

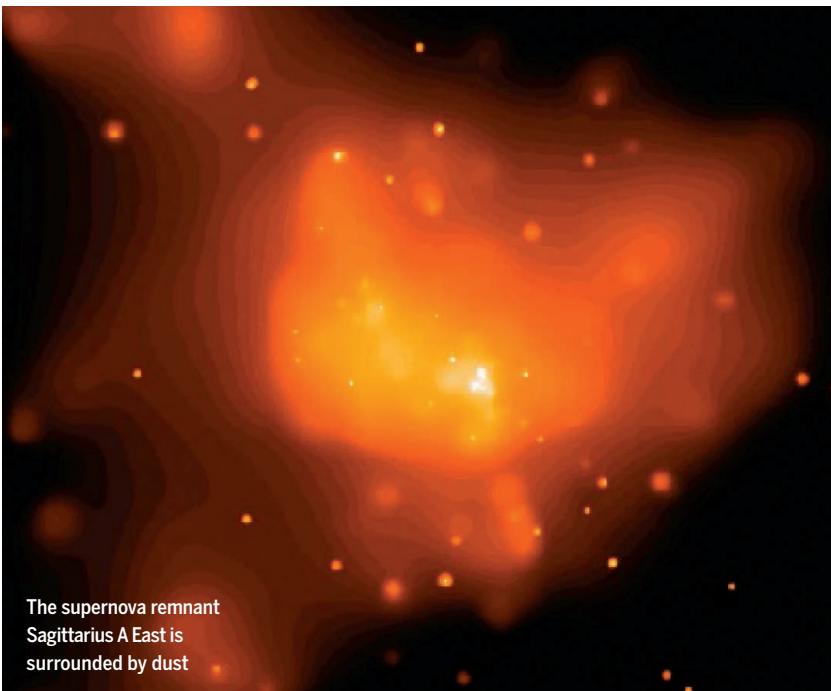
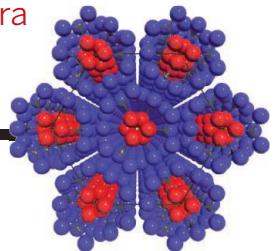
RESEARCH

IN SCIENCE JOURNALS

Edited by Caroline Ash

Supramolecular lattices from nanosized giant tetrahedra

Huang et al., p. 424



The supernova remnant Sagittarius A East is surrounded by dust

ISOTOPE GEOCHEMISTRY

What controls clumped isotopes?

Stable isotopes of a molecule can clump together in several combinations, depending on their mass. Even for simple molecules such as O₂, which can contain ¹⁶O, ¹⁷O, and ¹⁸O in various combinations, clumped isotopes can potentially reveal the temperatures at which molecules form. Away from equilibrium, however, the pattern of clumped isotopes may reflect a complex array of processes. Using high-resolution gas-phase mass spectrometry, Yeung et al. found that biological factors influence the clumped isotope signature of oxygen produced during photosynthesis (see the Perspective by Passey). Similarly, Wang et al. showed that away from equilibrium, kinetic

effects causing isotope clumping can lead to overestimation of the temperature at which microbially produced methane forms. — NW

Science, this issue p. 431; p. 428; see also p. 394

EBOLA VIRUS

Ebola virus vaccine candidate

As there are so few possibilities for drugs and vaccines that protect against the Ebola virus (EBOV), we need more options. Marzi et al. present initial studies in monkeys of a promising whole Ebola virus vaccine based on a defective form of the virus in which an essential viral gene (VP30) is knocked out. One or two doses of this defective virus, with or without further peroxide inactivation, protected against

a lethal challenge of EBOV. In limited immunological analyses, protection correlated with the production of antibodies to the EBOV envelope. — CA

Science, this issue p. 439

GENOME EDITING

Generating homozygous mutations

Loss-of-function mutations may only produce a mutant phenotype when both copies of the gene are mutated. Gantz and Bier developed a method they call mutagenic chain reaction (MCR) that autocatalytically produces homozygous mutations. MCR uses the initial mutated allele to cause a mutation in the allele on the opposing chromosome and thus the homozygosity of the trait. MCR

technology could have broad applications in diverse organisms. — BAP

Science, this issue p. 442

INFECTIOUS DISEASE

A genetic cause for severe influenza

Although chicken soup and plenty of rest get most kids through an influenza virus infection, some require hospitalization. Ciancanelli et al. report on one child who suffered severely from influenza because of null mutations in the gene for transcription factor IRF7. Cells isolated from this patient could not make enough secreted antiviral proteins, called interferons, to halt viral replication. The requirement for IRF7 seems quite specific, because this

patient recovers normally from other common childhood viral infections. — KLM
Science, this issue p. 448

BIOENGINEERING

Drug testing in a patient's own tumor cells

It is not easy to predict a patient's response to chemotherapy with animal models and cultured cells. The true test of responsiveness requires the evaluation of drug activity within human tumors. In two studies, Jonas *et al.* and Klinghoffer *et al.* have engineered devices that deliver microdoses of drugs directly into tumors. After treatment, the researchers removed the tumor tissue and quantified cancer-cell death. In both studies, the local response to chemotherapy matched the systemic response to known, as well as experimental, drugs. However, drug sensitivities were revealed that were not detected in cell culture. — MLF

Sci. Transl. Med. **7**, 284ra57 and 284ra58 (2015).

PROTEIN FOLDING

Ribosomes help careful protein folding

Protein assembly *in vitro* is useful for studying small molecules but is problematic for studying the assembly of larger, more complex proteins. Kim *et al.* analyzed the biogenesis of the mutation-prone nucleotide-binding domain of the cystic fibrosis conductance regulator (CFTR) (see the Perspective by Puglisi). Newly synthesized polypeptides emerged relatively slowly from the ribosome and folded through a modulated pathway that ensured correct protein folding. Some parts of the protein chain folded immediately upon synthesis, whereas other segments did so more slowly. It appears that acquiring the correct conformation for this complex

protein is partly guided by the ribosome itself. — SMH

Science, this issue p. 444; see also p. 399

CARDIAC PHYSIOLOGY

Keeping hearts at the right size

If left untreated, high blood pressure can lead to abnormally enlarged hearts (a condition called pathological hypertrophy) and heart failure. Inhibitors of protein kinase C (PKC) isoforms are in development for treating heart failure and some cancers. Withal *et al.* report that the inhibition of some PKC isoforms may exacerbate heart pathology. During development, mice lacking two related PKC isoforms, PKC δ and PKC ϵ , had abnormally large hearts and usually died in utero. Thus, drugs that inhibit PKC δ and PKC ϵ could trigger adverse cardiac side effects. — WW

Sci. Signal. **8**, ra39 (2015).

POLYMER CHEMISTRY

A more direct way to synthesize styrene

Foam cups, foam pellets, plastic cutlery: All are made of polystyrene, which in turn is made of styrene. The massive manufacturing scale of this commodity chemical places a premium on the efficiency of its synthesis. The current industrial route requires three steps to make styrene from benzene and ethylene. Vaughan *et al.* present a rhodium catalyst that achieves the coupling in a single step by using a recyclable copper salt as an oxidant. Although the catalyst is slow for industrial application, it demonstrates the viability of a more direct process. — JSY

Science, this issue p. 421



IN OTHER JOURNALS

Edited by Sacha Vignieri and Jesse Smith



The structured surface of a leaf beetle carapace diffracts light to produce spectacular colors

APPLIED OPTICS

A stretch to change color

The reflection of white light from structured surfaces often results in a spectacular display of color as the white light is split into its different wavelengths through diffraction. Structure gives rise to the intense iridescent colors that distinguish some members of the animal kingdom, such as beetles and butterflies. Human-made materials, such as DVD or CD surfaces, also diffract light into a rainbow. Zhu *et al.* combine surface structure with membrane flexibility to show that they can locally select the color of reflected light, as they stretch the membrane and change the periodicity of the structure. This technique could be used in a range of applications, including camouflage coatings, optical sensing and steering, and displays. — ISO

Optica **2**, 255 (2015).

CELLULAR BIOMECHANICS

The mechanics of cellular left and right

Cells need to know their own left and right in order to coordinate with neighboring cells in collective movement or embryonic

development. To do so, each cell has to establish left/right asymmetry. Tee *et al.* studied actin organization in human cells to understand underlying mechanisms, using fluorescence and electron microscopy and simulations. Actin fibers forming

ALSO IN SCIENCE JOURNALS

Edited by Caroline Ash

GALAXY EVOLUTION

Galaxies stripped down and evicted

It is easy to imagine that the relatively small and dense compact elliptical galaxies once had more to them. Especially when massive galactic neighbors are seen nearby that may have robbed them of their matter, but sometimes there are no such neighbors. Chilingarian and Zolotukhin have mined survey data to show that stripped-down galaxies are found in varied environments, with and without obvious disruptive companions. These isolated galaxies may still have been tidally stripped and then ejected from more crowded neighborhoods. — MMM

Science, this issue p. 418

RNA IMAGING

Multiplexed RNA imaging in single cells

The basis of cellular function is where and when proteins are expressed and in what quantities. Single-molecule fluorescence *in situ* hybridization (smFISH) experiments quantify the copy number and location of mRNA molecules; however, the numbers of RNA species that can be simultaneously measured by smFISH has been limited. Using combinatorial labeling with error-robust encoding schemes, Chen *et al.* simultaneously imaged 100 to 1000 RNA species in a single cell. Such large-scale detection allows regulatory interactions to be analyzed at the transcriptome scale. — VV

Science, this issue p. 412

SELF-ASSEMBLY

Creating unusual nanostructures

Self-assembly often occurs when dissimilar molecular fragments are forced together by covalent bonding. Surfactants

or block copolymers are two common examples. Huang *et al.* grafted four different nanoparticles, based on polyhedral oligomeric silsesquioxanes with slightly different compositions, onto a single tetrahedral core (see the Perspective by Yang). Depending on the type of nanoparticle, they assembled into a range of defined, ordered supramolecular lattices similar to a range of metal alloys. These include phases that have higher coordination numbers than usually found in the packing of spherical objects. — MSL

Science, this issue p. 424;
see also p. 396

RESEARCH FUNDING

Proof that peer review picks promising proposals

A key issue in the economics of science is finding effective mechanisms for innovation. A concern about research grants and other research and development subsidies is that the public sector may make poor decisions about which projects to fund. Despite its importance, especially for the advancement of basic and early-stage science, there is currently no large-scale empirical evidence on how successfully governments select research investments. Li and Agha analyze more than 130,000 grants funded by the U.S. National Institutes of Health during 1980–2008 and find clear benefits of peer evaluations, particularly for distinguishing high-impact potential among the most competitive applications. — BW

Science, this issue p. 434

NEUROBIOLOGY

SARM1-driven axon degeneration

Axons, the long protrusions of nerve cells, are programmed to self-destruct under certain conditions that occur during

development, stress, or disease states. Gerdts *et al.* outline a biochemical mechanism that controls such axon degeneration. The authors designed versions of SARM1 (sterile alpha and TIR motif—constraining 1) that could be activated or inhibited in cells. Their experiments showed that the activation of SARM1 was necessary and sufficient to cause axon destruction in cultured mouse neurons. SARM1-mediated destruction was associated with depletion of the metabolic cofactor NAD⁺ from cells. — LBR

Science, this issue p. 453

RIBOSOME

Force to unblock a clogged ribosome

The synthesis of proteins from mRNA by the ribosome is highly regulated. But newly synthesized protein chains can still block the ribosome exit tunnel and slow protein synthesis. Goldman *et al.* use optical tweezers to show that by pulling on the stuck protein chain, they can unblock a clogged exit tunnel (see the Perspective by Puglisi). *In vivo*, the folding of a nascent protein chain just outside the tunnel also generates enough force to unclog a block, indicating that ribosome-peptide interactions fine-tune protein synthesis. — GR

Science, this issue p. 457;
see also p. 399

VASCULAR DISEASE

Matrikine regulates endothelial leakage

The matrikine acetyl-proline-glycine-proline (N- α -PGP) regulates endothelial permeability in inflammatory disease. Xu *et al.* probed the loss of barrier function of the endothelium, which is characteristic of inflammation. Leakiness is stimulated by a signaling cascade initiated by the modified peptide N- α -PGP

and involving CXCR2. N- α -PGP is an extracellular matrix fragment generated by tissue damage. N- α -PGP induces vascular leaks, and its removal attenuates the lipopolysaccharide-induced leak. N- α -PGP is a novel matrikine and could be a new therapeutic target. — PLY

Sci. Adv. 10.1126/
sciadv.1500175 (2015).

EVOLUTIONARY BIOLOGY

Symbionts on the road to a single species

Most species of plants and animals live in symbiotic relationships with other organisms. For example, many insects rely on bacteria within their cells—endosymbionts—to supply nutrients, and pea and bean plants have essential relationships with bacteria called rhizobia to fix atmospheric nitrogen. Why do some symbioses progress to ever-tighter partnerships and ultimately form a single new organism? In a Perspective, Kiers and West explore the conditions that favor such major evolutionary transitions in individuality. One key factor is the mode of symbiont transmission; another is the ecological context of the symbiosis. Once a partnership has reached a level of high dependence and low conflict, genetic drift alone can lead to further integration. However, major transitions of this kind require strict conditions to be fulfilled and are thus rare. — JFU

Science, this issue p. 392

MULTILAYER ASSEMBLY

Thin-film fabrication

The deposition of thin films from multiple materials is essential to a range of materials fabrication processes. Layer-by-layer processes involve the sequential deposition of two or more materials that physically bond together. Richardson *et al.*

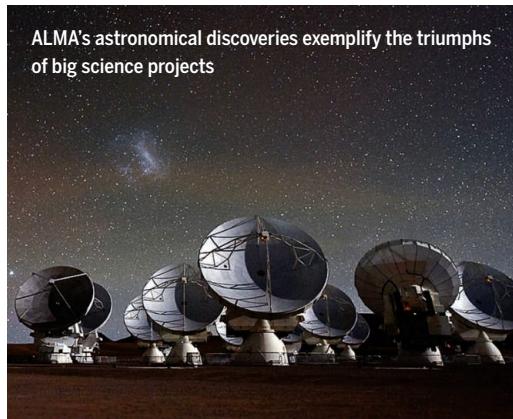
review some of the techniques and materials that are used to make thin films, including sequential dip coating, spraying, and electrochemical deposition. Despite the versatility of the methods and the range of materials that can be deposited, the techniques remain mostly confined to the lab because of challenges in industrial scaling. But because there is tremendous scope for fine-tuning the structure and properties of the multilayers, there is interest in broadening the use of these techniques. — MSL

Science, this issue p. 411

Big science is hard but worth it

Big science is hard. It is the throw-deep approach that pushes technology to the edge to achieve stunning breakthroughs that dramatically extend the frontiers of science, while inspiring the next generation of scientists and capturing the imagination of the public. The Large Hadron Collider (LHC) produced the Higgs boson, a particle that explains why all other particles have mass and whose name is now known around the world; the Atacama Large Millimeter and Submillimeter Array (ALMA) revealed some of the earliest galaxies, as well as a nearby planetary system that has a striking resemblance to our own; and for 25 years, the Hubble Space Telescope has dazzled us with its discoveries and iconic astronomical images. But before they became famous for discoveries, these and other big projects were infamous for their problems. What are the challenges for big science, and what does it take to succeed?

Four hundred years ago, Galileo urged “measure what is measurable” and “make measurable that which is not.” Until recently, the latter almost always involved a single scientist or small group inventing new instrumentation (and often still does). But as science has matured, advances now often require big teams and expensive facilities. The technological challenges are daunting, from operating a 100-ton, 27-km superfluid helium system at 1.9 kelvin at the LHC to preparing for the unfolding of the James Webb Space Telescope’s (JWST’s) 18-segment, 6.5-m mirror, 1.5 million km from Earth in 2018. Sociological, budgetary, organizational, and cultural issues can be even more difficult. Big-science time scales are longer, making it harder for graduate students to complete thesis work and for postdocs and assistant professors to achieve results that advance their careers in a timely way. Projects are so large and expensive that proper management is complex, and even small budget overruns have broad programmatic impact. Big science often involves partnering, and whether it is multiple funding agencies, several countries, or a public/private partnership, getting partners with varying cultures to act coherently is difficult.



ALMA’s astronomical discoveries exemplify the triumphs of big science projects

“It is the throw-deep approach that pushes technology to the edge...”

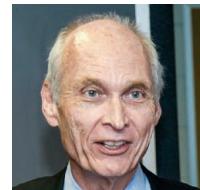
And yet big science can overcome these hurdles and be triumphant. From my experience at the U.S. National Science Foundation with big projects, including ALMA and the LHC, I see four key ingredients for success at this scale. All of the stakeholders, from the science community to the funders, must view the science as truly worthy of a big-science approach and must be committed to seeing things through during good times and bad. With one-of-a-kind, envelope-pushing projects, problems will

undoubtedly arise, but with the appropriate management structure, independent oversight, and project contingency funds, they can be identified quickly and solved. Partners must be committed to the project and to their well-defined, agreed-upon responsibilities and willing to delegate authority to a managing partner or to central project management. Last, but not least, all involved must be honest and realistic about both progress and problems and willing to make hard decisions, whether it be de-scoping, raising additional funds, or even cancellation.

Four years ago, the U.S. Congress came close to cancelling Hubble’s successor, the JWST. Although the details of how big projects suffer near-death experiences vary, JWST’s problems were not atypical: The budget had gotten out of control, the management structure was insufficient, and there was an all-around failure to acknowledge difficulties. Thankfully, the science community—not just astronomers—recognized the importance of game-changing projects such as JWST and rallied to its support. NASA made major management changes, and Congress committed to the needed funding, despite tough federal budgets. Today, JWST still faces substantial technical challenges as testing and systems integration begin, but it is on schedule and on budget with appropriate resources to address contingencies.

When we see the stunning images from JWST and hear about the breakthroughs, we will all be proud of what was accomplished and know that it was worth the hard work and big investment.

—Michael S. Turner



Michael S. Turner is the Rauner Distinguished Service Professor and director of the Kavli Institute for Cosmological Physics at the University of Chicago, Chicago, IL. E-mail: mturner@kicp.uchicago.edu

patient recovers normally from other common childhood viral infections. — KLM
Science, this issue p. 448

BIOENGINEERING

Drug testing in a patient's own tumor cells

It is not easy to predict a patient's response to chemotherapy with animal models and cultured cells. The true test of responsiveness requires the evaluation of drug activity within human tumors. In two studies, Jonas *et al.* and Klinghoffer *et al.* have engineered devices that deliver microdoses of drugs directly into tumors. After treatment, the researchers removed the tumor tissue and quantified cancer-cell death. In both studies, the local response to chemotherapy matched the systemic response to known, as well as experimental, drugs. However, drug sensitivities were revealed that were not detected in cell culture. — MLF

Sci. Transl. Med. **7**, 284ra57 and 284ra58 (2015).

PROTEIN FOLDING

Ribosomes help careful protein folding

Protein assembly *in vitro* is useful for studying small molecules but is problematic for studying the assembly of larger, more complex proteins. Kim *et al.* analyzed the biogenesis of the mutation-prone nucleotide-binding domain of the cystic fibrosis conductance regulator (CFTR) (see the Perspective by Puglisi). Newly synthesized polypeptides emerged relatively slowly from the ribosome and folded through a modulated pathway that ensured correct protein folding. Some parts of the protein chain folded immediately upon synthesis, whereas other segments did so more slowly. It appears that acquiring the correct conformation for this complex

protein is partly guided by the ribosome itself. — SMH

Science, this issue p. 444; see also p. 399

CARDIAC PHYSIOLOGY

Keeping hearts at the right size

If left untreated, high blood pressure can lead to abnormally enlarged hearts (a condition called pathological hypertrophy) and heart failure. Inhibitors of protein kinase C (PKC) isoforms are in development for treating heart failure and some cancers. Withal *et al.* report that the inhibition of some PKC isoforms may exacerbate heart pathology. During development, mice lacking two related PKC isoforms, PKC δ and PKC ϵ , had abnormally large hearts and usually died in utero. Thus, drugs that inhibit PKC δ and PKC ϵ could trigger adverse cardiac side effects. — WW

Sci. Signal. **8**, ra39 (2015).

POLYMER CHEMISTRY

A more direct way to synthesize styrene

Foam cups, foam pellets, plastic cutlery: All are made of polystyrene, which in turn is made of styrene. The massive manufacturing scale of this commodity chemical places a premium on the efficiency of its synthesis. The current industrial route requires three steps to make styrene from benzene and ethylene. Vaughan *et al.* present a rhodium catalyst that achieves the coupling in a single step by using a recyclable copper salt as an oxidant. Although the catalyst is slow for industrial application, it demonstrates the viability of a more direct process. — JSY

Science, this issue p. 421



Polystyrene pellets

IN OTHER JOURNALS

Edited by Sacha Vignieri and Jesse Smith



The structured surface of a leaf beetle carapace diffracts light to produce spectacular colors

APPLIED OPTICS

A stretch to change color

The reflection of white light from structured surfaces often results in a spectacular display of color as the white light is split into its different wavelengths through diffraction. Structure gives rise to the intense iridescent colors that distinguish some members of the animal kingdom, such as beetles and butterflies. Human-made materials, such as DVD or CD surfaces, also diffract light into a rainbow. Zhu *et al.* combine surface structure with membrane flexibility to show that they can locally select the color of reflected light, as they stretch the membrane and change the periodicity of the structure. This technique could be used in a range of applications, including camouflage coatings, optical sensing and steering, and displays. — ISO

Optica **2**, 255 (2015).

CELLULAR BIOMECHANICS

The mechanics of cellular left and right

Cells need to know their own left and right in order to coordinate with neighboring cells in collective movement or embryonic

development. To do so, each cell has to establish left/right asymmetry. Tee *et al.* studied actin organization in human cells to understand underlying mechanisms, using fluorescence and electron microscopy and simulations. Actin fibers forming



Perception of angry faces is shaped by stereotypes

SOCIAL PSYCHOLOGY

Stereotyping sticks

A person better remembers faces of people who are members of one's own group—as defined, for instance, by sex or race—than of those who belong to an outgroup. An angry expression might reduce this difference, because threatening stimuli capture one's attention. Alternatively, it might increase the difference were it to trigger stereotyping. In a careful study using white and black faces and undergraduates, Gwinn *et al.* show that both white and black students better remembered individual neutral black faces than angry black faces, whereas their memories for angry and neutral white faces were similar, consistent with a stereotypical association of black faces and threat in the United States. — GJC

J. Exp. Soc. Psychol. **58**, 1 (2015).

the cellular skeleton rearranged from a symmetric to an asymmetric pattern through interplay between two types of fibers. The fibers stretching along the cell edge swirled towards the center, whereas the radially assembled fibers tilted unidirectionally. The contractile stress and rotational growth of the fibers drove the motions, while an actin cross-linking protein controlled the clockwise or anticlockwise directionality. — MSM

Nat. Cell Biol. 10.1038/ncb3137 (2015).

MATERIALS SCIENCE

Ordering up just the right temperature

One typically thinks of glassy or amorphous materials as having little of the order or structure that is found in a crystal. However, the molecules of organic glasses deposited from the vapor phase, like those that form the active

layer in a light-emitting diode or solar cell, can show a preferential molecular orientation. Dalal *et al.* use high-throughput screening to show that the quality of this orientation can be tuned through control over the substrate temperature during deposition, which influences the molecular mobility at the surface. This ability may give a simple route to optimizing active layers in organic electronics, because the molecular orientation can affect light emission, charge mobility, and device lifetimes. — MSL

Proc. Nat. Acad. Sci. 10.1073/pnas.1421042112 (2015).

CELL DIFFERENTIATION

Differentiating blood stem cells

Hematopoietic stem cells (HSCs), which generate all blood cell types, contain hundreds of long noncoding RNAs (lncRNAs). Luo *et al.* identify 159 lncRNAs that are likely to be HSC-specific (lncHSCs) through deep RNA sequencing of primitive long-term HSCs from mouse bone marrow. As HSC function decreases with age, lncHSCs show reduced expression. Furthermore, like HSC-specific protein genes, promoters of lncRNAs are bound specifically by HSC transcription factors for regulated expression. Analysis of two specific lncRNAs reveals a role for lncHSC-1 in myeloid differentiation and lncHSC-2 involvement in HSC self-renewal and T cell differentiation. Analysis of the many other HSC-specific lncRNAs is likely to reveal varied mechanisms in regulating the self-renewal and differentiation of blood types. — BAP

Cell Stem Cell 10.1016/j.stem.2015.02.002 (2015).

NEURODEVELOPMENT

Catching it early

Communication difficulties associated with autism spectrum disorders (ASD) vary, making treatment more difficult. To anticipate language capability among those with ASD, Lombardo *et al.* used functional magnetic resonance imaging to analyze responses of toddlers to language during sleep. In individuals who later displayed poor language development, the superior temporal cortex responded differently than in individuals with better language development. Linkages between large-scale brain systems,

such as those driving emotion, reward, and motor function, were atypical in toddlers with poor language outcomes. For those with better language outcomes, brain systems were at first abnormal but later normalized. Given the variation in severity of ASD, this early identification could inform ongoing treatments. — PJH

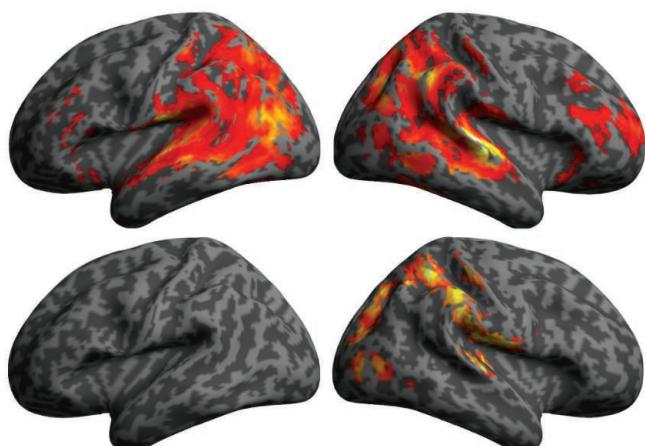
Neuron 10.1016/j.neuron.2015.03.023 (2015).

SUBSURFACE MICROBES

Generating archeal diversity

Retroelements—gene sequences that are introduced into DNA through an RNA intermediate—in subsurface archaea help increase gene and protein diversity. Paul *et al.* identified an archeal virus from a methane seep in a California borderlands basin that contains a complete and active diversity-generating retroelement (DGR). Furthermore, published metagenomes of two uncultivated nanoarchaea from the marine subsurface contain at least four DGRs, more than any other bacteria or virus known to date. Because anaerobic archaea in the subsurface grow so slowly, and thus have low background genome mutations, DGRs may be a widespread source of genetic diversity in these environments. — NW

Nat. Comm. 10.1038/ncomms7585 (2015).



Differences in language maps of young autistic brains predict future abilities

NMR Analysis Software

InsightMR software is designed for online NMR analysis of chemical processes. For the first time, a software platform enables the adjustment of experimental parameters based on real-time data during the analysis of chemical processes by NMR spectroscopy. InsightMR's user-friendly interface allows both expert and non-expert users to set up, monitor, and adjust key experimental parameters. As a result the software is the ideal solution for both industrial and academic scientists studying and optimizing organic reactions. Online NMR is a groundbreaking approach to reaction understanding and chemical process monitoring. InsightMR makes NMR accessible to users in a range of application areas, including synthetic organic chemistry, process chemistry, and process analytical technology. InsightMR provides a complete workflow from data acquisition to project report with the ability to export data in multiple formats for additional processing. The software seamlessly integrates with Bruker spectrometers providing a turnkey solution to investigate chemical reaction mechanisms and optimize reaction processes by NMR.

Bruker Corporation

For info: +49-6181-4384-100
www.bruker.com

NGS Analysis Software

The newest version of NextGENe provides a no-scripting, point-and-click analysis of next generation sequencing from all Illumina and Ion sequencing platforms. The latest version of the software includes a wide-ranging number of both application additions as well as program enhancements. New applications to this version of the software are: a new Somatic Mutation Comparison Tool, which

has been shown to reduce false positives when performing somatic mutation analysis by up to 70%; a HLA analysis module, which matches sample reads to a dictionary reference file with allele-specific sequences for the targeted HLA genes; and an enhanced copy-number variation analysis module that now accepts multiple controls. In addition to the above applications, the software now includes a User Management Module, which password protects unauthorized access to the software and provides parallel processing of multiple analyses, delivering up to an eightfold increase in processing speed.

Soft Genetics

For info: 888-791-1270
www.softgenetics.com



Lab Control Software

AVA Software is a major step-change in lab control software to enable scientists to work smarter and complete projects faster. AVA is very easy to use and can control and log data from a variety of manufacturer's laboratory equipment. It offers four software levels to suit the application and budget, from controlling a single stirring hotplate up to a multi-device jacketed reaction system. Scientists can use AVA to control a wide range of laboratory devices, including stirring hotplates, overhead stirrers, circulators, balances, pumps, pH meters, vacuum pumps, gas flow controllers, and temperature sensors; to eliminate time-consuming, manual data recording; to "track and repeat" experiments easily and accurately; to define safety limits and alarms for safe unattended operation; to visualize results for rapid analysis and reporting; to link devices for dual control, e.g., pump and balance for mass additions; and to control up to four experiments in parallel.

Radleys

For info: +44-(0)-1799-513320
www.radleys.com

Liquid Handling Software

With CyBio Overture, switching from manual to automated pipetting is easier than ever before. This is made possible via the protocol designer, which has an intuitive user interface based on a detailed representation of the liquid handling robot and all of the required sample carriers and accessories. Along with the liquid handling and workflows, the software also controls additional critical functions. This includes laboratory and equipment settings, the calculation of reaction mixtures, dead volumes, and well contents. CyBio Overture is able to calculate reaction mixtures and spare the user work that is both time-consuming and prone to error. The results are then automatically applied to the workflow. Using color coding, both the wells and the corresponding pipetting instructions can be recognized quickly. This makes the workflows easier to follow. CyBio Overture is able to analyze well contents rapidly as well.

Analytik Jena

For info: +49-3641-7770
www.cybio-ag.com

LIM System

LabWare's unique Enterprise Laboratory Platform brings together the capabilities of Laboratory Information Management Systems (LIMS) and Electronic Laboratory Notebooks (ELNs) in a single comprehensive solution. LabWare 7 is the most adaptable and functionally complete LIMS system available and offers a proven solution for any size laboratory. The Enterprise Laboratory Platform is scalable to suit every size organization and LabWare's customers range from single site, single user to global organizations with over 100 sites and thousands

of users. LabWare 7 and LabWare ELN may also be implemented independently, integrated into your organization's existing software ecosystem. Some of the significant features of LabWare 7 include: functional enhancements with more than 500 new capabilities now added to the system's already robust functional footprint, a new interactive dynamic user interface, and a new LabWare Design Studio which enables the creation and management of Dynamic Visual Workflows and Dashboards without the need for programming skills.

LabWare

For info: 302-658-8444
www.labware.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/products/newproducts.dtl 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.

39.6%

Fraction of foreign students in U.S. science and engineering graduate programs in 2013. That all-time record is due both to increases in temporary visas and declining domestic enrollment, according to the National Science Foundation.

IN BRIEF

Sikuliaq ready for duty



The RV *Sikuliaq* in the Bering Sea during ice and science trials.

The rumbling of the RV *Sikuliaq*'s engine was music to ocean scientists' ears this month, during a 23-day cruise to test how the newest addition to the U.S. oceanographic fleet handled icy seas. Starting from the island of Amaknak in Alaska, the ship crunched north into so-called ten-tenths sea ice—the name shiphands give to a sea ice coating that stretches to the horizon. The 80-meter-long, \$200 million *Sikuliaq* is not an icebreaker, but its hardened hull is rated to move through solid sea ice as thick as 0.8 meters—a rating that will allow scientists on board to access icy areas during the fall and spring, such as the southern parts of the Bering Sea. The *Sikuliaq* easily passed its various trials: Winches delivered sampling equipment in ice and water, and crews exited the ship on foot to take samples from surrounding ice. "She's a great ship, and a great addition to the U.S. science fleet," says chief scientist Carin Ashjian, a biological oceanographer from the Woods Hole Oceanographic Institution in Massachusetts. *Sikuliaq* begins full research operations in ice later this year. <http://scim.ag/Sikuliaq>

AROUND THE WORLD

Gene editing patent fight

OAKLAND, CALIFORNIA | The dispute over who owns the rights to a groundbreaking and lucrative gene-editing technique has reached the U.S. Patent and Trademark Office. The University of California (UC) last week asked the office to reconsider 10 patents awarded to synthetic biologist Feng Zhang of the Broad Institute in Cambridge, Massachusetts, for a method of precisely cutting or replacing letters in a DNA sequence that could lead to new treatments for genetic diseases. UC Berkeley biochemist Jennifer Doudna and collaborators first described the technique, often called CRISPR-Cas9, in a 2012 *Science* paper, but Zhang won the patents last year by submitting notebooks that suggest his discovery predated Doudna's. In what could be a long and costly battle, both sides will present evidence for the dates of their findings to a board of patent examiners.

