change of ~1.7 and ~4.8 in IgG IC_{50}, respectively (Fig. 4). On the other hand, all tested 2C4 viruses were 3BNC117-resistant (mean IC_{50} >20 μg/ml), yet they were ~6.5-fold more sensitive to week 24 IgG versus d0 IgG (fig. S7). In conclusion, viremic individuals receiving 3BNC117 produced antibodies to autologous viruses that were both sensitive and resistant to 3BNC117.

Although exceptional bNAbs to HIV-1 develop only sporadically in a fraction of infected individuals, most HIV-1–infected individuals develop some level of neutralization breadth (*1*–*4*). Here we show that 3BNC117 immunotherapy accelerates this process. This boost in heterologous

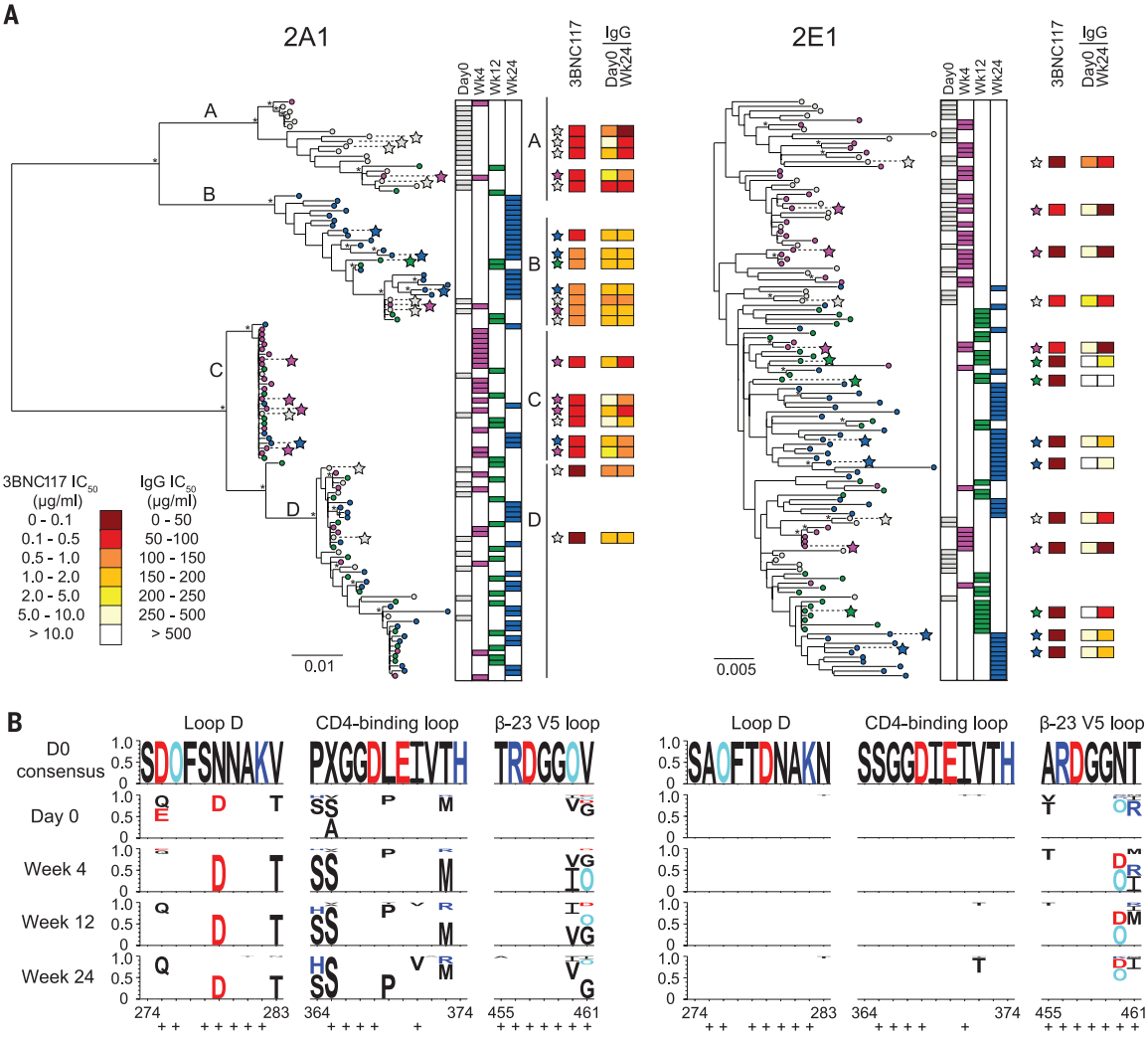


Fig. 4. Antibody responses to the evolving viral quasispecies. (**A**) Maximum-likelihood phylogenetic trees of single-genome–derived *env* gene sequences from patients 2A1 and 2E1 sampled on d0 and weeks 4, 12, and 24 after 3BNC117 infusion (left). Clades with bootstrap support ≥70% are indicated by a black asterisk and are arbitrarily named groups A to D in the case of patient 2A1. Bar graphs (middle) indicate the time points from which sequences in the tree are derived. Heat maps (right) show the 3BNC117 IC_{50}, d0 IgG IC_{50}, and week 24 IgG IC_{50} values against autologous pseudoviruses using *env* sequences as indicated by colored stars. Neutralization assays were performed in duplicates. (**B**) Sequence logo plots illustrating longitudinal amino acid changes in and around known 3BNC117 contact residues (*26*, *27*) in patients 2A1 and 2E1. Letters indicate deviations from the d0 consensus shown at the top, whereas white spaces indicate agreement with the d0 consensus. Colors indicate basic (dark blue) and acidic (red) residues, and a turquoise "O" is used instead of "N" to indicate a potential N-glycosylation site. Logo plots were generated using LASSIE (*28*). Plus symbols indicate 3BNC117 contact residues confirmed by two crystal structures (*26*, *27*).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

breadth occurs irrespective of demographic, virologic, or dosage factors and was associated with both transient and lasting changes to the viral quasispecies. It is of note that the neutralization improvements observed were modest in most individuals, potentially owing to the transient nature of therapy with a single antibody as well as the short time frame of observation.

Although the effect of 3BNC117 on neutralizing responses to heterologous HIV-1 viruses may seem surprising, antibodies to HIV-1 have been associated with enhanced immunity in infants born to HIV-1–infected mothers that have circulating antibodies to HIV-1 and macaques treated with monoclonal antibodies or neutralizing serum (*29*–*31*).

How passively administered antibodies to HIV-1 accelerate the emergence of bNAbs is not completely understood. One possibility is that 3BNC117 infusion selected for viral variants with altered antigenic properties, which in turn stimulated new B cell lineages (*23*–*25*, *32*–*34*). A second possibility is that immune complexes formed by 3BNC117 and circulating viruses act as potent immunogens, a phenomenon that is believed to be responsible for the enhanced $CD8^+$ T cell immunity to tumor antigens in individuals receiving monoclonal antibody–based immunotherapy (*35*–*37*).

Irrespective of the mechanism(s), the enhanced antibody response found in individuals receiving 3BNC117 therapy indicates that immunotherapy boosts host immunity to HIV-1. Moreover, the finding that antibody responses to heterologous tier 2 viruses develop in nearly all 3BNC117-treated individuals suggests that host genetics or a specific viral envelope sequence do not limit the development of neutralizing antibodies to HIV-1.

REFERENCES AND NOTES

1. F. Klein *et al.*, *Science* **341**, 1199–1204 (2013).
2. A. P. West Jr. *et al.*, *Cell* **156**, 633–648 (2014).
3. I. Mikell *et al.*, *PLOS Pathog.* **7**, e1001251 (2011).
4. P. Hraber *et al.*, *AIDS* **28**, 163–169 (2014).
5. X. Wu *et al.*, *Science* **329**, 856–861 (2010).
6. J. F. Scheid *et al.*, *Science* **333**, 1633–1637 (2011).
7. L. M. Walker *et al.*, *Nature* **477**, 466–470 (2011).
8. F. Klein *et al.*, *J. Exp. Med.* **209**, 1469–1479 (2012).
9. M. Bonsignori *et al.*, *J. Virol.* **86**, 4688–4692 (2012).
10. J. F. Scheid *et al.*, *Nature* **458**, 636–640 (2009).
11. F. Klein *et al.*, *Nature* **492**, 118–122 (2012).
12. J. A. Horwitz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16538–16543 (2013).
13. J. Pietzsch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15859–15864 (2012).
14. D. H. Barouch *et al.*, *Nature* **503**, 224–228 (2013).
15. M. Shingai *et al.*, *Nature* **503**, 277–280 (2013).
16. M. Caskey *et al.*, *Nature* **522**, 487–491 (2015).
17. S. Bournazos, J. V. Ravetch, *Immunol. Rev.* **268**, 88–103 (2015).
18. C.-L. Lu, *Science* **352**, 1001–1004 (2016).
19. Materials and methods are available as supplementary materials on *Science* Online.
20. X. Wei *et al.*, *Nature* **422**, 307–312 (2003).
21. A. deCamp *et al.*, *J. Virol.* **88**, 2489–2507 (2014).
22. S. G. Deeks *et al.*, *J. Virol.* **80**, 6155–6164 (2006).
23. H.-X. Liao *et al.*, *Nature* **496**, 469–476 (2013).
24. N. A. Doria-Rose *et al.*, *Nature* **509**, 55–62 (2014).
25. P. L. Moore, C. Williamson, L. Morris, *Trends Microbiol.* **23**, 204–211 (2015).
26. T. Zhou *et al.*, *Science* **329**, 811–817 (2010).
27. F. Klein *et al.*, *Cell* **153**, 126–138 (2013).
28. P. Hraber *et al.*, *Viruses* **7**, 5443–5475 (2015).
29. L. Goo, V. Chohan, R. Nduati, J. Overbaugh, *Nat. Med.* **20**, 655–658 (2014).
30. N. L. Haigwood *et al.*, *J. Virol.* **78**, 5983–5995 (2004).
31. C. T. Ng *et al.*, *Nat. Med.* **16**, 1117–1119 (2010).
32. P. L. Moore *et al.*, *Nat. Med.* **18**, 1688–1692 (2012).
33. F. Gao *et al.*, *Cell* **158**, 481–491 (2014).
34. J. N. Bhiman *et al.*, *Nat. Med.* **21**, 1332–1336 (2015).
35. T. T. Wang *et al.*, *Cell* **162**, 160–169 (2015).
36. S. Bournazos, D. J. DiLillo, J. V. Ravetch, *J. Exp. Med.* **212**, 1361–1369 (2015).
37. D. J. DiLillo, J. V. Ravetch, *Cell* **161**, 1035–1045 (2015).
38. S. J. Ratcliffe, J. Shults, *J. Stat. Softw.* **25**, 1–14 (2008).
39. P. B. Gilbert, A. J. Rossini, R. Shankarappa, *Biometrics* **61**, 106–117 (2005).
40. E. E. Giorgi, T. Bhattacharya, *Biometrics* **68**, 1323–1326 (2012).
41. W. Deng *et al.*, *Biotechniques* **48**, 405–408 (2010).

ACKNOWLEDGMENTS

We thank all study participants, who devoted time to our research. We thank the Rockefeller University Hospital Clinical Research Support Office, the nursing staff for patient care and recruitment, the clinical study group of the Infectious Disease Division at the University Hospital Cologne, and all members of the Nussenzweig lab for helpful discussions. We thank M. Schechter and C. Baro for technical assistance, P. Fast and H. Park for clinical monitoring, E. Gotschlich and B. Coller for input on study design, and P. Hraber for helping with LASSIE analyses. The data reported in this study are tabulated in the main paper and in the supplementary materials. Envelope single-genome sequencing data can be downloaded from GenBank (accession numbers KX027737 to KX028736). This work was supported in part by the Bill and Melinda Gates Foundation Collaboration for AIDS Vaccine Discovery, grants OPP1032144 (M.S.S.) and OPP1092074 and OPP1124068 (M.C.N); the Robertson Foundation to M.C.N.; the NIH Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, grants 1UM1 AI100663-01 (M.C.N) and 1UM1 AI00645 (B.H.H.); the University of Pennsylvania Center for AIDS Research Single Genome Amplification Service Center P30, grant AI045008 (B.H.H.); and NIH grants UM1AI068618 (M.J.M.), UM1AI069481 (M.J.M.), F30 AI112426 (E.F.K), and HIVRAD P01 AI100148 (P.J.B.). T.S. is supported by a Deutsche Forschungsgemeinschaft postdoctoral fellowship (grant SCHO 1612/1-1). F.K. is supported by the Heisenberg-Program of the Deutsche Forschungsgemeinschaft (grant KL 2389/2-1); the European Research Council (grant ERC-StG639961); and the German Center for Infection Research, partner site Bonn–Cologne, Cologne, Germany. M.B. is supported by the German National Academic Foundation. J.C.C.L. is supported by an award from the Conselho Nacional de Desenvolvimento Científico e Tecnológico "Ciencia sem Fronteiras" (grant 248676/2013-0). M.C.N. is a Howard Hughes Medical Investigator and an inventor on U.S Patent Application no. 14/118,496, filed by Rockefeller University related to 3BNC117.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/997/suppl/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S11
References (*42*–*52*)

17 December 2015; accepted 14 April 2016
Published online 5 May 2016
10.1126/science.aaf0972

HIV-1 ANTIBODIES

Enhanced clearance of HIV-1–infected cells by broadly neutralizing antibodies against HIV-1 in vivo

Ching-Lan Lu,[1,2] Dariusz K. Murakowski,[3]* Stylianos Bournazos,[4] Till Schoofs,[1] Debolina Sarkar,[3] Ariel Halper-Stromberg,[1] Joshua A. Horwitz,[1] Lilian Nogueira,[1] Jovana Golijanin,[1] Anna Gazumyan,[1] Jeffrey V. Ravetch,[4] Marina Caskey,[1] Arup K. Chakraborty,[3,5,6,7,8]† Michel C. Nussenzweig[1,9]†

Antiretroviral drugs and antibodies limit HIV-1 infection by interfering with the viral life cycle. In addition, antibodies also have the potential to guide host immune effector cells to kill HIV-1–infected cells. Examination of the kinetics of HIV-1 suppression in infected individuals by passively administered 3BNC117, a broadly neutralizing antibody, suggested that the effects of the antibody are not limited to free viral clearance and blocking new infection but also include acceleration of infected cell clearance. Consistent with these observations, we find that broadly neutralizing antibodies can target $CD4^+$ T cells infected with patient viruses and can decrease their in vivo half-lives by a mechanism that requires Fcγ receptor engagement in a humanized mouse model. The results indicate that passive immunotherapy can accelerate elimination of HIV-1–infected cells.

Broadly neutralizing antibodies (bNAbs) to HIV-1 can block acquisition and suppress viremia in chronically infected humanized mice and macaques (*1*, *2*). In humans, a single infusion of 3BNC117, a bNAb that targets the CD4-binding site on the HIV-1 envelope glycoprotein gp160, led to a rapid but transient reduction in viral loads by an average of 1.48 $\log_{10}$ copies per ml (*3*).

Antibodies differ from small-molecule drugs that interfere with viral replication in that antibodies have the potential to affect the half-lives of both free virus and infected cells. Indeed, antibodies accelerate the clearance of free virions from the blood of macaques (*4*) and induce killing of infected cells in vitro by Fcγ receptor (FcγR)–mediated mechanisms (*5*, *6*). However, the majority of infected cells die rapidly by apoptosis

Downloaded from http://science.sciencemag.org/ on May 19, 2016

or pyroptosis (*7*, *8*), and whether bNAbs can accelerate HIV-1–infected cell clearance in vivo has not been tested directly.

To examine the components that contribute to viral clearance in humans given a single infusion of 3BNC117, we adapted an existing model of HIV-1 viral dynamics (*3*, *9*, *10*). The model (*11*) (fig. S1) includes virus-producing infected cells, as well as transport of free plasma virus to lymphoid tissues (LTs) and vice versa. To this basic model, we added the feature that antibodies bind to virus particles, which leads to virus neutralization and loss of antibody. Measurements of the decline of antibody concentrations in healthy humans were fitted to a two-compartment model (*12*, *13*) to obtain the parameters characterizing the intrinsic antibody decay rates and transport between tissue and plasma over the time scale during which viral loads decay in patients treated with 3BNC117 (fig. S2). The rate of free virus neutralization was fitted to the virus kinetics in 19 patients (fig. S3), but we focused on patients showing an initial monophasic viral load decline (2B3, 2C1, 2C5, 2D3, 2E1, and 2E2), which tended to coincide with those receiving a higher antibody dose (*3*).

This model is unable to recapitulate the kinetics of viral load decline for any of the 3BNC117-treated viremic patients (Fig. 1, green; and fig. S3). If we fit the overall extent of viral load decrease, the rate of viral load decay is predicted to be too fast. Conversely, matching the initial rate of viral load decline results in insufficient overall reduction of the viral load. Thus, we adjusted our model to incorporate a mechanism that includes antibodies acting to clear infected cells and explored the revised model to see if it provided additional reduction of virus over a longer time scale (*11*). The rates of free-virus neutralization and infected cell clearance are fit to the measured plasma viral load. Including cell clearance substantially improves the fit to patient data (Fig. 1, purple; fig. S3; and table S3), because reducing the number of infected cells in tissues results in a second-order decay in the plasma viral load over a longer time scale.

Our modeling clearly shows that the patient data cannot be explained if 3BNC117 acts only to neutralize free virions and thus makes them incapable of infecting target cells. Including infected cell clearance improves the fit but does not quantitatively recapitulate the data; the reasons for these results are noted in the supplementary materials (*11*). Further evidence for such a mechanism is indicated by comparison of our modeling and clinical data for patients treated with a constant high level of entry-inhibitor drugs like maraviroc (*11*, *14*) (fig. S4).

To determine whether 3BNC117 can recognize the HIV-1 envelope (Env) trimer expressed on the surface of infected cells, we stained $CD4^+$ T cells infected with HIV-1_{YU2} or primary isolates obtained from patients 2C1, 2C5, 2D3, and 2E5 before they were infused with 3BNC117 (*3*). Consistent with their neutralizing activity in TZM-bl assays, 3BNC117, PG16, and 10-1074 specifically stained HIV-1_{YU2}–infected cells (fig. S5A). Also, 3BNC117 and 10-1074 recognized nearly all Gag^+ cells infected with primary isolates from patients 2C1, 2C5, 2D3, and 2E5; however, the mean fluorescence intensity of staining for Env was lower than for HIV-1_{YU2} (fig. S5B), possibly because of lower levels of Env on the surface of cells infected with patient viruses or variable levels of tetherin antagonism by HIV-1 accessory protein Vpu (*15*). We conclude that 3BNC117 recognizes the HIV-1 Env glycoprotein on the surface of infected $CD4^+$ T cells.

To examine whether bNAbs can accelerate clearance of infected cells in vivo, we performed adoptive transfer experiments using nonobese diabetic (NOD) mice with $Rag1^{-/-}Il2rg^{null}$ (NRG mice). To prevent spread of infection between human cells, animals and infected cell suspensions were treated with antiretroviral therapy (ART) before adoptive transfer. Anti–HIV-1 bNAbs or isotype control antibodies were administered 12 hours before infected cell transfer (fig. S6).