Rubber plants imperil biodiversity

XISHUANGBANNA, CHINA | A growing demand for rubber products—particularly tires—has fueled the conversion of more than 2 million hectares of forests and farms worldwide into rubber plantations in the past decade. That could exacerbate the extinction crisis in southern China and other parts of Southeast Asia, new research in *Conservation Letters* suggests. The researchers found that conversion of forest to rubber monoculture decreases the number of bird, bat, and insect species. Additionally, pesticide, herbicide, and sediment runoff affects aquatic life, and the loss of smaller trees and shrubs leads



Harvesting latex from the caoutchouc, or rubber tree.

to soil erosion and landslide risk. Some 84% of the world's existing 9.9 million hectares of rubber trees are in Southeast Asia, and global rubber consumption will likely grow 3.5% annually, the team found. But there may be hope: Sustainability certification schemes have reduced the negative impacts of oil palm and paper and pulp growing. A similar effort for rubber, the Sustainable Natural Rubber Initiative, launched its pilot phase in January. <http://scim.ag/rubberplants>

Anger over 'consensus center'

PERTH, AUSTRALIA | The Australian government has again angered the country's scientists by announcing it will contribute \$4 million toward a new center to be built at the University of Western Australia's Business School in collaboration with the Copenhagen Consensus Center, a Massachusetts-based think tank created and directed by controversial environmental author Bjørn Lomborg. Lomborg has crossed swords with scientists for his views on mitigating climate change. "In the face of deep [funding] cuts to ... scientific research organisations, it's an insult to Australia's scientific community," said Australia's Climate Council, a nonprofit science and outreach center established by scientists after the Australian government shuttered its climate change commission in 2013. The Australian Consensus Centre will focus on the economic implications of agriculture, aid, and global development, according to a university official, but not climate change.

NEWSMAKERS

Three Q's

In March of 1970, forestry graduate student **Doug Scott** of the University of Michigan helped create a massive, 5-day "Teach-in for the Environment." The teach-in was a precursor to the first Earth Day on 22 April, an environmental activism event founded by Wisconsin Senator Gaylord Nelson that drew 20 million people across the United States.

Q: Ecologist Barry Commoner, once called the "Paul Revere of ecology," was there; what role did he play?

A: Our kickoff event ... overflowed a 14,000-person basketball arena. The event started with the cast of *Hair* and had Nelson and the governor of Michigan. Commoner was the central speaker. [But] he was a college professor, not used to speaking in front of the hot lights where you couldn't see the audience.



Festival aims to make math fun

High-fives and wild laughter were plentiful at the United States' first National Math Festival. The festival, held on 18 April in Washington, D.C., aims to help kids and adults see the beauty and wonder of math, says David Eisenbud, director of the Mathematical Sciences Research Institute (MSRI) in Berkeley, California. "People don't really understand what mathematics is about, and if they understood it they would all like it as much as I do." MSRI co-organized the festival with the Institute for Advanced Study and in collaboration with the Smithsonian Institution. Children made Möbius strips, gaped at mathematical card tricks, and clutched balloon octahedra in lieu of balloon animals. In a race called the "Oobleck Olympics," teams competed to pour water out of jugs—sped up by swirling the bottle to create a vortex, as shown—while cornstarch and water on a speaker danced to the beat of music.

Q: What happened during the teach-in?

A: We basically took over the campus for 5 days. Hundreds of people turned out to help organize. Professors devoted class time to the topic of the environment.

Q: Why did science students lead the charge?

A: I suppose we were paying more attention to the impact that man was having on the environment. But the whole university got involved. The law school hosted a 2-day symposium about what was then the cutting-edge topic of environmental law. You had to be pretty dedicated to your studies to not know what was going on and join in.

Science funding head exits

The head of Portugal's science funding agency, biomedical researcher **Miguel Seabra**, stepped down this month amid

mounting criticism of his agency's policies. The Foundation for Science and Technology (FCT) angered researchers in January 2014 when it announced a sharp drop in state-funded Ph.D. and postdoctoral fellowships. Seabra also oversaw a controversial evaluation of the country's R&D units, announcing in June 2014 that 22% of the 322 evaluated units would lose their funding due to poor ratings, and another 26% would see their budgets reduced to "core funding." Critics slammed the evaluation process as neither robust nor transparent. Crystallographer Maria Arménia Carrondo, a former adviser to FCT's board, will take over from Seabra, the Ministry of Education and Science announced last week, but some fear her close connection to Seabra means that no drastic change is likely to happen.
<http://scim.ag/Seabra>



NEUROSCIENCE

Seeking tests for a contested brain disease

FDA warning to UCLA highlights struggle to diagnose chronic traumatic encephalopathy

By Emily Underwood

Tony Dorsett was always quick on his feet, nimbly evading many crushing collisions as a star running back for the Dallas Cowboys. Still, the former National Football League (NFL) player sustained numerous concussions during his 12-year career. Now, Dorsett believes he is paying for those hits.

Last year, several news outlets reported that after having his brain scanned at the University of California, Los Angeles (UCLA), Dorsett was told he shows signs of chronic traumatic encephalopathy (CTE), a neurodegenerative disease associated with repeated blows to the head. Along with several other former pro football players scanned at UCLA, the 61-year-old now attributes his failing memory and mood swings to CTE.

Dorsett's bombshell shocked the sports world—and alarmed many CTE researchers. Despite a widespread belief that the brain disorder is common among athletes in high-impact sports, there is no proven method of diagnosing CTE while a patient is alive, says Samuel Gandy, a neurologist at Mount Sinai Hospital in New York City. The medical literature lists only about 100 proven cases, all based on postmortem analysis of brain tissue. The scarcity of evidence, Gandy says, has fueled “controversy over what CTE is, and if it exists.” Meanwhile, high-stakes lawsuits, including ones filed by former players

against the NFL, have added to the pressure on researchers such as Gandy, who are struggling to come up with methods for diagnosing and tracking the disorder.

Last week, when CTE researchers met at a traumatic brain injury conference in Washington, D.C., to take stock of their fledgling field, they discussed some tantalizing leads. But hallway chatter centered on a new controversy. As first reported by the *Los Angeles Times* on 10 April, the U.S. Food and Drug Administration (FDA) in February ordered

“There are no diagnostic criteria for CTE in vivo.”

Douglas Smith, University of Pennsylvania

the doctor who delivered the bad news to Dorsett, UCLA psychiatrist Gary Small, to remove promotional language from the website of a company, TauMark, which has licensed his and colleagues' research into imaging a protein called tau. The company's site once displayed the slogan “Better Brain Diagnostics” and claimed that its PET scans could detect signs of CTE in living people. But those at the D.C. meeting agreed with FDA's conclusion that the company's message was misleading. “There are no diagnostic criteria for CTE in vivo,” declared neurologist Douglas Smith of the University

of Pennsylvania. “We need to clear the air.”

Only in the past month or so have researchers arrived at a consensus about what CTE looks like in postmortem brain tissue, Robert Stern, a neurologist at Boston University, told meeting attendees. Last month, eight neuropathologists convened to examine digital images of brain slices taken from people who had had a variety of neurodegenerative conditions, including Alzheimer's disease and suspected CTE. Blind to each sample's clinical diagnosis, the group identified two patterns that set CTE apart from other pathologies: clusters of a molecule called tau sequestered inside neurons that surround blood vessels in the brain; and clumps of tau in neurons and other brain cells at the bottoms of sulci, the folds that make up a human's wrinkly cortex. The group's findings, which will be presented at the American Academy of Neurology meeting in Washington, D.C., this week, are the first strong demonstration that CTE “is indeed a pathological disease that is unique,” Stern says.

Small and his UCLA colleagues say they have found similarly located deposits of tau in living patients using their noninvasive technique. In a series of papers, including one online on 6 April in the *Proceedings of the National Academy of Sciences (PNAS)*, they reported injecting former NFL athletes with a patented radioactive compound called FDDNP. The compound, originally designed to study β amyloid plaques in Alzheimer's

Former running back Tony Dorsett (33) was reportedly told that his football career had likely given him a neurodegenerative disease.

disease, binds to abnormal protein deposits in the brain and is visible under a PET scan. In the former players, it revealed a smattering of tau and amyloid protein deposits distributed in a pattern that resembles CTE pathology and can be reliably distinguished from the plaques and tangles found in Alzheimer's disease, the team reported.

Stern is not convinced, saying that the pattern of FDDNP-labeled areas that the UCLA team saw in the PET scans "is not consistent with the neuropathological findings" that came out of the consensus meeting in March. Because FDDNP binds to so many different substances in the brain, it is not an ideal compound for imaging tau's presence, adds Patrick Bellgowan, a program director at the National Institute of Neurological Disorders and Stroke (NINDS) in Bethesda, Maryland. The UCLA group stands by its findings, however, saying that FDDNP does not need to bind exclusively to tau in order to provide a reliable picture of CTE's progression and distinguish it from other neurodegenerative diseases.

The *PNAS* report also raised eyebrows because its authors include Robert Fitzsimmons, a personal injury lawyer who in 1999 represented the late Hall of Fame center Mike Webster in a disability lawsuit against the NFL. Although Fitzsimmons is not involved in the class action suits, which involve more than 4000 former players, Webster was the first NFL player to be diagnosed with CTE after his death, and many credit his case with launching the current suit. In a statement, the UCLA team said that Fitzsimmons, who is a director at TauMark, "had significantly contributed to the design of this study because of his broad experience with concussions and brain damage." (He and two other co-authors on the *PNAS* paper founded the Brain Injury Research Institute in 1996.)

Despite the field's bumpy progress, Stern is convinced that diagnostic tools for CTE in living people are within reach. At the

meeting, he presented preliminary PET data from former NFL players showing that T807, a compound that he says binds more specifically to tau than FDDNP does, detected deposits of the protein in the brain's cortical folds in a pattern similar to that seen in postmortem tissue. "I am confident that within the next five to ten years there will be highly accurate, clinically accepted, and FDA-approved methods to diagnose CTE during life," Stern wrote in an October 2014 affidavit to the lawsuit filed against the NFL by retired players.

The stakes are high for those players. Under the current settlement with the retired athletes, the NFL will compensate

only those diagnosed with Alzheimer's disease or a "neurocognitive disorder," Stern says. A player who could receive \$1 million if he has an Alzheimer's diagnosis, for example, might receive less than half that, or nothing at all, because CTE can't yet be definitively diagnosed. The settlement is now in final negotiations, but as written it can be revised to incorporate new diagnostic criteria for CTE only every 65 years, Stern says.

The rush to find new diagnostic tools makes it easy to forget that research into CTE "is just at the starting line," Smith cautions. Tau deposits may only be a shadow or aftereffect of head injury, and not a cause of symptoms, for example. Fundamentally, Smith says, "we need to explore what shifts you from a normal aging track to a neurodegenerative track."

To tackle that question, scientists must follow large groups of people with concussions and other head injuries, ideally until they die and their brain tissue can be examined, Bellgowan says. Research groups funded by the NFL and NINDS are already looking for markers of CTE in blood samples and brain tissue from thousands of people enrolled in an ongoing study funded by the National Institute on Aging, he says. And Bellgowan adds that NINDS is reviewing a fresh round of grant proposals aimed at detecting CTE and defining its progression. The clock is ticking on Stern's prediction. ■

EUROPE

Plan for E.U. research funds raises ire

European Parliament vows to shield science budget from stimulus package raid

By Tania Rabesandratana

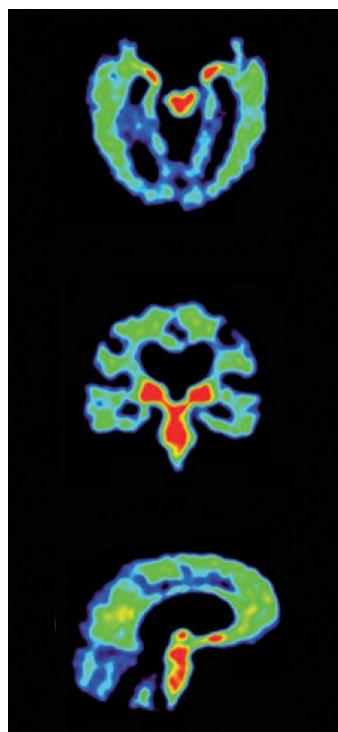
Abattle has erupted in Brussels over the European Commission's plan to raid research funds in a bid to boost Europe's lagging economy. Announced in November, the plan involves diverting €2.7 billion from the European Union's 2014 to 2020 research budget to create a new E.U. investment fund. Member states like the idea, but scientists protested—and now the European Parliament appears to have heard them. Earlier this week, it voted to oppose raiding the research budget, setting the stage for lengthy negotiations with the Council of Ministers, which represents member states.

European Commission chief Jean-Claude Juncker and Carlos Moedas, the European Union's research commissioner, have insisted that researchers have no cause for alarm. First, the commission says the money diverted for the investment fund, known as the European Fund for Strategic Investments (EFSI), represents "only 3.5%" of the overall budget of Horizon 2020, the European Union's 7-year research funding plan. Second, the commission claims that the money will not be lost to science: "On the contrary, this is money that will be used to attract much more important sums [from national governments and private investors] that will then be reinvested in innovation," the commission said in a statement.

Scientists and research organizations don't buy this argument. Universities won't be able to use the money that is diverted, says the European University Association: Instead of supporting research grants, the funds would become seed capital for loans that many public organizations cannot use because they are not allowed to borrow money. Scientists are particularly incensed that the European Research Council, which distributes individual grants for fundamental research, would lose €221 million.

Too much tau?

Brain scans of former NFL players highlight a disputed tag (red) for protein deposits.



Former running back Tony Dorsett (33) was reportedly told that his football career had likely given him a neurodegenerative disease.

disease, binds to abnormal protein deposits in the brain and is visible under a PET scan. In the former players, it revealed a smattering of tau and amyloid protein deposits distributed in a pattern that resembles CTE pathology and can be reliably distinguished from the plaques and tangles found in Alzheimer's disease, the team reported.

Stern is not convinced, saying that the pattern of FDDNP-labeled areas that the UCLA team saw in the PET scans "is not consistent with the neuropathological findings" that came out of the consensus meeting in March. Because FDDNP binds to so many different substances in the brain, it is not an ideal compound for imaging tau's presence, adds Patrick Bellgowan, a program director at the National Institute of Neurological Disorders and Stroke (NINDS) in Bethesda, Maryland. The UCLA group stands by its findings, however, saying that FDDNP does not need to bind exclusively to tau in order to provide a reliable picture of CTE's progression and distinguish it from other neurodegenerative diseases.

The *PNAS* report also raised eyebrows because its authors include Robert Fitzsimmons, a personal injury lawyer who in 1999 represented the late Hall of Fame center Mike Webster in a disability lawsuit against the NFL. Although Fitzsimmons is not involved in the class action suits, which involve more than 4000 former players, Webster was the first NFL player to be diagnosed with CTE after his death, and many credit his case with launching the current suit. In a statement, the UCLA team said that Fitzsimmons, who is a director at TauMark, "had significantly contributed to the design of this study because of his broad experience with concussions and brain damage." (He and two other co-authors on the *PNAS* paper founded the Brain Injury Research Institute in 1996.)

Despite the field's bumpy progress, Stern is convinced that diagnostic tools for CTE in living people are within reach. At the

meeting, he presented preliminary PET data from former NFL players showing that T807, a compound that he says binds more specifically to tau than FDDNP does, detected deposits of the protein in the brain's cortical folds in a pattern similar to that seen in postmortem tissue. "I am confident that within the next five to ten years there will be highly accurate, clinically accepted, and FDA-approved methods to diagnose CTE during life," Stern wrote in an October 2014 affidavit to the lawsuit filed against the NFL by retired players.

The stakes are high for those players. Under the current settlement with the retired athletes, the NFL will compensate

only those diagnosed with Alzheimer's disease or a "neurocognitive disorder," Stern says. A player who could receive \$1 million if he has an Alzheimer's diagnosis, for example, might receive less than half that, or nothing at all, because CTE can't yet be definitively diagnosed. The settlement is now in final negotiations, but as written it can be revised to incorporate new diagnostic criteria for CTE only every 65 years, Stern says.

The rush to find new diagnostic tools makes it easy to forget that research into CTE "is just at the starting line," Smith cautions. Tau deposits may only be a shadow or aftereffect of head injury, and not a cause of symptoms, for example. Fundamentally, Smith says, "we need to explore what shifts you from a normal aging track to a neurodegenerative track."

To tackle that question, scientists must follow large groups of people with concussions and other head injuries, ideally until they die and their brain tissue can be examined, Bellgowan says. Research groups funded by the NFL and NINDS are already looking for markers of CTE in blood samples and brain tissue from thousands of people enrolled in an ongoing study funded by the National Institute on Aging, he says. And Bellgowan adds that NINDS is reviewing a fresh round of grant proposals aimed at detecting CTE and defining its progression. The clock is ticking on Stern's prediction. ■

EUROPE

Plan for E.U. research funds raises ire

European Parliament vows to shield science budget from stimulus package raid

By Tania Rabesandratana

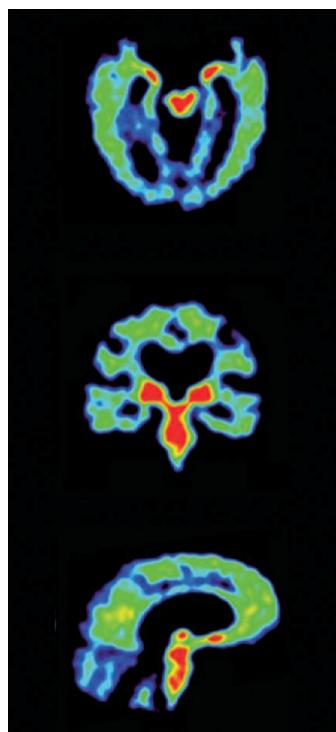
Abattle has erupted in Brussels over the European Commission's plan to raid research funds in a bid to boost Europe's lagging economy. Announced in November, the plan involves diverting €2.7 billion from the European Union's 2014 to 2020 research budget to create a new E.U. investment fund. Member states like the idea, but scientists protested—and now the European Parliament appears to have heard them. Earlier this week, it voted to oppose raiding the research budget, setting the stage for lengthy negotiations with the Council of Ministers, which represents member states.

European Commission chief Jean-Claude Juncker and Carlos Moedas, the European Union's research commissioner, have insisted that researchers have no cause for alarm. First, the commission says the money diverted for the investment fund, known as the European Fund for Strategic Investments (EFSI), represents "only 3.5%" of the overall budget of Horizon 2020, the European Union's 7-year research funding plan. Second, the commission claims that the money will not be lost to science: "On the contrary, this is money that will be used to attract much more important sums [from national governments and private investors] that will then be reinvested in innovation," the commission said in a statement.

Scientists and research organizations don't buy this argument. Universities won't be able to use the money that is diverted, says the European University Association: Instead of supporting research grants, the funds would become seed capital for loans that many public organizations cannot use because they are not allowed to borrow money. Scientists are particularly incensed that the European Research Council, which distributes individual grants for fundamental research, would lose €221 million.

Too much tau?

Brain scans of former NFL players highlight a disputed tag (red) for protein deposits.



In February, a group of Nobel laureates urged Juncker to reverse what they termed a “misguided and short sighted policy.” Cutting Horizon 2020 funds “send[s] a message that Europe is not the place to do high level science,” they wrote.

Juncker’s plan requires the approval of both the Parliament and the Council of Ministers. However, in a vote on 20 April, members of the European Parliament (MEPs) agreed to set up EFSI but opposed devoting Horizon 2020 funds to it. The Parliament “appears as the only E.U. institution defending Horizon 2020,” said Kurt Deketelaere, secretary-general of the League of European Research Universities, in a statement.

“It is completely unnecessary to transfer money from Horizon 2020,” says Kathleen van Brempt, a Belgian socio-democrat MEP who sits on the Parliament’s industry, research, and energy committee. Van Brempt instead wants EFSI to use unspent E.U. funds that are normally returned to national governments at the end of each year. “There are large enough leftovers to pay for the [EFSI] guarantee fund,” which the Parliament could watch over year after year, she says. However, national finance ministers like their annual E.U. refund, and so will certainly oppose van Brempt’s plan during the coming weeks of “trilogue” negotiations between the commission, Parliament, and member states. Juncker said he wants to finalize the plan before the summer.

Although the Parliament has often flexed its muscles this way since its powers were extended in 2009, the Council of Ministers usually has the upper hand in negotiations over the E.U. budget. “We can expect weeks of high political games,” Deketelaere says. “The question is who is going to give in.” The Parliament is showing a united front to protect research funds from cuts, but “I don’t know if that will hold until the end,” he adds. ■



European Commission chief Jean-Claude Juncker says scientists shouldn’t worry.

RESEARCH FUNDING

House science chief unveils contentious vision for science

Authorization bill would cut climate and energy research and reshape science agency policies

By Jeffrey Mervis and Adrian Cho

Representative Lamar Smith (R-TX) has never hidden his desire to reshape federal research policy—often over the objections of much of the scientific community—since he became chair of the House of Representatives science committee 2 years ago. Last week, he introduced legislation that lays out those plans in unprecedented detail, and the reaction was predictable. Although academic leaders say that some parts of the new, 189-page bill are better than previous versions, they believe it would seriously damage the U.S. research enterprise.

The bill not only sets out funding levels for several research agencies that in some cases depart sharply from those the Obama administration requested for 2016; it would also reshape key policies and priorities guiding those agencies. In particular, researchers complain that the bill (H.R. 1806), called the America COMPETES Reauthorization Act of 2015, would:

- Narrow the scope of research at the National Science Foundation (NSF) by designating some scientific disciplines as more important to the nation than others;
- Sharply reduce NSF’s authority to fund the social sciences and the geosciences;
- Restrict NSF’s ability to build large new scientific facilities by requiring the agency to follow new, controversial, accounting practices;
- Curtail climate change research at the Department of Energy (DOE);
- Block the government from using DOE research findings in writing regulations;
- And squeeze the budgets for DOE’s applied research program and its fledgling Advanced Research Projects Agency-Energy (ARPA-E).



Research on hurricane prediction could be affected by proposed geoscience cuts.

Smith doesn’t have the authority to impose that vision on Congress. The Senate has yet to draft a matching authorization bill, and the path to congressional passage is uncertain. But his committee’s oversight of several key scientific agencies means his ideas will play an important role in the debate.

The first America COMPETES Act, passed in 2007 and renewed in 2010, enjoyed bipartisan support from presidents George W. Bush and Barack Obama. In contrast, the new version has no Democratic co-sponsors, was not shown to Democratic committee members before it was unveiled, and has triggered a bitter partisan debate.

“The original American COMPETES Act was one of the crowning achievements of the science committee,” says Representative Eddie Bernice Johnson (D-TX), the top-ranked Democrat on the panel. “This bill is an America COMPETES bill in name only. It does nothing to further our scientific and innovation enterprise.”

Smith disagrees, saying the legislation “re-establishes the federal government’s primary scientific role to fund basic research [and] prioritizes taxpayer investments.” In offering additional spending—often at odds with what the administration has requested (see table, p. 381)—he cites the country’s need to catch up in supercomputing and particle physics and preserve its lead in other areas. Smith offsets those increases by cutting “later-stage” technology and commercialization programs that he believes “are more effectively pursued by the private sector.” The bill authorizes spending levels for the 2016 and 2017 fiscal years, with no increases in the second year.

Some science advocates object to Smith’s assertion that legislators can do a better job than the scientific community itself in identifying the most promising research areas. They are angry that he has proposed funding levels

In February, a group of Nobel laureates urged Juncker to reverse what they termed a “misguided and short sighted policy.” Cutting Horizon 2020 funds “send[s] a message that Europe is not the place to do high level science,” they wrote.

Juncker’s plan requires the approval of both the Parliament and the Council of Ministers. However, in a vote on 20 April, members of the European Parliament (MEPs) agreed to set up EFSI but opposed devoting Horizon 2020 funds to it. The Parliament “appears as the only E.U. institution defending Horizon 2020,” said Kurt Deketelaere, secretary-general of the League of European Research Universities, in a statement.

“It is completely unnecessary to transfer money from Horizon 2020,” says Kathleen van Brempt, a Belgian socio-democrat MEP who sits on the Parliament’s industry, research, and energy committee. Van Brempt instead wants EFSI to use unspent E.U. funds that are normally returned to national governments at the end of each year. “There are large enough leftovers to pay for the [EFSI] guarantee fund,” which the Parliament could watch over year after year, she says. However, national finance ministers like their annual E.U. refund, and so will certainly oppose van Brempt’s plan during the coming weeks of “trilogue” negotiations between the commission, Parliament, and member states. Juncker said he wants to finalize the plan before the summer.

Although the Parliament has often flexed its muscles this way since its powers were extended in 2009, the Council of Ministers usually has the upper hand in negotiations over the E.U. budget. “We can expect weeks of high political games,” Deketelaere says. “The question is who is going to give in.” The Parliament is showing a united front to protect research funds from cuts, but “I don’t know if that will hold until the end,” he adds. ■



European Commission chief Jean-Claude Juncker says scientists shouldn't worry.

RESEARCH FUNDING

House science chief unveils contentious vision for science

Authorization bill would cut climate and energy research and reshape science agency policies

By Jeffrey Mervis and Adrian Cho

Representative Lamar Smith (R-TX) has never hidden his desire to reshape federal research policy—often over the objections of much of the scientific community—since he became chair of the House of Representatives science committee 2 years ago. Last week, he introduced legislation that lays out those plans in unprecedented detail, and the reaction was predictable. Although academic leaders say that some parts of the new, 189-page bill are better than previous versions, they believe it would seriously damage the U.S. research enterprise.

The bill not only sets out funding levels for several research agencies that in some cases depart sharply from those the Obama administration requested for 2016; it would also reshape key policies and priorities guiding those agencies. In particular, researchers complain that the bill (H.R. 1806), called the America COMPETES Reauthorization Act of 2015, would:

- Narrow the scope of research at the National Science Foundation (NSF) by designating some scientific disciplines as more important to the nation than others;
- Sharply reduce NSF’s authority to fund the social sciences and the geosciences;
- Restrict NSF’s ability to build large new scientific facilities by requiring the agency to follow new, controversial, accounting practices;
- Curtail climate change research at the Department of Energy (DOE);
- Block the government from using DOE research findings in writing regulations;
- And squeeze the budgets for DOE’s applied research program and its fledgling Advanced Research Projects Agency-Energy (ARPA-E).

Smith doesn’t have the authority to impose that vision on Congress. The Senate has yet to draft a matching authorization bill, and the path to congressional passage is uncertain. But his committee’s oversight of several key scientific agencies means his ideas will play an important role in the debate.

The first America COMPETES Act, passed in 2007 and renewed in 2010, enjoyed bipartisan support from presidents George W. Bush and Barack Obama. In contrast, the new version has no Democratic co-sponsors, was not shown to Democratic committee members before it was unveiled, and has triggered a bitter partisan debate.

“The original American COMPETES Act was one of the crowning achievements of the science committee,” says Representative Eddie Bernice Johnson (D-TX), the top-ranked Democrat on the panel. “This bill is an America COMPETES bill in name only. It does nothing to further our scientific and innovation enterprise.”

Smith disagrees, saying the legislation “re-establishes the federal government’s primary scientific role to fund basic research [and] prioritizes taxpayer investments.” In offering additional spending—often at odds with what the administration has requested (see table, p. 381)—he cites the country’s need to catch up in supercomputing and particle physics and preserve its lead in other areas. Smith offsets those increases by cutting “later-stage” technology and commercialization programs that he believes “are more effectively pursued by the private sector.” The bill authorizes spending levels for the 2016 and 2017 fiscal years, with no increases in the second year.

Some science advocates object to Smith’s assertion that legislators can do a better job than the scientific community itself in identifying the most promising research areas. They are angry that he has proposed funding levels



Research on hurricane prediction could be affected by proposed geoscience cuts.

for each of NSF's seven directorates rather than giving the agency a top-level number and allowing it to distribute funding as appropriate. They are angrier still that he wants to cut more than \$100 million from two of those research directorates—geosciences and social sciences—shrinking the latter by more than half.

"I think it's ironic that the science committee would mark up this bill on Earth Day," says Sherri Goodman, president of the Consortium for Ocean Leadership in Washington, D.C., referring to a meeting this Wednesday of the panel. "Before they take such a drastic step, I hope they reconsider the adverse consequences it would have to the environment, the economy, and national security."

The bill also targets NSF's oversight of big new scientific facilities. Smith and other Republicans believe that NSF has been lax in this regard, and the bill requires the agency to "correct" any problems identified by an independent audit of a project's expected cost before starting construction. That language could seriously delay new projects, say NSF officials, who add that the bill also contains rules about the use of contingency funds that are at odds with existing federal policies.

Within DOE's Office of Science, the bill's impact would be limited to two of the office's six research programs. It would move more than \$60 million from biological and environmental research into fusion research. That money would presumably come out of DOE climate change efforts, as the bill would require DOE to eliminate any climate research deemed to overlap with what other federal agencies are doing.

When it comes to DOE's applied research efforts, however, the new COMPETES act takes out a hatchet. It calls for cutting spending on DOE's energy efficiency and renewable energy (EERE) program by 37% from its current level, to \$1.2 billion in 2016. In contrast, the White House has requested a 42% boost, to \$2.7 billion. Similarly, it would slash

the budget for ARPA-E, devoted to translating the best results from basic research to budding energy technologies, by 50%, to \$140 million, rather than increase it by the 16% the administration has sought. The House bill also zeros out numerous smaller applied research projects such as DOE's next generation lighting initiative, building standards program, and efforts to find additional uses for electric-car batteries.

Those changes are in line with the argument by many Republicans that private industry, not the federal government, should pick up the tab for translating the most promising basic research into commercial technologies, says Michael Lubell, a lobbyist with the American Physical Society (APS) in

Washington, D.C. But that assumption is unrealistic, he says. A 2008 APS study showed "that some federal money had to be provided because nobody else would do it," he notes.

Smith and his Republican colleagues do back government support for early-stage technologies outside the energy arena. Last December, in the final 2015 spending bill for all federal agencies, Smith championed a network of manufacturing innovation research centers, to be run by the National Institute of Standards and Technology, that would tap money now going to DOE's EERE. The original plan called for spending \$250 million over a decade on several new centers. But the COMPETES reauthorization would speed up that timetable, making \$150 million of the total available over the next 3 years.

Research funding levels aren't all that troubles Lubell and other science advocates. The bill also proclaims that "the results of any research, development, demonstration, or commercial application projects or activities of the [Energy] Department may not be used for regulatory assessments or determinations by Federal regulatory authorities." That clause, although it appears within a section on fossil energy research and development, would prevent the federal government from using any of the DOE research it paid for to inform policy. "That's absolutely bizarre," Lubell says.

The committee was scheduled to take up the bill on Wednesday, and Democrats are expected to offer a raft of amendments. Although few, if any, are likely to pass, science advocates are hoping that the Obama administration will object to enough of the bill's provisions to reject anything that manages to clear both houses of Congress. "At this point, this White House is poised to veto anything that threatens the president's priorities," Lubell says. "I don't really think there's a snowball's chance in hell of getting the president's signature on it." ■

COMPETEing visions for U.S. research

A bill drafted by House Republicans matches White House spending priorities for science in some areas but diverges greatly in others.

	2016 funding levels, in \$ billions		
	PRESIDENT'S REQUEST	COMPETES DRAFT	COMPETES COMPARED WITH REQUEST
COMPETES would boost:			
DOE fusion	0.420	0.488	16%
NSF biology	0.748	0.835	12%
NSF engineering	0.929	1.034	11%
NSF computer science	0.954	1.050	10%
NSF math/physical sciences	1.366	1.500	10%
COMPETES would reduce:			
NIST science	0.755	0.745	-1%
NSF overall	7.723	7.597	-2%
NSF education	0.962	0.866	-10%
DOE bio/environment	0.612	0.550	-10%
NSF geosciences	1.365	1.200	-12%
NIST	1.120	0.934	-17%
DOE renewables/efficiency	2.722	1.199	-56%
ARPA-E	0.325	0.140	-57%
NSF social/behavioral*	0.237	0.100	-58%
No change:			
NSF research account	6.186	6.186	0%
DOE Office of Science	5.340	5.340	0%
DOE advanced computing	0.621	0.621	0%
DOE Basic Energy Sciences	1.849	1.850	0%
DOE High Energy Physics	0.788	0.788	0%
DOE Nuclear Physics	0.625	0.625	0%

* Excludes funding for NSF's statistical agency



MEDICAL RESEARCH

Clinical trials get practical

Many clinical trials don't help doctors make decisions. A new breed of studies aims to change that

By Jennifer Couzin-Frankel, in
Philadelphia, Pennsylvania

With billions of dollars spent each year to test new drugs and devices, you would think clinical trials would help doctors treat the patient in front of them. But you would be wrong, says Robert Califf, a cardiologist at the U.S. Food and Drug Administration (FDA) in Silver Spring, Maryland. He highlights one salient example: Along with colleagues, Califf found that only 11% of the hundreds of guidelines for heart health were based on evidence from multiple randomized trials or meta-analyses. Many trials are too small or too poorly designed to tell us much, Califf says. Others don't address what doctors need to know, such as how one particular treatment stacks up against another. And even the biggest and best clinical trials tightly restrict who can sign up, casting doubt on their relevance to broad patient populations.

Last week, 200 statisticians, scientists, and physicians gathered at the University of Pennsylvania to hash out some options for repair. A leading remedy is a new breed of experiments: pragmatic clinical trials, which focus less on disease biology and more on helping doctors. They do not hew to a single design. Rather, pragmatic trials are guided by their end goal: informing practice. They might take all comers, focus on specific questions doctors want to answer, or streamline data collection to make an enormous trial more feasible. "These

trials will involve more women, more minorities, a range of incomes," says Monique Anderson, a cardiologist at Duke University in Durham, North Carolina.

But any new path raises questions, as last week's meeting made clear. They include how to approach informed consent ethically when trying to capture a broader swath of the population, whether the quality of data from sources like electronic health records is sufficient, and how easily results from these trials can be understood.

Broad enrollment is a key feature of pragmatic trials. Traditional trials tend to have strict admissions criteria, because drug companies believe the resulting data will be more precise, the drug will perform better, and it will be more likely to be approved. But in the real world, all sorts of patients take medication. "If you come into my emergency department, I can't say, 'You have one of these six things, I can't treat you,'" says Roger Lewis, an emergency medicine physician at the University of California, Los Angeles, who studies innovative trial designs. Once a treatment reaches a more diverse population—people with, say, diabetes or high blood pressure or obesity—there are often surprises about its safety and effectiveness.

In February, the Patient-Centered Outcomes Research Institute in Washington, D.C., announced \$64 million in funding for five pragmatic trials. And in 2012, the National Institutes of Health established the "Collaboratory" to study how pragmatic trials might play out. For now, pragmatic studies

One pragmatic clinical trial compares different approaches to dialysis. Studies like this will enroll a broader cohort, including more women and minorities.

tend to focus on health behaviors or compare available treatments, not test experimental drugs, although that could change.

Nine Collaboratory trials are under way. One tests whether patients on dialysis are more likely to survive and stay healthier if the dialysis treatment itself lasts longer. The study is randomizing about 400 dialysis centers around the country to either continue with their usual routine—dialysis typically ranges from about 3 to 5 hours in the United States—or administer it for at least 4.25 hours. Patients receive information about the trial at their clinic and a toll-free number to call if they have questions for the research team or wish to opt out.

An opt-out model is an option only for some of the lowest risk clinical trials: U.S. regulations require active informed consent for studies of experimental drugs. Because current pragmatic trials are comparing approaches doctors already use routinely, even ethicists agree that enrolling everyone, unless someone objects, is often reasonable.

Other challenges come in figuring out the best way to design pragmatic studies, interpret the results, and consider how much uncertainty is tolerable. For example, the Women's Health Initiative (WHI), which reported on randomized trials of hormone replacement therapy more than a decade ago, is now running a pragmatic trial that relies on Medicare data to tell it whether 26,000 women who receive mailings and coaching to encourage exercise are less likely to develop heart problems. It's a study so large that funding it would likely have been a stretch without plugging into health records. But there are disadvantages, too, explained Garnet Anderson, one of WHI's leaders and a biostatistician at the Fred Hutchinson Cancer Research Center in Seattle, Washington, at last week's meeting. Medicare data are designed for billing, not research, and the data have gaps. "What kind of errors can we tolerate?" she asked at the podium. "I don't know."

Regulators still need to consider whether and how pragmatic trials could be used to test experimental drugs. Califf, who recently left Duke University to become a deputy commissioner at FDA, says he plans to push for change. Although it's hard to argue against more practical trials—and most don't—companies may worry about retooling an entrenched model, and some scientists might be concerned about too heavy a focus on pragmatism and too little on basic biology. "I intend at FDA to fight that battle," Califf says, "and see how successful I can be." ■

ECOLOGY

Inbred wolf population on Isle Royale collapses

But other wolves adopt a new Lake Superior island

By Christine Mlot

Each January, before they fly to snow-bound Isle Royale in Lake Superior, ecologists Rolf Peterson and John Vucetich guess how many wolves they'll spot. This U.S. national park in Michigan is home to the world's longest running predator-prey study, of wolves and moose, but last year the wolves had dwindled to nine due to complications of inbreeding. This year, Peterson figured they'd likely find a mere seven wolves. But the island held only three, as the team from Michigan Technological University in Houghton announced last week. "The collapse of the wolves was beyond our expectation," Peterson says.

The three wolves included a pair, probably the last known to have reproduced, plus a smaller potential pup. The other wolves are presumed to have either died or left the island last year, in a reverse of how carnivores originally came to Isle Royale, when a bitter winter froze the channel to the mainland (*Science*, 24 May 2013, p. 919).

But even as the famed predator-prey study on Isle Royale appears to be on its last legs, other researchers may have caught the birth of a similar natural experiment. Across the lake in Canada, three mainland wolves crossed the ice to a smaller island with different prey and seem to have settled in, population ecologist Brent Patterson of Trent University in Peterborough, Canada, was slated to report at a meeting this week.

The wolves on Isle Royale once numbered as many as 50. But they have been mostly isolated for generations and the population has been overrun with spinal defects, likely from inbreeding. The possible pup seen this winter also displayed an abnormally short tail with raccoonlike stripes and a hunched back—perhaps due to the lack of genetic diversity. "It [didn't] look particularly healthy" in February, Vucetich says, and may be dead by now. He and Peterson collected frozen wolf scat for DNA analyses to identify the trio.

The fate of the island's other wolves is hard to pin down. One male wore a radio collar and was spotted dead; his carcass will be collected this spring for autopsy. If all the others died, the 70% mortality rate for the year would be the highest in the study's history, Peterson says. So the wolves may have simply left via the icy corridor to the mainland, seeking unrelated mates. When the channel froze last year (for only the second time in 16 years) scientists learned later that a female identified by her collar crossed to the mainland and was shot.

This year, an ice bridge formed again and for the first time scientists got a real-time view of how Isle Royale's wolves arrived. Genetic analyses have revealed after the fact that such immigrations and resulting matings must have happened several times during the study's 57 years. In February, Vucetich spotted two mainland wolves on the island. One, a female, fortuitously was fitted with a radio collar as part of a study by the Grand Portage, Minnesota, band of

Lake Superior Chippewa. The two visitors canvassed the southwest edge of Isle Royale for 5 days and then headed back to the mainland. The radio-collared Minnesota wolf next cruised two nearby small islands, perhaps seeking easier prey than a 400-kg moose. "That's how [wolves] get information," Peterson says. "They walk." Peterson thinks the newcomers were aware of the resident wolves, but that mating would be unlikely given the few residents and the pair bond of two of them.

Roughly 300 kilometers across the lake from Isle Royale, a trio of wolves might have found the Goldilocks option in the forests of 184-square-kilometer Michipicoten Island Provincial Park in Canada. Three mainland wolves appear to have colonized the island via ice and have probably bred. Based on aerial surveys, Patterson estimates that Michipicoten holds 250 to 300 "predator-naïve" woodland caribou, which are smaller than moose but bigger than deer. Working for the Ontario Ministry of Natural Resources and Forestry, Patterson and his colleagues radio-collared all three wolves in February and are preparing for a potential Isle Royale redux, Patterson was to announce this week at the Midwest Wolf Stewards meeting in Ashland, Wisconsin. As on Isle Royale, where wolves curbed the moose population and so helped preserve the vegetation, "the wolves are expected to have a profound impact," Patterson says.

Back on Isle Royale, it's no surprise that moose are booming. This year their numbers swelled by an estimated 200 to about 1250, despite the harsh winter. Wolves are so scarce that their effect on moose has been essentially nil for the past 4 years, according to Peterson.

Geneticists including Philip Hedrick of Arizona State University, Tempe, advocate introducing new breeding pairs of wolves to Isle Royale. "These animals might give the wolf population a new start and the remaining two adults might eventually contribute to this population," he says. But National Park Service officials say they plan to continue their hands-off stance for the moment. An environmental impact analysis and request for public input on how to manage the wolves, moose, and forest is to launch this spring. For now, Isle Royale's few wolves roam alone. ■

A tale of two islands

The iconic predator-prey study on Isle Royale may be ending, as only three wolves remain, including a pair and what may be their malformed pup (last in line in photo). But scientists have started a new study on Michipicoten Island, home to caribou and a trio of wolves recently arrived from the mainland.



Christine Mlot is a science writer based in Madison.

SCIENTIFIC PUBLISHING

NIH's peer review stands up to scrutiny

Analyses show better scored proposals produce more papers and citations

By Jeffrey Mervis

The debate over whether peer review can pick out the research most worthy of funding has heated up in the past decade as competition for federal dollars has become more intense.

Two new studies support claims that peer review works at the National Institutes of Health (NIH). But some who follow the peer-review debate say the papers' definition of success—three outcomes traditionally valued by the scientific community—ignores important factors, meaning that the debate is sure to continue.

One study, on page 434, examined the outcomes of 137,215 NIH research projects, or R01 grants awarded between 1980 and 2008. It found that grant proposals rated more highly by NIH study sections generated more publications and more citations than those that received lower scores. A second study, published online this month in *Research Policy*, found that the additional proposals funded after the agency received billions of dollars from the 2009 economic stimulus package garnered fewer publications and citations than the grants initially funded.

"Experts add value," says economist Danielle Li of the Harvard Business School in Boston, an author of the *Science* study. "It has something to do with their ability to see quality outcomes before they happen."

The head of NIH's massive grant-review enterprise, Richard Nakamura, agrees that the research appears to bolster the case for enlisting thousands of scientists as reviewers. But the data are hardly definitive, he says. The *Science* paper "says that, unlike what other studies have found, there is a relationship between scores and outcome measure if you look at enough grants," Nakamura says. "But it's a very noisy measure. And the debate over how to measure the outcome of grants remains very much alive."

The standard critique of peer review is that it works reasonably well in separating the wheat from the chaff, but that study section reviewers are less capable of making fine distinctions between two meritorious proposals. For the biomedical community served by NIH in particular, discontent with peer review has intensified over the past decade as success rates for project grants fell from one in three to nearly one

in six. The search for alternatives includes innovative ways of removing study sections from the allocation of grant funding (*Science*, 7 February 2014, p. 598).

Out of an endless number of possible outcome metrics with which to evaluate NIH's current reviewing process, Li and Leila Agha, who is at Boston University, chose perhaps the most conventional. They opted for the number of papers generated by a funded grant, how often those papers were cited by others, and what papers were among the most cited of the year. Their analysis showed that a proposal with a score one standard deviation above a second proposal resulted in 8% more publications, 17% more citations, and 24% more high-impact publications. (Factors such as an investigator's publication history, years since degree, and previous NIH funding cause the numbers to vary, but they remain statistically significant.)

The *Research Policy* study also gives NIH's peer-review system a pat on the back. Researchers from the Georgia Institute of Technology in Atlanta and Drexel University in Philadelphia treated 2775 awards NIH made with stimulus funding as a natural experiment, comparing them with 9779 regular grants. They found that those grants made after NIH lowered the pay line—the score dividing funded from unfunded proposals—generated fewer publications and citations than did projects that had survived the initial cut. The obvious implication: Reviewers knew what they were doing when they failed to fund the proposals the first time around.

Neither paper distinguishes between de novo applications—some 56% of the grants that Li and Agha examined—and renewals to continue work that NIH is already funding. That distinction is important, some scientists contend, because it's much easier to judge the value of research with a track record. "Peer review works very well in assessing past and present performance," says

Michael Lauer, head of cardiovascular science at NIH's National Heart, Lung, and Blood Institute in Bethesda, Maryland, whose work has questioned his institute's ability to pick the best research. "But it's much less good at making predictions" about whether an investigator's novel approach to a problem will bear fruit, he adds. Lauer notes that some NIH institutes are piloting approaches that bet on people rather than projects, a strategy that may place less importance on publications and citations.

Study section review scores are not the only way NIH decides how to invest scarce resources. Program managers must balance research portfolios across several fields, decide how large every award will be, and weigh whether funding someone with no other grants is likely to yield more science than adding to the pot of a well-heeled investigator. To capture those factors, Lauer says, he prefers to use return on investment—citation impact per million dollars spent—as a metric for research outcomes.

Nakamura says he worries about judging outcomes using publications and citations, because journal editors and authors have considerable control over those variables. But he's intrigued by another outcome metric in the *Science* paper that falls outside the traditional realm of academic science. Li and Agha found that the number of patents spawned by a funded grant correlated with proposal

scores—better reviews led to more patents. The relationship is not direct, however: The researchers counted patents that cited publications that in turn discuss other papers done under an NIH grant, not the grant that led to the patent.

Lauer hopes these new results will help NIH improve peer review by relying on vetted research rather than on anecdotes or gut instincts. "Their work allows us to talk about those options by using data rather than opinion," he says. ■

FEATURES

By Daniel Clery

At 25 years of age, the Hubble Space Telescope is still in its prime. Hubble's instruments are fully functional, and the orbiting observatory keeps cranking out new results. "It's at a peak of productivity," says Paul Hertz, director of NASA's astrophysics division.

Astronomers have published more than 13,000 papers based on Hubble data—"more than any other telescope in history," says Matt Mountain, president of the Association of Universities for Research in Astronomy in Washington, D.C., and former director of the Space Telescope Science Institute. It has helped measure the age and expansion of the universe, shown the ubiquity of supermassive black holes at the hearts of galaxies, watched comet Shoemaker-Levy 9 crash into Jupiter, and imaged some of the first galaxies. The cosmic photo album it has assembled includes some of the most memorable astronomical images ever.

Today, Hubble continues to do what it has always done: tirelessly snapping pictures of everything from nearby asteroids to the most distant active galactic nucleus. The best ground-based telescopes, with mirrors four times as wide as Hubble's and adaptive optics to correct for Earth's atmosphere, can now match Hubble's resolution, but Hubble can apply its keen eye over a wider field of view and still see much fainter objects. And it's never held up by clouds or sunrise. But what has really kept Hubble at the cutting edge is the five servicing missions by shuttle astronauts. Thanks to a series of repairs and upgrades—starting with a 1993 fix for its faulty mirror, which threatened to nip the whole mission in the bud—Hubble in 2015 is a better telescope than the one launched 25 years ago.

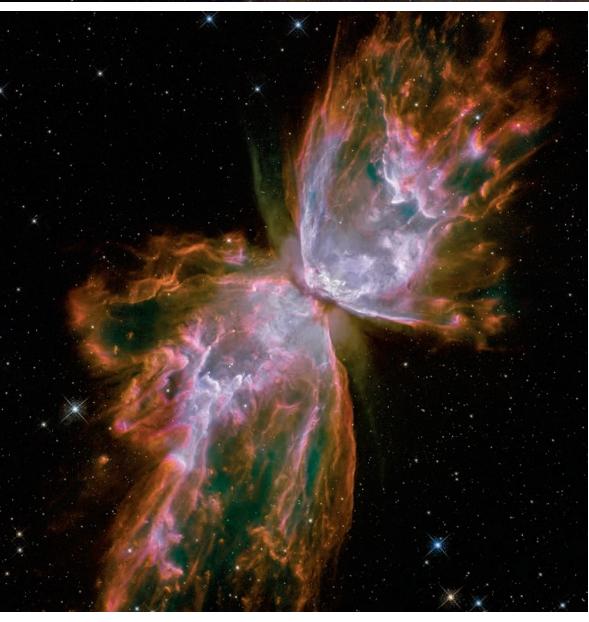
But all good things must end. With no space shuttle to service it, Hubble's demise is expected sometime next decade. Its instruments could fail, or its gyroscopes could conk out, leaving the observatory adrift. NASA remains confident that Hubble will survive to overlap for at least a few years with its successor, the James Webb Space Telescope (JWST), due for launch in 2018. Estimates of Hubble's orbital decay suggest that disposal, perhaps with the help of a remote-controlled spacecraft, won't be necessary before the 2030s.

As Hubble approaches those final years, researchers are turning their attention to what comes next. So far, the quest for a successor has faced rough going. JWST, the heir apparent, has been plagued by delays and cost overruns, sapping funds and support from other missions (see main story, p. 388). But even without those constraints, Hubble would be a tough act to follow. ■

Jewels from Hubble's trove. Clockwise from top left: Sombrero galaxy; newborn stars in N90 nebula; tower of cold gas and dust, Eagle nebula; hot gas jets, NGC 6302 (top); "Cat's Eye" nebula (bottom); pillar of gas and dust, Carina nebula.

PHOTOS: (CLOCKWISE FROM TOP LEFT) NASA/ESA AND THE HUBBLE HERITAGE TEAM (STSCI/AURA); NASA, ESA, AND THE HUBBLE HERITAGE TEAM (STSCI/AURA)-ESA/HUBBLE COLLABORATION; NASA, ESA, AND THE HUBBLE HERITAGE TEAM (STSCI/AURA); NASA, ESA, AND THE HUBBLE SM4 ERO TEAM; NASA, ESA, HEIC, AND THE HUBBLE HERITAGE TEAM (STSCI/AURA); NASA, ESA, M. LIVIO, AND THE HUBBLE 20TH ANNIVERSARY TEAM (STSCI)







Engineer at NASA's Marshall Space Flight Center inspects primary mirror segments destined for the Webb Space Telescope.

Downloaded from www.sciencemag.org on April 21, 2015

After Hubble

The Webb telescope's troubled history poses challenges for other contenders to replace the world's most popular space telescope

By Daniel Clery

Twenty years ago, when it was five, the Hubble Space Telescope took aim at a patch of sky so small it was almost empty of foreground stars and snapped the same image over and over for 10 days straight. When combined into a single detailed image, known as the Hubble Deep Field, the view transformed astronomers' knowledge of the early universe. The long total exposure, coupled with Hubble's unmatched resolution, revealed 3000 distant galaxies, some so red and faint that they dated back much of the way to the big bang. Hubble bettered the Deep Field in 2004 with the Hubble Ultra Deep Field and again in 2012 with the Hubble eXtreme Deep

Field, which showed some galaxies as they were 13.2 billion years ago—just 600 million years after the birth of the universe.

Now, as Hubble enters its final years, the astronomy community is wondering what instruments will deliver such revelations in the future. Hubble inaugurated an era of space-based Great Observatories, each probing a different part of the spectrum: the Compton Gamma Ray Observatory, the Chandra X-ray Observatory, and the infrared Spitzer Space Telescope. Each delivered its own revelations, from gamma ray bursts in the distant universe to far solar systems taking shape; all are now aging or defunct. NASA's one big plan for a follow-up space observatory, the James Webb Space Tele-

scope (JWST), has survived a near-death experience and is now on track for a launch in 3 years—but at a cost so steep, during a time of stagnant government funding, that it has squeezed out other missions or pushed them farther into the future.

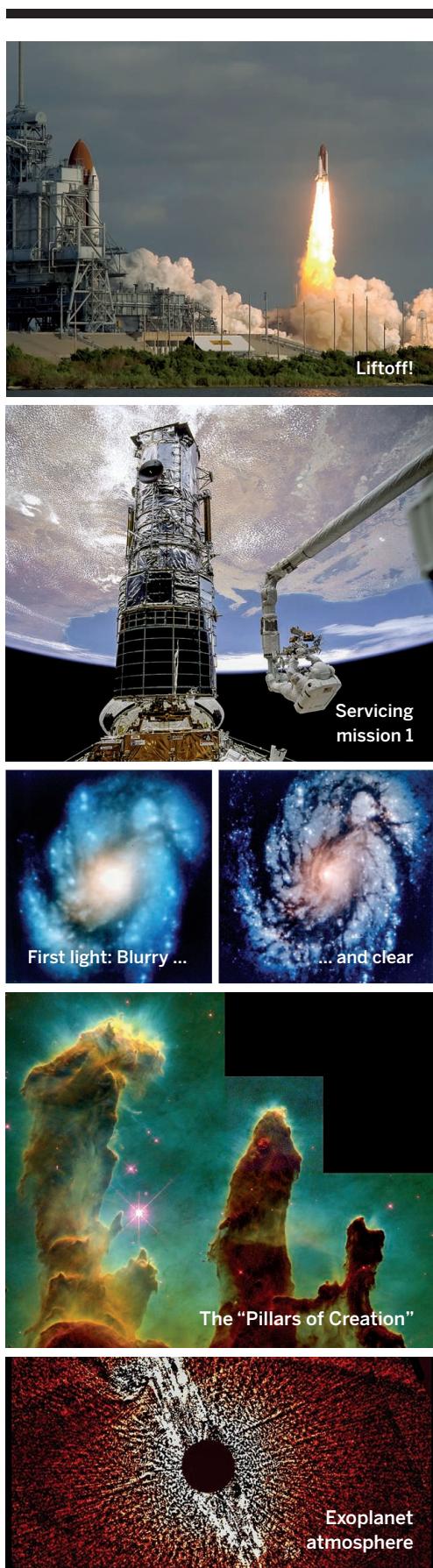
"It's been a struggle," says Michael Turner, director of the Kavli Institute for Cosmological Physics at the University of Chicago in Illinois. "This decade will be remembered for amazing discoveries and growth in the field, but new missions have not kept up with opportunities." As researchers begin to identify priorities for the next decade, they will have to make some tough decisions about where astronomy should focus its gaze.

IF ALL HAD GONE AS PLANNED, the Webb telescope would already be sending home its own dazzling views of the universe. In their 2000 decadal survey, part of a regular priority setting by U.S. astronomers, they rated the Next Generation Space Telescope—later named after former NASA Administrator James Webb—as their number one choice. JWST isn't a direct successor to Hubble: Whereas Hubble is sensitive to visible wavelengths plus small bands of ultraviolet and near-infrared, JWST ranges from orange and red visible light to mid-infrared.

That window on the cosmos will allow JWST to scrutinize light from some of the very first stars and galaxies, which the expansion of the universe has shifted into the infrared. Infrared light can also penetrate gas clouds that obscure areas where stars and galaxies form; some exoplanets and their atmospheres glow brightly in the infrared, as well. The Spitzer telescope, which ran out of liquid-helium coolant in 2009 and is now largely blind, made similar studies but was limited by its modest 85-centimeter mirror. JWST will capture infrared light with a segmented mirror 6.5 meters across. That increase “is an amazing leap from one generation to the next,” says JWST Director Eric Smith. “JWST will look at [star-forming] clouds with Hubble-like sharpness of vision.”

NASA initially estimated that JWST would cost \$1.6 billion and reach orbit in 2011. Design and technology development work began, but in 2005, spiraling costs forced a replanning of the project that pushed the launch back to 2013 and hiked the life cycle cost to \$4.5 billion, including hardware, launch, and 10 years’ operation. By 2010, the project was meeting all its technical goals, but schedule slips and budget overruns were building up again. An independent review criticized the project’s budgeting and management. In 2011, when NASA reported to Congress that the launch would likely slip to 2018 and the cost total more than \$8 billion, the House of Representatives appropriations committee responsible for science voted to cancel the program.

To save the mission, astronomers threw themselves into months of lobbying, letter writing, petition signing, and public outreach. “Cancellation would have been a disaster,” Turner says. “Great nations do great things, and this is exactly what the United States should be doing.” In a deal over the 2012 budget, Congress restored JWST funding but with conditions: Management of the project was reformed and the cost capped at \$8 billion. In return, lawmakers promised not to tinker with the budget each year. “They’ve kept up their end of the deal ... and we’re grateful for that,” Smith says.



A stellar career

The Hubble Space Telescope’s CV would fill volumes. Here are just a few milestones and a glimpse of the future:

24 April 1990

Hubble is launched

25 June 1990

First images are blurry because of a warped mirror

2–13 December 1993

Servicing mission 1 fixes mirror

13 January 1994

First clear images released

2 November 1995

“Pillars of Creation” photo released

15 January 1996

Hubble Deep Field images released

11–21 February 1997

Servicing mission 2

25 May 1999

Hubble measures expansion of the universe using Cepheid variable stars

13 November 1999

Hubble goes offline for about 2 months after four of its six gyroscopes fail

19–27 December 1999

Servicing mission 3A makes Hubble operational again

27 November 2001

First observation of an exoplanet atmosphere (HD 209458 b)

1–12 March 2002

Servicing mission 3B

1 February 2003

Space shuttle Columbia disintegrates on reentry; shuttle fleet grounded for 2.5 years

9 March 2004

Hubble Ultra Deep Field released

11–24 May 2009

Servicing mission 4

5 March 2015

Hubble discovers a supernova split into four images by a gravitational lens

2018?

Hubble’s successor, James Webb Space Telescope, to launch

~2020

Hubble stops operating

Today, JWST remains on schedule and within budget. Despite delays and unexpected complications, the project still has a 10- to 11-month schedule cushion in reserve. "That's more contingency than was planned for at this stage," Smith says. More than 99% of the spacecraft's hardware is either built or being made, but next comes the tricky task of putting it all together. This year, Smith says, the scope's mirror support structure, or backplane, will be delivered and the 18 hexagonal mirror sections—each 1.3 meters across—put in place. A row of three mirrors on each side of the backplane has to fold back like leaves of a folding table so that the mirror will fit into a rocket fairing, but the backplane must still hold the mirrors steady to within 1/10,000 the width of a human hair even while it is cooled to the -240°C of deep space.

The telescope will have to operate for at least 5 years in an orbit 1.5 million kilometers from Earth, beyond the reach of servicing missions like those that prolonged the life of Hubble, so engineers are planning a series of tests simulating the rigors of deep space. The biggest will come in 2017, when the entire telescope and its instruments will enter the cavernous Chamber A at NASA's Johnson Space Center in Houston, which can recreate the vacuum of space as well as the cycles of extreme heat and cold. This year, a working

model of the telescope will go through the same process three times. "Something could happen. You can never guarantee it won't," says Paul Hertz, director of NASA's astrophysics division. "Our job is to make sure that it doesn't happen."

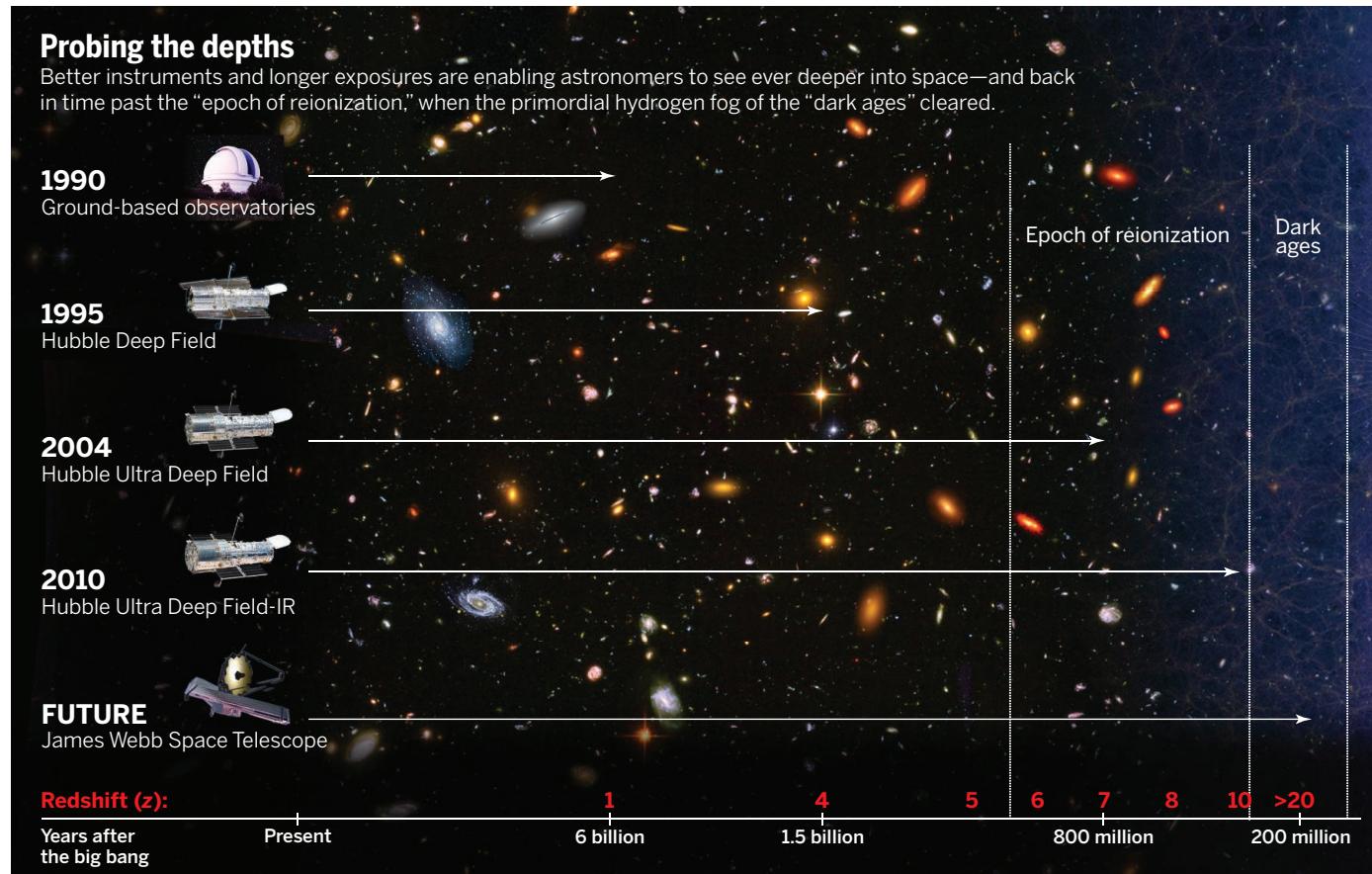
WHILE JWST HAS BEEN BROUGHT BACK from the brink, the project's delays and inflated cost have hindered other plans for major astronomy missions. "We always knew we couldn't start a new big mission until JWST was finished," Hertz says. One casualty is the top priority of the 2010 decadal survey: the Wide-Field Infrared Survey Telescope (WFIRST), an instrument designed to study the nature of dark energy, the mysterious force that appears to be accelerating the expansion of the universe.

"As soon as we released the [decadal survey] report, NASA told us about major cost problems with JWST," says Roger Blandford of Stanford University in Palo Alto, California, who headed the survey. "With JWST needing more spending to complete, there was less funding to spend than the [decadal] survey had assumed," Hertz admits. At the same time, the financial downturn squeezed NASA, causing the astrophysics budget to dip from up to \$1.5 billion a year during the last decade to about \$1.3 billion now. "We are

operating 10% lower than the historical average," Hertz says.

So WFIRST entered a phase of suspended animation. Hertz says the delay has had positive side effects. In 2012, the U.S. National Reconnaissance Office offered NASA two surplus 2.4-meter mirrors built for spy satellites. Now astronomers are reconfiguring WFIRST's design to accommodate a mirror more than a meter wider than originally planned and adding new capabilities, such as directly observing exoplanets. The result, they concluded earlier this year, will be a better telescope for almost no extra cost. Preparatory work continues, but "no formal decision has been made yet on when to go forward," Hertz says.