Either 3BNC117 alone ($P = 0.0012$) (Fig. 2A) or a combination of 3BNC117 and 10-1074 rapidly reduced the percentage of HIV-1_{YU2}–infected cells among $CD3^+CD8^-$ cells compared with that in mice treated with an isotype control ($P < 0.0001$) (Fig. 2B). Concomitant with reduction in the percentage of infected cells, cell-associated HIV-1 RNA levels were lower in bNAb-treated mice than in those treated with isotype controls ($P = 0.0054$) (Fig. 2B). These data indicate that bNAbs can accelerate clearance of HIV-1_{YU2}–infected cells in vivo.

[1]Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA. [2]Weill Cornell Medical College, New York, NY 10065, USA. [3]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Laboratory of Molecular Genetics and Immunology, The Rockefeller University, New York, NY 10065, USA. [5]Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA 02139, USA. [6]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [7]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [8]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [9]Howard Hughes Medical Institute.
*D.K.M. led the computational studies. **†Corresponding author. Email: nussen@rockefeller.edu (M.C.N.); arupc@mit.edu (A.K.C.)**

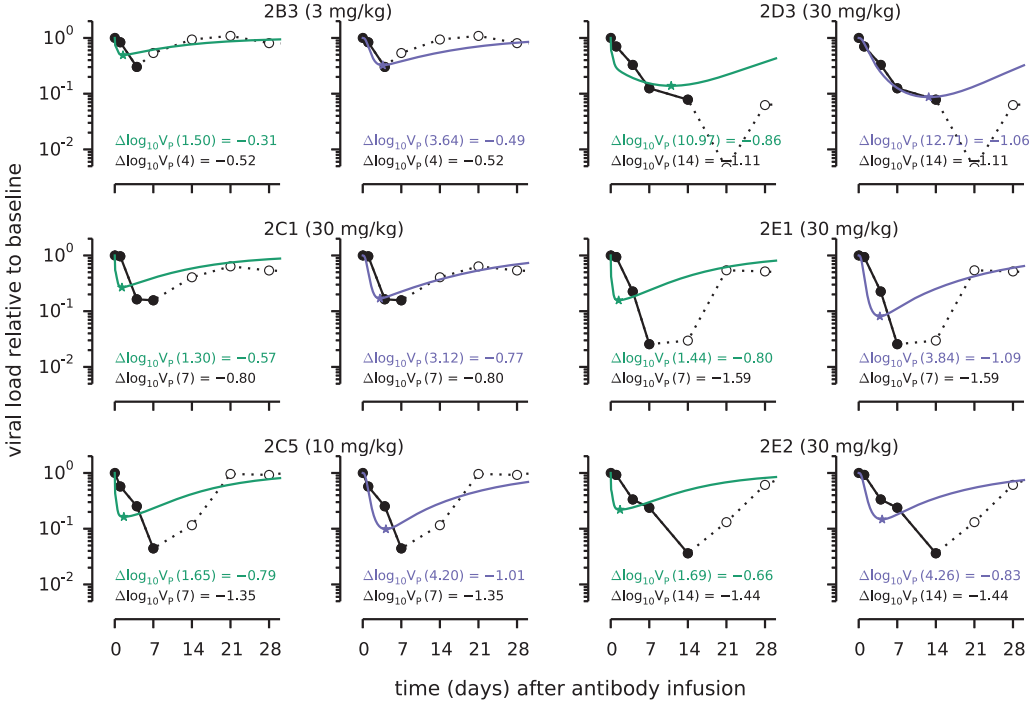


Fig. 1. Comparison of viral load measurements with best-fit model predictions. Viral load measurements (filled circles, solid black lines) with best-fit model predictions (solid colored lines). Each green line shows the predicted viral load over time, normalized by its initial amount, $V_P(t)/V_P(0)$, in a model in which antibody can only neutralize free virus particles. Each purple line shows a modified model in which antibody can also lead to clearance of infected cells. Only those patients with a day-1 viral load lower than baseline are shown. Open circles and dashed black lines represent data points that were not used for parameter estimation. Within each subfigure, we note the quantity $\Delta\log_{10}V_P(t_{min}) = \log_{10}[V_{P,min}/V_P(0)]$, i.e., the viral load at the nadir and the time in days at which this occurs for the data (black letters), as well as the predictions for a model with free virus clearance only (green) and for a model that also includes infected cell clearance (purple). This predicted minimum for each patient and model is denoted with a star.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

To determine whether HIV-1–specific antibodies can accelerate clearance of cells infected with primary HIV-1 isolates, we repeated the adoptive transfer experiment described above using human $CD4^+$ T cells infected with HIV_{2C1}, HIV_{2C5}, HIV_{2D3}, or HIV_{2E5}. As with HIV-1_{YU2}, bNAbs accelerated clearance of cells infected with patient isolates (Fig. 2C). We conclude that bNAbs can accelerate clearance of $CD4^+$ T cells infected with primary HIV-1 isolates.

To determine whether enhanced clearance of HIV-infected cells by antibodies depends on their ability to engage FcγR-expressing cells, we repeated the adoptive transfer experiments using bNAbs that carry mutations that specifically abrogate mouse FcγR binding [Arg substituted for Gly^{236} and for Leu^{328} (G236R/L328R or GRLR)] (*16*) (fig. S7). Although GRLR-bNAbs show normal levels of neutralizing activity in TZM-bl assays (*17*), they do not interact with cytotoxic or phagocytic cells and should therefore fail to accelerate clearance of HIV-1–infected cells in vivo. The frequency of infected cells remaining after treatment with GRLR-bNAbs was comparable to that found in mice receiving the isotype control (Fig. 3A) and significantly higher than that found in mice receiving wild-type bNAbs ($P < 0.0001$) (Fig. 3A). In addition to testing GRLR-bNAbs, we also blocked wild-type bNAb Fc-FcγR interactions using a combination of antibodies 2.4G2 and 9E9, targeting mouse FcγRs II/III and IV, respectively (fig. S7). Mice receiving FcγR-blocking antibodies failed to accelerate clearance of HIV-1–infected cells in response to bNAbs (Fig. 3B).

Human immunoglobulin-G1 (IgG1), the isotype of 3BNC117 and 10-1074, binds with highest affinity to mouse FcγRI and FcγRIV (mFcγRI and mFcγRIV) (*18*). These receptors are murine orthologs of human FcγRI and FcγRIII (hFcγRI and hFcγRIII) (*19*), which are expressed on human monocytes and natural killer (NK) cells that can perform antibody-dependent cell-mediated phagocytosis (ADCP) or cytotoxicity (ADCC) in vitro (*5*). To examine the role of mFcγRIV in clearance of HIV-1–infected cells in the NRG mouse, we blocked this receptor specifically with an mFcγRIV-specific monoclonal antibody (anti-mFcγRIV). Mice treated with bNAbs plus anti-mFcγRIV were comparable to isotype-treated mice (Fig. 3C), which indicated that mFcγRIV is essential for accelerating clearance of HIV-1_{YU2}–infected cells in vivo. We conclude that bNAbs accelerate infected cell clearance in NRG mice by a mechanism that requires mFcγRIV engagement.

We repeated adoptive transfer experiments in FcγR-humanized (hFcγR) mice that express only the human FcγRs (*20*). Similar to NRG mice, HIV-1_{YU2}–infected cells were reduced in bNAb-treated hFcγR mice ($P = 0.0064$) (fig. S8). These data suggest that bNAbs can use human FcγRs and not just murine FcγRs to accelerate clearance of HIV-1_{YU2}–infected cells in vivo.

To examine whether 3BNC117 accelerates clearance of HIV-1–infected cells in the context of chronic viral infection, we performed hemisplenectomy experiments in chronically infected

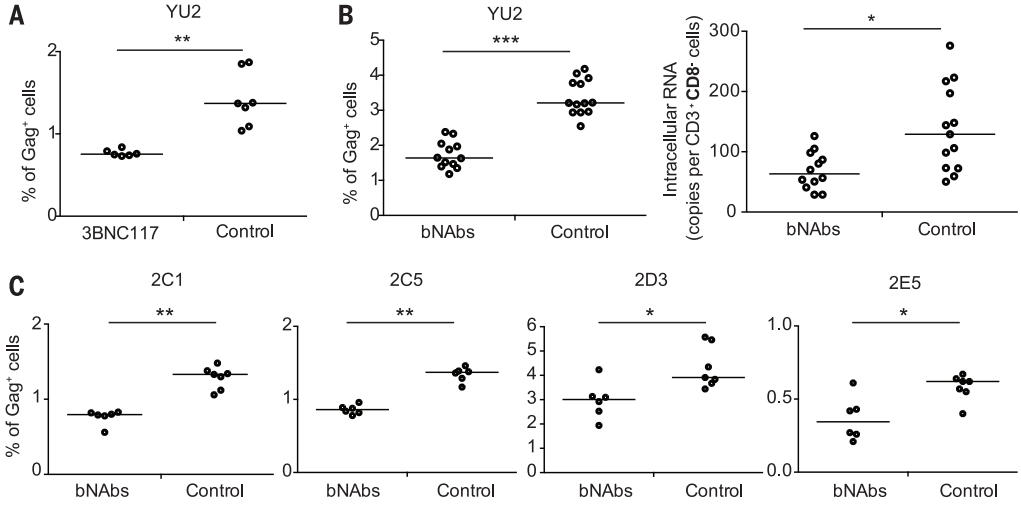


Fig. 2. bNAbs accelerate clearance of HIV-1–infected cells in vivo. (**A**) The percentage of Gag^+ cells among $CD3^+CD8^-$ cells in 3BNC117 (600 μg) or mice treated with isotype control 5 hours after HIV-1_{YU2}–infected cell transfer. (**B**) (Left) Percentage of Gag^+ cells among $CD3^+CD8^-$ cells in bNAbs (3BNC117 + 10-1074, 300 μg each) or mice treated with isotype controls 5 hours after HIV-1_{YU2}–infected cell transfer. (Right) Cell-associated HIV-1 RNA was measured in enriched human cells extracted from the spleen of mice 5 hours after transfer, plotted as the ratio of HIV-1 RNA to the number of $CD3^+CD8^-$ cells for each mouse. (**C**) Transfer experiments with cells infected by HIV_{2C1}, HIV_{2C5}, HIV_{2D3}, or HIV_{2E5}. Each dot represents one mouse. Lines represent median values. Data represent two to four independent experiments with a total of 6 to 13 mice per condition. *$P < 0.05$; **$P < 0.005$; ***$P < 0.001$; two-tailed Mann-Whitney U test.

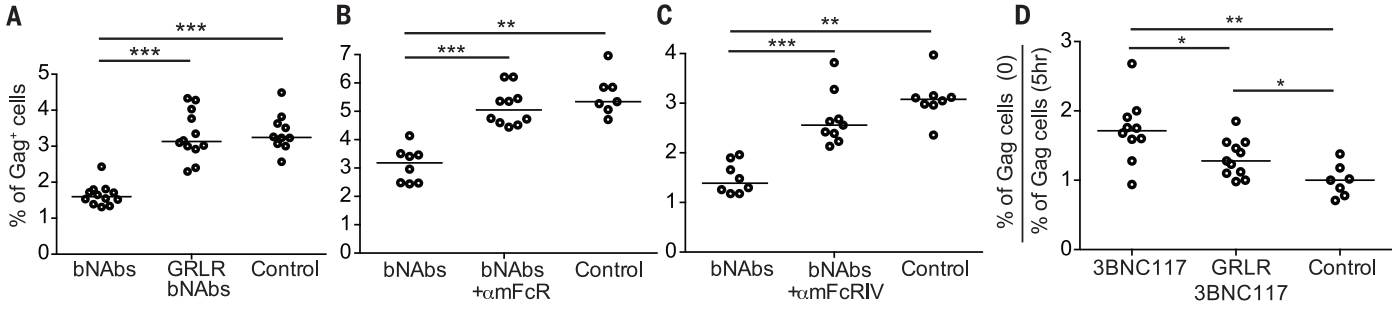


Fig. 3. FcγR engagement is required to facilitate clearance of HIV-1–infected cells. (**A**) Percentage of Gag^+ cells among $CD3^+CD8^-$ cells in bNAbs, GRLR-bNAbs, or mice treated with isotype controls. (**B**) NRG mice were injected with mouse antibodies blocking FcγRII, III, and IV or isotype controls 6 hours before injection of bNAbs. The percentage of Gag^+ cells among $CD3^+CD8^-$ cells is shown. (**C**) Percentage of Gag^+ cells among $CD3^+CD8^-$ cells in mice receiving FcγRIV-blocking antibody or isotype control. (**D**) Infected cell clearance in chronically HIV-1_{YU2}–infected hu-mice. The ratio of the percentage of Gag^+ cells among $CD3^+CD8^-$ cells before and 5 hours after 3BNC117, GRLR-3BNC117, or isotype control injection. Each dot represents one mouse. Lines represent median values. Data represent two to four independent experiments for each condition with a total of 7 to 12 mice per condition. *$P < 0.05$; **$P < 0.005$; ***$P < 0.001$; two-tailed Mann-Whitney U test.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

humanized mice (fig. S9). The frequency of HIV-1_{YU2}–infected cells before treatment was comparable in all groups of mice (fig. S10). As expected, mice treated with the Fc mutant antibody GRLR-3BNC117, which neutralizes HIV-1 but does not interact with effector cells, had lower frequencies of infected cells compared with isotype-treated controls (P = 0.0219) (Fig. 3D). However, mice treated with wild-type 3BNC117, which can mediate ADCC and ADCP, had substantially fewer infected cells compared with mice treated with GRLR-3BNC117 (P = 0.0167) (Fig. 3D). These data suggest that 3BNC117 can accelerate clearance of HIV-1_{YU2}–infected cells in the context of chronic viral infection.

In contrast to ART, antibodies have the potential to engage host immune cells in defense against the virus. They do so by binding to cell-free virions, which accelerates clearance and prevents their entry into target cells. Antibodies can also bind to HIV-1 Env on the surface of infected cells to induce ADCC or phagocytosis. Finally, immune complexes can activate antigen-presenting dendritic cells to elicit adaptive immune responses (*19*).

Both neutralizing and non-neutralizing antibodies support anti–HIV-1 ADCC activity in vitro (*5*, *21*), and Fc receptor binding is essential for optimal protection, postexposure prophylaxis, and therapy by bNAbs in animal models (*17*, *18*, *22*–*24*). Moreover, ADCC has been indirectly associated with both control of and protection against infection (*15*, *25*–*28*). However, the rapid death of HIV-1–infected cells has made it difficult to establish that antibodies can accelerate the clearance of infected cells in vivo. Our mathematical analysis of patient data, and the antibody-mediated reduction in infected cells seen in adoptive transfer experiments, establish that bNAbs alter the half-life of infected cells. This observation may help explain why postexposure prophylaxis with bNAbs is more effective than ART in hu-mice (*23*).

Experiments with human cells in mice cannot fully recapitulate the human host; nevertheless, these experiments establish that antibodies can accelerate clearance of infected cells in vivo and do so by an FcγR-dependent mechanism. The finding that antibodies can clear infected cells in vivo has important implications for therapies aimed at HIV prevention and viral reservoir reduction or elimination.

REFERENCES AND NOTES

1. H. Mouquet, *Trends Immunol.* **35**, 549–561 (2014).
2. A. P. West Jr. *et al.*, *Cell* **156**, 633–648 (2014).
3. M. Caskey *et al.*, *Nature* **522**, 487–491 (2015).
4. T. Igarashi *et al.*, *Nat. Med.* **5**, 211–216 (1999).
5. M. Kramski, M. S. Parsons, I. Stratov, S. J. Kent, *Curr. HIV Res.* **11**, 388–406 (2013).
6. Z. Euler, G. Alter, *AIDS Res. Hum. Retroviruses* **31**, 13–24 (2015).
7. G. Doitsh *et al.*, *Nature* **505**, 509–514 (2014).
8. A. S. Perelson *et al.*, *Nature* **387**, 188–191 (1997).
9. V. Müller, A. F. Marée, R. J. De Boer, *J. Virol.* **75**, 2597–2603 (2001).
10. R. J. De Boer, R. M. Ribeiro, A. S. Perelson, *PLOS Comput. Biol.* **6**, e1000906 (2010).
11. Supplementary analysis is available as supplementary materials on *Science* Online.
12. B. Joos *et al.*, *Antimicrob. Agents Chemother.* **50**, 1773–1779 (2006).
13. P. D. Gibaldi M, *Pharmacokinetics* (Marcel Dekker, New York, ed. 2, 1982).
14. G. Fätkenheuer *et al.*, *Nat. Med.* **11**, 1170–1172 (2005).
15. M. Kramski, I. Stratov, S. J. Kent, *AIDS* **29**, 137–144 (2015).
16. H. M. Horton *et al.*, *Blood* **116**, 3004–3012 (2010).
17. J. Pietzsch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15859–15864 (2012).
18. S. Bournazos *et al.*, *Cell* **158**, 1243–1253 (2014).
19. S. Bournazos, D. J. DiLillo, J. V. Ravetch, *J. Exp. Med.* **212**, 1361–1369 (2015).
20. P. Smith, D. J. DiLillo, S. Bournazos, F. Li, J. V. Ravetch, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6181–6186 (2012).
21. B. Su, C. Moog, *Front. Immunol.* **5**, 289 (2014).
22. A. J. Hessell *et al.*, *Nature* **449**, 101–104 (2007).
23. A. Halper-Stromberg *et al.*, *Cell* **158**, 989–999 (2014).
24. A. W. Boesch, E. P. Brown, M. E. Ackerman, *Immunol. Rev.* **268**, 296–310 (2015).
25. J. Pollara *et al.*, *Curr. HIV Res.* **11**, 378–387 (2013).
26. G. K. Lewis, *Immunology* **142**, 46–57 (2014).
27. J. R. Mascola, D. C. Montefiori, *Annu. Rev. Immunol.* **28**, 413–444 (2010).
28. B. F. Haynes *et al.*, *N. Engl. J. Med.* **366**, 1275–1286 (2012).

ACKNOWLEDGMENTS

We thank all study participants who devoted time to our research. We thank the Rockefeller University Hospital Clinical Research Support Office and nursing staff for patient care and recruitment. We thank T. Eisenreich for help with mouse colony management, K. Yao and S. Hinklein for technical help, P. Smith for FcγR humanized mice, CellDex Therapeutics for providing 3BNC117 and 10-1074, and all members of the Nussenzweig laboratory for helpful discussion and advice. The data reported in this paper are tabulated in the main paper and in the supplementary materials. This work was supported by Collaboration for AIDS Vaccine Discovery grant OPP1033115 (M.C.N. and J.V.R.). This work was also supported, in part, by grant 8 UL1 TR000043 from the National Center for Advancing Translational Sciences (NCATS); NIH Clinical and Translational Science Award (CTSA) program; NIH Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery (CHAVI-ID) 1UM1 AI100663-01 (M.C.N); Bill and Melinda Gates Foundation grants OPP1092074 and OPP1124068 (M.C.N); the Robertson Foundation to M.C.N.; German Research Foundation postdoctoral fellowship SCHO 1612/1-1 (T.S.); NIH grant F31 AI118555-01 (J.A.H.); the American Foundation for AIDS research (amfAR) Mathilde Krim Fellowship in Basic Biomedical Research (108977-57-RKVA) (S.B.); the Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard (A.K.C.); a NSF Graduate Research Fellowship under grant no. 1122374 (D.K.M.); and National Institute of Allergy and Infectious Diseases (NIH) grants AI100148-02 and AI081677-05 (M.C.N. and J.V.R.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. M.C.N. is a Howard Hughes Medical Institute Investigator. There are pending patent applications on the 3BNC117 and 10-1074 antibodies by Rockefeller University on which M.C.N. is an inventor. The patents are not licensed by any companies.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/1001/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 to S3
References (*29*–*37*)

21 December 2015; accepted 11 April 2016
Published online 5 May 2016
10.1126/science.aaf1279

DYNAMIC CYTOSKELETON

Accelerated actin filament polymerization from microtubule plus ends

Jessica L. Henty-Ridilla, Aneliya Rankova, Julian A. Eskin, Katelyn Kenny, Bruce L. Goode*

Microtubules (MTs) govern actin network remodeling in a wide range of biological processes, yet the mechanisms underlying this cytoskeletal cross-talk have remained obscure. We used single-molecule fluorescence microscopy to show that the MT plus-end–associated protein CLIP-170 binds tightly to formins to accelerate actin filament elongation. Furthermore, we observed mDia1 dimers and CLIP-170 dimers cotracking growing filament ends for several minutes. CLIP-170–mDia1 complexes promoted actin polymerization ~18 times faster than free–barbed-end growth while simultaneously enhancing protection from capping proteins. We used a MT-actin dynamics co-reconstitution system to observe CLIP-170–mDia1 complexes being recruited to growing MT ends by EB1. The complexes triggered rapid growth of actin filaments that remained attached to the MT surface. These activities of CLIP-170 were required in primary neurons for normal dendritic morphology. Thus, our results reveal a cellular mechanism whereby growing MT plus ends direct rapid actin assembly.