The casualties also include two other priorities of the 2010 survey, the International X-ray Observatory (IXO) and the Laser Interferometer Space Antenna (LISA), a detector of gravitational waves. Both were planned as collaborations with the European Space Agency (ESA), as Hubble itself is. "They seemed like dream collaborations. We'd learned how to work together," Turner says. But with NASA backpedaling on its commitment and ESA having funding troubles of its own, the Europeans didn't select either mission for the launch slots they were aiming for. ESA has since resurrected IXO as



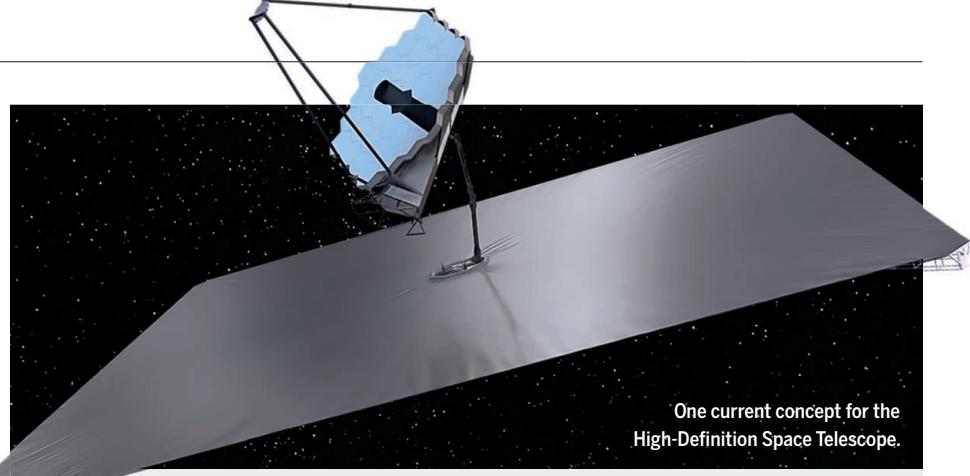
the Advanced Telescope for High-ENergy Astrophysics (ATHENA) and slated it for a 2028 launch. NASA is discussing joining the project as a minor partner. LISA remains in limbo: Both agencies are considering going it alone with scaled-down versions, but they are still talking about joining forces for a launch in the 2030s. “LISA will rise from the dead. The science needs to be done,” Turner says.

Astronomers continue to dream. Hertz, for example, has put forward a list of possible missions for the next decadal survey. They include a large telescope covering ultraviolet, visible, and near-infrared that would be a true successor to Hubble; a far-infrared telescope; a telescope for imaging exoplanets; and an x-ray observatory to follow ATHENA. This year, he will set up a few study teams to develop the ideas. “By the time of the decadal [survey], we’ll have well worked-out concepts,” Hertz says.

Others have also been laying groundwork. The Association of Universities for Research in Astronomy (AURA) in Washington, D.C., is studying a slightly different successor to Hubble: a large UV-visible light instrument it calls the High-Definition Space Telescope (see sidebar). HDST would both image galaxies in the early universe with exquisite detail and aim to obtain spectra from Earth-like exoplanets. “Everyone wants to find an Earth,” says Sara Seager of the Massachusetts Institute of Technology in Cambridge, co-leader of the study.

Whatever plans take shape, astronomers will surely have to fight hard to get their missions off the ground. More realistic costing and schedules, with plenty of contingencies, will be a must. But in the end, what saved JWST was the fact that its science goals are profound and a large number of scientists support it. “A mission has to appeal to a broad range of the community, and the American public has to buy into it,” says AURA President Matt Mountain. Achieving that broad constituency can be difficult, as the planners of HDST are finding as they try to marry the often diverging needs of exoplanet hunters and surveyors of the early universe.

Ultimately, the prospects for future great observatories may all hinge on a few nerve-racking weeks late in 2018 when ESA will launch JWST and transport it to a point 1.5 million kilometers from Earth. There it will unfurl its sunshield, piece together its mirror, and cool down its instruments. Although it’s all been practiced many times, it’ll be “7 weeks of terror,” Mountain says. “If they screw up, there is no way back, there’s no second chance,” he says. If, however, the telescope opens its eye on the universe and produces images as dazzling as Hubble’s, the future of orbital astronomy will also look brighter. ■



One current concept for the High-Definition Space Telescope.

Hubble on steroids

By Govert Schilling

As the Hubble Space Telescope’s decades-long mission winds down, astronomers are worried about the future of optical and ultraviolet astronomy from space. Hubble’s formal successor, the James Webb Space Telescope, is poised for launch in late 2018 (see main story, p. 388), but it will be an infrared observatory. So what about a *real* “son of Hubble”?

Astronomers started discussing plans for a “Very Large Space Telescope,” sporting an 8- to 10-meter mirror and possibly assembled in space, in 1989. “Those visionary ideas were probably ahead of their time,” says astronomer Garth Illingworth of the University of California’s Lick Observatory on Mount Hamilton. In June, a 17-person committee commissioned by the Washington, D.C.-based Association of Universities for Research in Astronomy (AURA) will try again. Their report will outline a plan for a High-Definition Space Telescope (HDST), so large and capable that it would make even the largest ground-based telescopes obsolete.

Given the budgetary and political woes of the Webb telescope, the proposal could still be ahead of its time. But it is designed to attract broad support from astronomers. “We want to make this a mission that works for everybody,” says committee Co-Chair Julianne Dalcanton of the University of Washington, Seattle. Earlier proposals favored either astrophysics—the study of galaxies and stars—or exoplanet research, fields that require very different equipment. HDST would serve both camps. “We need to get both on the same side,” says AURA’s president, Matt Mountain. “If they work separately, they don’t get anything.”

For astrophysicists, HDST would be “transformational,” says committee member Jason Tumlinson of the Space Telescope Science Institute in Baltimore, Maryland. Tumlinson says HDST will resolve structures a mere 300 light-years wide in every galaxy in the universe and will study star-forming regions no wider than Earth’s orbit.

When trained on exoplanets, says Tumlinson’s colleague Marc Postman, HDST could spectroscopically detect the atmospheric constituents, including possible biosignatures, of an Earth-like planet within 60 light-years or so. “With HDST, the path will be laid for characterizing Earth 2.0,” he says.

Postman envisions a 12-meter-class segmented, deployable mirror for HDST, comparable in design to the Webb telescope’s 6.5-meter segmented mirror. Wesley Traub of NASA’s Jet Propulsion Laboratory in Pasadena, California, prefers a cheaper, single-piece elliptical 8-by-3.5-meter mirror, to overcome potential problems with the segmented mirror design. Meanwhile, Illingworth would like the designers “to go bigger.” A 25-meter space telescope could be built for far less than twice the cost of a 12-meter instrument, he says, although it would also require the development of new launchers.

No one has formally estimated the cost of HDST. “It will be a multibillion-dollar project—that is about all one can say at this juncture,” Postman says. If astronomers support it in their next priority-setting exercise, the 2020 Astronomy and Astrophysics Decadal Survey, Hubble’s “true” successor could be on the launch pad in the mid-2030s, perhaps as a cooperative project with Europe and other countries. “I’m young enough to maybe get to use it,” Dalcanton says. ■

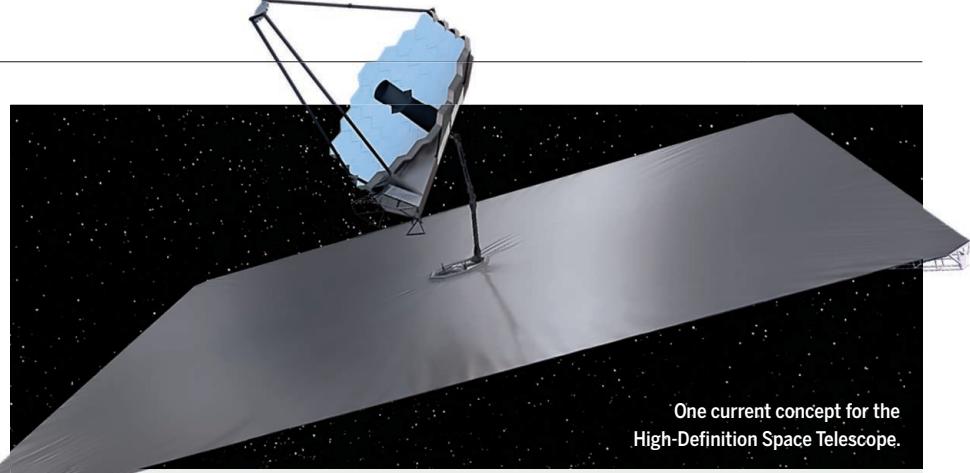
the Advanced Telescope for High-ENergy Astrophysics (ATHENA) and slated it for a 2028 launch. NASA is discussing joining the project as a minor partner. LISA remains in limbo: Both agencies are considering going it alone with scaled-down versions, but they are still talking about joining forces for a launch in the 2030s. “LISA will rise from the dead. The science needs to be done,” Turner says.

Astronomers continue to dream. Hertz, for example, has put forward a list of possible missions for the next decadal survey. They include a large telescope covering ultraviolet, visible, and near-infrared that would be a true successor to Hubble; a far-infrared telescope; a telescope for imaging exoplanets; and an x-ray observatory to follow ATHENA. This year, he will set up a few study teams to develop the ideas. “By the time of the decadal [survey], we’ll have well worked-out concepts,” Hertz says.

Others have also been laying groundwork. The Association of Universities for Research in Astronomy (AURA) in Washington, D.C., is studying a slightly different successor to Hubble: a large UV-visible light instrument it calls the High-Definition Space Telescope (see sidebar). HDST would both image galaxies in the early universe with exquisite detail and aim to obtain spectra from Earth-like exoplanets. “Everyone wants to find an Earth,” says Sara Seager of the Massachusetts Institute of Technology in Cambridge, co-leader of the study.

Whatever plans take shape, astronomers will surely have to fight hard to get their missions off the ground. More realistic costing and schedules, with plenty of contingencies, will be a must. But in the end, what saved JWST was the fact that its science goals are profound and a large number of scientists support it. “A mission has to appeal to a broad range of the community, and the American public has to buy into it,” says AURA President Matt Mountain. Achieving that broad constituency can be difficult, as the planners of HDST are finding as they try to marry the often diverging needs of exoplanet hunters and surveyors of the early universe.

Ultimately, the prospects for future great observatories may all hinge on a few nerve-racking weeks late in 2018 when ESA will launch JWST and transport it to a point 1.5 million kilometers from Earth. There it will unfurl its sunshield, piece together its mirror, and cool down its instruments. Although it’s all been practiced many times, it’ll be “7 weeks of terror,” Mountain says. “If they screw up, there is no way back, there’s no second chance,” he says. If, however, the telescope opens its eye on the universe and produces images as dazzling as Hubble’s, the future of orbital astronomy will also look brighter. ■



One current concept for the High-Definition Space Telescope.

Hubble on steroids

By Govert Schilling

As the Hubble Space Telescope’s decades-long mission winds down, astronomers are worried about the future of optical and ultraviolet astronomy from space. Hubble’s formal successor, the James Webb Space Telescope, is poised for launch in late 2018 (see main story, p. 388), but it will be an infrared observatory. So what about a *real* “son of Hubble”?

Astronomers started discussing plans for a “Very Large Space Telescope,” sporting an 8- to 10-meter mirror and possibly assembled in space, in 1989. “Those visionary ideas were probably ahead of their time,” says astronomer Garth Illingworth of the University of California’s Lick Observatory on Mount Hamilton. In June, a 17-person committee commissioned by the Washington, D.C.-based Association of Universities for Research in Astronomy (AURA) will try again. Their report will outline a plan for a High-Definition Space Telescope (HDST), so large and capable that it would make even the largest ground-based telescopes obsolete.

Given the budgetary and political woes of the Webb telescope, the proposal could still be ahead of its time. But it is designed to attract broad support from astronomers. “We want to make this a mission that works for everybody,” says committee Co-Chair Julianne Dalcanton of the University of Washington, Seattle. Earlier proposals favored either astrophysics—the study of galaxies and stars—or exoplanet research, fields that require very different equipment. HDST would serve both camps. “We need to get both on the same side,” says AURA’s president, Matt Mountain. “If they work separately, they don’t get anything.”

For astrophysicists, HDST would be “transformational,” says committee member Jason Tumlinson of the Space Telescope Science Institute in Baltimore, Maryland. Tumlinson says HDST will resolve structures a mere 300 light-years wide in every galaxy in the universe and will study star-forming regions no wider than Earth’s orbit.

When trained on exoplanets, says Tumlinson’s colleague Marc Postman, HDST could spectroscopically detect the atmospheric constituents, including possible biosignatures, of an Earth-like planet within 60 light-years or so. “With HDST, the path will be laid for characterizing Earth 2.0,” he says.

Postman envisions a 12-meter-class segmented, deployable mirror for HDST, comparable in design to the Webb telescope’s 6.5-meter segmented mirror. Wesley Traub of NASA’s Jet Propulsion Laboratory in Pasadena, California, prefers a cheaper, single-piece elliptical 8-by-3.5-meter mirror, to overcome potential problems with the segmented mirror design. Meanwhile, Illingworth would like the designers “to go bigger.” A 25-meter space telescope could be built for far less than twice the cost of a 12-meter instrument, he says, although it would also require the development of new launchers.

No one has formally estimated the cost of HDST. “It will be a multibillion-dollar project—that is about all one can say at this juncture,” Postman says. If astronomers support it in their next priority-setting exercise, the 2020 Astronomy and Astrophysics Decadal Survey, Hubble’s “true” successor could be on the launch pad in the mid-2030s, perhaps as a cooperative project with Europe and other countries. “I’m young enough to maybe get to use it,” Dalcanton says. ■

By Marion Ronit Munk and René Rückert

To work or not shouldn't be a question

We are a two-scientist couple, an Austrian and a German, both with experience working in the United States. So we read with great interest the Working Life story in which Michelle Gabriele Sandrian, an American, shared her experience working as a postdoc in Austria while starting a family. (You can read it in *Science* at <http://scim.ag/EuropePostdoc>.) Sandrian was happy to start a family in a place where she and her husband could take time off to care for their new baby. Sandrian and her family had a positive experience—we're glad about that—but that doesn't mean Austria's policies are good for women's careers.

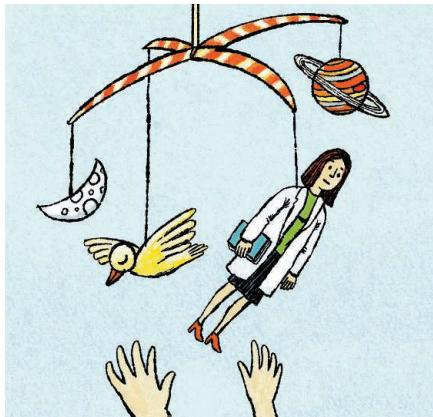
In Austria—Germany is much the same—many people don't have kids. Professional women tend not to have children, and women who do have children tend not to work or to work only part time. Men who have kids often have wives who stay at home.

Austria ranks 68th in equality of economic participation and opportunity in the World Economic Forum's *Global Gender Gap Report*, below Russia, Kazakhstan, and Brunei. In the United States, which ranks fourth, 3.2% of corporate CEOs are women—hardly impressive, but that's well above Europe's 2.4%, as reported in 2013 by the European Commission. As of 2012, when the data for the commission's report were collected, no major corporation in Austria had a woman as CEO. Not one.

Why are women leaders so rare in a society with such generous family-leave policies? Austria's conservative culture is partly to blame, but the blame should be shared by the policies. For women, the *Elternkarenz*, which pays parents to stay home for up to 3 years in Germany and up to 2 years in Austria, is the best career killer one could imagine. And child care for young children is hard to find and usually covers only part of the day. My sister working in Germany may have to give up a career in surgery because her day care recently decided to close at 3:30 p.m. She has no alternatives.

Long absences from work make women less competitive. They return to the office or the lab after a long leave to find that former colleagues have left them far behind.

These policies affect all women, not just those who take long leaves. During the paid absence of a young parent, no employer may hire a permanent replacement. Because women are much more likely than men to take long parental leaves, employers have a strong incentive to hire men. Europe's equal-employment laws make such discrimination



"Professional women tend not to have children."

illegal, but employers discriminate anyway, as many young women in science and medicine can attest.

How can the situation be improved? Women should continue to have sufficient time off, with pay, as they adjust to being mothers. But parental leaves need not last for years; Sandrian took just 5 months off before going back to work. Policies should encourage or require fathers to do their share; Norway, one of the most successful countries in Europe in terms of women's advancement, expects fathers to take off at least 10 weeks. More men taking more time off should reduce hiring bias and, as a bonus, encourage fathers to take on a greater share of the domestic burden, yielding benefits for women's careers for years to come.

Most importantly, mothers need to be able to return to work sooner without having to worry about their babies, so high-quality subsidized, flexible, full-day child care needs to be more widely available for younger children. Part-time work arrangements should be available, too, but the laws should strongly encourage men to take equal advantage.

If we wish to facilitate equal success for women in science, medicine, and other careers, changes in society are needed. Both women and men should demand policies that support mothers who choose to work and encourage their partners to take on their fair share of the domestic load. Time off is not the issue; we need to put policies in place that help women work. ■

Marion R. Munk is an M.D.-Ph.D. ophthalmologist with a background in immunology and clinical research. René Rückert is an M.D.-MBA immunologist. Both now live in Bern. For more on life and careers, visit ScienceCareers.org. Send your story to SciCareerEditor@aaas.org.

LETTERS

Edited by Jennifer Sills

A path to recruiting more STEM teachers

I STRONGLY DISAGREE with the implication in the News Feature "A classroom experiment" (J. Mervis, 6 February, p. 602) that recruiting new K-12 teachers in science, technology, engineering, and mathematics (STEM) is futile and that efforts to improve education should instead focus on in-service teachers. For disciplines such as physics and chemistry, fewer than half of all classrooms are led by a teacher with a degree in the subject (1). Without efforts to improve recruitment, we cannot effectively address the severe shortage of physical science teachers with deep content knowledge. Teachers with deep content knowledge in their subject, along with mastery of content-specific pedagogical skills, are the most important factor in ensuring excellence in STEM education (2).

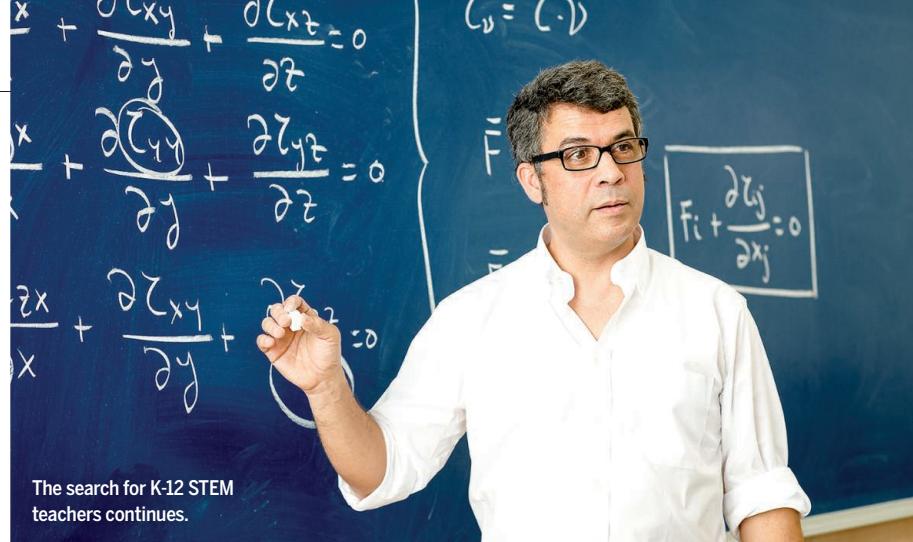
The American Physical Society, American Association of Physics Teachers, and dozens of universities across the country have worked during the past 15 years to address the severe shortage of physics teachers through the Physics Teacher Education Coalition (PhysTEC) project. PhysTEC sites establish a positive climate for teaching in the physics department; offer experiences that expose students to the rewards and challenges of teaching; build effective pathways for students to earn a degree and certification; and host a master teacher who mentors students. The result: PhysTEC sites have more than doubled the number of graduates well prepared to teach physics.

Moreover, in our experience, Noyce scholarships have provided critical financial support to PhysTEC students. Although scholarships alone are typically not enough to recruit new teachers, they are a key piece of the puzzle along with transforming disciplinary departments.

I was also troubled by the inaccurate statement that "5-year attrition rates rise to 50% or higher." Five-year retention rates for new teachers are about 70% (3), and retention rates for PhysTEC teachers are even higher (4). Propagating unfounded statements that confirm negative stereotypes about K-12 teaching is damaging and undermines efforts to improve it.

Monica Plisch

Director, Physics Teacher Education Coalition (PhysTEC); Associate Director of Education and



The search for K-12 STEM teachers continues.

Diversity, American Physical Society, College Park, MD 20740, USA. E-mail: plisch@aps.org

REFERENCES

1. J.G. Hill, "Education and Certification Qualifications of Departmentalized Public High School-Level Teachers of Core Subjects: Evidence from the 2007–08 Schools and Staffing Survey" (U.S. Department of Education, National Center for Education Statistics, Washington, DC, NCES 2011-317, 2011).
2. President's Council of Advisors on Science and Technology, "Prepare and Inspire: K-12 Education in Science, Technology, Engineering, and Math (STEM) for America's Future" (Executive Office of the President, Washington, DC, 2010); www.whitehouse.gov/administration/eop/ostp/pcast/docsreport.
3. R. Goldring, S. Taie, M. Riddles, "Teacher Attrition and Mobility: Results from the 2012–13 Teacher Follow-up Survey" (U.S. Department of Education, National Center for Education Statistics, Washington, DC, NCES 2014-077, 2014).
4. Physics Teacher Education Coalition, *PhysTEC News* 8, 1 (2014); www.phystec.org/publicity/files/fall14.pdf.

Editor's note

LAST OCTOBER, SCIENCE published an investigation into a dispute between physicist Ulf Leonhardt and the Centre for Optical and Electromagnetic Research (COER) at South China Normal University (SCNU) in Guangzhou ("Show me the money," 24 October 2014, p. 411). The story centered on Leonhardt's recruitment by COER as a part-time distinguished professor, funded by generous grants from the Chinese Central and Guangdong governments, and the premature end of his contract, when he charged that COER was not transparent about the sources of money and was spending it on research he did not know about or authorize. COER, for its part, accuses Leonhardt of bad faith and breach of contract. The dispute is acrimonious and ongoing.

Our goal in writing about the dispute was to illuminate the potential for misunderstanding and conflict inherent in these arrangements, which have become more common as China's science develops. The allegations reported in the article reflect the positions of the respective parties, and

Science does not endorse them. We regret any implication to the contrary.

Since the article appeared, we have learned that it contains several minor inaccuracies concerning the timing of events and the details of the grants and contracts. We have corrected these in the online version of the story. We have also clarified other language and removed an illustration that was open to misinterpretation. Since publishing the story, we have learned that before Leonhardt arrived in China, he received a draft of a presentation describing research planned for the Guangdong Leading Talent project. Leonhardt also took part in an oral defense in Beijing on March 21st, 2012, for the Guangdong Leading Talent project application. He and COER continue to disagree about how much advance knowledge he could have had about the research funded by his grants.

Tim Appenzeller

News Editor

A Chinese physics institute's defense

"SHOW ME THE money" (M. Hvistendahl, News Feature, 24 October 2014, p. 411) contains numerous false allegations, misinterpretations, and bias, which we must address.

Leonhardt claims he was unaware of the grant application contents. However, Leonhardt had a chance to revise a draft of the slides for the Guangdong Leading Talent Project and presented the slides to the committee. They detail the projects on transformation optics-based super-resolution imaging, cloaking, and absorbers for which funding was sought. Thus, Leonhardt should have been aware of the contents of the grant application. Casimir forces are briefly mentioned in

the Thousand Talents application because Leonhardt provided his Casimir force work (1) in a package of his selected papers when he asked the Centre for Optical and Electromagnetic Research (COER) to prepare the application for him. At that time no COER member worked on Casimir forces. Leonhardt claims he did not suggest Casimir forces until one year after August 2011, whereas the article claims that "Leonhardt planned to work on the theory of Casimir forces," which is self-contradictory.

The document provided to *Science* by Professor Jun Li explicitly connects the team slide (including potential collaborators) pictured in the article to the research directions of the Leading Talent defense, which did not include the Casimir force.

Our finances have been strictly in accordance with the approved budget and grant regulations. 790,043 RMB (\$128,462) was spent on research equipment; 208,920 RMB (\$33,971) was spent on travel for Leonhardt, his partner Jana Silberg, and their invited guests; and the unused 4,001,037 RMB (\$650,575) was returned to the grant authority. Normal salary is taxed and transferable abroad, whereas tax-free subsidies are intended to be used in China. Leonhardt made several unsuccessful attempts to

transfer subsidy money abroad while in Guangzhou from 12 to 18 January 2013. On 28 January 2013, Leonhardt e-mailed to COER, "Concerning Jana's payment, please arrange for a regular salary and give me the account number where I can return the 0.5M." This return would not have affected the net income Leonhardt would have received if he completed his 5-year contract. This 500,000 RMB (\$81,301) remains frozen in a South China Normal University (SCNU) bank account awaiting government instruction. 1,566,667 RMB (\$254,743) in pre-tax salary and housing subsidy remains in Leonhardt's possession, and he shows no intention of returning any of it. Since Leonhardt was in mainland China for only 57 days while under contract and did not fulfill the contract's terms, SCNU is taking legal action to recover a substantial portion of this money.

While we appreciate the over 20 changes, the modified article online still misrepresents our center's voice, most notably Professor Sailing He's quotes. They are out of context and based on the journalist's notes, which He did not have a chance to review and sign. When He said that Leonhardt "doesn't need to care about the details," he was referring to whether

the income was categorized as subsidies or salary.

As noted in the article, Leonhardt has waged a campaign through many Chinese agencies to defame COER. No Chinese agency has responded, because COER did nothing wrong and Leonhardt's allegations are groundless.

Chinese culture values trust between people. Unfortunately, Leonhardt betrayed our trust. We have learned some hard lessons and proposed reforms to related government agencies. SCNU has always strictly implemented relevant provisions of the national Thousand Talents Program and Guangdong Leading Talent Project, accepted supervision from upper-level departments, and thanks the community for support. SCNU will continue to improve management and service, and always welcome highly talented individuals who actively promote teaching and research development at the University.

Langping He

Deputy Dean, Centre for Optical and Electromagnetic Research and Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, Guangdong, 510006, China.
E-mail: Iphe@scnu.edu.cn

REFERENCE

- U. Leonhardt, T.G. Philbin, *New J. Phys.* **9**, 254 (2007).

LETTERS

Edited by Jennifer Sills

A path to recruiting more STEM teachers

I STRONGLY DISAGREE with the implication in the News Feature "A classroom experiment" (J. Mervis, 6 February, p. 602) that recruiting new K-12 teachers in science, technology, engineering, and mathematics (STEM) is futile and that efforts to improve education should instead focus on in-service teachers. For disciplines such as physics and chemistry, fewer than half of all classrooms are led by a teacher with a degree in the subject (1). Without efforts to improve recruitment, we cannot effectively address the severe shortage of physical science teachers with deep content knowledge. Teachers with deep content knowledge in their subject, along with mastery of content-specific pedagogical skills, are the most important factor in ensuring excellence in STEM education (2).

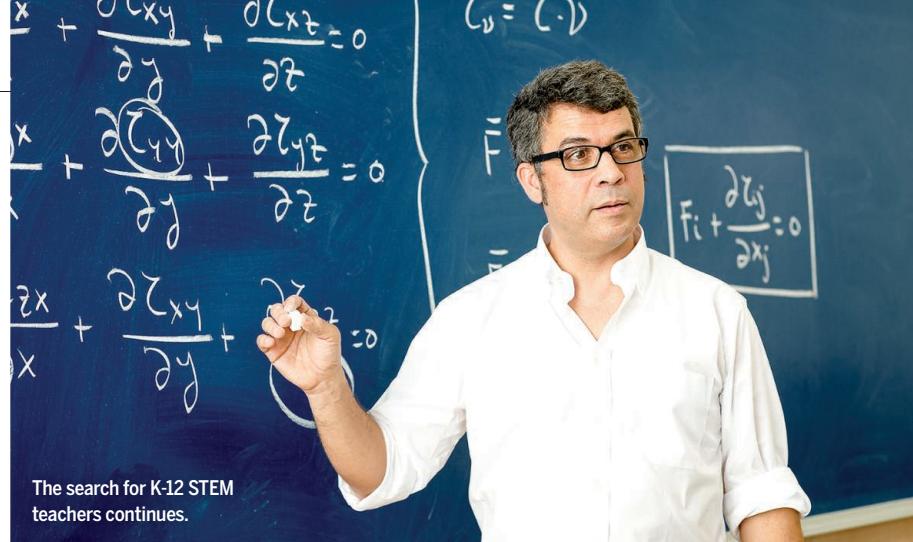
The American Physical Society, American Association of Physics Teachers, and dozens of universities across the country have worked during the past 15 years to address the severe shortage of physics teachers through the Physics Teacher Education Coalition (PhysTEC) project. PhysTEC sites establish a positive climate for teaching in the physics department; offer experiences that expose students to the rewards and challenges of teaching; build effective pathways for students to earn a degree and certification; and host a master teacher who mentors students. The result: PhysTEC sites have more than doubled the number of graduates well prepared to teach physics.

Moreover, in our experience, Noyce scholarships have provided critical financial support to PhysTEC students. Although scholarships alone are typically not enough to recruit new teachers, they are a key piece of the puzzle along with transforming disciplinary departments.

I was also troubled by the inaccurate statement that "5-year attrition rates rise to 50% or higher." Five-year retention rates for new teachers are about 70% (3), and retention rates for PhysTEC teachers are even higher (4). Propagating unfounded statements that confirm negative stereotypes about K-12 teaching is damaging and undermines efforts to improve it.

Monica Plisch

Director, Physics Teacher Education Coalition (PhysTEC), Associate Director of Education and



The search for K-12 STEM teachers continues.

Diversity, American Physical Society, College Park, MD 20740, USA. E-mail: plisch@aps.org

REFERENCES

1. J.G. Hill, "Education and Certification Qualifications of Departmentalized Public High School-Level Teachers of Core Subjects: Evidence from the 2007–08 Schools and Staffing Survey" (U.S. Department of Education, National Center for Education Statistics, Washington, DC, NCES 2011-317, 2011).
2. President's Council of Advisors on Science and Technology, "Prepare and Inspire: K-12 Education in Science, Technology, Engineering, and Math (STEM) for America's Future" (Executive Office of the President, Washington, DC, 2010); www.whitehouse.gov/administration/eop/ostp/pcast/docsreports.
3. R. Goldring, S. Taie, M. Riddles, "Teacher Attrition and Mobility: Results from the 2012–13 Teacher Follow-up Survey" (U.S. Department of Education, National Center for Education Statistics, Washington, DC, NCES 2014-077, 2014).
4. Physics Teacher Education Coalition, *PhysTEC News* 8, 1 (2014); www.phystec.org/publicity/files/fall14.pdf.

Editor's note

LAST OCTOBER, SCIENCE published an investigation into a dispute between physicist Ulf Leonhardt and the Centre for Optical and Electromagnetic Research (COER) at South China Normal University (SCNU) in Guangzhou ("Show me the money," 24 October 2014, p. 411). The story centered on Leonhardt's recruitment by COER as a part-time distinguished professor, funded by generous grants from the Chinese Central and Guangdong governments, and the premature end of his contract, when he charged that COER was not transparent about the sources of money and was spending it on research he did not know about or authorize. COER, for its part, accuses Leonhardt of bad faith and breach of contract. The dispute is acrimonious and ongoing.

Our goal in writing about the dispute was to illuminate the potential for misunderstanding and conflict inherent in these arrangements, which have become more common as China's science develops. The allegations reported in the article reflect the positions of the respective parties, and

Science does not endorse them. We regret any implication to the contrary.

Since the article appeared, we have learned that it contains several minor inaccuracies concerning the timing of events and the details of the grants and contracts. We have corrected these in the online version of the story. We have also clarified other language and removed an illustration that was open to misinterpretation. Since publishing the story, we have learned that before Leonhardt arrived in China, he received a draft of a presentation describing research planned for the Guangdong Leading Talent project. Leonhardt also took part in an oral defense in Beijing on March 21st, 2012, for the Guangdong Leading Talent project application. He and COER continue to disagree about how much advance knowledge he could have had about the research funded by his grants.

Tim Appenzeller

News Editor

A Chinese physics institute's defense

"SHOW ME THE money" (M. Hvistendahl, News Feature, 24 October 2014, p. 411) contains numerous false allegations, misinterpretations, and bias, which we must address.