Tight coordination between the microtubule (MT) and actin cytoskeletons is required for fundamental processes such as directed cell migration, neuronal arborization, and phagocytosis (*1*–*3*). In many of these settings, the growth of MT plus ends into actin-rich cortical regions triggers changes in actin assembly–based functions (*4*, *5*). In plant cells, after washout of the actin depolymerizing drug latrunculin B, new actin polymerization occurs from MT plus ends (*6*). In fission yeast, MT plus ends direct formin-mediated actin cable assembly during "new end take off" (*7*, *8*). In other systems, MT plus ends play an instrumental role in steering actin-based motility of neuronal growth cones during neurite outgrowth and axonal guidance (*9*–*11*). Thus, it has long been hypothesized that molecular cues associated with MT plus ends directly govern

Department of Biology, Brandeis University, 415 South Street, Waltham, MA 02454, USA.
***Corresponding author. Email: goode@brandeis.edu**

Downloaded from http://science.sciencemag.org/ on May 19, 2016

localized actin assembly; however, no such mechanism has been defined.

Formins directly stimulate both the nucleation and elongation phases of actin filament assembly, and they perform essential roles in constructing diverse cellular actin structures, including actin cables, stress fibers, filopodia, phagocytic cups, and cytokinetic rings (*3*). The dimeric formin homology 2 (FH2) domain nucleates actin filaments and remains processively attached to their growing barbed ends as they elongate, whereas the adjacent FH1 domains recruit profilin-bound actin monomers to accelerate elongation (*12*, *13*). The mammalian formin mDia1 supports one of the fastest known rates of actin filament elongation, at 55 subunits s^{-1} μM^{-1} (*13*).

CLIP-170 localizes to MT plus ends in cells via interactions with the MT end-binding protein EB1 but also localizes to actin-rich cortical zones (*5*, *14*–*19*). In addition, CLIP-170 binds and colocalizes with mDia1 in macrophages, where both proteins are required for phagocytosis (*5*), suggesting that their in vivo functions are tied together. The budding yeast protein Smy1 directly inhibits the yeast formin Bnr1, targeting its FH2 domain (*20*) via formin elongation effector domain (FEED) motifs (*21*). We discovered a FEED-like sequence in human CLIP-170, located in its central coiled-coil–rich region (Fig. 1A and fig. S1A). Thus, we purified full-length human CLIP-170 and used total internal reflection fluorescence (TIRF) microscopy to test its possible effects on mDia1-mediated actin assembly.

Mammalian CLIP-170 is expressed as three alternatively spliced isoforms (CLIP-170-1, CLIP-170-2, and CLIP-170-3) (fig. S1A). All splicing occurs in a specific region of CLIP-170 [amino acids (aa) 457 to 502], and the FEED motif, which is present in all three isoforms, borders the spliced region (fig. S1A). We purified all three isoforms of CLIP-170 as full-length proteins and tested their effects on mDia1 (FH1-FH2-tail)–mediated actin assembly. The CLIP-170 isoforms had no effect on actin assembly alone, but each one substantially accelerated the rate of mDia1-mediated actin filament elongation (Fig. 1, B to E; fig. S1, B to E; and movies S1 and S2). In constrast, these isoforms showed statistically insignificant effects on formin-mediated actin nucleation (fig. S1F). A closer examination of the distribution of filament elongation rates in reactions containing both mDia1 and CLIP-170 revealed three subpopulations with distinct rates (Fig. 1E and movie S2). The slowest had a mean elongation rate of 10.1 ± 1.3 (SE) subunits s^{-1} μM^{-1}, similar to the rate of free–barbed-end growth (*22*). The second subpopulation had a faster rate of 58.8 ± 8.8 subunits s^{-1} μM^{-1}, similar to filament elongation mediated by mDia1 alone (Fig. 1E and movie S2). The third was substantially accelerated, elongating at 179.1 ± 39.1 subunits s^{-1} μM^{-1}, or ~18 times faster than free–barbed-end growth in the absence of formins. This accelerated elongation occurred exclusively in reactions containing mDia1 and CLIP-170 and was observed at different actin concentrations (fig. S2). The fraction of filaments in the reaction undergoing accelerated elongation scaled with increasing concentrations of CLIP-170 (Fig. 1E). Because 25 nM CLIP-170 supported the strongest effects, this concentration was used for all further tests. CLIP-170 also increased the elongation rates of filaments assembled by four other formins (mDia2, Daam1, INF1, and INF2) (Fig. 1F and fig. S3). Thus, the regulatory effects of CLIP-170 extend to multiple formins, not just mDia1.

Accelerated elongation required profilin and the FH1 and FH2 domains of mDia1 (Fig. 1G). These are precisely the same requirements for mDia1-accelerated elongation without CLIP-170 (Fig. 1G)

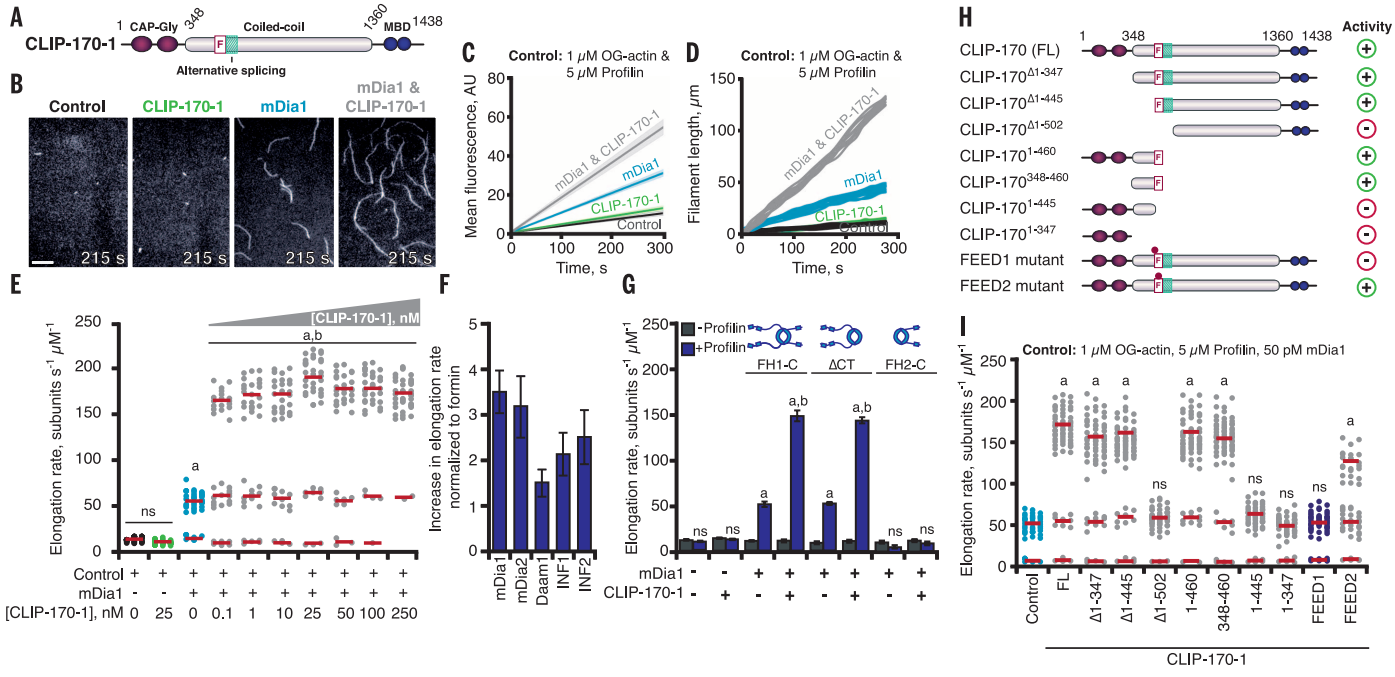


Fig. 1. CLIP-170 accelerates formin-mediated actin filament elongation. (**A**) CLIP-170 domains. F, FEED sequence; MBD, metal-binding domain. (**B**) Images from TIRF assays containing 50 pM mDia1 and/or 25 nM CLIP-170-1, taken 215 s after initiation of actin assembly. Scale bar, 20 μm. (**C**) Kinetics of total actin polymer mass (fluorescence intensity) accumulation averaged from multiple ($n \geq 3$) fields of view. Linear fits plotted with 95% confidence intervals are shown as shaded areas. AU, arbitrary units; OG-actin, Oregon Green actin. (**D**) Representative filament length traces (10 per condition) from TIRF movies. (**E**) Distributions of elongation rates from TIRF reactions, as in (B), for different concentrations of CLIP-170-1. Distributions are shown from one of three independent experiments (n = 50 filaments each). The red bars show mean elongation rates for subpopulations, measured from all filaments in three separate experiments (n = 150 filaments). (**F**) Fold increase in mean formin-mediated elongation rate stimulated by 25 nM CLIP-170-1. Error bars indicate SE. (**G**) Mean elongation rates from TIRF reactions containing different mDia1 constructs with or without profilin. (**H**) CLIP-170-1 constructs that enhance (+) or fail to enhance (−) the rate of mDia1-mediated elongation. FL, full length; F, FEED sequence; teal box, alternatively spliced region; red dots, FEED1 [^{445}VEEE/AAAA^{448} (V, Val; E, Glu; A, Ala)] and FEED2 [^{450}ITKGDLE/AAAAAAA^{456} (I, Ile; T, Thr; K, Lys; G, Gly; D, Asp; L, Leu)] mutants. (**I**) Distributions of elongation rates measured as in (E) for different CLIP-170-1 constructs. Reactions contained 1 μM globular actin (G-actin) (10% OG-labeled; 0.2% biotin-actin), ± 5 μM profilin, ± 50 pM mDia1, 50 pM mDia2, 50 pM Daam1, 100 nM INF1 or 100 nM INF2, ± 25 nM full-length CLIP-170-1 (WT, FEED1 mutant, or FEED2 mutant). Statistical differences in (E), (G), and (I): ns, not significantly different from control; a, compared with control (actin and profilin) (P < 0.05); b, compared with formin control (actin, profilin, and formin) (P < 0.05).

Downloaded from http://science.sciencemag.org/ on May 19, 2016

(*13*, *23*). Thus, CLIP-170 may be functioning by further improving the mechanism of elongation that is already used by formins. We mapped the activities in CLIP-170 to specific domains (Fig. 1, H and I). Because all three CLIP-170 isoforms had similar effects on mDia1, we mapped activities in the longest isoform, CLIP-170-1. All CLIP-170-1 fragments containing the FEED sequence supported accelerated elongation, and a short fragment (aa 348 to 460) encompassing the FEED sequence was sufficient, whereas the N-terminal MT- and EB1-binding region (aa 1 to 347) (*15*) was dispensable. Next, we generated two mutant alleles (alanine substitutions) in the FEED sequence of full-length CLIP-170-1 (fig. S1A). The FEED1 mutant abolished accelerated elongation, and the FEED2 mutant reduced the rate of accelerated elongation (Fig. 1I). Thus, CLIP-170–mediated effects on mDia1 are FEED-dependent.

A number of possible mechanisms could explain the effects of CLIP-170 on mDia1, including formation of stable or transient CLIP-170–mDia1 complexes at growing barbed ends of filaments. To address this, we used multiwavelength single-molecule TIRF microscopy to directly observe fluorescently labeled SNAP-tagged CLIP-170-1 and mDia1 molecules interacting with actin filaments during their assembly. 649-mDia1 and untagged mDia1 are dimeric and have indistinguishable activities (*24*). Step photobleaching analysis showed that 549-CLIP-170-1 molecules were also predominantly dimers (fig. S4), as expected (*15*). 649-mDia1 molecules processively tracked the barbed ends of filaments (Fig. 2A and movie S3, left panel) (*24*), and 549-CLIP-170-1 showed rare interactions with filaments. However, in reactions containing both 649-mDia1 and 549-CLIP-170-1, these proteins tracked the growing barbed end together (Fig. 2, A to E, and movie S3, right panel). 549-CLIP-170-1 and untagged CLIP-170-1 stimulated indistinguishable rates of formin-dependent accelerated elongation (Fig. 2B).

Accelerated elongation occurred only when 649-mDia1 and 549-CLIP-170-1 were observed together at the barbed end (300 out of 300 events) (Fig. 2, B to E, and movie S4). In our 3-min observation window, CLIP-170–dissociation events were rare (only 2 out of 100 barbed ends tracked), which suggests that CLIP-170-1–mDia1 barbed-end tracking complexes are long-lived and highly processive. In a few instances, we observed formation and dissolution of the mDia1–CLIP-170 complex (Fig. 2, D and E). The filament shown initially grew at the free–barbed-end rate (with no mDia1 or CLIP-170-1 associated); upon association of 649-mDia1, growth transitioned to a faster rate; and, subsequently, when 549-CLIP-170-1 joined the formin on the barbed end, growth immediately jumped to the further-enhanced rate. After several minutes, 549-CLIP-170-1 dissociated, and growth fell back to the mDia1-supported rate. When 649-mDia1 dissociated, growth returned to the free–barbed-end rate. Thus, when CLIP-170 joins mDia1 at the filament end, it forms a processive barbed-end tracking complex that enhances elongation. Single-molecule experiments also showed that CLIP-170-1 binds anchored mDia1, even in the absence of actin, and that binding is competitively displaced by an excess CLIP-170-1 fragment (aa 348 to 460) (Fig. 2F).

A hallmark of formins is their ability to protect growing barbed ends from capping protein

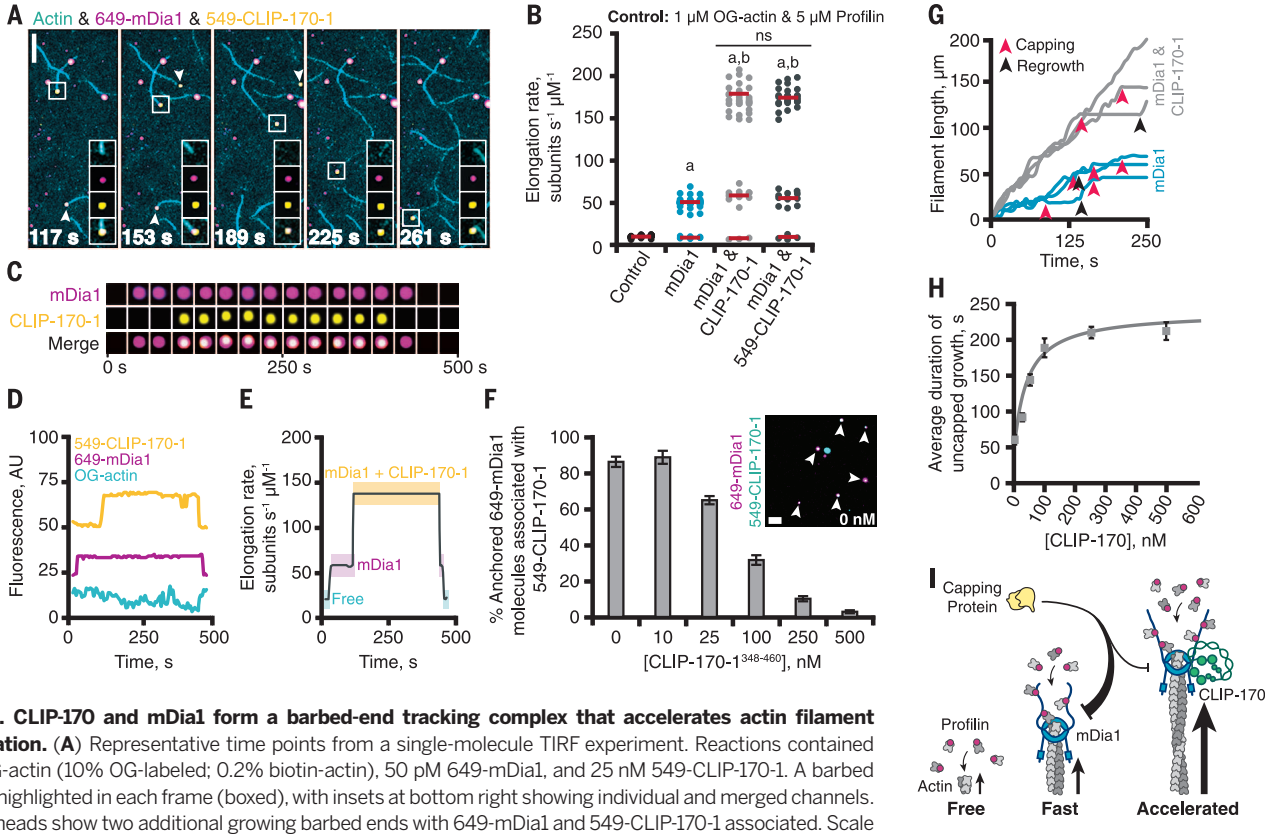


Fig. 2. CLIP-170 and mDia1 form a barbed-end tracking complex that accelerates actin filament elongation. (**A**) Representative time points from a single-molecule TIRF experiment. Reactions contained 1 μM G-actin (10% OG-labeled; 0.2% biotin-actin), 50 pM 649-mDia1, and 25 nM 549-CLIP-170-1. A barbed end is highlighted in each frame (boxed), with insets at bottom right showing individual and merged channels. Arrowheads show two additional growing barbed ends with 649-mDia1 and 549-CLIP-170-1 associated. Scale bar, 10 μm. Insets, 5 μm by 5 μm. (**B**) Effects of 25 nM CLIP-170-1 or 549-CLIP-170-1 on the rate of mDia1-mediated actin filament elongation. Reactions were as described in (A). Statistical differences: ns, not significantly different from control; a, compared with control (actin and profilin) ($P < 0.05$); b, compared with formin control (actin, profilin, and formin) ($P < 0.05$). (**C**) Formation and dissociation of a CLIP-170–mDia1 complex at a barbed end. (**D**) Fluorescence intensity profiles for each channel, showing formation and dissociation of the CLIP-170–mDia1 complex in (C). (**E**) Elongation rates correlate with arrival and dissociation of mDia1 and/or CLIP-170 at the barbed end. (**F**) Single-molecule colocalization of anchored 649-mDia1 and soluble full-length 549-CLIP-170-1 at different concentrations of unlabeled competitor fragment CLIP-170-1$^{348-460}$. Data were averaged from three fields of view in each of three independent experiments. Error bars indicate SE. The inset shows a representative field of view from a reaction with no competitor. Scale bar, 5 μm. (**G**) Representative filament traces from TIRF movies, conditions as in (A) except for the addition of 3 nM CP. Capping events (red arrowheads) and regrowth events (black arrowheads) are highlighted. (**H**) CLIP-170-1 enhances the duration of mDia1-mediated elongation in the presence of CP. Error bars indicate SE. (**I**) Cartoon of CLIP-170 joining mDia1 at the barbed end and increasing the rate of elongation and duration of growth in the presence of CP.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

(CP) (*25*–*27*). To test whether CLIP-170 influences this function of mDia1 (*26*, *28*), we spiked in different concentrations of CP early in bulk assays. CP slowed mDia1-mediated actin assembly in a dose-dependent manner (fig. S5A). These results are consistent with recent studies showing that CP and mDia1 can simultaneously interact with barbed ends and each catalyze dissociation of the other (*29*, *30*). CLIP-170-1 attenuated the effects

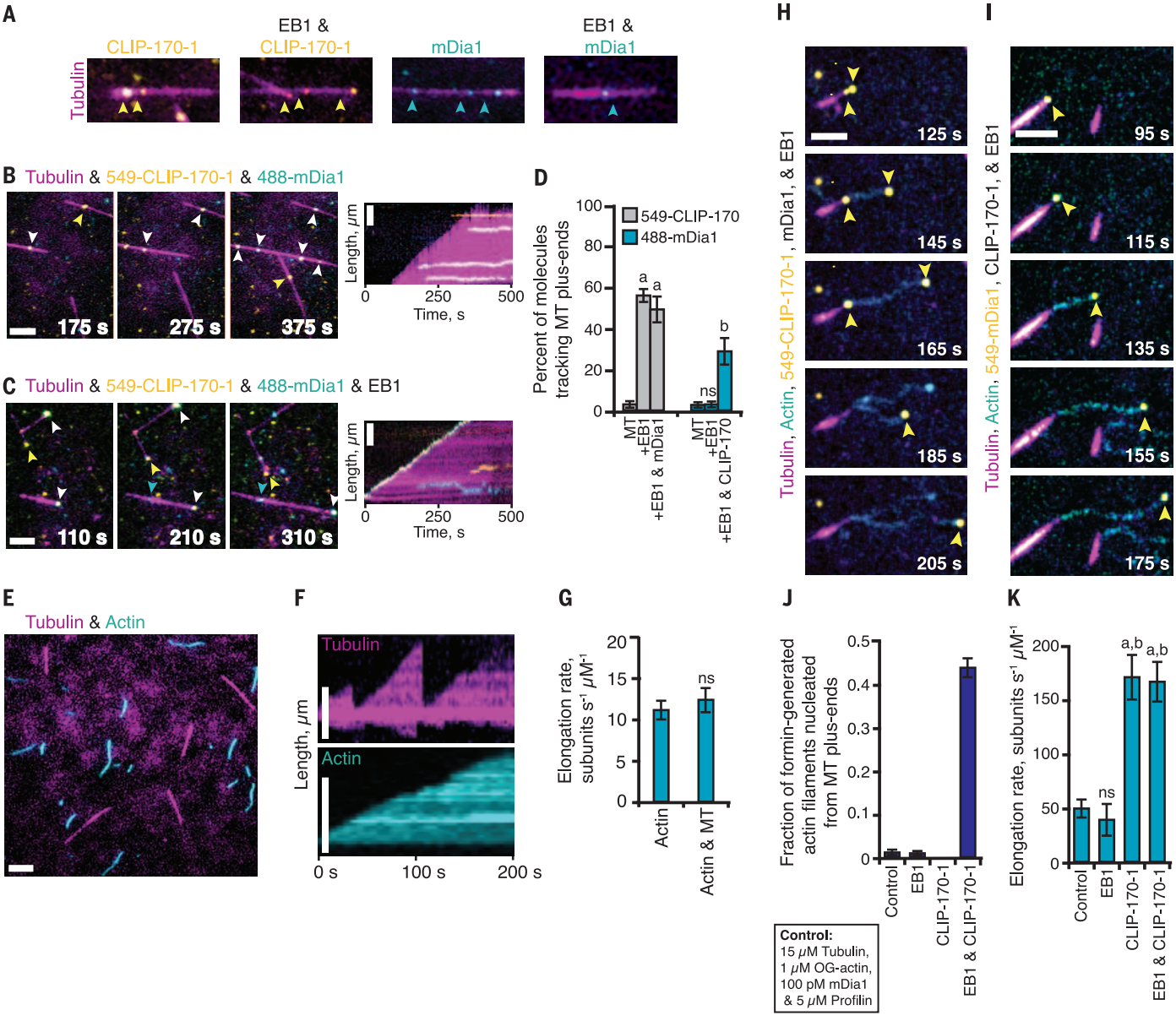


Fig. 3. CLIP-170–mDia1 complexes are recruited to MT plus ends by EB1 and stimulate actin polymerization from the MT surface. (**A**) TIRF movies show that 549-CLIP-170-1 binds MT sides and tracks MT plus ends only in the presence of EB1 (*17*). 488-mDia1 binds MT sides, with or without EB1, and does not track MT plus ends. Yellow arrowheads indicate CLIP-170 molecules; blue arrowheads indicate mDia1 molecules. (**B** and **C**) 549-CLIP-170-1 and 488-mDia1 colocalize on MT sides (B) and track MT plus ends together specifically in the presence of EB1 (C). White arrowheads indicate the presence of both CLIP-170 and mDia1 on MT plus ends. Scale bars, 5 μm. Reactions in (A) to (C) contain 15 μM tubulin (30% AlexaFluor649-labeled), biotinylated guanosine monophosphate–CPP MT seeds, and variable components (25 nM 549-CLIP-170-1, 100 pM 488-mDia1, and 500 nM EB1). (**D**) Percentage of molecules tracking MT plus ends from reactions in (A) to (C). Data were averaged from three experiments ($n > 100$ molecules). Statistical differences: ns, not different from control; a, compared with CLIP-170-1 ($P < 0.05$); b, compared with mDia1 ($P < 0.05$). (**E**) Co-reconstitution of MTs undergoing dynamic instability and polymerization of actin filaments. Reactions contained all of the same components as in (A) to (C), plus 1 μM G-actin (10% OG-labeled; 0.2% biotin-actin) and 5 μM profilin. Scale bar, 5 μm. (**F**) Kymographs of MT and actin dynamics. Scale bars, 5 μm. (**G**) The rate of actin filament elongation does not change significantly in the presence and absence of MTs. Data were averaged from three experiments ($n = 50$ actin filaments). (**H**) With the addition of unlabeled EB1 and mDia1, 549-CLIP-170-1 molecules (yellow arrowheads) were recruited to MT plus ends, where they triggered assembly of actin filaments that grew at the accelerated rate. Scale bar, 5 μm. (**I**) Similar observations as in (H), except using 549-mDia1 and unlabeled CLIP-170-1. Note that, in this panel, the yellow arrowheads indicate 549-mDia1 (rather than CLIP-170). Scale bar, 5 μm. (**J**) Fraction of formin-generated actin filaments that grew from MT plus ends in reactions as in (H) and (I). Data were from three experiments ($n = 50$ filaments per condition). (**K**) Actin elongation rates from reactions as in (H) and (I). Data were from three experiments ($n = 50$ filaments per condition). Statistical differences: ns, not different from control; a, compared with actin alone or control ($P < 0.05$); b, compared with EB1 control ($P < 0.05$). Error bars in (D) and (G) to (K), SE.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

of CP on mDia1 in these assays (fig. S5B). In TIRF microscopy assays, CLIP-170-1 increased the average elongation rate of mDia1-nucleated filaments in the presence of CP (fig. S5C). Similar effects were observed for each CLIP-170-1 construct that supported accelerated elongation (fig. S5D). CLIP-170-1 also increased the duration of accelerated growth in the presence of CP (Fig. 2, G and H, and fig. S5C) by up to a factor of 4. CP effects were sometimes reversed, as indicated by an abrupt return to growth at mDia1-supported rates, but these events were rare (n < 10 occurrences for each condition out of 300 capping events analyzed). On the other hand, CLIP-170 did not significantly affect the average duration of capping before growth resumed (fig. S5E). Thus, not only does CLIP-170 increase the rate of mDia1-mediated actin filament elongation by a factor of 3 to 4, it also prompts a fourfold increase in growth duration in the presence of CP (Fig. 2I).

CLIP-170 and mDia1 each bind MTs and EB1 (*17*, *31*–*34*). We thus investigated CLIP-170 and mDia1 interactions, alone and together, with dynamic MTs in the presence and absence of EB1 (Fig. 3A). 549-CLIP-170-1 interacted with MT sides and was recruited to growing MT ends by EB1 (Fig. 3A; fig. S6, A and B; and movie S5), as expected (*17*). 488-mDia1 bound to MT sides, both in the presence and absence of EB1, but was not recruited to MT ends (Fig. 3A; fig. S6, C and D; and movie S6). In the absence of EB1, 488-mDia1 colocalized with 549-CLIP-170-1 on MT sides (Fig. 3B and movie S6). However, in the presence of both EB1 and 549-CLIP-170-1, 488-mDia1 was recruited to the MT plus ends (Fig. 3C and movie S6). These observations demonstrate a hierarchal recruitment scheme wherein EB1 recruits CLIP-170, which recruits mDia1 to growing MT ends.

We developed a co-reconstitution TIRF system that enabled imaging of MTs undergoing dynamic instability and actin filaments polymerizing simultaneously (Fig. 3, E to G, and movie S7). In the presence of EB1, CLIP-170, and mDia1 (but not in reactions lacking any of these components), we observed coordination of growing MT ends with formation of rapidly polymerizing actin filaments. With unlabeled EB1 and mDia1, 549-CLIP-170-1 molecules were dynamically recruited to growing MT plus ends, where they triggered formation of actin filaments that polymerized from the MT surface (Fig. 3H and movie S8). Similar observations were made with unlabeled EB1, unlabeled CLIP-170-1, and 549-mDia1 (Fig. 3I and movie S9). About half of the actin filaments assembled by mDia1 originated from MT plus ends (Fig. 3J); in each case, 549-CLIP-170-1 or 549-mDia1 tracked the growing barbed ends, which polymerized at the accelerated rate away from the MT (Fig. 3K). Actin filament pointed ends remained attached to the MT surface for 45 ± 10.3 s, on average, until they spontaneously detached or were released by a MT catastrophe event.

To test the importance of these activities in a more physiological setting, we used a short hairpin RNA (shCLIP-170) to interfere with expression of all three CLIP-170 isoforms in rat primary cortical neurons (Fig. 4, A and B). Dendritic elaboration in these neurons depends on tight coordination between the MT and actin cytoskeletons (*11*, *16*, *35*). Depletion of CLIP-170 (validated in N2A cells; fig. S7) led to markedly reduced complexity of neuronal processes, as previously reported (*16*). This phenotype was rescued by a full-length wild-type (WT) CLIP-170-1 construct resistant to silencing (Fig. 4, A and B), but not the FEED1 mutant (Fig. 4,

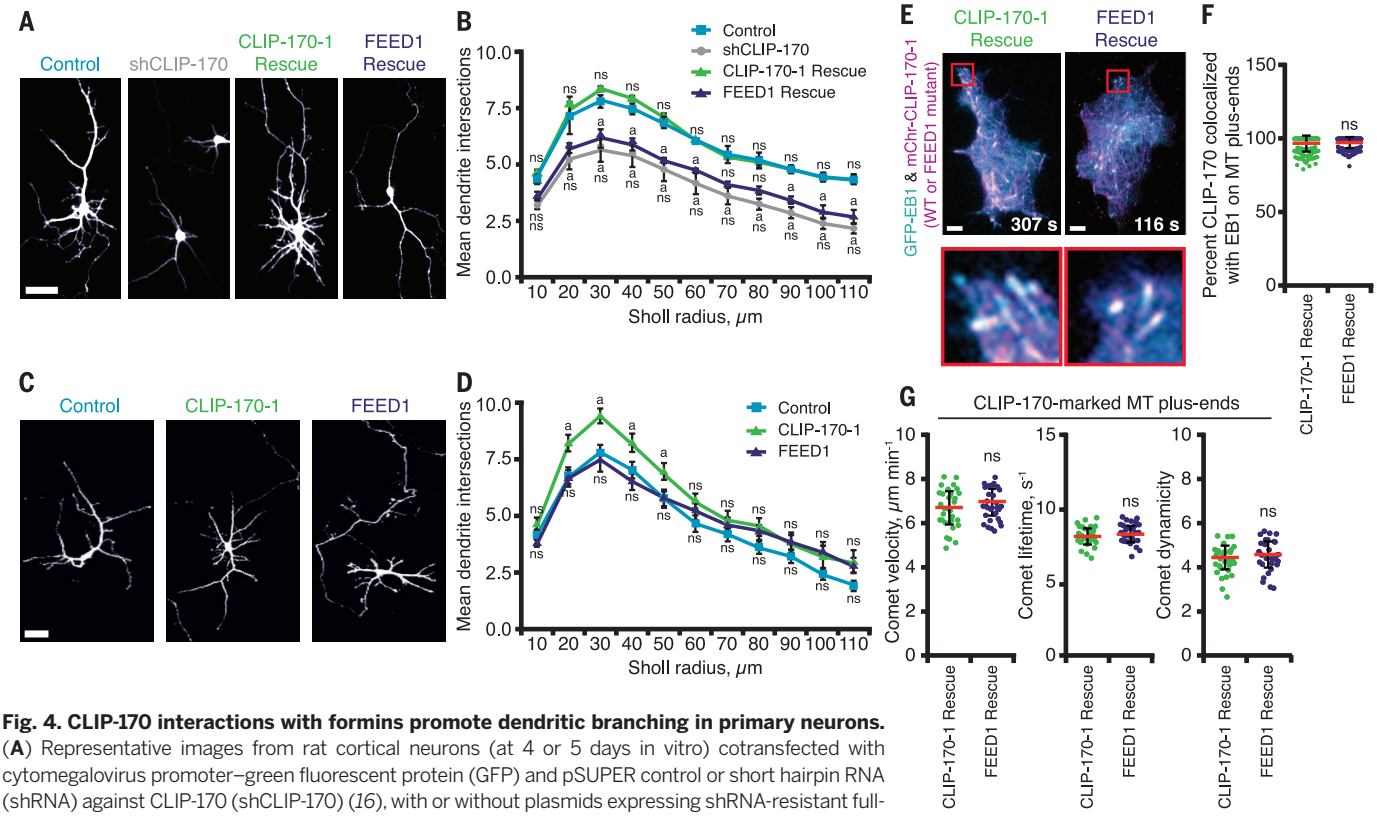


Fig. 4. CLIP-170 interactions with formins promote dendritic branching in primary neurons. (**A**) Representative images from rat cortical neurons (at 4 or 5 days in vitro) cotransfected with cytomegalovirus promoter–green fluorescent protein (GFP) and pSUPER control or short hairpin RNA (shRNA) against CLIP-170 (shCLIP-170) (*16*), with or without plasmids expressing shRNA-resistant full-length CLIP-170-1 rescue (WT or FEED1 mutant). Scale bar, 5 μm. (**B**) Quantification of dendritic branching complexity by Sholl analysis (n = 90 to 120 neurons, from two or more independent experiments). Statistical differences: ns, not significantly different from control; a, compared with scramble shRNA control (P < 0.05). There was never a significant difference between the FEED1 rescue and the shCLIP-170 knockdown alone. Error bars indicate SE. (**C**) Representative images from rat cortical neurons, as in (A), expressing full-length CLIP-170-1 (WT or FEED1 mutant) without silencing endogenous CLIP-170. Scale bar, 5 μm. (**D**) Quantification of dendritic branching complexity by Sholl analysis as in (B), except data are from n = 60 neurons in two independent experiments. Statistical differences: ns, not significantly different from control; a, compared with scramble shRNA control (P < 0.05). Error bars indicate SE. (**E**) Localization of mCherry–CLIP-170-1 (WT or FEED1 mutant) rescue constructs in N2A cells in which endogenous CLIP-170 was silenced. Scale bars, 10 μm. (**F**) Quantification of mCherry–CLIP-170-1 colocalization with EB1-GFP at MT plus ends from cells as in (E). Error bars indicate SD. (**G**) Quantitative tracking of mCherry–CLIP-170-1 (WT or FEED1 mutant) comets on growing MT plus ends. Analysis included comet velocity (or MT growth rate), comet lifetime (MT growth duration), and comet "dynamicity" [a general readout of MT dynamics (*37*)]. Error bars indicate SD.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

A and B). Conversely, increased expression of WT CLIP-170-1 over endogenous CLIP-170 led to elevated dendritic complexity, as previously shown (*16*), whereas expression of mutant CLIP-170-1 did not (Fig. 4, C and D). In N2A cells, WT and FEED1 mutant CLIP-170-1 were expressed at similar levels (fig. S7). Live imaging showed that they localized to MT plus ends similarly (Fig. 4, E and F; fig. S8, A to C; and movie S10) and that the mutant did not alter MT dynamics (Fig. 4G; fig. S8, D and E; and movies S10 and S11). Thus, CLIP-170 interactions with formins play an important role in coordinating MT and actin dynamics to regulate neuronal process formation.

Here we have shown that CLIP-170 interacts tightly with formins to substantially increase both the rate of actin filament elongation and the duration of elongation in the presence of CP. CLIP-170 is part of a mechanism that enables growing MT plus ends to trigger rapid assembly of actin filaments in vitro, directly linking MT and actin dynamics. This mechanism was consistent in a physiological setting, where EB1 and CLIP-170 colocalized on MT plus ends, as well as with previous reports that growing MT plus ends survey the actin-rich cortex (*10*) and that ~10% of mDia1 puncta in cells colocalize with MT plus ends (*32*). In neurons, CLIP-170 interactions with formins were required for proper dendritic branching. Similar mechanisms may explain the colocalization and cofunctioning of CLIP-170 and mDia1 in phagocytic cup formation (*5*) and reduced actin-based protrusive activity in neuronal growth cones after CLIP-170 silencing (*18*, *19*, *35*, *36*).