Leonhardt claims he was unaware of the grant application contents. However, Leonhardt had a chance to revise a draft of the slides for the Guangdong Leading Talent Project and presented the slides to the committee. They detail the projects on transformation optics-based super-resolution imaging, cloaking, and absorbers for which funding was sought. Thus, Leonhardt should have been aware of the contents of the grant application. Casimir forces are briefly mentioned in

the Thousand Talents application because Leonhardt provided his Casimir force work (1) in a package of his selected papers when he asked the Centre for Optical and Electromagnetic Research (COER) to prepare the application for him. At that time no COER member worked on Casimir forces. Leonhardt claims he did not suggest Casimir forces until one year after August 2011, whereas the article claims that "Leonhardt planned to work on the theory of Casimir forces," which is self-contradictory.

The document provided to *Science* by Professor Jun Li explicitly connects the team slide (including potential collaborators) pictured in the article to the research directions of the Leading Talent defense, which did not include the Casimir force.

Our finances have been strictly in accordance with the approved budget and grant regulations. 790,043 RMB (\$128,462) was spent on research equipment; 208,920 RMB (\$33,971) was spent on travel for Leonhardt, his partner Jana Silberg, and their invited guests; and the unused 4,001,037 RMB (\$650,575) was returned to the grant authority. Normal salary is taxed and transferable abroad, whereas tax-free subsidies are intended to be used in China. Leonhardt made several unsuccessful attempts to

transfer subsidy money abroad while in Guangzhou from 12 to 18 January 2013. On 28 January 2013, Leonhardt e-mailed to COER, "Concerning Jana's payment, please arrange for a regular salary and give me the account number where I can return the 0.5M." This return would not have affected the net income Leonhardt would have received if he completed his 5-year contract. This 500,000 RMB (\$81,301) remains frozen in a South China Normal University (SCNU) bank account awaiting government instruction. 1,566,667 RMB (\$254,743) in pre-tax salary and housing subsidy remains in Leonhardt's possession, and he shows no intention of returning any of it. Since Leonhardt was in mainland China for only 57 days while under contract and did not fulfill the contract's terms, SCNU is taking legal action to recover a substantial portion of this money.

While we appreciate the over 20 changes, the modified article online still misrepresents our center's voice, most notably Professor Sailing He's quotes. They are out of context and based on the journalist's notes, which He did not have a chance to review and sign. When He said that Leonhardt "doesn't need to care about the details," he was referring to whether

the income was categorized as subsidies or salary.

As noted in the article, Leonhardt has waged a campaign through many Chinese agencies to defame COER. No Chinese agency has responded, because COER did nothing wrong and Leonhardt's allegations are groundless.

Chinese culture values trust between people. Unfortunately, Leonhardt betrayed our trust. We have learned some hard lessons and proposed reforms to related government agencies. SCNU has always strictly implemented relevant provisions of the national Thousand Talents Program and Guangdong Leading Talent Project, accepted supervision from upper-level departments, and thanks the community for support. SCNU will continue to improve management and service, and always welcome highly talented individuals who actively promote teaching and research development at the University.

Langping He

Deputy Dean, Centre for Optical and Electromagnetic Research and Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, Guangdong, 510006, China.
E-mail: Iphe@scnu.edu.cn

REFERENCE

- U. Leonhardt, T.G. Philbin, *New J. Phys.* **9**, 254 (2007).

LETTERS

Edited by Jennifer Sills

A path to recruiting more STEM teachers

I STRONGLY DISAGREE with the implication in the News Feature "A classroom experiment" (J. Mervis, 6 February, p. 602) that recruiting new K-12 teachers in science, technology, engineering, and mathematics (STEM) is futile and that efforts to improve education should instead focus on in-service teachers. For disciplines such as physics and chemistry, fewer than half of all classrooms are led by a teacher with a degree in the subject (1). Without efforts to improve recruitment, we cannot effectively address the severe shortage of physical science teachers with deep content knowledge. Teachers with deep content knowledge in their subject, along with mastery of content-specific pedagogical skills, are the most important factor in ensuring excellence in STEM education (2).

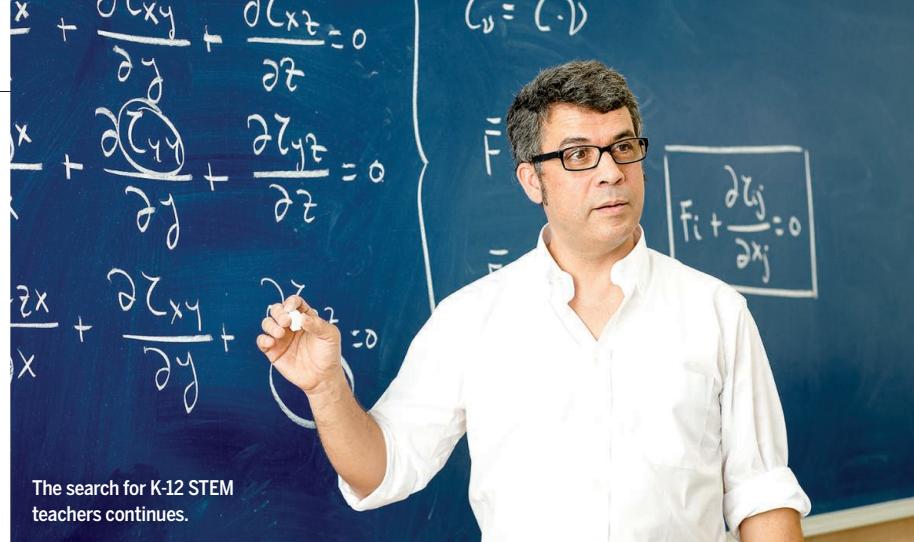
The American Physical Society, American Association of Physics Teachers, and dozens of universities across the country have worked during the past 15 years to address the severe shortage of physics teachers through the Physics Teacher Education Coalition (PhysTEC) project. PhysTEC sites establish a positive climate for teaching in the physics department; offer experiences that expose students to the rewards and challenges of teaching; build effective pathways for students to earn a degree and certification; and host a master teacher who mentors students. The result: PhysTEC sites have more than doubled the number of graduates well prepared to teach physics.

Moreover, in our experience, Noyce scholarships have provided critical financial support to PhysTEC students. Although scholarships alone are typically not enough to recruit new teachers, they are a key piece of the puzzle along with transforming disciplinary departments.

I was also troubled by the inaccurate statement that "5-year attrition rates rise to 50% or higher." Five-year retention rates for new teachers are about 70% (3), and retention rates for PhysTEC teachers are even higher (4). Propagating unfounded statements that confirm negative stereotypes about K-12 teaching is damaging and undermines efforts to improve it.

Monica Plisch

Director, Physics Teacher Education Coalition (PhysTEC); Associate Director of Education and



The search for K-12 STEM teachers continues.

Diversity, American Physical Society, College Park, MD 20740, USA. E-mail: plisch@aps.org

REFERENCES

1. J.G. Hill, "Education and Certification Qualifications of Departmentalized Public High School-Level Teachers of Core Subjects: Evidence from the 2007–08 Schools and Staffing Survey" (U.S. Department of Education, National Center for Education Statistics, Washington, DC, NCES 2011-317, 2011).
2. President's Council of Advisors on Science and Technology, "Prepare and Inspire: K-12 Education in Science, Technology, Engineering, and Math (STEM) for America's Future" (Executive Office of the President, Washington, DC, 2010); www.whitehouse.gov/administration/eop/ostp/pcast/docsreport.
3. R. Goldring, S. Taie, M. Riddles, "Teacher Attrition and Mobility: Results from the 2012–13 Teacher Follow-up Survey" (U.S. Department of Education, National Center for Education Statistics, Washington, DC, NCES 2014-077, 2014).
4. Physics Teacher Education Coalition, *PhysTEC News* 8, 1 (2014); www.phystec.org/publicity/files/fall14.pdf.

Editor's note

LAST OCTOBER, SCIENCE published an investigation into a dispute between physicist Ulf Leonhardt and the Centre for Optical and Electromagnetic Research (COER) at South China Normal University (SCNU) in Guangzhou ("Show me the money," 24 October 2014, p. 411). The story centered on Leonhardt's recruitment by COER as a part-time distinguished professor, funded by generous grants from the Chinese Central and Guangdong governments, and the premature end of his contract, when he charged that COER was not transparent about the sources of money and was spending it on research he did not know about or authorize. COER, for its part, accuses Leonhardt of bad faith and breach of contract. The dispute is acrimonious and ongoing.

Our goal in writing about the dispute was to illuminate the potential for misunderstanding and conflict inherent in these arrangements, which have become more common as China's science develops. The allegations reported in the article reflect the positions of the respective parties, and

Science does not endorse them. We regret any implication to the contrary.

Since the article appeared, we have learned that it contains several minor inaccuracies concerning the timing of events and the details of the grants and contracts. We have corrected these in the online version of the story. We have also clarified other language and removed an illustration that was open to misinterpretation. Since publishing the story, we have learned that before Leonhardt arrived in China, he received a draft of a presentation describing research planned for the Guangdong Leading Talent project. Leonhardt also took part in an oral defense in Beijing on March 21st, 2012, for the Guangdong Leading Talent project application. He and COER continue to disagree about how much advance knowledge he could have had about the research funded by his grants.

Tim Appenzeller

News Editor

A Chinese physics institute's defense

"SHOW ME THE money" (M. Hvistendahl, News Feature, 24 October 2014, p. 411) contains numerous false allegations, misinterpretations, and bias, which we must address.

Leonhardt claims he was unaware of the grant application contents. However, Leonhardt had a chance to revise a draft of the slides for the Guangdong Leading Talent Project and presented the slides to the committee. They detail the projects on transformation optics-based super-resolution imaging, cloaking, and absorbers for which funding was sought. Thus, Leonhardt should have been aware of the contents of the grant application. Casimir forces are briefly mentioned in

the Thousand Talents application because Leonhardt provided his Casimir force work (1) in a package of his selected papers when he asked the Centre for Optical and Electromagnetic Research (COER) to prepare the application for him. At that time no COER member worked on Casimir forces. Leonhardt claims he did not suggest Casimir forces until one year after August 2011, whereas the article claims that "Leonhardt planned to work on the theory of Casimir forces," which is self-contradictory.

The document provided to *Science* by Professor Jun Li explicitly connects the team slide (including potential collaborators) pictured in the article to the research directions of the Leading Talent defense, which did not include the Casimir force.

Our finances have been strictly in accordance with the approved budget and grant regulations. 790,043 RMB (\$128,462) was spent on research equipment; 208,920 RMB (\$33,971) was spent on travel for Leonhardt, his partner Jana Silberg, and their invited guests; and the unused 4,001,037 RMB (\$650,575) was returned to the grant authority. Normal salary is taxed and transferable abroad, whereas tax-free subsidies are intended to be used in China. Leonhardt made several unsuccessful attempts to

transfer subsidy money abroad while in Guangzhou from 12 to 18 January 2013. On 28 January 2013, Leonhardt e-mailed to COER, "Concerning Jana's payment, please arrange for a regular salary and give me the account number where I can return the 0.5M." This return would not have affected the net income Leonhardt would have received if he completed his 5-year contract. This 500,000 RMB (\$81,301) remains frozen in a South China Normal University (SCNU) bank account awaiting government instruction. 1,566,667 RMB (\$254,743) in pre-tax salary and housing subsidy remains in Leonhardt's possession, and he shows no intention of returning any of it. Since Leonhardt was in mainland China for only 57 days while under contract and did not fulfill the contract's terms, SCNU is taking legal action to recover a substantial portion of this money.

While we appreciate the over 20 changes, the modified article online still misrepresents our center's voice, most notably Professor Sailing He's quotes. They are out of context and based on the journalist's notes, which He did not have a chance to review and sign. When He said that Leonhardt "doesn't need to care about the details," he was referring to whether

the income was categorized as subsidies or salary.

As noted in the article, Leonhardt has waged a campaign through many Chinese agencies to defame COER. No Chinese agency has responded, because COER did nothing wrong and Leonhardt's allegations are groundless.

Chinese culture values trust between people. Unfortunately, Leonhardt betrayed our trust. We have learned some hard lessons and proposed reforms to related government agencies. SCNU has always strictly implemented relevant provisions of the national Thousand Talents Program and Guangdong Leading Talent Project, accepted supervision from upper-level departments, and thanks the community for support. SCNU will continue to improve management and service, and always welcome highly talented individuals who actively promote teaching and research development at the University.

Langping He

Deputy Dean, Centre for Optical and Electromagnetic Research and Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, Guangdong, 510006, China.
E-mail: Iphe@scnu.edu.cn

REFERENCE

- U. Leonhardt, T.G. Philbin, *New J. Phys.* **9**, 254 (2007).

BOOKS et al.



Red knots and other shorebirds ready for the final leg of their annual spring migration.

COASTAL ECOLOGY

Migration, interrupted

The imperiled existence of the rufa red knot

By Nigel A. Clark

In order for the shorebirds known as red knots to survive their annual migration from the Southern Hemisphere to the species' Arctic breeding grounds, they rely on a series of sites along the way to provide enough food for their stay and to fuel their onward flight. If just one of these sites is compromised, the species itself will fail.

Unfortunately for the knots, many of these vital sites are vast, intertidal flats that humans have come to see as areas ripe for development and exploitation. This is especially true for the main spring stopover sites that knot populations on each continent rely on, including areas in the Delaware Bay in the Americas, the Wadden Sea in Europe, and the Yellow Sea in East Asia. The marvel of their long migration and the beauty of their synchronously weaving flocks have inspired many to try to ensure that there is a place on the planet for the knots for years to come.

In *The Narrow Edge*, Deborah Cramer follows the migration route of the American rufa knot from the southern tip of South America to the high Arctic, where the ground thaws just long enough for them to raise their chicks before winter sets

in again. The book is perhaps more about people than birds, as Cramer tries to understand how and why humans and birds have come into conflict. She explores what motivates developers, fishermen, and conservationists and explores the ways scientists are working to understand how we can live together on an increasingly crowded planet. Throughout the book, Cramer explores the history that has led to the tenuous existence of the knot, as well as the tangled ecological web that it, and we, are part of.

The book is written from a conservationist's viewpoint, but Cramer also tries to understand the driving forces that, often inadvertently, make it more difficult for the knots to survive. For example, she recounts how, until recently, the salt marshlands in Argentina's Río Gallegos estuary were being sold off to developers and used as a municipal waste dump. As the population of the nearby city expanded over the past half-century, the number of knots observed each year in the estuary dropped precipitously. Now, as a result of the efforts of scientists Silvia Ferrari and Carlos Albrieu, Río Gallegos has two protected areas where further development of the marsh is prohibited. The city has also made an asset of the marsh, creating a visitor center and bird-viewing areas to educate locals and tourists about the value of conservation efforts. This has come too late for the knots, which no longer stop there, but has proved vital for

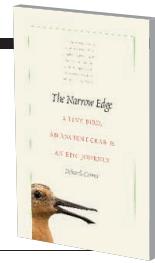
The Narrow Edge

A Tiny Bird, an Ancient Crab, and an Epic Journey

Deborah Cramer

Yale University Press, 2015.

303 pp.



many other species of shorebirds. Perhaps, in time, the knots will return.

Many knots that migrate from the Americas make their last stop to refuel in the Delaware Bay before heading to the Arctic. Here, they gorge on horseshoe crab eggs that are washed out of the sand, where they are laid in May of each year. The crabs and, hence, the birds have declined over the past two decades. It would be easy to blame modern fishermen for overexploiting the crabs, but Cramer shows that the tendency toward aggressive harvesting is consistent with historical practices. During the mid-to-late 1800s, horseshoe crabs were fed to pigs or ground into fertilizer in specially built factories. More than a million crabs were harvested from a single mile of beach in 1857, and over 4 million were taken from all of Delaware Bay in 1880. By the late 1800s, the horseshoe crab population in the Delaware Bay was almost nonexistent. It slowly recovered over the next century but crashed again when fishermen began collecting the crabs to use as bait for conch and eel fishing in the 1990s. With a history of exploitation, it is not surprising that the fishermen saw the crabs as a resource to be taken advantage of, rather than a population to be managed sustainably. Since 1998, there have been progressively more severe harvest restrictions on the crabs but, with a 10-year life cycle, we are just beginning to see signs of recovery.

If the horseshoe crab had gone extinct, the knots would not have been the only ones to suffer. In the early 1950s, the physician Frederick Bang discovered that the crab's blue blood clots in the presence of Gram-negative bacteria. By the mid-1980s, limulus amebocyte lysate (LAL), made from the blood of horseshoe crabs, had become an invaluable aid in the fight against infection in humans.

By focusing on the plight of one flagship species and the people who have dedicated their lives to understanding how to protect it, this book shows how conservation efforts are critical to maintaining coastal biodiversity. It also offers important lessons and strategies that may be implemented for the protection and preservation of other species.

10.1126/science.aaa8157

EXHIBITION

Hard evidence

Exploring the origins and evolution of forensic science

By Andrew Robinson

The rise of forensic science is said to date from the 19th century: the era of a reliable chemical test for arsenic poisoning, Michael Faraday's explanation of combustion, the introduction of fingerprinting, and Arthur Conan Doyle's fictional detective, Sherlock Holmes. But its beginnings actually lie far earlier, as chronicled in the wide-ranging Wellcome exhibition *Forensics: The Anatomy of Crime*.

The exhibit includes a Chinese handbook for coroners, published in 1247, that is still regarded as one of the crucial volumes on postmortem methods. *The Washing Away of Wrongs* by Song Ci includes the absorbing case of a man stabbed to death by the roadside. After comparing the slash marks on the body with his own experimental slash marks on a cow's carcass using various blades, the coroner concluded that the murder weapon was a sickle. He lined up all of the neighborhood's 70 adults with their sickles—free of any obvious blood traces—at their feet. Within seconds, a fly landed on the local moneylender's sickle, then another,

and another. According to the handbook, the moneylender now "knocked his head on the ground" and confessed to the murder. In the words of crime novelist Val McDermid, who wrote the gripping nonfiction book that accompanies the Wellcome exhibition (1), he had been betrayed by "the insect informers humming quietly at his feet." McDermid's book is based on revealing interviews with practicing experts in all fields of forensic science, some of whom also feature in the exhibition's video interviews.

Violent crime and the motives of killers have always fascinated the public, even those who avoid crime fiction. Indeed, the English word morgue derives from the French word *morguer*, meaning roughly "to peer." The Paris Morgue, an institution established in the early 1800s, was open to the public, who could peer through windows at unidentified corpses. The hope was that this would promote the identification of the unknown, but it quickly became a popular destination, attracting foreign visitors and even mothers with young children. A century later, the morgue was closed to idle gazers "out of concern for public morality," notes a caption that accompanies one of the illustrations on display.

The exhibition necessarily eschews human remains, except for some skulls and bones, including some that were exhumed

Forensics:
The Anatomy of Crime

Wellcome Collection,
London.
26 February to 21 June 2015.
<http://wellcomecollection.org/forensics>



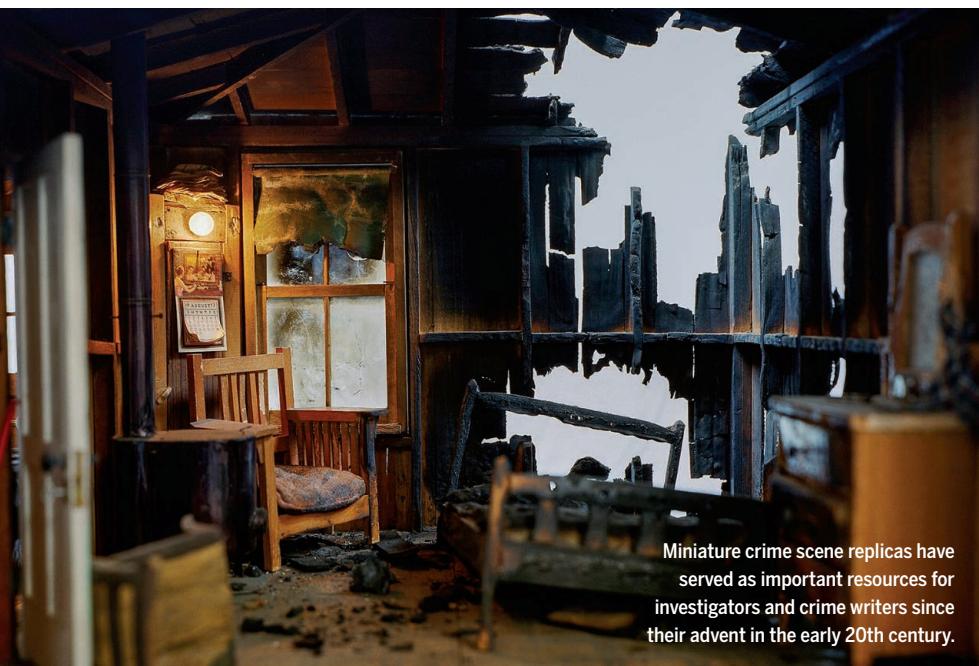
from mass graves during recent massacres. Yet there is plenty to instruct, discomfort, appall, and entertain the visitor. For example, in a deliberately dark corner of "The Morgue," next to an illuminated ceramic postmortem table, one can don a pair of headphones and listen to the soundtrack of a real human autopsy recently conducted in murder-ridden Mexico. This part of the exhibition is not for the squeamish.

One of the most thought-provoking exhibits is far from the most gruesome, although it does have something of the appeal of a Hitchcock horror film. At the start of the exhibition, one may peer into the rooms of what looks like a large doll's house, in which a murder has been committed. The house is one of a series of models known as "Nutshell Studies of Unexplained Death," created in the 1930s and 1940s by Frances Glessner Lee, the wealthy Chicago heiress who founded the Harvard School of Legal Medicine in 1931. Both police officers and crime writers (including Erle Stanley Gardner, the creator of the Perry Mason series of detective stories) have long used Nutshell models as a training tool. In 2012, they became the subject of a documentary film, *Of Dolls and Murder* (2), a clip from which is shown next to the model. Disquieting music from the film can be heard throughout the early portion of the exhibition.

At the end, three near-life-sized contemporary photographs by the artist Taryn Simon serve as a reminder of one of the benefits of forensic science. Shot as part of the Innocence Project, an initiative dedicated to exonerating wrongfully convicted individuals and reforming the criminal justice system in the United States, each image shows a falsely imprisoned man posing after his release at the scene of his alleged crime. In the United States alone, more than 300 convicted people have been exonerated by DNA evidence and released from incarceration since the organization's inception in 1992. Without sensitive and sophisticated forensic investigations, there would likely be more such miscarriages of justice.

REFERENCES

- V. McDermid, *Forensics: The Anatomy of Crime* (Profile Books Ltd., London, 2014).
- S. Marks, director. *Of Dolls and Murder* (2012).



Miniature crime scene replicas have served as important resources for investigators and crime writers since their advent in the early 20th century.

Ecosystem services lost to oil and gas in North America

Net primary production reduced in crop and rangelands

By Brady W. Allred,^{1,*} W. Kolby Smith,^{1,2}
Dirac Twidwell,³ Julia H. Haggerty,⁴
Steven W. Running,¹ David E. Naugle,¹
Samuel D. Fuhlendorf⁵

Advanced technologies in oil and gas extraction coupled with energy demand have encouraged an average of 50,000 new wells per year throughout central North America since 2000. Although similar to past trends (see the graph, this page), the space and infrastructure required for horizontal drilling and high-volume hydraulic fracturing are transforming millions of hectares of

POLICY the Great Plains into industrialized landscapes, with drilling projected to continue (1, 2). Although this development brings economic benefits (3) and expectations of energy security, policy and regulation give little attention to trade-offs in the form of lost or degraded ecosystem services (4). It is the scale of this transformation that is important, as accumulating land degradation can result in continental impacts that are undetectable when focusing on any single region (5). With the impact of this transformation on natural systems and ecosystem services yet to be quantified at broad extents, decisions are being made with few data at hand (see the graph, this page).

We provide a first empirical analysis to advance beyond common rhetoric and speculation of oil and gas development (6), combining high-resolution satellite data of vegetation dynamics with industry data and publicly available data of historical and present-day oil and gas well locations for central North America. In addition to this broad-scale assessment of satellite-derived net primary production (NPP), a fundamental measure of a region's ability to provide ecosystem services (7), we also evaluate patterns of land-use change and

water use. Before this work, little has been done in examining these types of data and their relations with ecosystem services at broad scales.

ECOSYSTEM SERVICE TRADE-OFFS. NPP is the amount of carbon fixed by plants and accumulated as biomass. It is a fundamental and supporting ecosystem service that is the basis for all life on Earth (8). As such, the dynamics of NPP affect regional ability to provide a host of other essential ecosystem services (e.g., food production, biodiversity, wildlife habitat), which makes it a robust metric for broad evaluation of ecosystem services. Oil and gas activity reduces NPP through direct removal of vegetation to construct oil pads, roads, and so on.

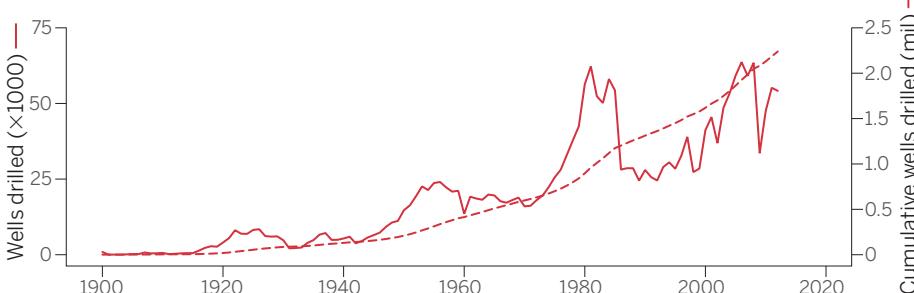
These satellite-derived measurements of

of carbon per year, we convert to equivalent biomass-based measurements to provide context and discussion.

We estimate that vegetation removal by oil and gas development from 2000 to 2012 reduced NPP by ~4.5 Tg of carbon or 10 Tg of dry biomass across central North America (see the chart on page 402, left). The total amount lost in rangelands is the equivalent of approximately five million animal unit months (AUM; the amount of forage required for one animal for 1 month), which is more than half of annual available grazing on public lands managed by the U.S. Bureau of Land Management (BLM). The amount of biomass lost in croplands is the equivalent of 120.2 million bushels of wheat, ~6% of the wheat produced in 2013 within the region and 13% of the wheat exported by the United States (see SM for equivalency calculations).

The loss of NPP is likely long-lasting and potentially permanent, as recovery or reclamation of previously drilled land has not kept pace with accelerated drilling (SM). This is not surprising because current reclamation practices vary by land ownership and governing body, target only limited portions of the energy landscape, require

Oil and gas wells drilled in central Canada and USA



The number of oil and gas wells drilled within central provinces of Canada and central U.S. states 1900–2012. Canadian provinces: Alberta, Manitoba, and Saskatchewan. U.S. states: Colorado, Kansas, Montana, Nebraska, New Mexico, North Dakota, Oklahoma, South Dakota, Texas, Utah, and Wyoming. See SM.

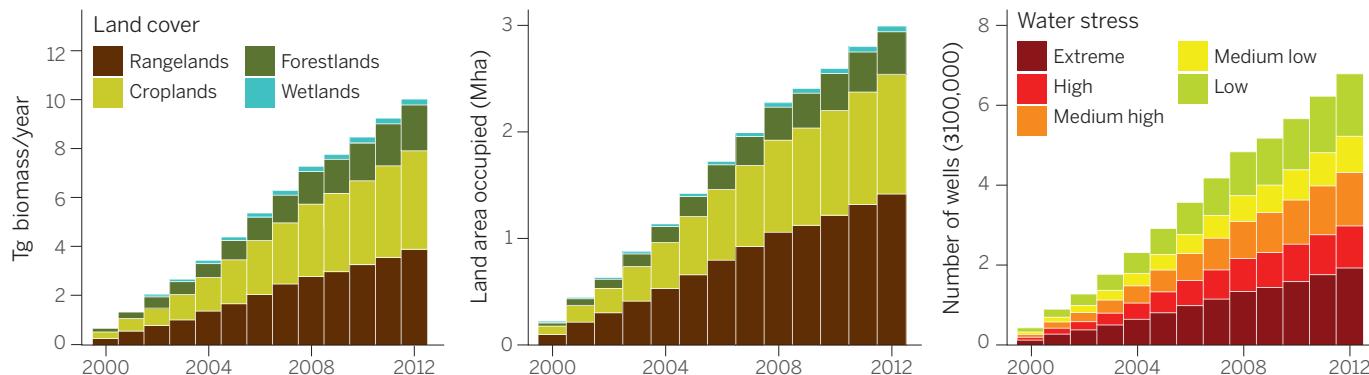
NPP began in 2000 and are produced annually; they capture interannual dynamics (9). To match the spatial scale of NPP measurement (~1 km²), we determined annual density of oil and gas activity at the same resolution and estimated annual loss of NPP relative to such densities. Direct loss of vegetation resulting from oil and gas activity was validated at medium and fine spatial scales (~250 m² and 30 m², respectively) by examining vegetation and disturbance trends before and after drilling [see supplementary materials (SM)]. We categorized annual reductions in NPP relative to land cover type (e.g., cropland and rangeland). As NPP is measured in grams

substantial funding and implementation commitments, and are often not initiated until the end of life of a well (10). Barring changes from existing trends and practices, it is likely that NPP loss and its effects (i.e., further loss of forage) will continue to parallel drilling trends and, potentially, may create unforeseen conflicts among agriculture, conservation, and energy.

Additional ecosystem functions, including wildlife habitat and landscape connectivity, are arguably as important as NPP. We estimate that the land area occupied by well pads, roads, and storage facilities built from 2000 to 2012 is ~3 million ha, the equivalent land area of three Yellowstone

¹College of Forestry and Conservation, University of Montana, Missoula, MT 59812, USA. ²Institute on the Environment, University of Minnesota, St. Paul, MN 55108, USA. ³Department of Agronomy and Horticulture, University of Nebraska—Lincoln, Lincoln, NE 68583, USA. ⁴Department of Earth Sciences, Montana State University, Bozeman, MT 59717, USA. ⁵Department of Natural Resource Ecology and Management, Oklahoma State University, Stillwater, OK 74078, USA.
^{*}Corresponding author. brady.allred@umontana.edu

Oil and gas development impacts on ecosystem services



Cumulative impacts of oil and gas development on ecosystem services in central North America 2000–2012. (Left) Reduction in NPP (biomass), per land cover type. (Middle) Land area occupied, per land cover type. (Right) Number of wells in water-stressed regions (22). See SM.

National Parks (see the chart, middle). Although small in comparison with the total land area of the continent, this important land use is not accounted for (11) and creates additional pressures for conserving rangelands and their ecosystem functions. The distribution of this land area has negative impacts: increasing fragmentation that can sever migratory pathways, alter wildlife behavior and mortality, and increase susceptibility to ecologically disruptive invasive species (6, 12). As competition for arable land intensifies because of food and bioenergy demand (13), oil and gas may further expand into native rangelands.

The hydraulic fracturing technology underlying the current expansion of oil and gas drilling in the region has profound implications for hydrological, water-quality, and water-use regimes. High-volume hydraulic fracturing uses 8000 to 50,000 m³ of water per well for the initial fracturing event (14), which results in 7187 to 33,903 million m³ for wells drilled across this region during 2000 to 2012 (see SM). Nearly half of wells drilled in this time period occurred in already highly or extremely water-stressed regions (see the chart, right). As refracturing becomes more common to yield greater production, oil and gas development adds to an already fraught competition among agriculture, aquatic ecosystems, and municipalities for water resources, in addition to concerns of water quality (15).

AVOIDING BROAD-SCALE LOSS. The capacity for insight into land-use decisions has improved substantially since the last major episode of widespread land-use change across the Great Plains. In the early 20th century, rapid agricultural expansion and widespread displacement of native vegetation reduced the resilience of the region to drought, ul-

timately contributing to the Dust Bowl of the 1930s. It took catastrophic disruption of livelihoods and economies to trigger policy reforms that addressed environmental and social risks of land-use change.

Fortunately, data and information are now far less of a barrier in understanding and addressing continental and cumulative impacts. However, the scale and focus of most land-use decision-making discourages comprehensive assessment of trade-offs implied in oil and gas development (16, 17). Recent planning efforts by U.S. federal management agencies demonstrate potential to balance demand for energy development with the need to protect other values (18, 19), but the scope is limited to lands under federal jurisdiction. About 90% of oil and gas infrastructure in this region occurs on private land (United States only; see SM). Provinces, states, and municipalities that permit the majority of oil and gas development lack the capacity and mandate to address continental or regional consequences that transcend political boundaries; this lack leads to fragmented and piecemeal policies (16, 20).