REFERENCES AND NOTES

1. P. Forscher, S. J. Smith, *J. Cell Biol.* **107**, 1505–1516 (1988).
2. O. C. Rodriguez *et al.*, *Nat. Cell Biol.* **5**, 599–609 (2003).
3. M. A. Chesarone, A. G. DuPage, B. L. Goode, *Nat. Rev. Mol. Cell Biol.* **11**, 62–74 (2010).
4. T. M. Svitkina *et al.*, *J. Cell Biol.* **160**, 409–421 (2003).
5. E. Lewkowicz *et al.*, *J. Cell Biol.* **183**, 1287–1298 (2008).
6. A. Sampathkumar *et al.*, *Plant Cell* **23**, 2302–2313 (2011).
7. S. G. Martin, W. H. McDonald, J. R. Yates III, F. Chang, *Dev. Cell* **8**, 479–491 (2005).
8. S. G. Martin, S. A. Rincón, R. Basu, P. Pérez, F. Chang, *Mol. Biol. Cell* **18**, 4155–4167 (2007).
9. D. M. Suter, P. Forscher, *J. Neurobiol.* **44**, 97–113 (2000).
10. W. C. Salmon, M. C. Adams, C. M. Waterman-Storer, *J. Cell Biol.* **158**, 31–37 (2002).
11. C. H. Coles, F. Bradke, *Curr. Biol.* **25**, R677–R691 (2015).
12. M. Chesarone, C. J. Gould, J. B. Moseley, B. L. Goode, *Dev. Cell* **16**, 292–302 (2009).
13. D. R. Kovar, E. S. Harris, R. Mahaffy, H. N. Higgs, T. D. Pollard, *Cell* **124**, 423–435 (2006).
14. A. Akhmanova *et al.*, *Cell* **104**, 923–935 (2001).
15. K. C. Slep, R. D. Vale, *Mol. Cell* **27**, 976–991 (2007).
16. D. Neukirchen, F. Bradke, *J. Neurosci.* **31**, 1528–1538 (2011).
17. R. Dixit *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 492–497 (2009).
18. J.-H. Weng *et al.*, *Nat. Chem. Biol.* **9**, 636–642 (2013).
19. R. Beaven *et al.*, *Mol. Biol. Cell* **26**, 1491–1508 (2015).
20. M. Chesarone-Cataldo *et al.*, *Dev. Cell* **21**, 217–230 (2011).
21. J. A. Eskin, A. Rankova, A. B. Johnston, S. L. Alioto, B. L. Goode, *Mol. Biol. Cell* **27**, 828–837 (2016).
22. J. R. Kuhn, T. D. Pollard, *Biophys. J.* **88**, 1387–1402 (2005).
23. C. J. Gould *et al.*, *Curr. Biol.* **21**, 384–390 (2011).
24. D. Breitsprecher *et al.*, *Science* **336**, 1164–1168 (2012).
25. J. B. Moseley *et al.*, *Mol. Biol. Cell* **15**, 896–907 (2004).
26. S. H. Zigmond *et al.*, *Curr. Biol.* **13**, 1820–1823 (2003).
27. E. S. Harris, F. Li, H. N. Higgs, *J. Biol. Chem.* **279**, 20076–20087 (2004).
28. M. A. Wear, J. A. Cooper, *Trends Biochem. Sci.* **29**, 418–428 (2004).
29. J. P. Bombardier *et al.*, *Nat. Commun.* **6**, 8707 (2015).
30. S. Shekhar *et al.*, *Nat. Commun.* **6**, 8730 (2015).
31. P. Pierre, J. Scheel, J. E. Rickard, T. E. Kreis, *Cell* **70**, 887–900 (1992).
32. Y. Wen *et al.*, *Nat. Cell Biol.* **6**, 820–830 (2004).
33. J. Gaillard *et al.*, *Mol. Biol. Cell* **22**, 4575–4587 (2011).
34. F. Bartolini *et al.*, *J. Cell Biol.* **181**, 523–536 (2008).
35. L. Swiech *et al.*, *J. Neurosci.* **31**, 4555–4568 (2011).
36. C. Erck *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7853–7858 (2005).
37. K. T. Applegate *et al.*, *J. Struct. Biol.* **176**, 168–184 (2011).

ACKNOWLEDGMENTS

We thank S. Paradis for guidance on experiments using neurons, D. Breitsprecher for pioneering the MT-actin co-reconstitution system, J. Gelles for guidance on single-molecule analysis, S. Jansen for guidance with cell culture, H. Higgs for providing INF1 and INF2 proteins, and L. Cassimeris for providing pGFP-EB1. This research was supported by NIH grant GM083137 to B.L.G. and Brandeis NSF Materials Research Science and Engineering Center grant 142038. J.L.H.-R. was supported in part by a fellowship from the Leukemia and Lymphoma Society and in part by NIH training grant T32NS007292. J.L.H.-R. and B.L.G. designed the experiments and wrote the manuscript, J.L.H.-R. performed the experiments and data analysis, J.A.E. performed data analysis, A.R. built reagents and performed preliminary experiments, and K.K. performed overexpression analysis in neurons and assisted in other neuronal work. We declare no conflicts of interest. The supplementary materials contain additional data.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/1004/suppl/DC1
Materials and Methods
Figs. S1 to S9
References (*38*–*48*)
Movies S1 to S11

29 December 2015; accepted 5 April 2016
10.1126/science.aaf1709

GENE EVOLUTION

Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals

Xun Lan[1,3]* and Jonathan K. Pritchard[1,2,3]*

Gene duplication is a fundamental process in genome evolution. However, most young duplicates are degraded by loss-of-function mutations, and the factors that allow some duplicate pairs to survive long-term remain controversial. One class of models to explain duplicate retention invokes sub- or neofunctionalization, whereas others focus on sharing of gene dosage. RNA-sequencing data from 46 human and 26 mouse tissues indicate that subfunctionalization of expression evolves slowly and is rare among duplicates that arose within the placental mammals, possibly because tandem duplicates are coregulated by shared genomic elements. Instead, consistent with the dosage-sharing hypothesis, most young duplicates are down-regulated to match expression levels of single-copy genes. Thus, dosage sharing of expression allows for the initial survival of mammalian duplicates, followed by slower functional adaptation enabling long-term preservation.

Gene duplications are a major source of new genes and ultimately of new biological functions (*1*). However, recently arisen gene duplicates tend to be functionally redundant and thus susceptible to loss-of-function mutations that degrade one of the copies into a pseudogene. The average half-life of new primate duplicates has been estimated at just 4 million years (*2*). This raises the question of what evolutionary forces govern the persistence of young duplicates.

Various models have been proposed to understand why some duplicate pairs do survive over long evolutionary time scales (*3*). Dosage-balance models focus on the importance of maintaining correct stoichiometric ratios in gene expression between different genes (*4*–*6*) and likely explain how gene copies are maintained after whole-genome duplication (WGD), because subsequent gene losses would disrupt dosage balance (*6*, *7*).

Alternatively, functional partitioning of duplicates can occur, either by neofunctionalization (one copy gains new functions) or subfunctionalization (the copies divide the ancestral functions between them). The duplication-degeneration-complementation (DDC) model proposes that complementary degeneration of regulatory elements causes the two copies to be expressed in different tissues, such that both copies are required to provide the overall expression of the ancestral gene (*8*). Similarly, neofunctionalization of expression could lead to one gene copy gaining function in a tissue where the parent gene was not expressed. Functional divergence may also occur at the protein level (*9*), but this is thought to be a slow process, with initial divergence more often occurring through changes in gene regulation (*10*).

It is currently unclear which factors are most important for long-term survival of gene duplications in mammals, where most duplications arise

[1]Department of Genetics, Stanford University, Stanford, CA, USA. [2]Department of Biology, Stanford University, Stanford, CA, USA. [3]Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA.
***Corresponding author. Email: xlan@stanford.edu (X.L.); pritch@stanford.edu (J.K.P.)**

Downloaded from http://science.sciencemag.org/ on May 19, 2016

through segmental duplications or retrotranspositions that increase copy numbers of just one or a few genes. These small-scale duplications most likely disrupt overall dosage balance and should thus favor gene loss rather than preservation.

We therefore set out to investigate whether gene expression data across tissues in human and mouse support either model of duplicate preservation. We analyzed RNA-sequencing (RNA-seq) data from 10 individuals for each of 46 diverse human tissues collected by the Genotype-Tissue Expression (GTEx) project (*11*) and replicated our main conclusions using RNA-seq from 26 diverse mouse tissues (*12*).

We developed a computational pipeline to identify duplicate gene pairs in the human genome (*13*). After excluding annotated pseudogenes, we identified 1444 high-confidence reciprocal best-hit duplicate gene pairs with >80% alignable coding sequence and >50% average sequence identity. We used synonymous divergence, d_S, as a proxy for divergence time, while noting that divergence of some gene pairs may be affected by nonallelic homologous gene conversion in young duplicates. Additional analyses using the phylogenetic distribution of duplicates to refine date estimates were highly concordant with results based on d_S alone (figs. S5 to S7). We estimate that d_S for duplicates that arose at the time of the human-mouse split averages ~0.45 and that most pairs with d_S > ~0.7 predate the origin of the placental mammals (figs. S3 and S4). Thus, most of our analysis focuses on duplicates that likely arose within the mammalian lineage and postdate the early vertebrate whole-genome duplications.

Accurate measurement of expression in gene duplicates can be challenging if RNA-seq reads map well to both gene copies. Mapping may also be biased if the two copies have differential homology with other genomic locations. To overcome these challenges, we estimated expression ratios using only paralogous positions for which reads from both copies would map uniquely to the correct gene (*13*). This approach is related to a method for measuring allele-specific expression (*14*). These strict criteria mean that some very young genes are excluded from our expression analyses as unmappable, but, for the remaining genes, simulations show that our pipeline yields highly accurate, unbiased estimates of expression ratios (fig. S1).

This read-mapping pipeline allowed us to classify duplicates into categories on the basis of their coexpression patterns (*13*). First, within each pair, we classified the gene with higher overall expression as the "major" gene and its partner as the "minor" gene. We then defined a gene pair as potentially sub- or neofunctionalized if both the major and minor copy are significantly more highly expressed than the other in at least one tissue each (at least a twofold difference and $P < 0.001$ with paired *t* test) (Fig. 1A). We refer to pairs with consistent asymmetry as asymmetrically expressed duplicates (AEDs) if the major gene is significantly more highly expressed in at least 1/3 of tissues where either gene is expressed and not expressed at a significantly lower level than its partner in any tissue (Fig. 1B). The remaining duplicates were classified as having no difference, although many of these pairs show weaker levels of asymmetry.

Few duplicate pairs show evidence of sub- or neofunctionalization of expression (Fig. 2, A to C). Moreover, most gene pairs with such patterns are very old, dating to before the emergence of the placental mammals: For duplicates with d_S < 0.7, just 15.2% of duplicates are classified as potentially sub- or neofunctionalized in expression. Given that even modest variation in expression profiles across tissues would meet our criteria for subfunctionalization, the fraction of truly subfunctionalized duplicates may be even lower.

We also found similar levels of potential subfunctionalization in a mouse data set (*12*) that, unlike GTEx, includes fetal tissues (fig. S14). We examined whether subfunctionalization might instead be occurring through differential splicing of exons; however, we found little evidence for this (fig. S20). Last, we hypothesized that subfunctionalization might be more prevalent in gene pairs with higher tissue specificity (because they likely have more tissue-specific enhancers), but this is not the case (fig. S13).

Although relatively scarce, the genes identified as potentially subfunctionalized exhibit systematic differences from other duplicates. First, subfunctionalized gene pairs are expected to be under stronger selective constraint than genes without diverged expression, because the two copies are not functionally redundant. Consistent with this, we find that putatively subfunctionalized genes tend to have a higher fraction of rare variants in human polymorphism data (*15*) ($P = 2 \times 10^{-5}$ for missense mutations; Kolmogorov-Smirnov test) (Fig. 2D). Second, we hypothesized that if subfunctionalized genes have distinct

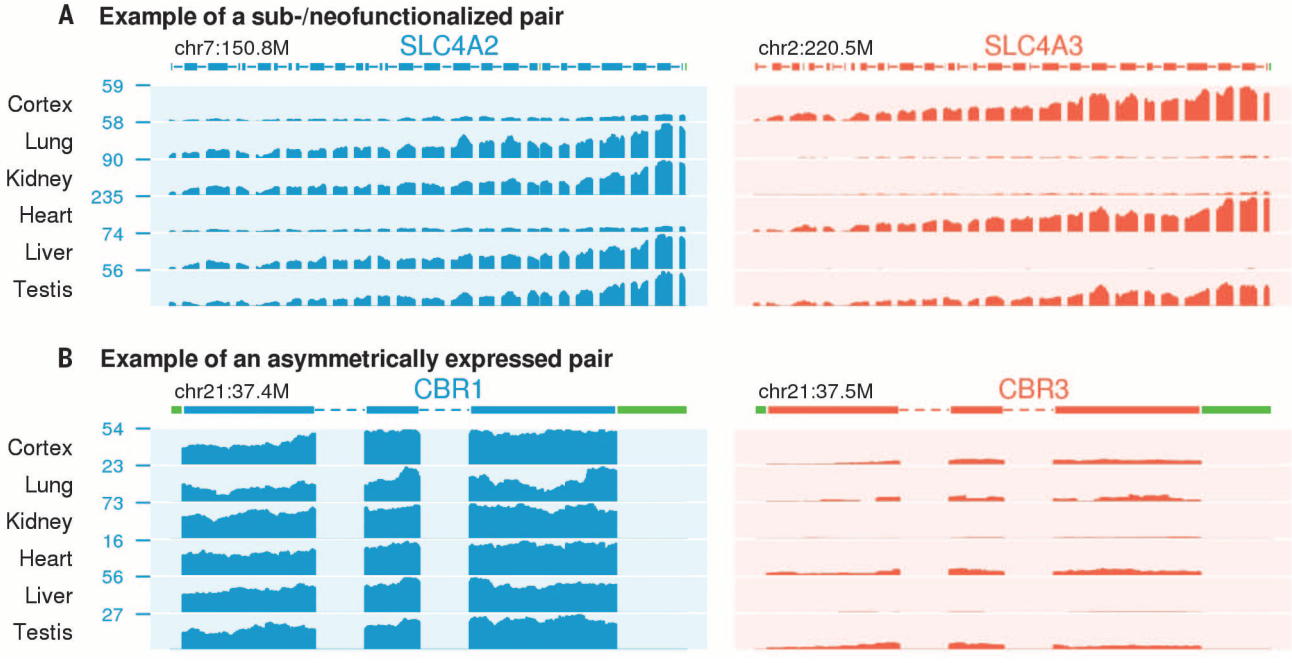


Fig. 1. Expression profiles of duplicate genes. (**A**) A gene pair whose expression profile is consistent with sub- or neofunctionalization: i.e., each gene is significantly more highly expressed than the other in at least one tissue. (**B**) An asymmetrically expressed gene pair. Expression of CBR1 exceeds expression of CBR3 in all tissues. Introns shortened for display purposes. The *y* axis shows read depth per billion mapped reads. Green regions in the gene models are unmappable.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

functions, then they may be associated with distinct genetic diseases. Examining a database of gene associations with disease (*16*), we found a correlation between the degree of expression subfunctionalization and the number of diseases reported for only one member of the gene pair ($P = 5 \times 10^{-12}$, controlling for relevant covariates; Wald test) (Fig. 2E and table S3).

In sharp contrast to the expectations of subfunctionalization, many duplicate pairs exhibit systematically biased expression, as seen in some species after whole-genome duplication (*17*). Across all duplicate pairs, the mean expression of the less-expressed gene is 40% that of its duplicate (Fig. 2, B and C) ($P \sim 0$, relative to a model with no true asymmetry). Among duplicates that likely arose within the placental mammals ($d_S < 0.7$), 52.6% of duplicate pairs are AEDs, compared with just 15.2% that are potentially subfunctionalized. As might be expected, the minor genes at AEDs show evidence of reduced selective constraint relative to their duplicate partners, both within the human population (fig. S23) and between species (fig. S21). Furthermore, in gene pairs with asymmetric expression, the minor genes tend to be associated with significantly fewer diseases ($P = 8 \times 10^{-7}$; Wald test) (Fig. 2E). Nonetheless, despite their reduced importance, minor genes are not dispensable: 97% of minor genes have $d_N/d_S < 1$, a hallmark of protein-coding constraint (fig. S21).

Together, these results show that subfunctionalization of expression evolves slowly. However, we noticed much higher rates of sub- or neofunctionalization for duplicates located on different chromosomes, compared with duplicates in tandem ($P = 5 \times 10^{-23}$; Fisher's exact test) (fig. S24). We thus wanted to understand whether separation of duplicates enables subfunctionalization or whether the higher rate simply reflects the greater age of separated duplicates. Most duplicates arise as segmental duplications (*18*) and are close together in the genome: 87% of young gene pairs ($d_S < 0.1$) are on the same chromosome (Fig. 3A). Duplicates may subsequently become separated as the result of chromosomal rearrangements; however, this is a slow process. It is not until $d_S = 0.6$ that half of gene duplicates are found on different chromosomes.

Even controlling for duplicate age, however, there is a strong signal that genomic separation is a key factor enabling expression divergence (Fig. 3B). Separated duplicates have roughly 50% lower correlation of expression across tissues:

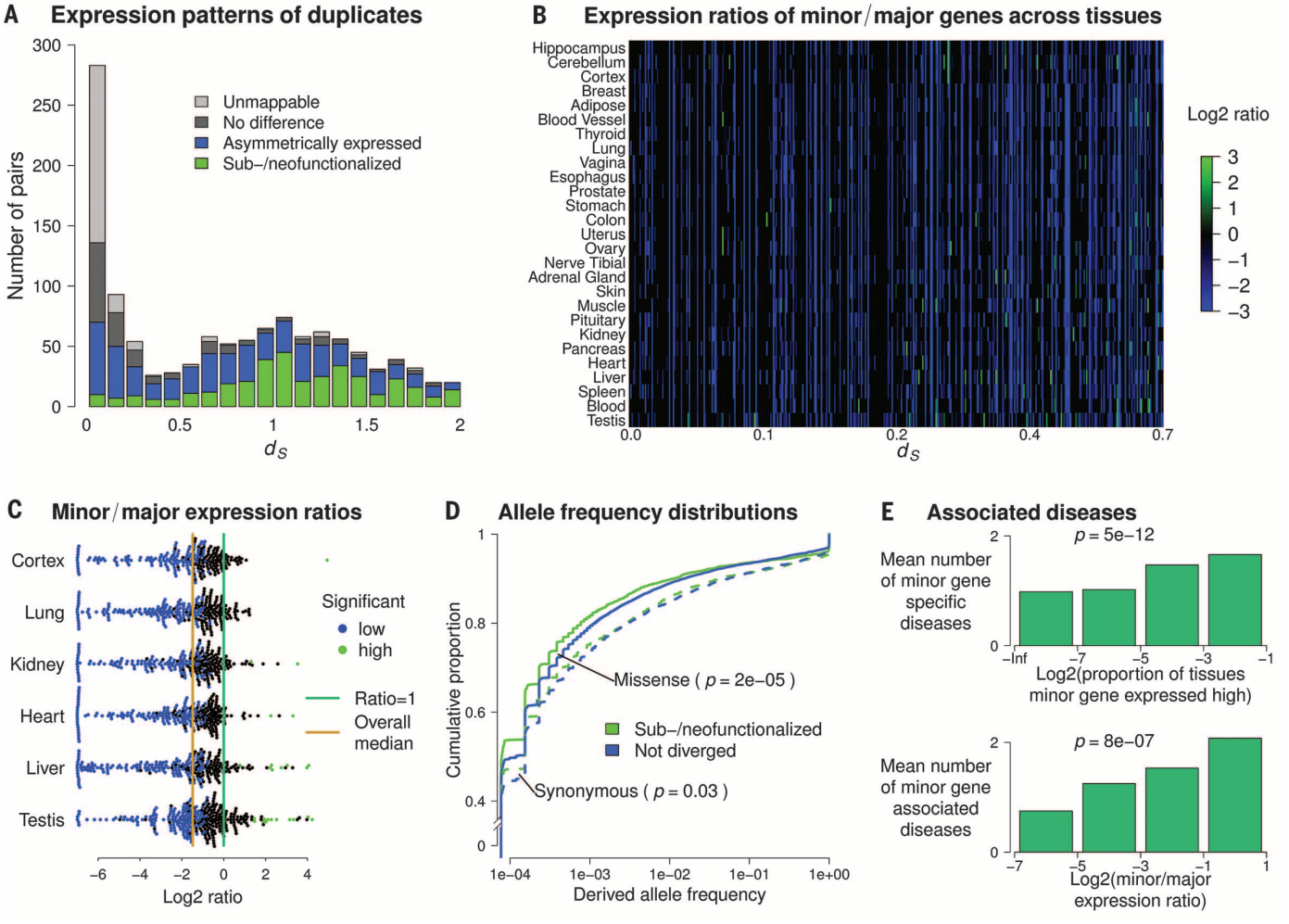


Fig. 2. Properties of subfunctionalized genes. (**A**) Classification of gene pairs by expression patterns. For context, note that duplicates arising at the human-mouse split would have $d_S \sim 0.45$. (**B**) Heat map of expression ratios for duplicate pairs. For each duplicate pair (plotted in columns), the ratios show the tissue-specific expression level of the minor gene relative to its duplicate. Green indicates evidence for subfunctionalization; consistently blue columns indicate AEDs. Black indicates tissue ratios not significantly different from 1 ($P > 0.001$). (**C**) Distributions of expression ratios in different tissues (minor genes/major genes). Ratios significantly >1 marked in green. (**D**) Frequency spectra of human polymorphism data (*15*) for synonymous and nonsynonymous variants in subfunctionalized duplicates (green) and duplicates without significant expression differences (black). The plots show cumulative derived allele frequencies at segregating sites. The lines that climb more steeply (subfunctionalized genes) have a higher fraction of rare variants, indicating stronger selective constraint. (**E**) Disease burden of minor genes is highly correlated with degree of subfunctionalization (top) and overall expression relative to major genes (bottom). Data in (B), (C), and (D) are for $d_S < 0.7$.

Downloaded from http://science.sciencemag.org/ on May 19, 2016

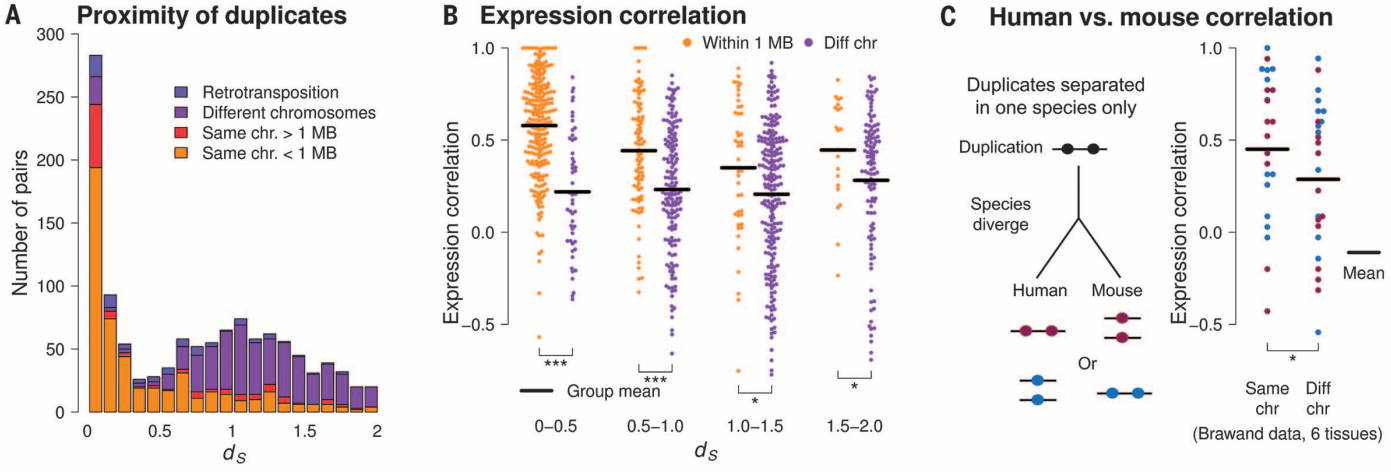
$P = 3 \times 10^{-30}$, controlling for age by d_S in a multiple regression model (table S4 and fig. S26); $P = 6 \times 10^{-18}$, controlling for age by phylogenetic distribution (fig. S7). Further, we see the same effect in a paired test of duplicates that are separated in human but not mouse, or vice versa (Fig. 3C). Notably, duplicate age itself is a much weaker predictor ($P = 2 \times 10^{-6}$ for d_S) than is genomic separation ($P = 3 \times 10^{-30}$) (table S4). [In contrast to correlation across tissues, the asymmetry of mean expression is uncorrelated with whether the duplicates are on the same chromosome or not ($P = 0.9$, controlling for d_S; Wald test).]

These results echo previous observations that, in general, genes that are close in the genome tend to be coregulated, with correlated expression (*19*) and often shared expression quantitative trait loci (eQTLs) (*20*). This effect is yet stronger for duplicates: Gene expression is more correlated for tandem duplicates than for singleton neighbors ($P = 10^{-19}$; *t* test) (Fig. 3D), and duplicates share eQTLs at higher rates than matched singletons ($P = 6 \times 10^{-4}$ and 5×10^{-4} in two data sets; Fisher's exact test) (*13*, *20*, *21*). Further, duplicates show higher connectivity by whole-genome chromosome conformation capture (Hi-C) (*22*), including higher numbers of promoter-promoter links than neighboring singletons (Fig. 3E) (mean effect size = 1.7-fold, $P = 3 \times 10^{-6}$; Wald test) (*13*). Promoter-promoter links may reflect a tendency of coregulated genes to be transcribed simultaneously within transcription factories (*23*). In contrast, duplicates on different chromosomes show no evidence of Hi-C linkage. In summary, we hypothesize that tandem duplicates tend to be highly coregulated and that genomic separation is a key factor enabling independent evolution.

Thus far, our results argue that expression subfunctionalization evolves slowly, in large part because tandem duplicates tend to be coregulated. An alternative explanation for the initial survival of duplicates is that they are both necessary to produce the required expression dosage (*6*). However, in contrast to whole-genome duplications, the small-scale duplications that are typical in mammals would initially disrupt dosage of the duplicated genes relative to all other genes. Thus, if dosage sharing is important in mammals, this would suggest that after tandem duplication, the duplicates should rapidly evolve reduced expression. Subsequent loss of either gene would then cause a deficit of expression and be deleterious.

To evaluate this, we analyzed the expression of human duplicates that arose since the human-macaque split, using RNA-seq data from six tissues in human and macaque (Fig. 4A) (*13*, *24*). Indeed, there is a very clear signal that both human copies tend to evolve reduced expression, such that the median summed expression of the human duplicates is close to the expression of the singleton orthologs in macaque (median expression ratio 1.11; this is significantly less than the 2:1 expression ratio expected on the basis of copy number, $P = 3 \times 10^{-7}$; *t* test). Interestingly, polymorphic duplicates also show partial down-regulation, whereas the youngest fixed duplicates are about as down-regulated as older pairs, suggesting that reduced expression occurs rapidly (fig. S19). In contrast, we find no evidence for coding adaptation in these relatively young duplicate pairs (fig. S16). Thus, dosage sharing may be a frequent first step in the preservation of tandem duplicates. However, although dosage sharing evolves quickly, it is notable that duplicate genes remain less conserved than singleton genes over long evolutionary time scales ($d_S \leq 0.7$, or roughly the age of placental mammals) (Fig. 4B and fig. S22).

We propose that down-regulation is a key first step enabling the initial survival of duplicates, followed by dosage sharing, as suggested for WGDs (Fig. 4C) (*6*). In this view, the early survival of young duplicates is a race between down-regulation to achieve dosage balance versus mutational degradation of one copy. If dosage balance is achieved, then the relative expression levels of



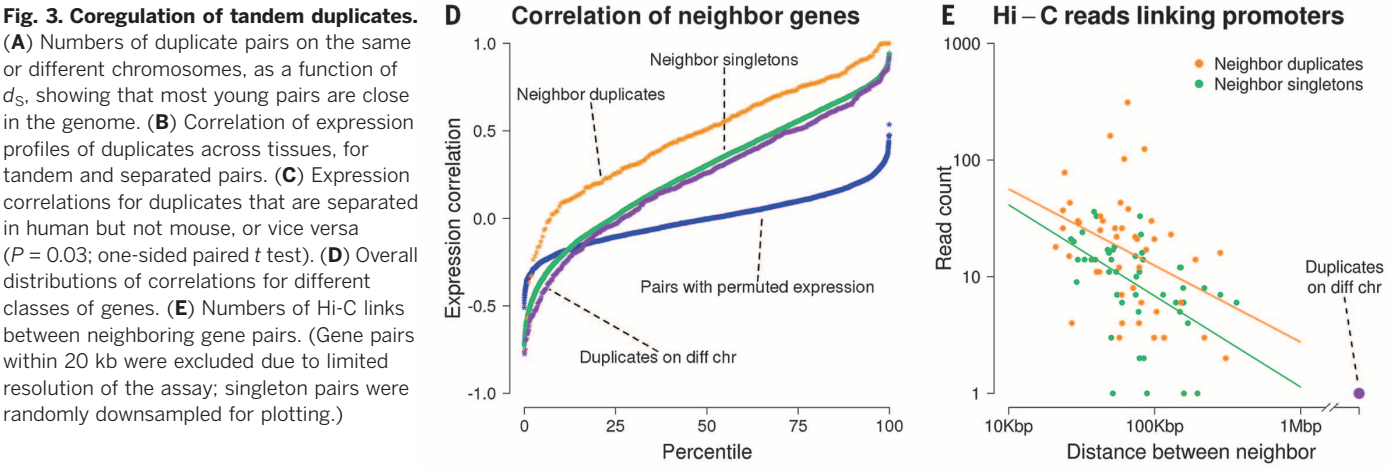


Fig. 3. Coregulation of tandem duplicates. (**A**) Numbers of duplicate pairs on the same or different chromosomes, as a function of d_S, showing that most young pairs are close in the genome. (**B**) Correlation of expression profiles of duplicates across tissues, for tandem and separated pairs. (**C**) Expression correlations for duplicates that are separated in human but not mouse, or vice versa ($P = 0.03$; one-sided paired *t* test). (**D**) Overall distributions of correlations for different classes of genes. (**E**) Numbers of Hi-C links between neighboring gene pairs. (Gene pairs within 20 kb were excluded due to limited resolution of the assay; singleton pairs were randomly downsampled for plotting.)

Downloaded from http://science.sciencemag.org/ on May 19, 2016

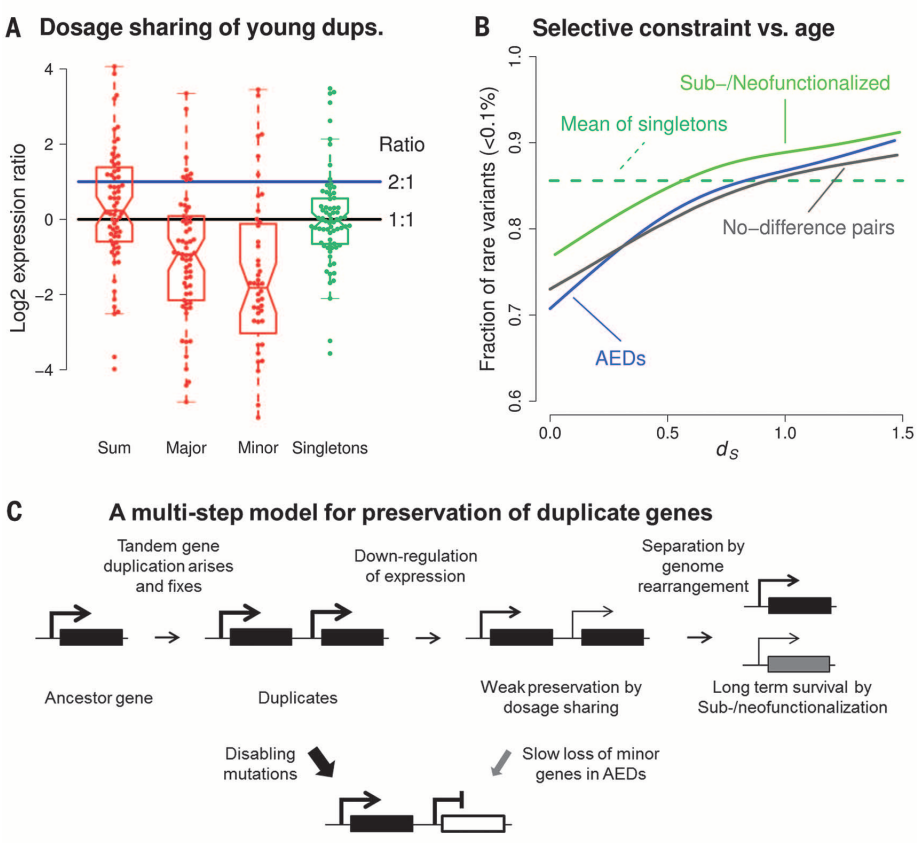


Fig. 4. Long-term survival of duplicate genes. (**A**) Expression levels of young duplicates compared to their macaque orthologs in six tissues (*24*), for human duplicates that are single-copy genes in macaque. Sum shows the summed expression of both duplicates, relative to expression of the macaque orthologs in the same tissues. "Major" and "Minor" show corresponding ratios for major and minor genes separately, classified using GTEx data. The green data show a random set of singleton orthologs. Each tissue-gene expression ratio is plotted separately. (**B**) The strength of purifying selection in humans increases with duplicate age. The fraction of rare missense variants in a large human data set (*15*) is used as a proxy for the strength of purifying selection. (**C**) Conceptual model of duplicate gene evolution. Other transitions not explicitly shown would occur at lower but nonzero rates.

the two genes evolve slowly as a random walk due to constraint on their combined expression (*7*, *25*). Both copies tend to evolve under reduced constraint, especially for minor genes of AEDs. Genomic separation frees expression of the duplicates to evolve independently and may also encourage protein adaptation, potentially leading to true functional differentiation and long-term survival. In summary, we find that subfunctionalization of expression evolves slowly in mammals due to coregulation of tandem duplicates and that rapid evolution of dosage sharing may be the most frequent first step to duplicate preservation.

REFERENCES AND NOTES

1. G. C. Conant, K. H. Wolfe, *Nat. Rev. Genet.* **9**, 938–950 (2008).
2. M. Lynch, J. S. Conery, *J. Struct. Funct. Genomics* **3**, 35–44 (2003).
3. H. Innan, F. Kondrashov, *Nat. Rev. Genet.* **11**, 97–108 (2010).
4. A. Stoltzfus, *J. Mol. Evol.* **49**, 169–181 (1999).
5. W. Qian, B. Y. Liao, A. Y. Chang, J. Zhang, *Trends Genet.* **26**, 425–430 (2010).
6. G. C. Conant, J. A. Birchler, J. C. Pires, *Curr. Opin. Plant Biol.* **19**, 91–98 (2014).
7. J. F. Gout, M. Lynch, *Mol. Biol. Evol.* **32**, 2141–2148 (2015).
8. A. Force *et al.*, *Genetics* **151**, 1531–1545 (1999).
9. C. R. Baker, V. Hanson-Smith, A. D. Johnson, *Science* **342**, 104–108 (2013).
10. I. Wapinski, A. Pfeffer, N. Friedman, A. Regev, *Nature* **449**, 54–61 (2007).
11. GTEx Consortium, *Science* **348**, 648–660 (2015).
12. T. Babak *et al.*, *Nat. Genet.* **47**, 544–549 (2015).
13. See supplementary materials and methods on *Science* Online.
14. B. van de Geijn, G. McVicker, Y. Gilad, J. K. Pritchard, *Nat. Methods* **12**, 1061–1063 (2015).
15. W. Fu *et al.*, *Nature* **493**, 216–220 (2013).
16. K. Peng *et al.*, *Nucleic Acids Res.* **41**, D553–D560 (2013).
17. E. W. Ganko, B. C. Meyers, T. J. Vision, *Mol. Biol. Evol.* **24**, 2298–2309 (2007).
18. J. A. Bailey *et al.*, *Science* **297**, 1003–1007 (2002).
19. A. T. Ghanbarian, L. D. Hurst, *Mol. Biol. Evol.* **32**, 1748–1766 (2015).
20. A. Battle *et al.*, *Genome Res.* **24**, 14–24 (2014).
21. T. Lappalainen *et al.*, *Nature* **501**, 506–511 (2013).
22. S. S. Rao *et al.*, *Cell* **159**, 1665–1680 (2014).
23. A. Feuerborn, P. R. Cook, *Trends Genet.* **31**, 483–490 (2015).
24. D. Brawand *et al.*, *Nature* **478**, 343–348 (2011).
25. K. Y. Popadin *et al.*, *Am. J. Hum. Genet.* **95**, 660–674 (2014).

ACKNOWLEDGMENTS

This work was funded by NIH grants ES025009 and MH101825 and by the Howard Hughes Medical Institute. We thank H. Fraser for prepublication access to data (*12*) and H. Fraser, A. Fu, A. Harpak, Y. I. Li, D. Petrov, P. C. Phillips, M. Przeworski, A. Stoltzfus, and the anonymous reviewers for comments or discussion. J.K.P. is on advisory boards for 23andMe and DNAnexus, with stock options in both.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6288/1009/suppl/DC1
Materials and Methods
Supplemental Text
Figs. S1 to S27
Tables S1 to S5
Supplementary Files S1 to S3
References (*26*–*79*)

9 November 2015; accepted 2 April 2016
10.1126/science.aad8411

Downloaded from http://science.sciencemag.org/ on May 19, 2016

WHO WILL WIN THE PRIZE

The result will be unveiled in June.
The winners will receive:

★A Grand Prize of $25,000 and a Runner-Up Prize of $5,000 will be awarded.
★The Grand Prize Winning Essay will be published in *Science*;
a brief abstract of the Runner-Up Essay will be published in *Science*.

The 2016 award ceremony will be held in San Francisco, on 23rd June, 2016.

Stem cells and regenerative medicine is the new frontier in life sciences. Boyalife, *Science*, and *Science Translational Medicine* jointly establish a global award to recognize and reward scientists in the fields of stem cells and regenerative medicine with a focus on developing cell-based treatments for cancer, degenerative disorders, immunotherapy and stem cells transplantation.

BOYALIFE Science & Science Translational Medicine Award | in Stem Cells and Regenerative Medicine

Learn more about the Prize, visit :
http://www.sciencemag.org/prizes/boyalife

BOYALIFE For Better Health

Science AAAS

Science Translational Medicine AAAS

400

BIO-RAD'S DROPLET DIGITAL™ PCR SYSTEMS

Over 400 Peer-Reviewed Droplet Digital PCR (ddPCR™) Publications*

From detection of rare mutations and cancer biomarkers to quantification of gene editing events and miniscule viral loads, the QX100™ and QX200™ Droplet Digital PCR Systems have been used to redefine the limits of absolute nucleic acid quantification. With over 400 peer-reviewed publications, ddPCR platforms have outperformed other digital PCR systems by several orders of magnitude. The third-generation QX200™ AutoDG™ System now brings automation and scalability to digital PCR.

Visit **bio-rad.com/info/400science** for the publication list and to learn more.

BIO-RAD

* Based on PubMed data, March 2016.