Decision-makers and scientists must work together to ensure that the best available information guides development of policies at the water-energy-food nexus (21). Traditional laws and regulations may have limited application, as oil and gas can be exempt from key environmental regulations (20), or such regulations isolate features of systems—e.g., a single species—while failing to capture interrelated impacts. Active synthesis and consolidation of data will improve accessibility and monitoring. Integration of these data into land-use planning and policy across scales and jurisdictions is necessary to achieve energy policies that minimize ecosystem service losses.

REFERENCES AND NOTES

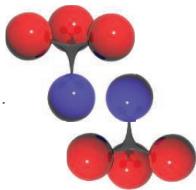
- International Energy Agency, "World energy outlook 2013" (IEA, Paris, 2013).
- U.S. Energy Information Administration, "International energy outlook" (EIA, Department of Energy, Washington, DC, 2013).
- J. G. Weber, *Energy Econ.* **34**, 1580 (2012).
- N. Butt *et al.*, *Science* **342**, 425 (2013).
- D. P. C. Peters *et al.*, *Front. Ecol. Environ.* **6**, 229 (2008).
- D. Sontag, R. Gebeloff, *New York Times*, 22 November 2014; www.nytimes.com/interactive/2014/11/23/us/north-dakota-oil-boom-downside.html.
- N. F. Jones, L. Pejchar, *PLOS ONE* **8**, e81391 (2013).
- Millennium Ecosystem Assessment, *Ecosystems and Human Well-Being* (Island Press, Washington, DC, 2005).
- M. Zhao, S. W. Running, *Science* **329**, 940 (2010).
- A. L. Mitchell, E. A. Casman, *Environ. Sci. Technol.* **45**, 9506 (2011).
- M. A. Drummond, R. Auch, "Land cover trends in the United States Great Plains" (U.S. Geological Survey, 2012); http://bit.ly/1PppT6a.
- J. M. Northrup, G. Wittemyer, *Ecol. Lett.* **16**, 112 (2013).
- W. K. Smith, C. C. Cleveland, S. C. Reed, N. L. Miller, S. W. Running, *Environ. Sci. Technol.* **46**, 3536 (2012).
- A. Vengosh, R. B. Jackson, N. Warner, T. H. Darrah, A. Kondash, *Environ. Sci. Technol.* **48**, 8334 (2014).
- R. B. Jackson *et al.*, *Annu. Rev. Environ. Resour.* **39**, 327 (2014).
- B. G. Rabe, *Environ. Sci. Technol.* **48**, 8369 (2014).
- J. C. Jacquet, *Environ. Sci. Technol.* **48**, 8321 (2014).
- Secretary Salazar on wildlife protections in the petroleum reserve in Alaska [press release] (BLM, 2013); www.blm.gov/wo/st/en/info/newsroom/2013/february/nr_02_21_2013.html.
- "Interior Department final rule to support safe, responsible hydraulic fracturing activities on public and tribal lands" [press release] (U.S. Department of the Interior, Washington, DC, 2015); http://on.doi.gov/1HbC57V.
- B. Warner, J. Shapiro, *Publius* **43**, 474 (2013).
- Food and Agriculture Organization of the United Nations, "The water-energy-food nexus" (FAO, Rome, 2014); www.fao.org/hr/water/docs/FAO_nexus_concept.pdf
- F. M. Gassert *et al.*, Aqueduct Global Maps 2.0, Working paper (World Resources Institute, Washington, DC, 2013); http://bit.ly/1y9HSb5.

ACKNOWLEDGMENTS

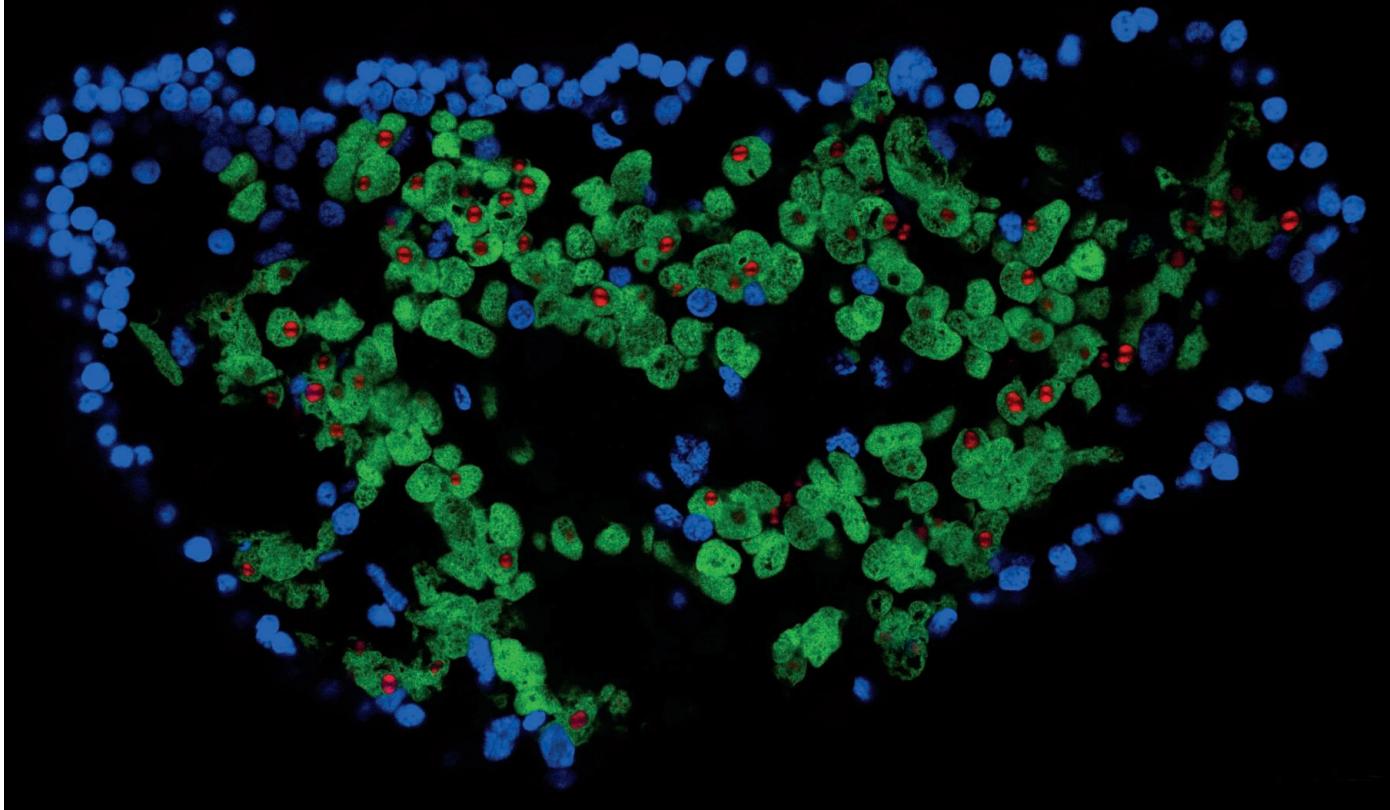
We thank A. Ballantyne and M. Nie for comments. This research was conducted with support from NASA Earth Science Division.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/401/suppl/DC1



PERSPECTIVES



Mutual dependence. Formerly free-living individuals can become physically, genetically, and metabolically integrated, such as shown in the cross section of *Paracatenula* flatworms colonized by intracellular *Candidatus Riegeria* symbionts (green). Host nuclei are shown in blue and storage compounds in red.

EVOLUTIONARY BIOLOGY

Evolving new organisms via symbiosis

When and how do symbiotic partnerships become new, integrated organisms?

By E. Toby Kiers¹ and Stuart A. West²

Sybiotic partnerships are a major source of evolutionary innovation. They have driven rapid diversification of organisms, allowed hosts to harness new forms of energy, and radically modified Earth's nutrient cycles. The application of next-generation sequencing and advanced microscopic techniques has revealed not only the ubiquity of symbiotic partnerships, but the extent to

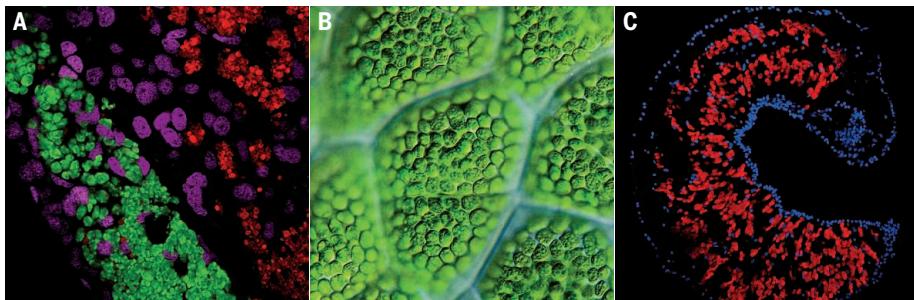
which partnerships can become physically, genetically, and metabolically integrated (1). When and why does this integration of once free-living organisms happen?

Many insects harbor endosymbionts—bacteria that live within the host's cells (see the figure, panel A). Although separate organisms, they function as a metabolic unit. Such dependency can drive extreme genomic integration of host and symbiont at many levels. For example, species of mealy bugs depend on bacterial endo-

symbionts for nutrient provisioning, and the endosymbiont can in turn harbor its own endosymbiont (2). Patterns of symbiont within symbiont dependencies occur across animals, as well as plant hosts, which use photosynthesizing plastids as a source of energy (see the figure, panel B) (3). Endosymbionts can even speciate within their hosts, as has been found in *Cicada* insects (4).

One potential outcome of host-symbiont integration is a reduction in symbiont

Major transitions



Not major transitions



Major transitions. (A to C) Symbiotic partnership resulting in major transitions in individuality: (A) endosymbionts *Candidatus Hodgkinia* (red) and *Candidatus Sulcia* (green) in a cicada host, with insect cell nuclei shown in magenta; (B) photosynthetic plastids of fern cells; (C) *Paracatenula* flatworm with intracellular *C. Riegeria* symbionts (red) and host nuclei (blue). (D to F) Symbiotic partnership where major transitions in individuality have not taken place: (D) giant marine tubeworm, *Riftia*, which is obligately dependent on nutritional symbionts gained during its larval stage; (E) nodules of legume *Lathyrus japonicus* housing N₂-fixing rhizobial symbionts; (F) bioluminescent bobtail squid.

genome size. A species of leaf hopper, *Macrostelus quadrilineatus*, harbors the endosymbiont *Nasutia deltocephalinicola*, the smallest bacterial genome sequenced to date (5). Similarities in genome size, coding capacity, and the ability to import proteins produced by the host has driven a debate as to whether such endosymbionts are functionally equivalent to host organelles (6). Organelles evolved when free-living proteobacterial and cyanobacterial ancestors became incorporated into host cells, ultimately forming the mitochondria and plastids that power eukaryotes. Although the debate has largely focused on the genetic and cellular differences between endosymbionts and organelles, it raises a larger evolutionary question: When do we expect symbiotic partnerships to evolve into new, integrated organisms?

Maynard Smith and Szathmáry's major transitions approach provides a framework for addressing this question (7). It focuses on cases where groups of individuals that could previously replicate independently cooperate to form a new, more complex organism. Examples include the formation of chromosomes from the simplest replicators, of the eukaryotic cell from archaeal

and eubacterial cells, and of multicellular organisms through cellular cooperation. By asking what conditions favor these transitions, the framework emphasizes comparisons across different levels of biological organization (8).

The elegance of the major transitions framework is its simplicity. It argues that the same problem—how to overcome the selfish interests of individuals to form mutually dependent cooperative groups—has arisen and been solved at several crucial moments in history across all orders of life. At the same time, it recognizes that transitions in individuality are rare and require strict conditions: Partner interests need to be aligned and the benefits of more integrated cooperation must lead to mutual dependence.

How can group conflict be eliminated and loss of autonomy become favorable? Symbiotic partnerships involve two levels of potential conflict: between symbiont and hosts and among symbionts sharing a host. Hosts have evolved sophisticated mechanisms to manage their microbes, but the resulting symbioses do not necessarily eliminate group conflict (9). Furthermore, although repression of competition is necessary, it is not sufficient to drive a major evolutionary transition, which requires mutual dependence. This can happen when symbionts access new forms of energy for

their hosts, as with mitochondria, and the fitness of partners becomes so intertwined that obligate dependency is favored (10).

Some symbioses are good candidates for major transitions. For example, *Paracatenula* flatworms are colonized by symbionts that harness chemical sources of energy (see the lead photo, page 392). The hosts have evolved such a level of symbiont dependency that they have lost mouths and digestive tracts (see the figure, panel C). In turn, the symbionts have reduced genomes and are passed directly from parent to offspring (vertical transmission) (11). As with some insect endosymbioses (see the figure, panel A), selection has created organism-level adaptations, such as complementary genome modifications, that suggest successful and ongoing major transitions.

To understand when and why new organisms evolve via symbiosis, it is useful to also look at cases where major transitions have not been made, for example when there are asymmetries in dependence. The giant marine tubeworm *Riftia* lacks a digestive system as an adult and depends on a nutritional symbiont gained during the larval stage (12) (see the figure, panel D). Whereas the host has evolved a highly specialized organ to house this symbiont, the bacterial partner retains a free-living stage, is transmitted horizontally, and has not experienced major genome reduction. This suggests that although the partnership provides benefits, there is sufficient conflict or favorable options outside the host to select against symbiont integration.

Other examples where major transitions have not been made include legumes, which are provided with nitrogen by their rhizobia symbionts (see the figure, panel E), and squid, which obtain light for camouflage from bioluminescent bacteria (see the figure, panel F). Both hosts acquire their bacterial symbionts directly from the environment, relying on complex signaling and coordinated molecular pathways to initiate symbiotic development. Yet despite strong coevolutionary histories, partners retain autonomy as individuals (13, 14). The major transitions framework suggests that this is because either strict mutual dependence is not beneficial or there is sufficient conflict between partners or among symbionts. Thus, even when coevolution results in intricate cross-talk and specialized structures to house symbionts, this coordination does not necessarily imply a major transition to a new level of organism.

Major transitions research suggests that the mode of transmission is key to which symbioses form new organisms. When the bacterial partner is acquired directly from the environment, as in the giant tubeworm,

¹Institute of Ecological Sciences, Vrije Universiteit, 1081 HV Amsterdam, Netherlands. ²Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. E-mail: toby.kiers@vu.nl

squid, and legume symbioses, the host is working simultaneously with multiple genotypes. This allows hosts to exploit a wider range of environmental conditions, but also increases the potential for conflict among competing symbionts (15). To help manage conflict, specific structures in hosts (such as crypts in squid) effectively separate different symbiont genotypes into individual chambers (13). In contrast, vertical transmission of symbionts from parent to offspring, as in *Paracatenula* flatworms, can lead to a high relatedness between the symbionts within a host (15), linking the fitness of the symbiont to host performance, reducing conflict, and selecting for integration.

Ecological context can also help to predict when and why transitions have been made (8). When benefits of partnerships vary with environmental context, mutual dependency is less likely to evolve. For example, in the legume-rhizobia symbiosis, the reliance on the bacterial partner decreases in high-nitrogen environments (14). Here, strict dependency can be costly, making a major transition to a single, higher-level organism precarious and unlikely if benefits are not absolute.

The advantage of the major transitions framework is that it emphasizes different questions from the mechanistic ones currently being asked. Interplay between evolutionary theory and genomic research will allow us to understand the evolution of organismal complexity within a single, unified framework. ■

REFERENCES AND NOTES

1. J. Archibald, *One Plus One Equals One* (Oxford Univ. Press, Oxford, 2014).
2. F. Husnik et al., *Cell* **153**, 1567 (2013).
3. P.J. Keeling, *Annu. Rev. Plant Biol.* **64**, 583 (2013).
4. J.T. Van Leuven, R.C. Meister, C. Simon, J.P. McCutcheon, *Cell* **158**, 1270 (2014).
5. G.M. Bennett, N.A. Moran, *Genome Biol. Evol.* **5**, 675 (2013).
6. J.P. McCutcheon, P.J. Keeling, *Curr. Biol.* **24**, R654 (2014).
7. J. Maynard Smith, E. Szathmáry, *The Major Transitions in Evolution* (Oxford Univ. Press, Oxford, 1995).
8. A.F.G. Bourke, *Principles of Social Evolution* (Oxford Univ. Press, Oxford, 2011).
9. J.L. Sachs, R.G. Skophamer, J.U. Regus, *Proc. Natl. Acad. Sci. U.S.A.* **108** (suppl. 2), 10800 (2011).
10. N. Dubilier, C. Bergin, C. Lott, *Nat. Rev. Microbiol.* **6**, 725 (2008).
11. H.R. Gruber-Vodicka et al., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12078 (2011).
12. A.D. Nussbaumer, C.R. Fisher, M. Bright, *Nature* **441**, 345 (2006).
13. M. McFall-Ngai, *PLOS Biol.* **12**, e1001783 (2014).
14. S.S. Porter, E.L. Simms, *Ecol. Lett.* **17**, 1121 (2014).
15. M. Bright, S. Bulgheresi, *Nat. Rev. Microbiol.* **8**, 218 (2010).

ACKNOWLEDGMENTS

We thank N. Ruby, N. Dubilier, J. Mills, and J. McCutcheon for critical insights and inspiration. Supported by the European Research Council and the Netherlands Organisation for Scientific Research.

GEOCHEMISTRY

Biogeochemical tales told by isotope clumps

Molecules with two or more heavy isotopes provide insights into diverse biological and geological phenomena

By Benjamin H. Passey

How do you take a dinosaur's temperature, reconstruct the elevation histories of Earth's great mountain ranges, probe the workings of photosynthesis, and confirm biological origins of a greenhouse gas? Increasingly, the answer lies in clumps. Clumped isotope geochemistry (1) is the latest branch of stable isotope geochemistry, the field that illuminated the Pleistocene glacial-interglacial cycles (2), the oxygenation of Earth's atmosphere some 2.5 billion years ago (3), and the enigmatic presence of grass in hominid diets (4). In this issue, Yeung *et al.* (page 431) (5) and Wang *et al.* (page 428) (6) describe clumped isotope effects that, among other things, can serve as tracers of biological versus abiological origins of gases.

Traditional isotope geochemistry is concerned with ratios of heavy to light isotopes. In contrast, clumped isotope geochemistry looks at occurrences of two or more heavy isotopes in the same molecule, such as ^{18}O , $^{13}\text{C}^{18}\text{O}^{16}\text{O}$, and $^{13}\text{CH}_3\text{D}$. In its decade of existence, clumped isotope geochemistry has focused on the thermodynamic aspects—particularly the temperatures of molecular and mineral synthesis—recorded in isotopic clumps. Hence, dinosaur body temperatures have been estimated from clumped carbonate ions ($^{13}\text{C}^{18}\text{O}^{16}\text{O}_2^-$) in fossil teeth (7), and altitudes of ancient mountain ranges from $^{13}\text{C}^{18}\text{O}^{16}\text{O}_2^-$ in fossil soil carbonates (8).

A simple game helps to understand isotopic clumping. Here, we are not interested in the total amount of isotopic clumping, which simply scales with the isotopic composition of the substance. Rather, we seek the deviation from amounts predicted by chance alone. The game is rolling the dice, and the goal is snake eyes—one pip facing up on both dice. The probability of snake eyes for regular six-sided dice is $1/6 \times 1/6 = 1/36$, or 2.77%. The probability changes with the number of sides on each dice, becoming $1/4$ for two-sided dice and $1/10^6$ for 100-sided dice. Changing the number of sides is like changing the isotopic composition (9). To win the game, we have to beat chance, and

changing the number of sides will not help.

So far, clumped isotope geochemistry has been concerned with gaming the system on the positive side: coming up with more snake-eyes (or heavy isotope clumps) than predicted by chance alone. This situation is actually preferred by thermodynamics: under conditions of chemical equilibrium, levels of clumping will be slightly higher than levels predicted by chance alone. The isotopic "dice" are weighted by thermodynamics (see the figure). This enrichment in clumps, signified by Δ and on the order of a few parts per thousand, increases as molecular synthesis temperatures decrease (10, 11).

Yeung *et al.* and Wang *et al.* now show that the system can also be gamed in the opposite direction: coming up with fewer snake-eyes

"How can molecules seemingly evade both chance and thermodynamics? The answer appears to relate to the biological assembly of molecules from nonidentical substrate binding sites under irreversible conditions."

(or fewer heavy isotope clumps) than predicted by thermodynamics and sometimes fewer than predicted by chance statistics. How can molecules seemingly evade both chance and thermodynamics? The answer appears to relate to the biological assembly of molecules from nonidentical substrate binding sites under irreversible conditions.

Photosynthesis generates O_2 by combining oxygen atoms from two water molecules. The oxygen-evolving complex of Photosystem II has two water-binding sites that are thought to be nonidentical. It is therefore plausible that each site has a dif-

Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD 21218, USA.
E-mail: bhpassey@jhu.edu

squid, and legume symbioses, the host is working simultaneously with multiple genotypes. This allows hosts to exploit a wider range of environmental conditions, but also increases the potential for conflict among competing symbionts (15). To help manage conflict, specific structures in hosts (such as crypts in squid) effectively separate different symbiont genotypes into individual chambers (13). In contrast, vertical transmission of symbionts from parent to offspring, as in *Paracatenula* flatworms, can lead to a high relatedness between the symbionts within a host (15), linking the fitness of the symbiont to host performance, reducing conflict, and selecting for integration.

Ecological context can also help to predict when and why transitions have been made (8). When benefits of partnerships vary with environmental context, mutual dependency is less likely to evolve. For example, in the legume-rhizobia symbiosis, the reliance on the bacterial partner decreases in high-nitrogen environments (14). Here, strict dependency can be costly, making a major transition to a single, higher-level organism precarious and unlikely if benefits are not absolute.

The advantage of the major transitions framework is that it emphasizes different questions from the mechanistic ones currently being asked. Interplay between evolutionary theory and genomic research will allow us to understand the evolution of organismal complexity within a single, unified framework. ■

REFERENCES AND NOTES

1. J. Archibald, *One Plus One Equals One* (Oxford Univ. Press, Oxford, 2014).
2. F. Husnik et al., *Cell* **153**, 1567 (2013).
3. P.J. Keeling, *Annu. Rev. Plant Biol.* **64**, 583 (2013).
4. J.T. Van Leuven, R.C. Meister, C. Simon, J.P. McCutcheon, *Cell* **158**, 1270 (2014).
5. G.M. Bennett, N.A. Moran, *Genome Biol. Evol.* **5**, 675 (2013).
6. J.P. McCutcheon, P.J. Keeling, *Curr. Biol.* **24**, R654 (2014).
7. J. Maynard Smith, E. Szathmáry, *The Major Transitions in Evolution* (Oxford Univ. Press, Oxford, 1995).
8. A.F.G. Bourke, *Principles of Social Evolution* (Oxford Univ. Press, Oxford, 2011).
9. J.L. Sachs, R.G. Skophamer, J.U. Regus, *Proc. Natl. Acad. Sci. U.S.A.* **108** (suppl. 2), 10800 (2011).
10. N. Dubilier, C. Bergin, C. Lott, *Nat. Rev. Microbiol.* **6**, 725 (2008).
11. H.R. Gruber-Vodicka et al., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12078 (2011).
12. A.D. Nussbaumer, C.R. Fisher, M. Bright, *Nature* **441**, 345 (2006).
13. M. McFall-Ngai, *PLOS Biol.* **12**, e1001783 (2014).
14. S.S. Porter, E.L. Simms, *Ecol. Lett.* **17**, 1121 (2014).
15. M. Bright, S. Bulgheresi, *Nat. Rev. Microbiol.* **8**, 218 (2010).

ACKNOWLEDGMENTS

We thank N. Ruby, N. Dubilier, J. Mills, and J. McCutcheon for critical insights and inspiration. Supported by the European Research Council and the Netherlands Organisation for Scientific Research.

10.1126/science.aaa9605

GEOCHEMISTRY

Biogeochemical tales told by isotope clumps

Molecules with two or more heavy isotopes provide insights into diverse biological and geological phenomena

By Benjamin H. Passey

How do you take a dinosaur's temperature, reconstruct the elevation histories of Earth's great mountain ranges, probe the workings of photosynthesis, and confirm biological origins of a greenhouse gas? Increasingly, the answer lies in clumps. Clumped isotope geochemistry (1) is the latest branch of stable isotope geochemistry, the field that illuminated the Pleistocene glacial-interglacial cycles (2), the oxygenation of Earth's atmosphere some 2.5 billion years ago (3), and the enigmatic presence of grass in hominid diets (4). In this issue, Yeung et al. (page 431) (5) and Wang et al. (page 428) (6) describe clumped isotope effects that, among other things, can serve as tracers of biological versus abiological origins of gases.

Traditional isotope geochemistry is concerned with ratios of heavy to light isotopes. In contrast, clumped isotope geochemistry looks at occurrences of two or more heavy isotopes in the same molecule, such as ^{18}O , $^{13}\text{C}^{18}\text{O}^{16}\text{O}$, and $^{13}\text{CH}_3\text{D}$. In its decade of existence, clumped isotope geochemistry has focused on the thermodynamic aspects—particularly the temperatures of molecular and mineral synthesis—recorded in isotopic clumps. Hence, dinosaur body temperatures have been estimated from clumped carbonate ions ($^{13}\text{C}^{18}\text{O}^{16}\text{O}_2^-$) in fossil teeth (7), and altitudes of ancient mountain ranges from $^{13}\text{C}^{18}\text{O}^{16}\text{O}_2^-$ in fossil soil carbonates (8).

A simple game helps to understand isotopic clumping. Here, we are not interested in the total amount of isotopic clumping, which simply scales with the isotopic composition of the substance. Rather, we seek the deviation from amounts predicted by chance alone. The game is rolling the dice, and the goal is snake eyes—one pip facing up on both dice. The probability of snake eyes for regular six-sided dice is $1/6 \times 1/6 = 1/36$, or 2.77%. The probability changes with the number of sides on each dice, becoming $1/4$ for two-sided dice and $1/10^6$ for 100-sided dice. Changing the number of sides is like changing the isotopic composition (9). To win the game, we have to beat chance, and

changing the number of sides will not help.

So far, clumped isotope geochemistry has been concerned with gaming the system on the positive side: coming up with more snake-eyes (or heavy isotope clumps) than predicted by chance alone. This situation is actually preferred by thermodynamics: under conditions of chemical equilibrium, levels of clumping will be slightly higher than levels predicted by chance alone. The isotopic "dice" are weighted by thermodynamics (see the figure). This enrichment in clumps, signified by Δ and on the order of a few parts per thousand, increases as molecular synthesis temperatures decrease (10, 11).

Yeung et al. and Wang et al. now show that the system can also be gamed in the opposite direction: coming up with fewer snake-eyes

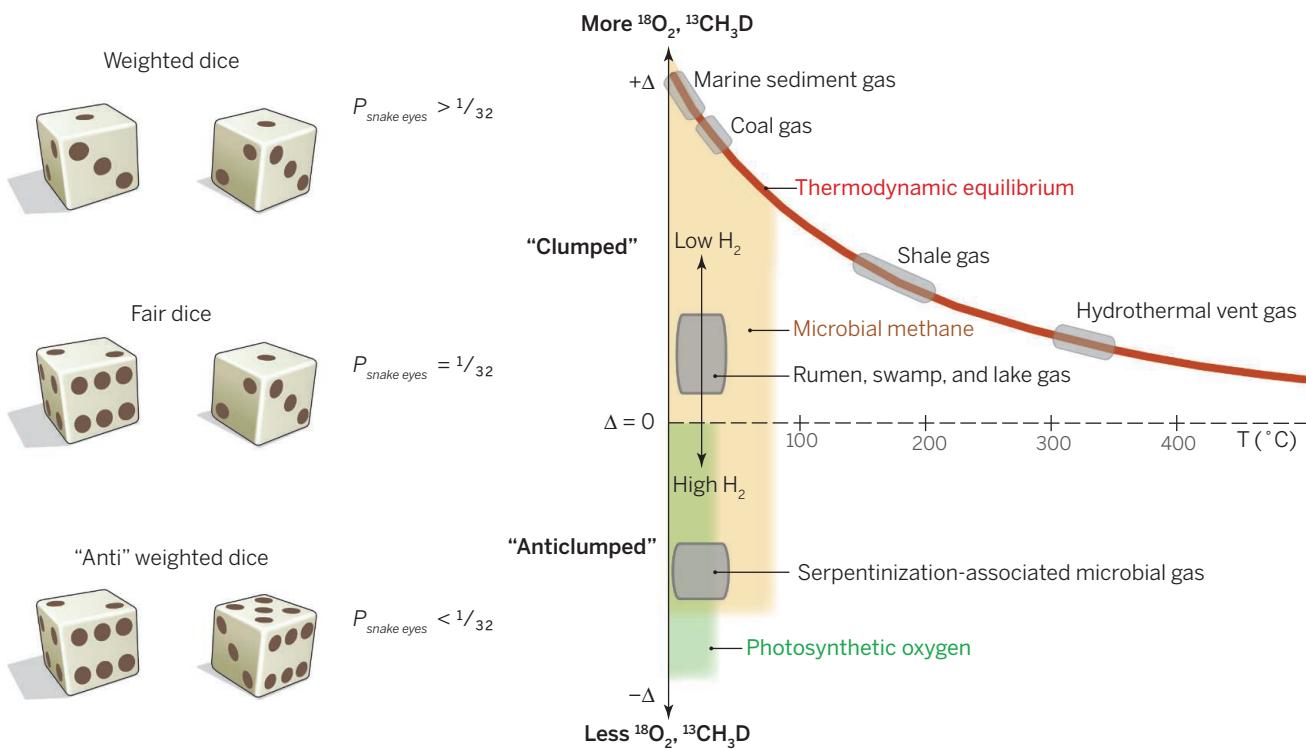
"How can molecules seemingly evade both chance and thermodynamics? The answer appears to relate to the biological assembly of molecules from nonidentical substrate binding sites under irreversible conditions."

(or fewer heavy isotope clumps) than predicted by thermodynamics and sometimes fewer than predicted by chance statistics. How can molecules seemingly evade both chance and thermodynamics? The answer appears to relate to the biological assembly of molecules from nonidentical substrate binding sites under irreversible conditions.

Photosynthesis generates O_2 by combining oxygen atoms from two water molecules. The oxygen-evolving complex of Photosystem II has two water-binding sites that are thought to be nonidentical. It is therefore plausible that each site has a dif-

Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD 21218, USA.
E-mail: bhpassey@jhu.edu

Isotope clumping and anticlumping in oxygen and methane



More than a roll of the dice. The appearance of more than one rare, heavy isotope in a molecule (a heavy isotope “clump”) is not random. Instead, the isotopic dice are weighted. Clumps are strongly favored under conditions of chemical equilibrium, whereas kinetic processes related to biosynthesis may have lower preference for clumps and in some cases may select against clumps, leading to “anticlumped” products. Natural methane gas shows both behaviors (6), whereas the isotope content of photosynthetic oxygen is random or anticlumped (i.e., below chance levels) (5).

ferent affinity for ^{18}O relative to common ^{16}O (or ^{17}O relative to ^{16}O). Yeung *et al.* predict that if this is true, the amounts of $^{18}\text{O}_2$ and $^{17}\text{O}^{18}\text{O}$ in photosynthetically produced O_2 will always be lower than levels predicted by chance statistics. This astonishing outcome is true regardless of whether both sites prefer ^{18}O , both sites prefer ^{16}O , or one site prefers ^{16}O and one site prefers ^{18}O . The authors investigate this prediction by studying the evolution of clumped O_2 in a sealed terrarium filled with water hyacinths. The experiment did not go exactly as envisioned: The plants eventually died and were replaced by an “algae-dominated ecosystem.” However, there was a clear lowering of $^{18}\text{O}_2$ and $^{17}\text{O}^{18}\text{O}$ toward the “sub-chance” levels (negative Δ) predicted by the model for photosynthetic O_2 .