AAAS 2017 ANNUAL MEETING

FEBRUARY 16-20 • BOSTON

SERVING SOCIETY THROUGH

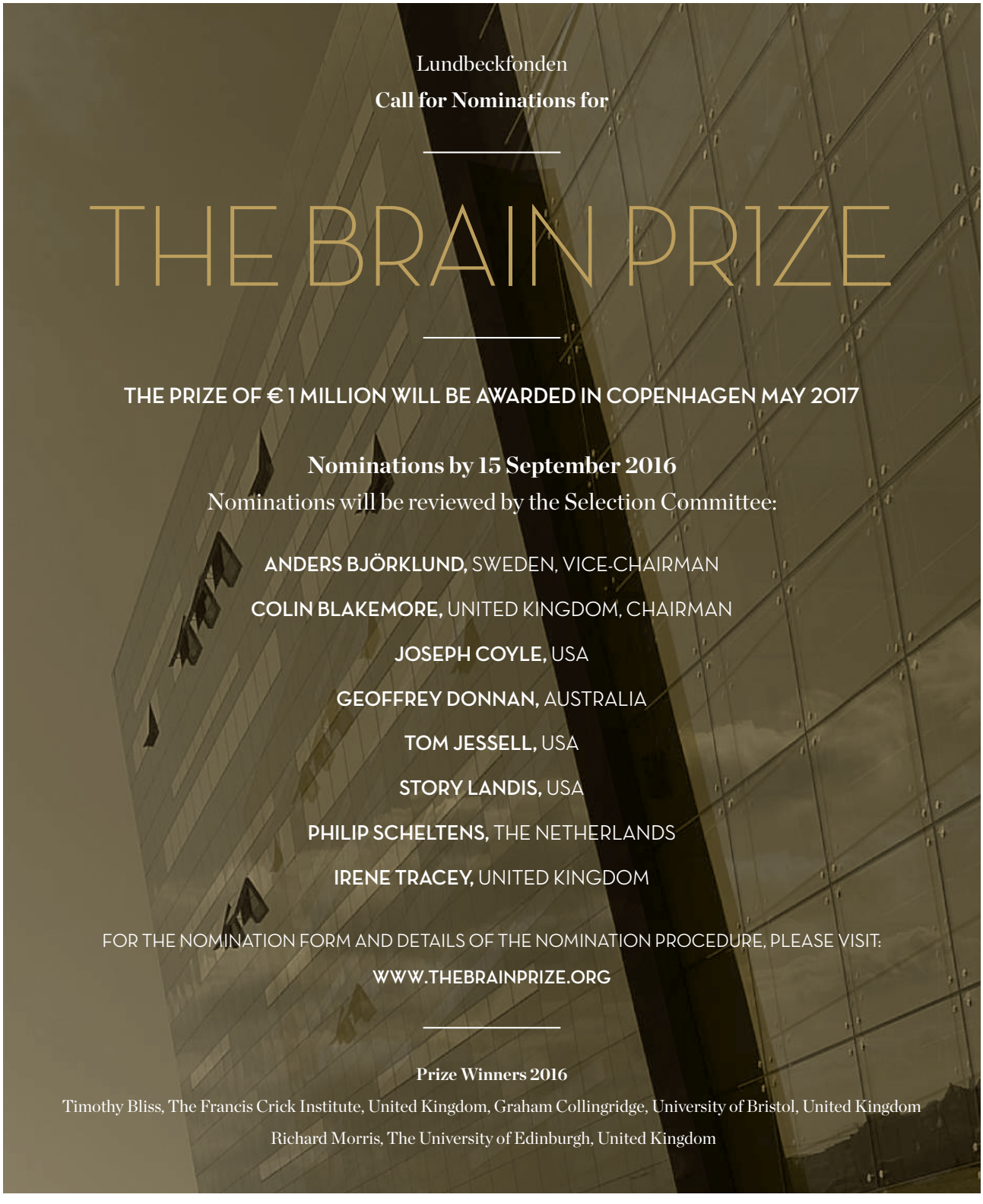
SCIENCE POLICY

Join Us in Boston

The AAAS Annual Meeting is interdisciplinary and inclusive. Thousands of leading scientists, engineers, educators, policymakers, and journalists from around the world will gather to discuss recent developments in science and technology.

Registration opens August 2016.

aaas.org/meetings



Lundbeckfonden

Call for Nominations for

THE BRAIN PRIZE

THE PRIZE OF € 1 MILLION WILL BE AWARDED IN COPENHAGEN MAY 2017

Nominations by 15 September 2016

Nominations will be reviewed by the Selection Committee:

ANDERS BJÖRKLUND, SWEDEN, VICE-CHAIRMAN

COLIN BLAKEMORE, UNITED KINGDOM, CHAIRMAN

JOSEPH COYLE, USA

GEOFFREY DONNAN, AUSTRALIA

TOM JESSELL, USA

STORY LANDIS, USA

PHILIP SCHELTENS, THE NETHERLANDS

IRENE TRACEY, UNITED KINGDOM

FOR THE NOMINATION FORM AND DETAILS OF THE NOMINATION PROCEDURE, PLEASE VISIT:

WWW.THEBRAINPRIZE.ORG

Prize Winners 2016

Timothy Bliss, The Francis Crick Institute, United Kingdom, Graham Collingridge, University of Bristol, United Kingdom

Richard Morris, The University of Edinburgh, United Kingdom

THE BRAIN PRIZE

The Brain Prize recognizes and rewards outstanding contributions to European neuroscience, from basic to clinical

AAAS *Travels*

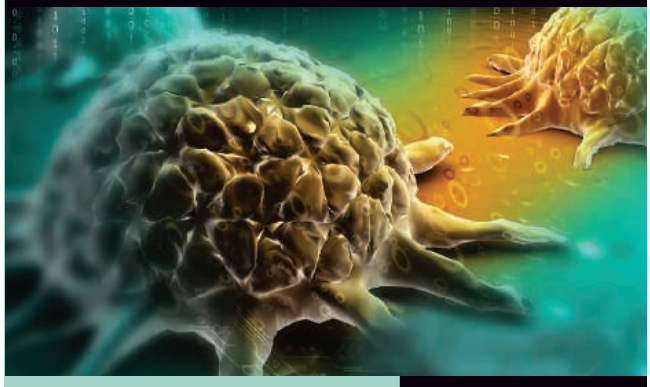
NEW ZEALAND
November 18–December 3, 2016

Join one of the top naturalists Lloyd Esler to see some of the most spectacular scenery in the world in New Zealand's national parks. From kiwis to kakas, temperate rainforests to glacial valleys, New Zealand possesses a rich heritage. The geologic forces are as dynamic as anywhere on earth, and can be seen in the fascinating geyserfields in Rotorua, in the plunging cliffs at Milford Sound, and the spectacular peaks and glaciers of the Southern Alps. $4,995 pp + air.

For a detailed brochure, call (800) 252-4910
All prices are per person twin share + air

BETCHART EXPEDITIONS Inc.
17050 Montebello Rd, Cupertino, CA 95014
Email: AAASInfo@betchartexpeditions.com
www.betchartexpeditions.com

Join our live webinar!



Key signaling pathways in cancer

Links to developmental biology

To register, go to:
bit.ly/ScienceWebinarJune2

Thursday, June 2, 2016
1 p.m. Eastern,
10 a.m. Pacific,
6 p.m. UK,
7 p.m. CEST

Science
AAAS

Sponsored by Cell Signaling Technology

Be Among the First to Publish in *Science Robotics*

NOW ACCEPTING MANUSCRIPTS
ScienceRobotics.org

Science Robotics is a unique journal created to help advance the research and development of robotics for all environments. *Science Robotics* will provide a much-needed central forum to share the latest technological discoveries and to discuss the field's critical issues.

Join in the excitement for the Fall 2016 debut!

Science Robotics
AAAS

Inflammation Panels

To understand the role that inflammation plays in diseases such as cancer, heart disease, and diabetes, researchers must understand the complex interplay of many inflammatory proteins. With the Bio-Plex Pro Human Inflammation Panels, complex immune responses can be profiled in a single assay for the first time. The kits detect and quantify key inflammation biomarkers from the tumor necrosis factor (TNF), type 1 and 2 interferon (IFN), matrix metalloproteinase (MMP), and interleukin-10 (IL-10)/regulatory T cell (Treg) cytokine families. By studying up to 37 inflammatory proteins in one assay, researchers save time, money, and increase sample throughput. Results can be generated in as little as three to four hours. Premixed, ready-to-use screening kits are available in 37-plex and 24-plex formats. A Treg-pathway-focused 12-plex assay kit is also available. Researchers have the added flexibility of ordering singleplex or custom kits for more focused investigations.

Bio-Rad
For info: 800-424-6723
www.bio-rad.com/inflamassay

In-Vacuum X-Ray Cameras

The PI-MTE in-vacuum X-ray camera line now features specially designed higher-sensitivity charge-coupled devices (CCDs). As the smallest commercially available scientific cameras engineered for in-vacuum use, the PI-MTE series provides all of the advanced features and functions needed to tackle demanding soft X-ray spectroscopy and imaging applications. Every PI-MTE delivers highly flexible multiple-speed readout capabilities, and allows researchers to optimize Princeton Instruments' renowned ultralow-noise and wide-dynamic-range performance for their specific soft X-ray applications. These thermoelectrically cooled cameras, which can support CCDs offering spatial resolution of up to 4 megapixels, are utilized at labs and synchrotrons worldwide. Compact PI-MTE cameras measure 4.57 in. (116.0 mm) × 2.36 in. (60.0 mm) × 3.15 in. (80.0 mm) and weigh just 2.5 lbs. (1.13 kg). The complete system includes an ST-133 controller, Princeton Instruments' optional LightField or WinX software, and in-vacuum electrical and cooling interface technologies that permit uninterrupted camera operation inside a vacuum chamber without compromising performance.

Princeton Instruments
For info: 978-268-0327
www.princetoninstruments.com

Collagen Gel

PureCol EZ Gel is a ready-to-use Type I collagen solution that forms a firm 3D gel simply upon warming to 37°C. The product consists of purified Type I collagen at a concentration of 0.5% (~5 mg/ml), Dulbecco's modified Eagle's medium (DMEM)/F-12 medium, and a mixture of L-glutamine and dipeptide (L-alanine-L-glutamine) to provide a long-lasting L-glutamine source for cell culture. PureCol EZ Gel is ideal for providing a firm 3D gel, and can also be used in the preparation of a thin layer for culturing cells. PureCol EZ Gel is provided in a user-friendly, ready-to-use sterile packaging. PureCol EZ Gel is the easiest way to form a 3D collagen gel.

Advanced BioMatrix
For info: 800-883-8220
www.advancedbiomatrix.com

3D Cell Invasion Assay

The Oris 3D Embedded Invasion Assay is a unique product offering quantifiable real-time analysis of cell invasion. This new cell invasion assay is designed to enable cell movement through an extracellular matrix (ECM) in three dimensions, thereby getting you as close as possible to an in vivo experiment using in vitro conditions. Based upon patented "exclusion zone" technology, which uses cells embedded in a collagen matrix, Oris 3D Embedded Invasion Assays provide a simple, quantitative, and efficient way to monitor cell invasion and morphology changes in a more physiologically relevant microenvironment. With no artificial membranes obscuring cell visualization, cell invasion rates and morphology changes can be reproducibly and quantifiably measured as end-point assays or in real-time. Designed in an industry-standard 96-well plate format, the assay can be used with microscopes, plate readers, and high-content imaging systems, thereby delivering significant reductions in handling time for setup and analysis.

AMS Biotechnology
For info: +44-(0)-1235-828200
www.amsbio.com

RSPO1 Protein-Producing Cell Lines

HA-R-Spondin-Fc293T cells can be used to produce purified Roof plate specific Spondin-1 (R-Spondin 1 or RSPO1) and RSPO1-conditioned media. RSPO1 is the critical ingredient used in the maintenance and proliferation of mouse and human organoid progenitor stem cells. These cells can be passaged indefinitely and can produce unlimited amounts of RSPO1 protein, allowing laboratories conducting research with organoids to save time and money. The RSPO1 cells are authenticated by Trevigen and tested to be free from mouse or human pathogenic agents. More than 90% of drugs entering clinical trials fail. Typically compounds that look promising after in vitro screening fail after in vivo analysis. By introducing the first commercially available HA-R-Spondin-Fc293T cell line to the market, Trevigen makes a positive step toward the adoption of 3D cultures in drug screening and other cell-based research applications.

Trevigen
For info: 301-216-2800
www.trevigen.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

Don't miss the debut of *Science Immunology*.

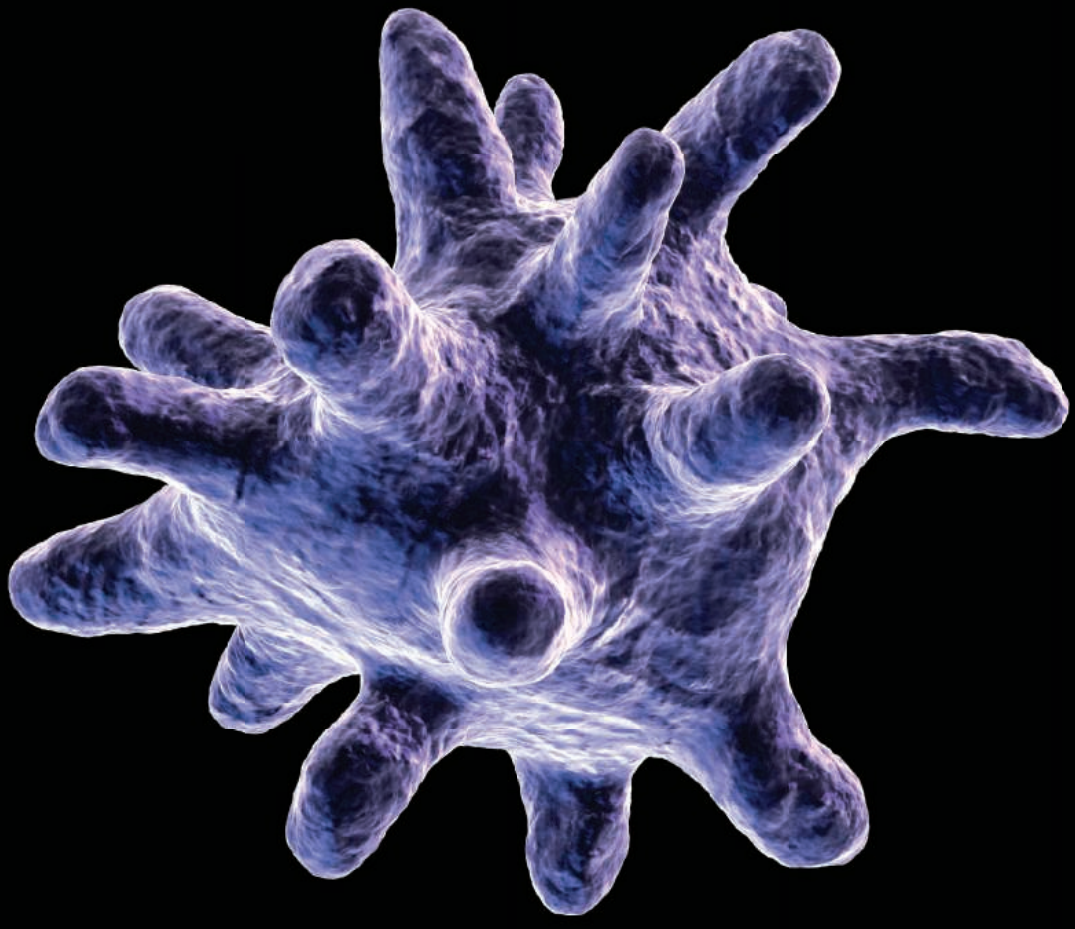


Image: Eraxion / iStockPhoto

NOW ACCEPTING PAPERS

Science is expanding its reach into immunology—now offering the newest online-only, weekly journal in the *Science* family of publications. *Science Immunology* will provide original, peer-reviewed research articles that report critical advances in all areas of immunological research, including studies that provide insight into the human immune response in health and disease.

Be a part of the *Science Immunology* debut issue publishing Summer 2016!

Submit your manuscript today at **ScienceImmunology.org.**

Science Immunology

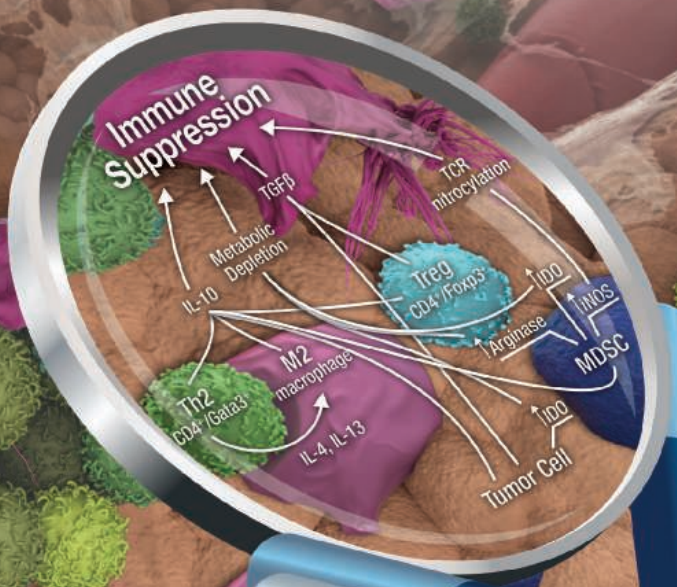
AAAS

CELL SIGNALING TECHNOLOGY

Deciphering Cancer

Tools to Interpret the Tumor-Immune Cell Crosstalk

PD-1
IDO
PD-L1
LAG3
PD-L2
ICOS
TIM-3
CD3
Arginase-1
CD8
VISTA



Visit our website to request our **Tumor Immunology Poster**

www.cellsignal.com/immunetumor

For Research Use Only. Not For Use In Diagnostic Procedures.
© 2016 Cell Signaling Technology, Inc. Cell Signaling Technology and CST are trademarks of Cell Signaling Technology, Inc.

16PADIMMU0017ENG

Cell Signaling TECHNOLOGY®

Science Careers Advertising

For full advertising details, go to ScienceCareers.org and click For Employers, or call one of our representatives.

Tracy Holmes
Worldwide Associate Director
Science Careers
Phone: +44 (0) 1223 326525

THE AMERICAS

E-mail: advertise@sciencecareers.org
Fax: +1 (202) 289 6742

Tina Burks
Phone: +1 (202) 326 6577

Nancy Toema
Phone: +1 (202) 326 6578

Online Job Posting Questions
Phone: +1 (202) 312 6375

EUROPE / INDIA / AUSTRALIA / NEW ZEALAND / REST OF WORLD

E-mail: ads@science-int.co.uk
Fax: +44 (0) 1223 326532

Sarah Lelarge
Phone: +44 (0) 1223 326527

Kelly Grace
Phone: +44 (0) 1223 326528

Online Job Posting Questions
Phone: +44 (0) 1223 326528

JAPAN

Katsuyoshi Fukamizu (Tokyo)
E-mail: kfukamizu@aaas.org
Phone: +81 3 3219 5777

Hiroyuki Mashiki (Kyoto)
E-mail: hmashiki@aaas.org
Phone: +81 75 823 1109

CHINA / KOREA / SINGAPORE / TAIWAN / THAILAND

Ruolei Wu
E-mail: rwu@aaas.org
Phone: +86 186 0082 9345

All ads submitted for publication must comply with applicable U.S. and non-U.S. laws. *Science* reserves the right to refuse any advertisement at its sole discretion for any reason, including without limitation for offensive language or inappropriate content, and all advertising is subject to publisher approval. *Science* encourages our readers to alert us to any ads that they feel may be discriminatory or offensive.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

ScienceCcareers.org

POSITIONS OPEN

ENDOWED CHAIR IN BIOMEDICAL RESEARCH AS AN ASSOCIATE OR FULL PROFESSOR OF VIROLOGY IN SAN DIEGO

The Viral Information Institute is a new interdisciplinary research center at San Diego State University (SDSU) built upon the pioneering work in virology of a diverse group of collaborative faculty. The institute aims to understand the impact of viruses on human health and the environment through cutting-edge research.

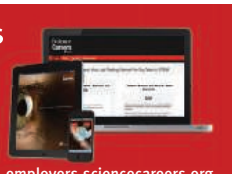
The Department of Biology at SDSU is recruiting an individual as either an associate professor or a full professor promoted to full within the last 5 years to be the Conrad Prebys Endowed Chair in Biomedical Research as part of the Viral Information Institute.

The successful candidate will have a demonstrated record of research accomplishments and funding and will employ state-of-the-art approaches to study the role of viruses in the human microbiome. The candidate will also have a strong record in grant writing and extramural support, as well as a demonstrated capacity for collaborating, mentoring, and teaching.

Applications should be submitted via Interfolio at website: **http://apply.interfolio.com/34893**. Review of applications will begin 01 August 2016, and will continue until the position is filled. Incomplete applications are not guaranteed full consideration. For more information see website: **http://www.bio.sdsu.edu/jobs/**.

SDSU is a Title IX, Equal Opportunity Employer.

Post your jobs
Fast and Easy
Science Careers
employers.sciencecareers.org



RUTGERS
UNIVERSITY

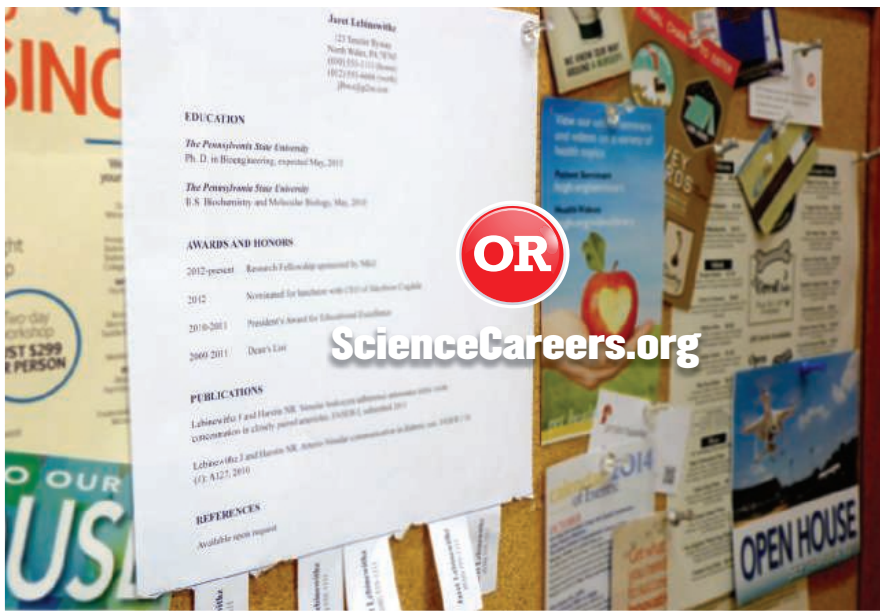
Assistant Professor/Associate Professor

The Neuroscience and Cell Biology Department, Rutgers-Robert Wood Johnson Medical School seeks a full-time, non-tenure track faculty member who can teach medical and graduate students gross anatomy, embryology and histology, with the potential for future promotion to Anatomy Course Director.

Qualifications include PhD, MD, or equivalent degree, at least 3 years experience teaching human gross anatomy and evidence of scholarly activities.

Application materials should include current curriculum vitae, cover letter, teaching philosophy, and three professional references. Applicants should respond to **Zamina Deen**, **abdelhza@rutgers.edu**

Rutgers, The State University of New Jersey, is an Equal Opportunity/Affirmative Action Employer. Qualified applicants will be considered for employment without regard to race, creed, color, religion, sex, sexual orientation, gender identity or expression, national origin, disability status, genetic information, protected veteran status, military service or any other category protected by law. As an institution, we value diversity of background and opinion, and prohibit discrimination or harassment on the basis of any legally protected class in the areas of hiring, recruitment, promotion, transfer, demotion, training, compensation, pay fringe benefits, layoff, termination or any other terms and conditions of employment.



Conduct your job search the easy way.

Target your job search using relevant resources on **ScienceCareers.org**.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

ScienceCareers online @sciencecareers.org

Special Job Focus:

Microbiology

Issue date: June 10

Book ad by May 24 to guarantee space

Ads accepted until June 3 if space allows

For recruitment in science, there's only one *Science*.

What makes *Science* the best choice?

- Read and respected by 400,00 readers around the globe
- 75% of readers read *Science* more often than any other journal
- Your ad dollars support AAAS and its programs, which strengthens the global scientific community.

Why choose this Microbiology Focus for your advertisement?

- Relevant ads lead off the career section with special microbiology banner
- Bonus distribution to American Society for Microbiology, June 16–20, Boston, MA.

Expand your exposure. Post your print ad online to benefit from:

- Link on the job board homepage directly to microbiology jobs
- Dedicated landing page for jobs in microbiology
- Additional marketing driving relevant job seekers to the job board.

Deliver your message to a global audience of targeted, qualified scientists.

129,574
subscribers in print every week

49,637
unique active job seekers searching for **microbiology** positions in 2015

37,252
applications submitted for **microbiology** positions in 2015

SCIENCECAREERS.ORG

Produced by the *Science*/AAAS Custom Publishing Office.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

To book your ad: **advertise@sciencecareers.org**

The Americas
+202 326 6582

Japan
+81 3 3219 5777

Europe/RoW
+44(0)1223 326500

China/Korea/Singapore/Taiwan
+86 186 0082 9345

The 19th Video Recruitment of High-level Overseas Talents for China's Top Universities

The editorial department of CHISA, China Education and Research Network and China Education Online will jointly sponsor this recruitment activity with more than twenty Chinese universities and top disciplines. Those who are intend to participate Chang Jiang Scholars Program and other high-level talents recruitment are warmly invited.

Holding Date : 7:00-24:00 29th May, 2016 (GMT+8)

Recruitment Requirement: Overseas Scholars and Doctors

Participating Universities :

上海工程技术大学 Shanghai University of Engineering Science

上海师范大学 SHANGHAI NORMAL UNIVERSITY

郑州大学 ZhengZhou University

广东外语外贸大学 GUANGDONG UNIVERSITY OF FOREIGN STUDIES

江西财经大学 JIANGXI UNIVERSITY OF FINANCE AND ECONOMICS

西南大学 SOUTHWEST UNIVERSITY

上海电力学院 SHANGHAI UNIVERSITY OF ELECTRIC POWER

广东海洋大学 GUANGDONG OCEAN UNIVERSITY

上海立信会计学院 SHANGHAI LIXIN UNIVERSITY OF COMMERCE

上海海事大学 SHANGHAI MARITIME UNIVERSITY

东南大学 SOUTHEAST UNIVERSITY

郑州轻工业学院 Zhengzhou University of Light Industry

湖南商学院 HUNAN UNIVERSITY OF COMMERCE

陕西师范大学 SHANXI NORMAL UNIVERSITY

西安交通大学 XI'AN JIAOTONG UNIVERSITY

西南科技大学 Southwest University of Science and Technology

东北林业大学 NORTHEAST FORESTRY UNIVERSITY

哈尔滨工程大学 HARBIN ENGINEERING UNIVERSITY

上海电机学院 SHANGHAI DIANJI UNIVERSITY

西南交通大学 Southwest Jiaotong University

第三军医大学

北京理工大学前沿交叉科学研究院 Beijing Institute of Technology Advanced Research Institute for Multidisciplinary Science

Position Requirements :

Mechanical Engineering, Information and Communication Engineering, Power System and Automation, Environmental Science and Engineering, New Energy Materials, Traffic Engineering, Electronic Science, Biology, Cytobiology, Clinical Medicine, Thermal Energy and Power Engineering, Civil Engineering, Instrument Science and Technology, Food Science and Engineering, Mathematics, Geography, Economics, Psychology, The Science of Law, Foreign Linguistics and Applied linguistics, Pedagogy, Art, Management Science, Marxist Philosophy.

Participation Approaches:

1. Please send your resume to wubh@eol.cn
2. Visit the website http://www.edu.cn/cv. Each university has his own chat room and you can communicate with recruiter face to face. [7:00-24:00 29th May, 2016 (GMT+8)]

For more information, Please contact : 86-10-62603770 , 86-15300215485

Science Careers online @sciencecareers.org
OPPORTUNITIES IN CHINA

10 ways that *Science* Careers can help advance your career

1. Register for a free online account on **ScienceCareers.org**.
2. Search thousands of job postings and find your perfect job.
3. Sign up to receive e-mail alerts about job postings that match your criteria.
4. Upload your resume into our database and connect with employers.
5. Watch one of our many webinars on different career topics such as job searching, networking, and more.
6. Download our career booklets, including Career Basics, Careers Beyond the Bench, and Developing Your Skills.
7. Complete an interactive, personalized career plan at "my IDP."
8. Visit our Career Forum and get advice from career experts and your peers.
9. Research graduate program information and find a program right for you.
10. Read relevant career advice articles from our library of thousands.

Visit **ScienceCareers.org** today — all resources are free

ScienceCareers

FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

WE ARE SEEKING OUTSTANDING RESEARCHERS IN MOLECULAR MEDICINE

The **Wallenberg Centres for Molecular Medicine** are key elements in a national effort to reposition Sweden as a world-leading life science nation. The initiative was taken by the Knut and Alice Wallenberg Foundation, and is a joint venture with the Universities and University Hospitals of Gothenburg, Lund, Umeå and Linköping. SciLifeLab in Stockholm and Uppsala serves as a research partner and unique core facility for the four Centres.

Through repeated calls in the upcoming years with **tenure track research positions – Wallenberg Molecular Medicine Fellows** – we will recruit translationally oriented **research groups**. The groups will be centered on internationally recruited young scientists of outstanding potential and funded at a globally competitive level through **very generous starting packages** with the possibility of promotion to Senior Lecturer within four years. Each of these groups will synergize with pre-existing excellent research environments as well as strong clinical collaborators, promoting ground-breaking research in molecular and translational medicine. Together, we will rise to future challenges within molecular life science in order to improve human health.

Note that the application deadline at the four Centres may vary.

www.wcmm.se

UMEÅ UNIVERSITY:

Up to four tenure track research positions are advertised in the areas of **cancer, infection biology, metabolism/diabetes and neuroscience**. The positions are at Associate Senior Lecturer (Assistant Professor) level. The successful candidates will work in association with strong research environments and have access to excellent research infrastructures including unique collections of longitudinal samples in existing biobanks. **www.wcmm.umu.se**

LINKÖPING UNIVERSITY:

With a **medicine-technology interface** and extending from a **biomedical engineering** profile, onwards to **basic biomedical** and **clinical research**, we plan to currently recruit up to four tenure track faculty members, at the level of Assistant/Associate Professor. The successful candidates will be expected to develop a dynamic, extramurally funded, internationally recognized research program, and to integrate with existing strengths in medical technology, materials science and bioengineering. **www.liu.se/wcmm**

LUND UNIVERSITY:

We are seeking one young scientist each within the area **Regenerative Medicine** (Regeneration, Replacement, and Repair) in: 1) the Respiratory system, 2) Diabetes, and 3) the Hematopoietic system. Successful candidates will work in a translational environment, including clinical scientists from Skåne University Hospital, and have access to cutting-edge molecular biology and imaging facilities at one of the largest Universities in Scandinavia. Tenure track positions at the level of Associate Senior Lecturer (Assistant Professor) are announced. **www.med.lu.se/wcmm**

UNIVERSITY OF GOTHENBURG:

Two tenure track research positions as Associate Senior Lecturer (Assistant Professor) are currently advertised in Molecular Medicine with a focus on **Respiratory Diseases** and **Life Science Chemistry** respectively. Successful candidates will become part of strong research environments with excellent infrastructure and enjoy close collaboration with AstraZeneca and clinical/translational research platforms at one of the largest University hospitals in northern Europe. **www.wcmtm.gu.se**

Umeå University · LINKÖPING UNIVERSITY · LUND UNIVERSITY · UNIVERSITY OF GOTHENBURG

WALLENBERG CENTRES FOR MOLECULAR MEDICINE SWEDEN | Knut och Alice Wallenbergs Stiftelse

Advance your career with expert advice from *Science* Careers.

Download Free Career Advice Booklets!
ScienceCareers.org/booklets

Featured Topics:

- Networking
- Industry or Academia
- Job Searching
- Non-Bench Careers
- And More

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

ScienceCareers online @sciencecareers.org

ScienceCareers online @sciencecareers.org

myIDP: A career plan customized for you, by you.

For your career in science, there's only one Science

my IDP
ScienceCareers

Recommended by leading professional societies and the NIH

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.
- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track.
- Articles and resources to guide you through the process.
- Options to save materials online and print them for further review and discussion.
- Ability to select which portion of your IDP you wish to share with advisors, mentors, or others.
- A certificate of completion for users that finish myIDP.

Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers In partnership with:

advancing the life sciences 1912-2012 FASEB Federation of American Societies for Experimental Biology

UCSF University of California San Francisco

Medical College of Wisconsin

Burroughs Wellcome Fund

Science & Diplomacy

Science & Diplomacy provides an open access forum for rigorous thought, analysis, and insight to serve stakeholders who develop, implement, or teach all aspects of science and diplomacy. Learn more about the latest ideas in science diplomacy and receive regular updates by following @SciDip on Twitter, liking the quarterly's page on Facebook (www.facebook.com/sciencediplomacy), and registering for free at **www.sciencediplomacy.org/user/register**.

WWW.SCIENCEDIPLOMACY.ORG

Science & Diplomacy is published by the Center for Science Diplomacy of the American Association for the Advancement of Science (AAAS), the world's largest general scientific society.

Science & Diplomacy

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

ETH zürich

Assistant Professor of RNA Biology and Disease

→ The Department of Biology (www.biol.ethz.ch) at ETH Zurich invites applications the above-mentioned assistant professorship.

→ It will be embedded within one of the following institutes depending on the research topic: the Institute of Biochemistry (www.bc.biol.ethz.ch), the Institute of Molecular Biology and Biophysics (www.mol.biol.ethz.ch) or the Institute of Molecular Health Sciences (www.imhs.biol.ethz.ch). In addition, the assistant professor will become a junior principle investigator in the NCCR "RNA and Disease" (www.nccr-rna-and-disease.ch). ETH Zurich and the NCCR "RNA and Disease" offer an excellent environment for collaborative projects as well as access to various technology platforms.

→ This assistant professorship will take responsibility for building a vibrant and internationally visible research programme in the area of RNA Biology and Disease, using chemical, structural, biochemical, cell biological, systems or genetics approaches to generate fundamental new knowledge of how RNA or RNA-Protein complexes contribute to homeostatic mechanisms at a molecular level. The successful candidate will be expected to acquire competitive third party funding and will contribute to teaching activities at the Department of Biology. An outstanding research track record and proven teaching skills are required. The new assistant professor will be expected to teach undergraduate level courses (German or English) and graduate level courses (English).

→ This assistant professorship has been established to promote the careers of younger scientists. The initial appointment is for four years with the possibility of renewal for an additional three-year period. Towards the end of this appointment, a full professor position at ETH Zurich will be created in the area "RNA and Disease", for which the assistant professor is eligible to apply.

→ **Please apply online at www.facultyaffairs.ethz.ch**

→ Applications should include a curriculum vitae, a list of publications, a statement of future research and teaching interests, and three of your most important achievements. The letter of application should be addressed **to the President of ETH Zurich, Prof. Dr. Lino Guzzella. The closing date for applications is 31 July 2016**.
ETH Zurich is an equal opportunity and family friendly employer and is further responsive to the needs of dual career couples. We specifically encourage women to apply.

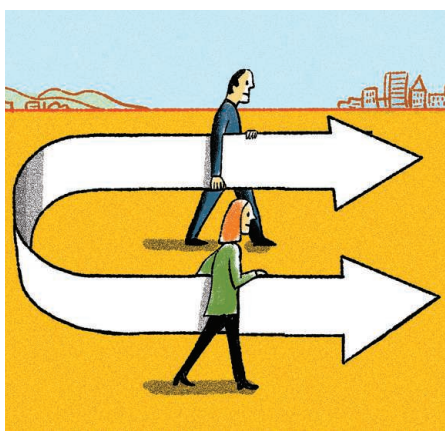
ScienceCareers online @sciencecareers.org

By **Paul Rogge** *and* **Emily Regier**

A problem by any other name

For many in the academic community, the term "two-body problem" brings to mind not the physics of two separate but interacting particles, but the calculus of two committed partners seeking careers in the same location. For us, it elicits regrets and visions of what might have been if we had questioned the term's implication that relationships are uniquely incompatible with academia and had, instead, followed our hearts and each other. Our fear of the two-body problem led us to abandon our academic dreams in pursuit of a stable life as a professional couple, but today we regret that decision. Realizing that stability may not be the touchstone of a satisfying career, we are now rethinking what is important to us and reclaiming our ambitions as a dual academic couple.

After Paul got his Ph.D. in materials science, he considered several postdoc positions. But most last just 2 or 3 years, and he knew that he might need to relocate for a second (or third) temporary position before landing a tenure-track professorship. Would Emily, a lawyer, have to take a new bar exam and rebuild her practice in state after state? She had considered returning to school to study the equality issues that had so excited her as a law student, but this too seemed incompatible with a long-term relationship and especially a family, which we both wanted.



"We are ... reclaiming our ambitions as a dual academic couple."

This is how we settled on a new life in Boise, a livable place where we thought we could put down roots. Paul took a job working with corporate inventors and writing patent applications. Emily took one more bar exam and planned to join a local law firm; if all went well, she could start her own in a few years. This is the compromise we made to avoid the two-body problem.

Although we have found a certain amount of satisfaction in our quieter life, we sorely miss dedicating ourselves to work that matters to us in a deeper sense. We have settled down, but our dreams haven't. Over the past year, we have also come to appreciate how hard it is for partners to stay in one place and maintain two ambitious careers, academic or not. As we have seen at our firms and their clients, corporate jobs come and go as activist investors and a fickle economy yield unpredictable results. And losing a specialized job in all but the most major cities may force a couple to confront the challenges of co-locating once again.

Our decision to forgo an academic future was at least in part due to the term "two-body problem," which both points to and perpetuates the erroneous but widely held attitude that academics in committed relationships face a uniquely intractable dilemma. Encountering a less negative term might have helped us accept the risks—and see the rewards—of pursuing our academic passions. Rather than reinforcing the idea that long-term relationships are incompatible with academia, a new term should recognize academic partners' admirable endeavor to maintain both their careers and their commitment to each other. We propose the "two-variable equation." The partners define the equation that best suits their relationship, career goals, and long-range plans. The aim is not to fix a nagging problem, but to optimize each individual's potential to flourish.

For our part, over the past year, we have reevaluated our goals and priorities and redefined our two-variable equation. We have realized that intellectual curiosity, a willingness to push boundaries, and a conviction that our work matters are critical to our identities, both individual and shared. The upshot of these epiphanies has been a push to reclaim our career goals: This fall, Emily will begin a Ph.D. program in political science in Philadelphia, Pennsylvania, where Paul is seeking a postdoc.

We know that the road ahead will not be easy and that we will need to continue optimizing our two-variable equation, but we are excited about taking this step together. Our only certainty at this point is that we will take a passionate path—one we are proud of and would be proud to tell our children about. ■

Paul Rogge has a Ph.D. from the University of California, Berkeley. Emily Regier has a J.D. from Harvard Law School. Send your story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

Downloaded from http://science.sciencemag.org/ on May 19, 2016

Published by AAAS