Wang *et al.* study clumped methane ($^{13}\text{CH}_3\text{D}$) from sources as diverse as cow rumens, deep sea hydrothermal vents, and the Marcellus Shale. In agreement with a study published last year (12), the geological gases had positive Δ values reflective of isotopic equilibrium at the temperatures of their sources. In contrast, methane from cow rumens, swamps, lakes, and lab cultures had

low levels of clumping which, if interpreted in the framework of thermodynamic equilibrium, would suggest exceedingly high temperatures of methane synthesis—typically greater than 200°C. Such temperatures are nonsense, and clearly the thermodynamic equilibrium model is not applicable to this system. Prompted by these findings, Wang *et al.* developed a model of microbial methanogenesis that predicts discrimination against the heavy clumped species ($^{13}\text{CH}_3\text{D}$) relative to molecules with single heavy isotopes ($^{12}\text{CH}_3\text{D}$, $^{12}\text{CH}_4$). This discrimination increases as the degree of irreversibility increases, which is proportional to the availability of the limiting substrate, here H_2 . Thus, the model predicts lower Δ values under higher H_2 and hence higher rates of synthesis.

The clumped isotope anomalies will help place much-needed constraints on biogeochemical sources, sinks, and budgets of O_2 and CH_4 , perhaps even over glacial-interglacial cycles from gases trapped in ice cores. Similar effects are possible in other biogenic gases like nitrous oxide and ethane. Clumping in O_2 from human-made oxygen-evolving systems (13) could help to elucidate reaction mechanisms. And far away on Mars, the Curiosity rover has detected transient whiffs of methane (14). If the laser spectrometer of Wang *et al.* could be adapted for missions to Mars and beyond, it would be a major boon to the search for extraterrestrial life. ■

REFERENCES AND NOTES

1. J. M. Eiler, *Annu. Rev. Earth Planet. Sci.* **41**, 411 (2013).
2. C. Emiliani, *J. Geol.* **63**, 538 (1955).
3. J. Farquhar, H. Bao, M. Thiemens, *Science* **289**, 756 (2000).
4. M. Sponheimer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10513 (2013).
5. L.Y. Yeung, J.L. Ash, E.D. Young, *Science* **348**, 431 (2015).
6. D.T. Wang *et al.*, *Science* **348**, 428 (2015).
7. R.A. Eagle *et al.*, *Science* **333**, 443 (2011).
8. C.N. Garzione *et al.*, *Earth Planet. Sci. Lett.* **393**, 173 (2014).
9. A die representing terrestrial oxygen would have ~260 sides, with one side being ^{17}O , five being ^{18}O , and the rest ^{16}O . The probability of ^{18}O – ^{18}O snake eyes would be $\sim 4 \times 10^{-6}$. An isotopically enriched oxygen sample might have six ^{18}O sides, and the probability of ^{18}O – ^{18}O would then be 5×10^{-6} .
10. Z. Wang, E.A. Schauble, J.M. Eiler, *Geochim. Cosmochim. Acta* **68**, 4779 (2004).
11. E.A. Schauble, P. Ghosh, J.M. Eiler, *Geochim. Cosmochim. Acta* **70**, 2510 (2006).
12. D.A. Stolper *et al.*, *Science* **344**, 1500 (2014).
13. A.M. Angeles-Boza, J.P. Roth, *Inorg. Chem.* **51**, 4722 (2012).
14. C.R. Webster *et al.*, *Science* **347**, 415 (2015).

10.1126/science.aab1604



Biogeochemical tales told by isotope clumps

Benjamin H. Passey

Science **348**, 394 (2015);

DOI: 10.1126/science.aab1604

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 23, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/394.full.html>

A list of selected additional articles on the Science Web sites related to this article can be found at:

<http://www.sciencemag.org/content/348/6233/394.full.html#related>

This article cites 12 articles, 5 of which can be accessed free:

<http://www.sciencemag.org/content/348/6233/394.full.html#ref-list-1>

This article appears in the following subject collections:

Geochemistry, Geophysics

http://www.sciencemag.org/cgi/collection/geochem_phys

SELF-ASSEMBLED MATERIALS

Supramolecular lattices from tetrahedral nanobuilding blocks

The location of hydrophilic and hydrophobic regions on tetrahedral nanoparticles determines how they pack together into a variety of structures

By Shu Yang

The self-assembly and packing of nano-scale particles that have nonspherical shapes is only beginning to be explored. Recent computer simulations have suggested that hard, convex polyhedra can pack more densely than spheres, leading to complex structures of liquid crystals, plastic crystals, crystals, and disordered phases (1). In particular, aperiodic quasicrystalline and crystalline phases occur in the packing of tetrahedra, the simplest polyhedron (2). On page 424 of this issue, Huang *et al.* report the precise synthesis of rigid giant tetrahedra consisting of hydrophobic and hydrophilic polyhedral oligomeric silsesquioxane (POSS) cages (3). They assemble a diverse range of highly ordered supramolecular lattices by exploiting the location of the hydrophilic POSS cage substituent and the number of substitutions, together with the interplay of crystallization and the collective hydrogen-bonding interactions.

The A15 phase was discovered in metal alloys with A_3B composition (4). As seen on the right side of the figure, it is a cubic lattice consisting of six A units (pale red spheres) in Kasper polyhedra with 14-coordination and two B units (dark red spheres) in icosahedra with 12-coordination. Its space group is $Pm\bar{3}n$ and it exhibits a two-dimensional 4⁴-square tiling pattern. Relative to other superlattices, the A15 phase is more loosely packed and can better accommodate distortion away from ideal sphere packing.

Closely packed face-centered cubic (fcc) and hexagonal close-packed (hcp) structures are common from packing of hard spheres, but soft spheres often form non-close-packed lattices, including body-centered cubic (bcc), body-centered orthorhombic (bco), and the diamond lattices. Recently, the formation of the A15 phase for soft spheres has been suggested by theory (5, 6) and demonstrated experimentally for packing of “deformable” soft spheres assembled from lyotropic lipids, conical fan-shaped dendrimers (7), and asymmetric block copolymers (8) (see the figure, top and middle left, respectively).

Typically, these building blocks have a hard core that favors lattices with maximum packing fraction and a soft corona that can be deformed to maximize entropy and minimize interfacial contact. Nonetheless, these soft spheres do not constitute a single-particle system. Often a mixture of bcc lattices, Frank-Kasper σ phases, and A15 phases is observed, because block copolymer and dendrimer particles are not identical but are polydisperse—they have a distribution of chain lengths. The large number of chain conformations and the soft nature of polymer chains also impede

denoted T_8R_8 (where each Si is connected with three oxygens to form a T unit). The resulting compounds have the general formula $(RSiO_{3/2})_8$.

Creating giant tetrahedra with precise partitioning of hydrophobic and hydrophilic POSS cages requires site-selective functionalization and regioselective multifunctionalization of the apex by the POSS cages. First, the hydrophobic POSS cages with seven isobutyl groups (BPOSS) and one alkyne group are coupled to tetrakis(4-azidophenyl)methane via the azide-alkyne [3+2] cycloaddition reaction. Because these “click” reactions offer near-quantitative yield, giant tetrahedra with one, two, or three unreacted azide groups are obtained by controlling the stoichiometric feed ratio of the azide versus alkyne groups. In the second step, hydrophilic POSS cages with either hydroxyl or carboxylic acid groups are coupled to the giant tetrahedron via a thiol-ene reaction. Although vinyl groups are often incompatible with hydroxyl or carboxylic groups, they possess orthogonal reactivities in click chemistry, where the reaction is site-selective.

Thus, it is possible to sequentially introduce hydrophilic POSS cages one by one on the giant tetrahedra, much like putting together Lego blocks. Such unprecedented control of shape and topology is not possible in polymer systems. Incorporation of different reactivities leads to competing interactions that drive self-assembly—that is, collective hydrogen-bonding interactions between the hydrophilic POSS cages and the crystallization of BPOSS cages. Meanwhile, replacing the BPOSS cage with a hydrophilic POSS cage on the tetrahedron lowers the molecular symmetry.

When cast from solution on mica, crystallization of BPOSS cages dominates. To minimize interfacial energy, frustrated supramolecular lamellae are formed with two layers of BPOSS and one interdigitated layer of the hydrophilic POSS cages. The preformation of lamellae is critical to the

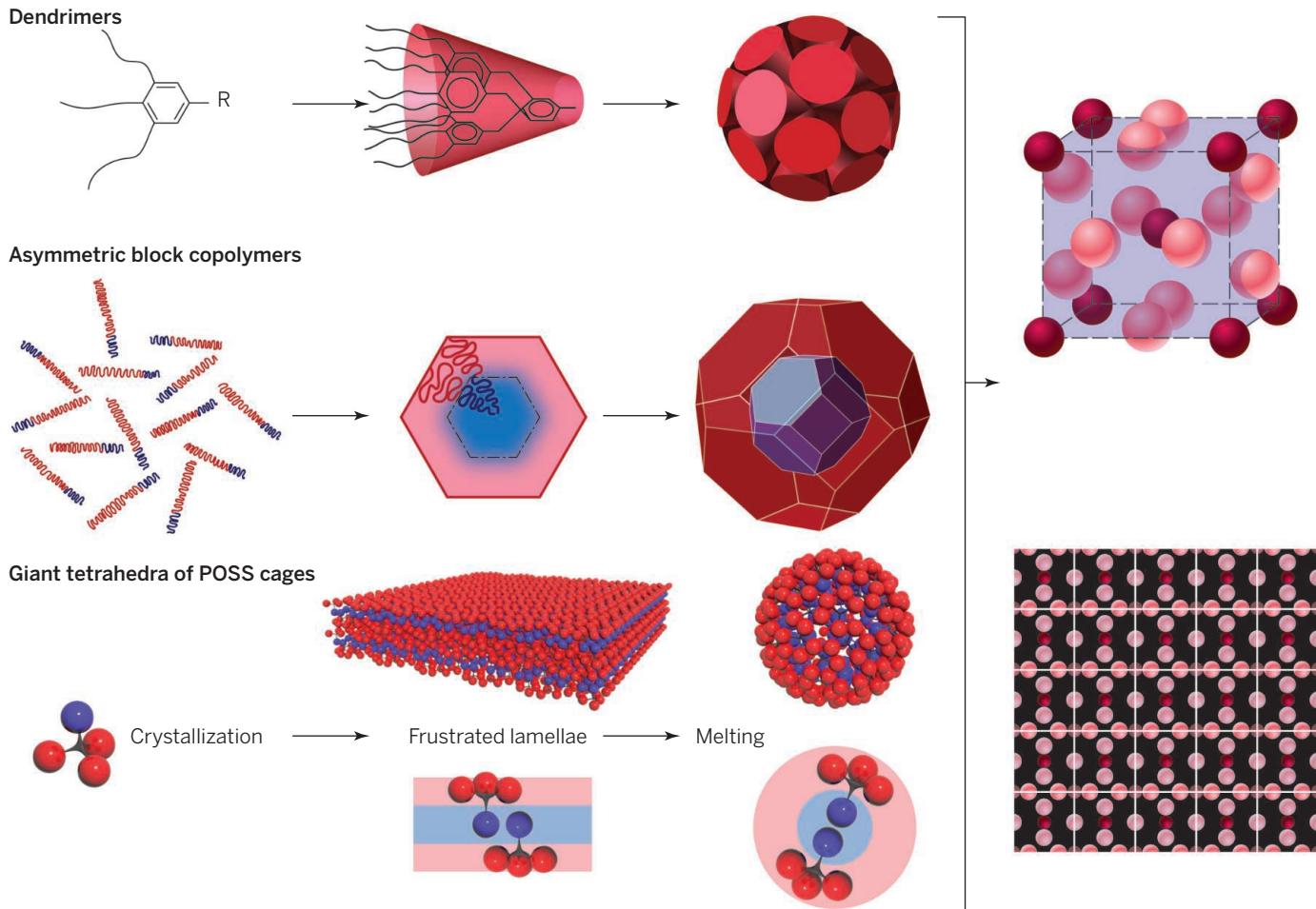
It will be interesting to extend the chemical strategy presented by Huang et al. to construct other types of giant polyhedra and investigate their self-assembly.

precise control of the molecular structures and their assemblies.

Molecular nanoparticles (MNPs), including fullerenes such as C_{60} , POSS derivatives, and polyoxometalates, represent a new class of nanobuilding blocks. Bridging the length scale between polymers and small molecules, MNPs can be synthesized with explicit chemical composition and precisely controlled size, shape, symmetry, and function. They are rigid, incompressible, and impenetrable. MNPs have been regarded as “nanoatoms” or “giant molecules.”

Among them, POSS derivatives (9) have enjoyed rapid growth in research and applications because of their well-defined structures, high interaction parameters, and tunable chemistry of the periphery groups. Hydrolysis of $RSiX_3$ (where X is a halide and R can be alkyl, aryl, alkoxy, epoxy, hydroxyl, carboxylic acid, amine, or vinyl groups), followed by condensation, leads to the spontaneous formation of cage-like POSS,

Department of Materials Science and Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA.
E-mail: shuyang@seas.upenn.edu



Different parts for self-assembly of A15 supramolecular lattices. On the right side of the figure, the A15 lattice, which assembles for metal alloys from two different types of atoms, is shown as a three-dimensional rendering at the top and as a 4⁴-square tiling pattern at the bottom. Different routes to this same type of lattice are shown on the left side for dendrimers at the top, asymmetric block copolymers in the middle, and giant tetrahedral POSS cages, as elaborated by Huang *et al.*, at the bottom.

latter superlattice formation. Upon melting, hydrophilic POSS cages form spherical aggregates while BPOSS layers scroll to form the shell. The core-shell spheres spontaneously pack into the A15 supramolecular lattice as a result of balancing the molecular symmetry and molecular interactions (see the figure, lower left).

Design of giant nanoatoms from precisely controlled organic and hybrid materials also has caught the eye of polymer scientists. For example, Ober and co-workers demonstrated sub-50 nm patterns using glass-forming, low-molecular-weight organic compounds as photoresists (10). Relative to traditional polymeric resists, molecular glass resists offer smaller feature size and line-edge roughness as well as higher dissolution contrast, which is critical to driving miniaturization of electronic devices on the sub-100 nm scale.

The precise synthesis of Lego-like POSS cages opens up an entirely new paradigm of self-assembly in which lattices can be assembled one-by-one and on demand. The rich chemistry that can be introduced on

POSS cages, the versatility and regioselectivity enabled by click chemistry, and the possible structural variations of POSS derivatives will be of great interest to materials scientists. Besides T₈-type POSS cages, there are less stable T₆, T₁₀, and T₁₂ POSS cages. It will be interesting to extend the chemical strategy presented by Huang *et al.* to construct other types of giant polyhedra and investigate their self-assembly. Meanwhile, the coupling reactions and the molecular interactions can be fine-tuned to assemble other superlattices. Furthermore, small-molecule fluorophores, metal complexes, conjugated molecules, and peptides can be introduced into the nanocavity of the POSS cage or its periphery for imaging, sensing, catalysis, charge-carrier transport, and biological applications (9). Also, the T₈ core may not act as a traditional silica-like insulator.

Finally, it will be interesting to disperse these shape-persistent tetrahedra and their superlattices into a polymer matrix or a liquid crystal medium, or to place them on a patterned surface to create truly hierarchi-

cal assemblies. They can also be tethered to a polymer chain to construct a block copolymer analog, so as to take advantage of their rich phase separation behaviors, or to form complexation with other polyhedra to create Janus superlattices. The ability to build nano-Legos with precise geometrical factors and tunable molecular interactions offers a powerful yet versatile tool to create ever more complex functional materials. ■

REFERENCES

- P. F. Damasceno, M. Engel, S. C. Glotzer, *Science* **337**, 453 (2012).
- A. Hajj-Akbari *et al.*, *Nature* **462**, 773 (2009).
- M. Huang *et al.*, *Science* **348**, 424 (2015).
- F. C. Frank, J. S. Kasper, *Acta Crystallogr.* **11**, 184 (1958).
- P. Ziherl, R. D. Kamien, *J. Phys. Chem. B* **105**, 10147 (2001).
- G. M. Grason, B. A. DiDonna, R. D. Kamien, *Phys. Rev. Lett.* **91**, 058304 (2003).
- V. S. K. Balagurusamy, G. Ungar, V. Percec, G. Johansson, *J. Am. Chem. Soc.* **119**, 1539 (1997).
- S. Lee, C. Leighton, F. S. Bates, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 17723 (2014).
- D. B. Cordes, P. D. Lickiss, F. Rataboul, *Chem. Rev.* **110**, 2081 (2010).
- J. Y. Dai *et al.*, *Chem. Mater.* **18**, 3404 (2006).

PHYSIOLOGY

Calcilytics for asthma relief

A promiscuous calcium receptor holds promise as a therapeutic target for asthma

By Raymond B. Penn

For decades, patients suffering from allergic asthma—the symptoms of which range from wheezing and shortness of breath to extreme difficulty in breathing—generally have been treated with agents that reduce airway inflammation, such as corticosteroids, or reverse airway constriction, such as β_2 -adrenergic receptor agonists. Unfortunately, up to 55% of asthmatics have suboptimal control with these drugs (1). Moreover, safety concerns regarding the use of β_2 agonists have resulted in more stringent guidelines for their clinical use (2). Thus, there is pressing need to de-

velop new and safe therapies. A study by Yarova *et al.* (3) points to a protein that controls the constriction and proliferation of smooth muscle cells lining the bronchial airway. Moreover, there is already a drug that inhibits this protein.

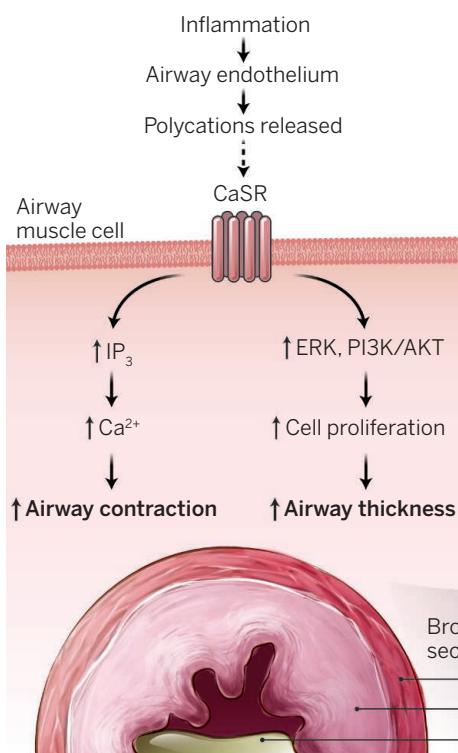
The calcium-sensing receptor (CaSR) is best known for its role in regulating calcium homeostasis. Expressed in the parathyroid gland, it surveys the concentration of circulating calcium. Calcium binds to and activates the CaSR, triggering intracellular signals that suppress the release of parathyroid hormone. This reduces calcium release from stores such as bone. However, the CaSR is promiscuous in its activation by other molecules, including polyvalent cations, amino acids, and virus elements. Indeed, Yarova *et al.* show that the CaSR uses this robust sensing ability to control airway hyperresponsiveness

and inflammation in asthma. Importantly, inhibitory small molecules previously developed to antagonize the CaSR and treat hypocalcemia mitigate key features of allergic asthma.

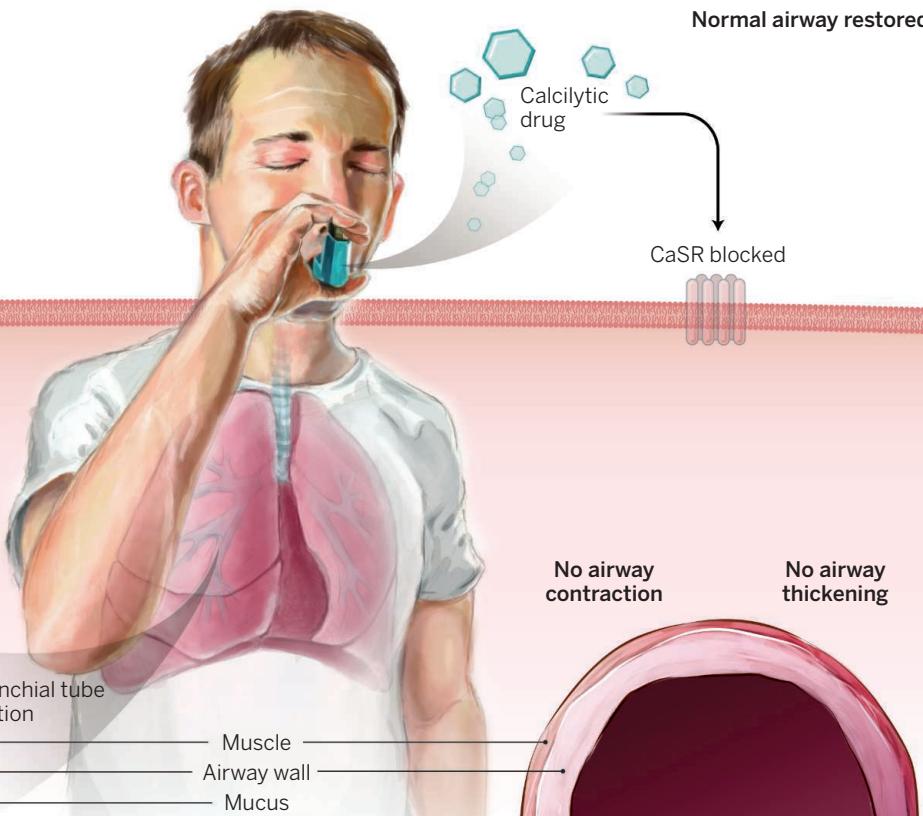
Yarova *et al.* examined bronchial airway tissues and cells from a mouse model of asthma and biopsies from asthma patients, and demonstrate that calcium, as well as various agents (including eosinophilic cationic protein and L-arginine-derived polyamines) whose concentrations are elevated in the bronchial airway of asthma patients during allergic inflammation, can stimulate CaSR. In bronchial airway tissue excised from the mouse asthma model, as well as in the intact animal, this stimulation resulted in airway smooth muscle contraction, and enhanced the contractile effect of bronchoconstrictors such as acetylcholine or histamine. The authors confirmed the CaSR-dependent nature of these effects in both human bronchial airway smooth muscle cells with reduced expression of CaSR and in bronchial airway tissue and cells isolated from genetically modified mice lacking CaSR. In addition, when mice with allergic lung inflammation inhaled CaSR antagonists (calcilytics), the asthma phenotype was reversed. Calcilytics blocked

Department of Medicine, Division of Pulmonary and Critical Care Medicine, Center for Translational Medicine, Jane and Leonard Korman Lung Center, Thomas Jefferson University, Philadelphia, PA, USA. E-mail: raymond.penn@jefferson.edu

During asthma symptoms



Asthma relief. Numerous agents activate the CaSR expressed on airway smooth muscle, as well as on certain inflammatory cells and airway epithelia (not shown). CaSR activation in airway smooth muscle promotes contractile hyperresponsiveness and possibly airway remodeling. CaSR activity also facilitates inflammatory cell infiltration into the lung (not shown). Calcilytics block the CaSR, and may represent an inhalable agent to treat asthma. IP₃, inositol 1,4,5-trisphosphate; ERK, extracellular signal-regulated kinase; PI3K, phosphatidylinositol 3-kinase.



the CaSR, prevented airway hyperresponsiveness and, surprisingly, inflammation. The latter effect could be due to expression of the CaSR not only in resident lung cells (airway smooth muscle and epithelia) but also in invading inflammatory cells (eosinophils, macrophages).

Interestingly, Yarova *et al.* show that inflammation increases CaSR expression in human and mouse. That is consistent with the findings that show a prominent role of the CaSR in affecting airway contractility under asthmatic conditions.

Thus, the CaSR emerges as an exciting potential asthma therapeutic target (see the figure). Recently, both bitter tastant receptors and chloride channels have been proposed as drug targets in asthma (4). Activated bitter tastant receptors relax airway smooth muscle and dilate the airway in human and mouse, but the mechanism is not yet clear (5). Chloride channel blockers also act on airway smooth muscle and limit the ability of agonists such as acetylcholine to stimulate contraction (6). However, calcilytics have certain advantages that favor success as a therapy for asthma. They are small molecules that are readily deliverable by inhalation, thus limiting potential problematic systemic effects. Their efficacy is favored by the ability to target multiple cell types and mechanisms that contribute to the asthma phenotype. Perhaps most importantly, the safety of three different oral calcilytics as osteoporosis drugs is suggested in phase I clinical studies (7, 8), and the calcilytic NPSP795 is currently in phase II clinical trials for the treatment of autosomal dominant hypocalcemia (9). Results from these studies may help limit the regulatory hurdles that an inhaled calcilytic as an asthma therapy will ultimately face. The ability to assess the presence of CaSR activators in the lung could enable identification of those asthmatics most likely to respond to calcilytics, in line with the goal of personalized asthma treatment (10). ■

REFERENCES

1. S.P. Peters *et al.*, *J. Allergy Clin. Immunol.* **119**, 1454 (2007).
2. V.E. Ortega, S.P. Peters, *Curr. Opin. Pharmacol.* **10**, 246 (2010).
3. P.L. Yarova *et al.*, *Sci. Transl. Med.* **284**, ra60 (2015).
4. E.A. Townsend *et al.*, *J. Allergy* **2012**, 321949 (2012).
5. D.A. Deshpande *et al.*, *Nat. Med.* **16**, 1299 (2010).
6. J. Danielsson *et al.*, *Am. J. Physiol. Lung Cell. Mol. Physiol.* **307**, L273 (2014).
7. <https://clinicaltrials.gov/ct2/show/NCT01802892?term=ronacaleret&rank=1>
8. <https://clinicaltrials.gov/ct2/show/NCT00417261?term=AXT914&rank=1>
9. <https://clinicaltrials.gov/ct2/show/NCT02204579?term=npsp795&rank=1>
10. J. Lötvall *et al.*, *J. Allergy Clin. Immunol.* **127**, 355 (2011).

10.1126/science.aab2173

PROTEIN SYNTHESIS

The delicate dance of translation and folding

RNA translation and protein folding affect each other during protein synthesis

By Joseph D. Puglisi

The past decade has seen rapid growth in our knowledge of how proteins are synthesized in cells. This includes the basic step of how transfer RNAs (tRNAs) decode messenger RNAs (mRNAs) with high fidelity and speed, how the ribosome moves along mRNA from codon to codon during translation, and how synthesis of the corresponding polypeptide chain is initiated and terminated at specific points on the mRNA (1). Structures of the ribosomal particles—megadalton RNA-protein assemblies—have provided detailed molecular views of the active sites for mRNA decoding and peptide bond formation, and suggested pathways for movement of ligands, factors, and the ribosomal subunits themselves. Obscured in this rich cache of knowledge is the fate of the protein product. How does it fold during mRNA translation and how might protein folding affect translation itself? These questions are addressed by elegant biophysical and biochemical approaches reported by Goldman *et al.* (2) and Kim *et al.* (3) on pages 457 and 444, respectively, of this issue, adding to a growing appreciation of cotranslational protein folding (4–6).

During translation, a nascent protein is synthesized from its amino to carboxyl terminus. The growing polypeptide passes through a 100 Å conduit in the large ribosomal subunit (the ribosomal exit tunnel) to emerge into solution. The exit tunnel is narrow, and can only accommodate α -helical conformations of the polypeptide chain. Thus, the nascent protein can only fold into tertiary structure once the appropriate regions of the polypeptide chain are exposed to solution. The folding of small protein domains is spontaneous and rapid in vitro, with folding time scales of 10^{-6} to 10^{-3} s. For larger proteins, those with complex topologies, or multidomain proteins, folding can take longer and require the assistance of protein-folding chaperones. Because translation occurs at rates of 1 to 20 amino acids per second (depending on organism and conditions), folding and translation often occur simultaneously on a

translating ribosome (cotranslational folding). Many chaperones and processing factors bind to nascent chains as they emerge from the ribosomal exit tunnel (7). It is thus paramount to understand the relationship among protein synthesis rates and protein-folding processes.

This interplay is underscored by the existence of sequences within a nascent polypeptide that cause translating ribosomes to pause or stall (8). Stalling sequences were discovered during translation of SecM from *Escherichia coli*. Ribosomes translating pro-

"...how might protein folding affect translation itself?"

teins with the SecM sequence, an 18-amino acid stretch, pause and stall when this sequence is located within the ribosomal exit tunnel near the peptidyl tRNA and narrow constrictions in the tunnel (9). Interactions between the peptide and ribosomal elements slow and block subsequent steps in the elongation process (10) at a precise point, during which a paused translation complex docks to a secretion tunnel that uses energy from the hydrolysis of adenosine 5'-triphosphate to pull the protein through the secretion tunnel. This suggests that force applied to the ribosomal nascent chain can relieve stalling of translation. Several *in vivo* measurements have supported this hypothesis (11, 12).

Applying elegant single-molecule approaches, Goldman *et al.* demonstrate explicitly that force on a polypeptide chain can relieve translational stalling. They used optical trapping technology, whereby molecules attached to larger beads are trapped and manipulated by a strong laser beam, allowing measurements and changes in the force applied to a biological system. Goldman *et al.* prepared stalled, translating ribosomes containing nascent chains labeled

Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA.
E-mail: puglisi@stanford.edu



Calcilytics for asthma relief
 Raymond B. Penn
Science **348**, 398 (2015);
 DOI: 10.1126/science.aab2173

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 23, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/398.full.html>

This article cites 7 articles, 1 of which can be accessed free:

<http://www.sciencemag.org/content/348/6233/398.full.html#ref-list-1>

This article appears in the following subject collections:
 Medicine, Diseases

<http://www.sciencemag.org/cgi/collection/medicine>

the CaSR, prevented airway hyperresponsiveness and, surprisingly, inflammation. The latter effect could be due to expression of the CaSR not only in resident lung cells (airway smooth muscle and epithelia) but also in invading inflammatory cells (eosinophils, macrophages).

Interestingly, Yarova *et al.* show that inflammation increases CaSR expression in human and mouse. That is consistent with the findings that show a prominent role of the CaSR in affecting airway contractility under asthmatic conditions.

Thus, the CaSR emerges as an exciting potential asthma therapeutic target (see the figure). Recently, both bitter tastant receptors and chloride channels have been proposed as drug targets in asthma (4). Activated bitter tastant receptors relax airway smooth muscle and dilate the airway in human and mouse, but the mechanism is not yet clear (5). Chloride channel blockers also act on airway smooth muscle and limit the ability of agonists such as acetylcholine to stimulate contraction (6). However, calcilytics have certain advantages that favor success as a therapy for asthma. They are small molecules that are readily deliverable by inhalation, thus limiting potential problematic systemic effects. Their efficacy is favored by the ability to target multiple cell types and mechanisms that contribute to the asthma phenotype. Perhaps most importantly, the safety of three different oral calcilytics as osteoporosis drugs is suggested in phase I clinical studies (7, 8), and the calcilytic NPSP795 is currently in phase II clinical trials for the treatment of autosomal dominant hypocalcemia (9). Results from these studies may help limit the regulatory hurdles that an inhaled calcilytic as an asthma therapy will ultimately face. The ability to assess the presence of CaSR activators in the lung could enable identification of those asthmatics most likely to respond to calcilytics, in line with the goal of personalized asthma treatment (10). ■

REFERENCES

1. S. P. Peters *et al.*, *J. Allergy Clin. Immunol.* **119**, 1454 (2007).
2. V. E. Ortega, S. P. Peters, *Curr. Opin. Pharmacol.* **10**, 246 (2010).
3. P. L. Yarova *et al.*, *Sci. Transl. Med.* **284**, ra60 (2015).
4. E. A. Townsend *et al.*, *J. Allergy* **2012**, 321949 (2012).
5. D. A. Deshpande *et al.*, *Nat. Med.* **16**, 1299 (2010).
6. J. Danielsson *et al.*, *Am. J. Physiol. Lung Cell. Mol. Physiol.* **307**, L273 (2014).
7. <https://clinicaltrials.gov/ct2/show/NCT01802892?term=ronacalceret&rank=1>
8. <https://clinicaltrials.gov/ct2/show/NCT00417261?term=AXT914&rank=1>
9. <https://clinicaltrials.gov/ct2/show/NCT02204579?term=npsp795&rank=1>
10. J. Lötvall *et al.*, *J. Allergy Clin. Immunol.* **127**, 355 (2011).

10.1126/science.aab2173

PROTEIN SYNTHESIS

The delicate dance of translation and folding

RNA translation and protein folding affect each other during protein synthesis

By Joseph D. Puglisi

The past decade has seen rapid growth in our knowledge of how proteins are synthesized in cells. This includes the basic step of how transfer RNAs (tRNAs) decode messenger RNAs (mRNAs) with high fidelity and speed, how the ribosome moves along mRNA from codon to codon during translation, and how synthesis of the corresponding polypeptide chain is initiated and terminated at specific points on the mRNA (1). Structures of the ribosomal particles—megadalton RNA-protein assemblies—have provided detailed molecular views of the active sites for mRNA decoding and peptide bond formation, and suggested pathways for movement of ligands, factors, and the ribosomal subunits themselves. Obscured in this rich cache of knowledge is the fate of the protein product. How does it fold during mRNA translation and how might protein folding affect translation itself? These questions are addressed by elegant biophysical and biochemical approaches reported by Goldman *et al.* (2) and Kim *et al.* (3) on pages 457 and 444, respectively, of this issue, adding to a growing appreciation of cotranslational protein folding (4–6).

During translation, a nascent protein is synthesized from its amino to carboxyl terminus. The growing polypeptide passes through a 100 Å conduit in the large ribosomal subunit (the ribosomal exit tunnel) to emerge into solution. The exit tunnel is narrow, and can only accommodate α -helical conformations of the polypeptide chain. Thus, the nascent protein can only fold into tertiary structure once the appropriate regions of the polypeptide chain are exposed to solution. The folding of small protein domains is spontaneous and rapid in vitro, with folding time scales of 10^{-6} to 10^{-3} s. For larger proteins, those with complex topologies, or multidomain proteins, folding can take longer and require the assistance of protein-folding chaperones. Because translation occurs at rates of 1 to 20 amino acids per second (depending on organism and conditions), folding and translation often occur simultaneously on a

translating ribosome (cotranslational folding). Many chaperones and processing factors bind to nascent chains as they emerge from the ribosomal exit tunnel (7). It is thus paramount to understand the relationship among protein synthesis rates and protein-folding processes.

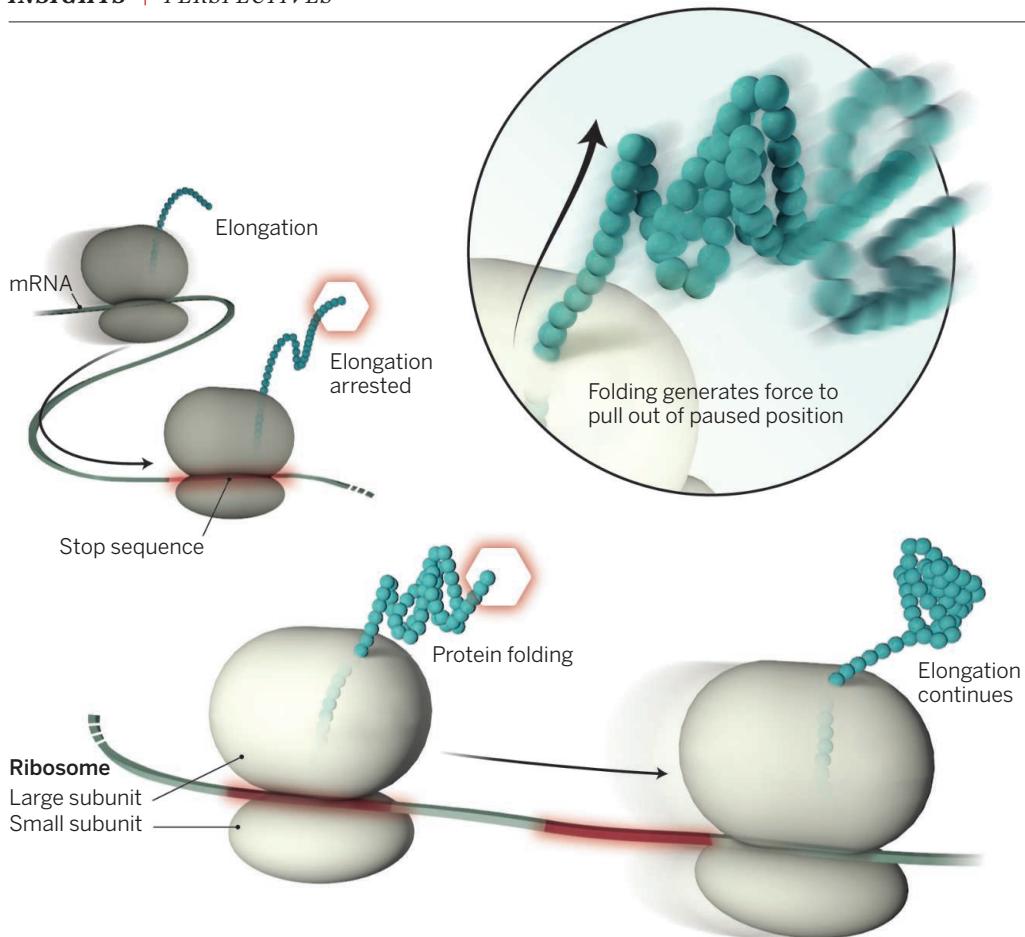
This interplay is underscored by the existence of sequences within a nascent polypeptide that cause translating ribosomes to pause or stall (8). Stalling sequences were discovered during translation of SecM from *Escherichia coli*. Ribosomes translating pro-

“...how might protein folding affect translation itself?”

teins with the SecM sequence, an 18-amino acid stretch, pause and stall when this sequence is located within the ribosomal exit tunnel near the peptidyl tRNA and narrow constrictions in the tunnel (9). Interactions between the peptide and ribosomal elements slow and block subsequent steps in the elongation process (10) at a precise point, during which a paused translation complex docks to a secretion tunnel that uses energy from the hydrolysis of adenosine 5'-triphosphate to pull the protein through the secretion tunnel. This suggests that force applied to the ribosomal nascent chain can relieve stalling of translation. Several in vivo measurements have supported this hypothesis (11, 12).

Applying elegant single-molecule approaches, Goldman *et al.* demonstrate explicitly that force on a polypeptide chain can relieve translational stalling. They used optical trapping technology, whereby molecules attached to larger beads are trapped and manipulated by a strong laser beam, allowing measurements and changes in the force applied to a biological system. Goldman *et al.* prepared stalled, translating ribosomes containing nascent chains labeled

Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA.
E-mail: puglisi@stanford.edu



Quit stalling. The ribosome is shown as two subunits. As the SecM sequence (red) in the mRNA is translated and the amino acid residues fill the ribosomal exit tunnel, translation stalls. The model shows how folding of the protein just adjacent to the ribosomal exit tunnel portal can generate a directional force that allows translation to resume.

with biotin. A ribosome was linked, through a tether, to a static micropipette, and the biotinylated nascent chain was attached to a polystyrene bead. The stalled nascent chain was designed to have an amino-terminal calmodulin domain, which has a well-characterized folding behavior. At a fixed small applied force, the calmodulin domain fluctuates between folded and unfolded. This oscillation yields a clear signal as the polystyrene bead moves away and toward the center of the trap following the unfolding and refolding of the protein. Restarting of translation releases the tether as protein synthesis resumes, allowing an indirect monitoring of translation and force on the nascent chain.

Goldman *et al.* show that increasing the force on the nascent chain overcomes the stalling induced by the SecM sequence. The authors then used a protein whose folding-unfolding transition as a function of force has been well characterized (Top7), to see whether folding of a protein near the ribosome could generate a directional pulling force similar to that of an optical trap. Using an *in vivo* assay for synthesis of fluorescent protein that can only occur if a SecM stall

is relieved, Goldman *et al.* show that having a folding domain near the ribosomal exit tunnel could exert force on the stalled ribosomal nascent chain such that translation can resume (see the figure). To relieve a SecM-induced stall in translation, at least 10 pN of force—equivalent to that generated by folding of a small domain—must be applied to a nascent chain.

The results of Goldman *et al.* show how protein folding can modulate translational dynamics, but can translation rates and occupancy of nascent chains within the ribosome also control protein-folding pathways? A growing body of experiments has probed this question both *in vivo* and *in vitro*. Kim *et al.* use ensemble fluorescence resonance energy transfer (FRET) to probe distances between an amino-terminal fluorescent protein and dyes attached chemically to distinct residues in carboxyl-terminal domains of the cystic fibrosis transmembrane conductance regulator, whose misfolding is linked to the disease. Kim *et al.* measured ensemble FRET for nascent chains that were either attached to ribosomes (not translating) or released free in solution, and found that late folding events that involved interaction

between the carboxy-terminal part of the polypeptide chain with an amino-terminal domain of the chain were slowed. This delay was tuned by the occupancy of that strand within the tunnel, and by stabilized folding of domains beyond the ribosome, through interaction with the ribosome. These results, and those of Goldman *et al.*, show how the ribosome itself can guide folding pathways. Kim *et al.* also show how codon usage could tune translation rates and occupancy of nascent chains within the exit tunnel, consistent with prior predictions from multiple studies.

The results of Goldman *et al.* and Kim *et al.* demonstrate clearly the coupling of protein folding and translation, but much work remains. Real-time dynamic data (13) would reveal directly how folding and translational steps are linked. Structural studies are needed, with advances in cryo-electron microscopy in particular (14), to show how a ribosome interacts with a nascent chain to stall translation or guide folding pathways. An additional layer of biological importance and complexity is the cluster of factors that bind to a nascent polypeptide as it exits

a ribosome. These factors process or modify the chain, chaperone its folding, or guide the nascent protein to a cellular target (7). Defining their roles dynamically, mechanistically, and structurally will enrich our view of translation. ■

REFERENCES

- V. Ramakrishnan, *Cell* **159**, 979 (2014).
- D. H. Goldman *et al.*, *Science* **348**, 457 (2015).
- S. J. Kim *et al.*, *Science* **348**, 444 (2015).
- J. L. Chaney, P. L. Clark, *Annu. Rev. Biophys.* 10.1146/annurev-biophys-060414-034333 (2015).
- E. P. O'Brien, P. Cirayam, M. Vendruscolo, C. M. Dobson, *Acc. Chem. Res.* **47**, 1536 (2014).
- C. J. Bustamante, C. M. Kaiser, R. A. Maillard, D. H. Goldman, C. A. Wilson, *Annu. Rev. Biophys.* **43**, 119 (2014).
- F. Gloger, A. H. Becker, G. Kramer, B. Bukau, *Curr. Opin. Struct. Biol.* **24**, 24 (2014).
- K. Ito, S. Chiba, *Annu. Rev. Biochem.* **82**, 171 (2013).
- J. Gumbart, E. Schreiner, D. N. Wilson, R. Beckmann, K. Schulten, *Biophys. J.* **103**, 331 (2012).
- A. Tsai, G. Kornberg, M. Johansson, J. Chen, J. D. Puglisi, *Cell Rep.* **7**, 1521 (2014).
- F. Cymer, N. Ismail, G. von Heijne, *FEBS Lett.* **588**, 1930 (2014).
- N. Ismail, R. Hedman, N. Schiller, G. von Heijne, *Nat. Struct. Mol. Biol.* **19**, 1018 (2012).
- T. R. Noriega, J. Chen, P. Walter, J. D. Puglisi, *eLife* **3**, e04418 (2014).
- X. Li *et al.*, *Nat. Methods* **10**, 584 (2013).

10.1126/science.aab2157



The delicate dance of translation and folding
Joseph D. Puglisi
Science **348**, 399 (2015);
 DOI: 10.1126/science.aab2157

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 23, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/399.full.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/348/6233/399.full.html#related>

This article **cites 14 articles**, 3 of which can be accessed free:

<http://www.sciencemag.org/content/348/6233/399.full.html#ref-list-1>

This article appears in the following **subject collections**:

Biochemistry

<http://www.sciencemag.org/cgi/collection/biochem>

Documenting border conflicts from afar

Civilian researchers find early signs of cross-border violence in satellite images

By Ginger Pinholster

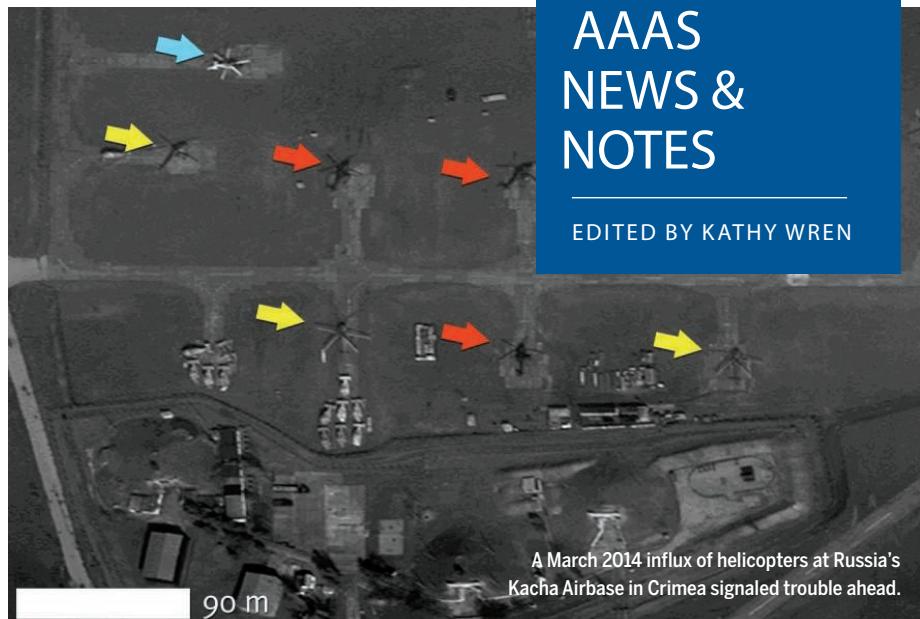
By early April 2014, despite denials by Russian government officials, signs of a coordinated Russian military campaign to annex Ukraine's Crimean Peninsula were falling into place. On 10 April, NATO released commercial satellite images showing a buildup of Russian forces at Novocherkassk and other towns along the border with Ukraine. A Russian military official claimed that the photos showed military drills from August 2013, but AAAS analysis of satellite images from the same time period and locations reinforced NATO's case.

Several weeks earlier, that AAAS research effort had also confirmed reports of military activity in the Ukrainian port city of Sevastopol. The satellite imagery showed, for example, military vehicles parked in depots throughout the area and a tugboat towing a floating chain to control access to the entrance of Sevastopol Bay.

Civilian researchers have long used satellite images to investigate threats to human rights and the environment, but the AAAS study, which was part of a broader effort to document cross-border conflicts, represents a new direction for this type of analysis, said Susan Wolfinbarger, director of the AAAS Geospatial Technologies Project. In addition to revealing destruction to cultural sites or natural resources, satellite images can help "pierce the fog of war" by confirming or overturning reports about what is happening on the ground during conflicts, said Jonathan Drake, a AAAS senior program associate.

This research is "hitting at a huge and very important question," said Noel Dickover, senior program officer at the PeaceTech Lab, which is affiliated with the United States Institute of Peace (USIP). "What if peace-builders had access to geospatial information in near-real time to look at incidents of violent conflicts before they arise? How would that change our world?"

Wolfinbarger and her colleagues spoke at an 11 March event at USIP, which funded seven cross-border conflict studies by AAAS. The researchers compared images



A March 2014 influx of helicopters at Russia's Kacha Airbase in Crimea signaled trouble ahead.

from before and during the escalation of the conflicts, which took place in regions roiled by the end of colonialism in Africa and South/Southeast Asia, and by the fall of the Soviet Union. Together, these images revealed key warning signs of impending war: the movement of materials and troops, as well as the appearance of new infrastructure and military camps.

Evidence of troop movements could be seen in satellite images from the majority of the conflicts studied, said AAAS Program Associate Eric Ashcroft. In Africa, for example, a new road with access to the disputed border between Djibouti and Eritrea seemed to signal the buildup of military activity before a 2008 conflict. Military mobilization of troops was more directly visible before the 2001–02 conflict between India and Pakistan. The sudden emergence of new infrastructure also marked the start of a conflict at the Cambodia-Thailand border in 2008. And, shortly after a 2011 referendum made South Sudan an independent state, "you could see a large amount of materiel moving through the space, presumably to the front lines" at the Sudanese border, Ashcroft said.

For now, the high-resolution satellite images needed for this type of analysis aren't easily available. The first commercial satellite was launched in 1999, and now there are eight offering image resolutions of one meter or less. Costs can be prohibitive, however: a single, small image can cost \$250 or more, and to analyze each location requires (at minimum) an image captured before and after a conflict or event of interest. Commercial satellites also do not acquire images of the whole globe continuously.

In the future, the emergence of smaller "microsatellites" that continuously scan

AAAS NEWS & NOTES

EDITED BY KATHY WREN

the planet may potentially support more automated analysis of regions over time, Ashcroft noted. Those images are consistently captured from a 90-degree angle, he explained, and so they can be precisely aligned on top of each other.

As more images become available for civilian analysis, said Dickover, "you can envision scenarios in the very near future where ad hoc groups would have the ability to crowd-fund the cost of the imagery, and you already see in the humanitarian space where they are doing the analysis themselves. Imagine if that happens to the peace-building world...the implications are fairly vast."

The AAAS Geospatial Technologies Project has posted its seven case studies and other instructional documents for researchers who want to leverage satellite-image analysis, at <http://aaas.org/geotech/borders>. The group has also published an ethics statement about satellite imaging in cultural sites of conflict. ■

AAAS launches new Public Engagement Institute

The newly announced Alan I. Leshner Leadership Institute for Public Engagement will empower cohorts of scientists and engineers to communicate effectively about key issues at the intersection of science and society. As of 10 April 2015, more than 130 gifts and pledges had been committed, exceeding the initial goal of \$500,000 to fund the Institute in its first 5 years. The first class of fellows, planned to convene in 2016, will focus on climate change, with subsequent topics to be determined by a AAAS advisory committee. Give online at www.supportaaas.org/AlanLeshnerFund.

REVIEW SUMMARY

MULTILAYER ASSEMBLY

Technology-driven layer-by-layer assembly of nanofilms

Joseph J. Richardson, Mattias Björnholm, Frank Caruso*

BACKGROUND: Over the past few decades, layer-by-layer (LbL) assembly of thin films has been of considerable interest because of its ability to exert nanometer control over film thickness and its extensive choice of usable materials for coating planar and particulate substrates.

ON OUR WEB SITE

The choice of materials allows for responsive and functional thin films to be engineered for various applications, including catalysis, optics, energy, membranes, and biomedicine. Furthermore, there is now a growing realization that the assembly technologies substantially affect the physicochemical properties and, ultimately, the performance of the thin films.

ADVANCES: Recent advances in LbL assembly technologies have explored different driving forces for the assembly process when compared with the diffusion-driven kinetics of classical LbL assembly, where a substrate is immersed in a polymer solution. Examples of different assembly technologies that are now available include: dipping, dewetting, roll-to-roll, centrifugation, creaming, calculated-saturation, immobilization, spinning, high gravity, spraying, atomization, electrodeposition, magnetic assembly, electrocoupling, filtration, microfluidics, and fluidized beds. These technologies can be condensed into five broad categories to which automation or robotics can also be applied—namely, (i) immersive, (ii) spin, (iii) spray, (iv) electromagnetic, and (v) fluidic assembly. Many of

these technologies are still new and are actively being explored, with research shedding light on how the deposition technologies and the underlying driving forces affect the formation, properties, and performance of the films, as well as the ease, yield, and scale of the processing.

OUTLOOK: Layer-by-layer assembly has proven markedly powerful over the past two decades and has had a profound interdisciplinary effect on scientific research. Scaling up the process is crucial for furthering real-world applications, and moving forward, an understanding of how to carefully select assembly methods to harness the specific strengths of different technologies has the potential to be transformative. Comprehensive comparisons between the technologies still need to be conducted, especially in regard to coating particulate substrates, where comparisons are limited but crucial for advancing fundamental research and practical applications. ■

Australian Research Council (ARC) Centre of Excellence in Convergent Bio-Nano Science and Technology, and the Department of Chemical and Biomolecular Engineering, The University of Melbourne, Parkville, Victoria 3010, Australia.

*Corresponding author. E-mail: fcaruso@unimelb.edu.au
Cite this article as J. J. Richardson et al., *Science* 348, aaa2491 (2015). DOI: 10.1126/science.aaa2491

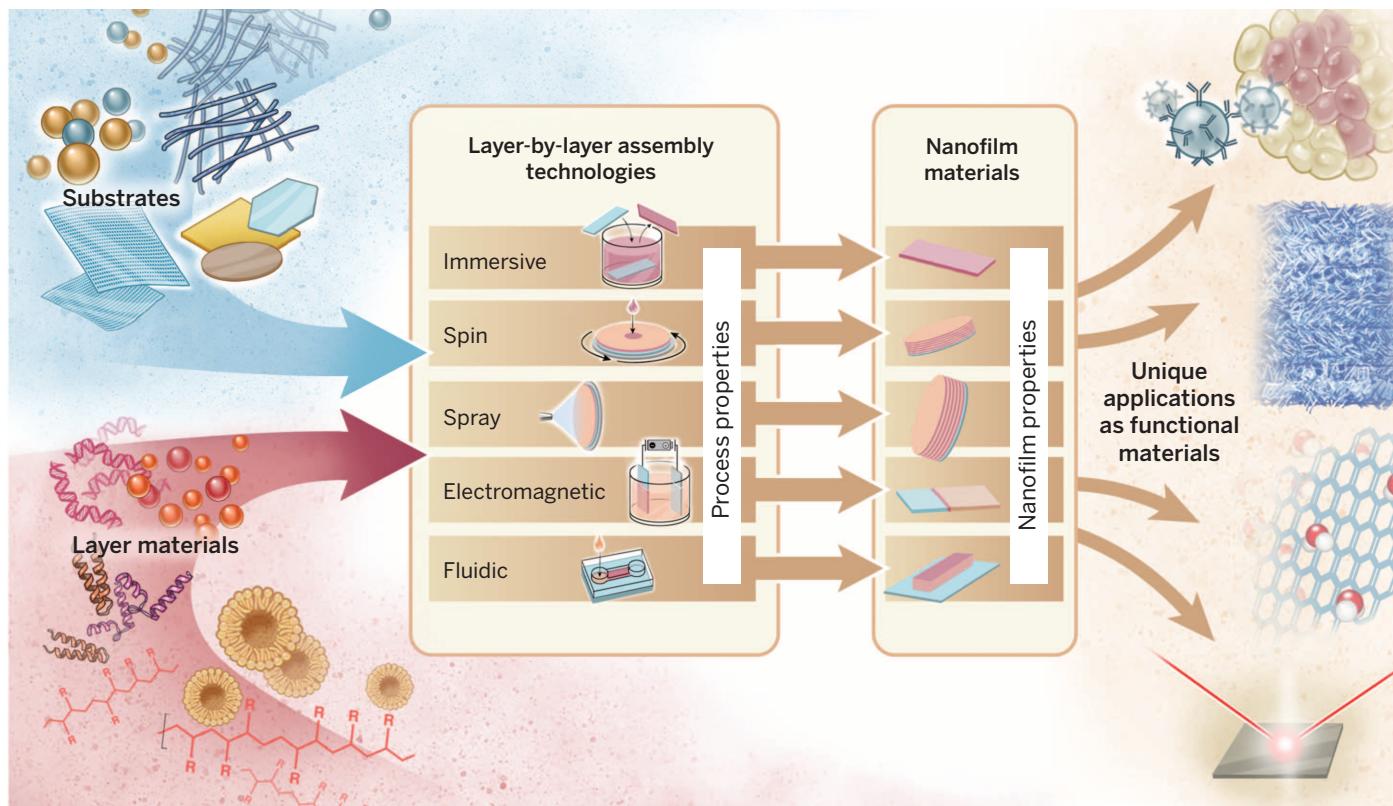


ILLUSTRATION CREDIT: ALISON E. BURKE AND CASSIO LYNN

Layer-by-layer assembly of nanofilms for preparing functional materials. The properties and performance of the resulting films depend on the substrate and layer material choices, as well as the assembly technology.

REVIEW

MULTILAYER ASSEMBLY

Technology-driven layer-by-layer assembly of nanofilms

Joseph J. Richardson, Mattias Björnholm, Frank Caruso*

Multilayer thin films have garnered intense scientific interest due to their potential application in diverse fields such as catalysis, optics, energy, membranes, and biomedicine. Here we review the current technologies for multilayer thin-film deposition using layer-by-layer assembly, and we discuss the different properties and applications arising from the technologies. We highlight five distinct routes of assembly—immersive, spin, spray, electromagnetic, and fluidic assembly—each of which offers material and processing advantages for assembling layer-by-layer films. Each technology encompasses numerous innovations for automating and improving layering, which is important for research and industrial applications. Furthermore, we discuss how judicious choice of the assembly technology enables the engineering of thin films with tailor-made physicochemical properties, such as distinct-layer stratification, controlled roughness, and highly ordered packing.

The performance of functional materials is governed by their ability to interact with surrounding environments in a well-defined and controlled manner. Whether harnessing photons or electrons, separating out gas molecules or solutes, or responding to biomolecules or organisms, the environment-material interface is essential in determining the performance of the materials in various applications. Coating technologies provide the means to control the surface of a material, thus creating composite materials where the interface and the bulk of the material can, to a large extent, be engineered and controlled independently.

Layer-by-layer (LbL) assembly is a prevalent method for coating substrates with functional thin films. Following early studies that reported multilayer assembly (1, 2), it is only in the past two decades that the field has witnessed considerable growth (3). Generally, LbL assembly is a cyclical process in which a charged material is adsorbed onto a substrate, and after washing, an oppositely charged material is adsorbed on top of the first layer. This constitutes a single bilayer with a thickness generally on the order of nanometers, and the deposition process can then be repeated until a multilayer film of desired thickness has been assembled (3). For certain applications the substrate can then be removed, yielding freestanding macroscopic films, such as membranes (4), or freestanding micro- or nanoscopic films, such as hollow capsules (5, 6). Although electrostatic interactions remain widely used in facilitating formation of the films, other molecular interactions (e.g., covalent, hydrogen-bonding, host-guest) are now well established

for LbL assembly, with diverse materials (e.g., polymers, proteins, lipids, nucleic acids, nanoparticles, suprastructures) used as film constituents (7). The simplicity, versatility, and nanoscale control that LbL assembly provides make it one of the most widely used technologies for coating both planar and particulate substrates in a diverse range of fields, including optics, energy, catalysis, separations, and biomedicine (Fig. 1A).

The widespread use of LbL assembly in fields with different standard tools and procedures—as well as the different processing requirements associated with substrates such as porous membranes, particles, and biological matter—has led to the development of a number of LbL assembly technologies. Examples include dipping (2), dewetting (8), roll-to-roll (9), centrifugation (10), creaming (11), calculated saturation (12), immobilization (13), spinning (14), high gravity (15), spraying (16), atomization (17), electrodeposition (18), magnetic assembly (19), electrocoupling (20), filtration (21), fluidics (22), and fluidized beds (23). These different methods have often been treated as “black boxes,” where the main focus has been on what materials are used (the input) for assembling the thin films (the output), with little focus placed on the actual assembly method. However, there is now a growing realization that the assembly method not only determines the process properties (such as the time, scalability, and manual intervention) but also directly affects the physicochemical properties of the films (such as the thickness, homogeneity, and inter- and intralayer film organization), with both sets of properties linked to application-specific performance (Fig. 1B).

Unpacking the “black box”

The basis of LbL assembly is the sequential exposure of a substrate to the materials that will compose the multilayer films. The assembly tech-

nologies used to assemble such films form five distinct categories, namely: (i) immersive, (ii) spin, (iii) spray, (iv) electromagnetic, and (v) fluidic assembly (Fig. 2). These assembly technologies affect both the process properties and the resultant material properties (Table 1), and therefore careful choice of the assembly method can be crucial for successful application of the assembled films. Furthermore, two main themes can be identified for current developments in assembly technologies: The first is the move away from random diffusion-driven kinetics for layer deposition, and the second is the advancement from manual assembly toward automated systems.

Immersive assembly

Immersive LbL assembly, sometimes referred to as “dip assembly,” is the most widely used method and the standard that newer technologies are often compared against. Immersive assembly is typically performed by manually immersing a planar substrate into a solution of the desired material (2, 24, 25), followed by three washing steps to remove unbound material (26). Particulate substrates can also be layered using immersion; however, the washing and deposition steps are generally broken up by centrifugation to pellet the particles (5, 6, 10). Early studies on using particles for depositing planar multilayers noted that, theoretically, any material capable of having a surface charge (such as metals, nonmetals, organics, and inorganics) could be applied for growing multilayers if suitable conditions are used (2, 24, 27). Further, it was also reported that the thickness of each layer corresponds to the thickness of the particles being adsorbed (24, 28). Immersive assembly allows for more homogeneous films [when using either particle (27) or polymer multilayers (3)] in comparison with non-LbL assembly technologies such as gas deposition and nucleation deposition, making LbL assembly widely used for thin-film formation.

Improvements in immersive assembly include speeding up the process by shifting the deposition kinetics away from random diffusion toward faster kinetics, such as those arising from dewetting (8), and by automating labor-intensive steps with robotic immersion machines (24, 26, 29, 30). The colloids used for planar assembly in early studies required only 1 min of immersion for each adsorption step (31); however, for immersive assembly using polymers, the substrate is ideally immersed for ~15 min for sufficient layer deposition (25, 26). To reduce the assembly time for polymers and to allow for the deposition of low-surface charge and/or small-contact area materials, solutions doped with organic solvents (e.g., dimethylformamide) can be used to eliminate the need for rinsing and drying steps through the process of dewetting (8). Dewetting leads to a ~30-fold reduction in assembly time because the adsorption process is no longer governed by diffusion but by evaporation and dewetting. Another interesting move away from random diffusion utilizes polymer solutions that are constantly stirred by magnetic stirrer bars, which allows for robust layers to be deposited within

Australian Research Council (ARC) Centre of Excellence in Convergent Bio-Nano Science and Technology, and the Department of Chemical and Biomolecular Engineering, The University of Melbourne, Parkville, Victoria 3010, Australia.
*Corresponding author. E-mail: fcaruso@unimelb.edu.au

tens of seconds after immersion (32). Instead of speeding up the adsorption process by using different adsorption kinetics, handling times can be decreased by automating the process (9, 24, 26, 27, 29, 30, 33). One approach to automation uses a quartz crystal microbalance (QCM) as a substrate, allowing for layering to be controlled with a computer-monitored feedback loop (30). QCM enables the layering process to be based on a fixed mass of adsorbed material rather than fixed immersion times. Furthermore, the feedback loop allows for precise and reproducible control over the film growth and allows for linear film growth to be engineered from polymer combinations that give nonlinear film growth using fixed times (30). For fixed-time immersive assembly, computer-programmed automated slide stainers can be retrofitted for automated multilayer assembly, allowing for agitation and solution exchange during washing steps (24, 29). A similar, although custom-built, computer-programmed machine can deposit ~1000 layers of charged colloids onto particulate substrates (substrates ~100 μm in diameter) (27).

Although automation decreases manual involvement, it does not substantially reduce the overall assembly time, which is why some efforts have focused on combining faster deposition kinetics with automated systems. For example, one commercially available robot uses a rotating slide holder to speed up the assembly process (26). This rotation allows for a 3- to 10-fold reduction in adsorption times and allows for thicker films to be prepared using higher rotation speeds. Roll-to-roll assembly also allows for layering to be performed faster (by 5- to 10-fold), through the use of flexible substrates (9). The immersion time and speed of the rolling process play a large role in determining the film properties, and the drying conditions, wettability, and substrate movement speed require optimization to produce films with similar properties to standard immersive assembly (9). A further improvement to roll-to-roll assembly uses a nip-roll technique to prevent excess solution from cross-contaminating the system, resulting in more homogenous coatings than immersive assembly (34).

Immersive assembly can be performed on particulate substrates that are too small to sediment quickly or physically move between solutions, such as micro- and nanoparticles. The most common technology for immersive assembly on particulate substrates is performed by adding polymer solution to dispersed dense particulate substrates, pelleting the particles with centrifugation, removing the supernatant, washing multiple times with a similar pelleting process, and then repeating the steps for multilayer growth (5, 6, 10). This is generally time-consuming and labor-intensive due to the centrifugation steps, and particles dense and large enough to be pelleted are required. However, by using particulate substrates lighter than water (e.g., emulsions), creaming and skimming cycles can be applied for washing steps (11), although centrifugation can also be used to speed up the flotation and creaming process (35), with lighter emulsions

capable of creaming in a matter of minutes rather than hours (36). The use of emulsions as templates results in thicker films compared with using solid templates, probably due to

the surfactants used for emulsion stabilization (36).

The major driving force behind the development of immersive assembly technologies for

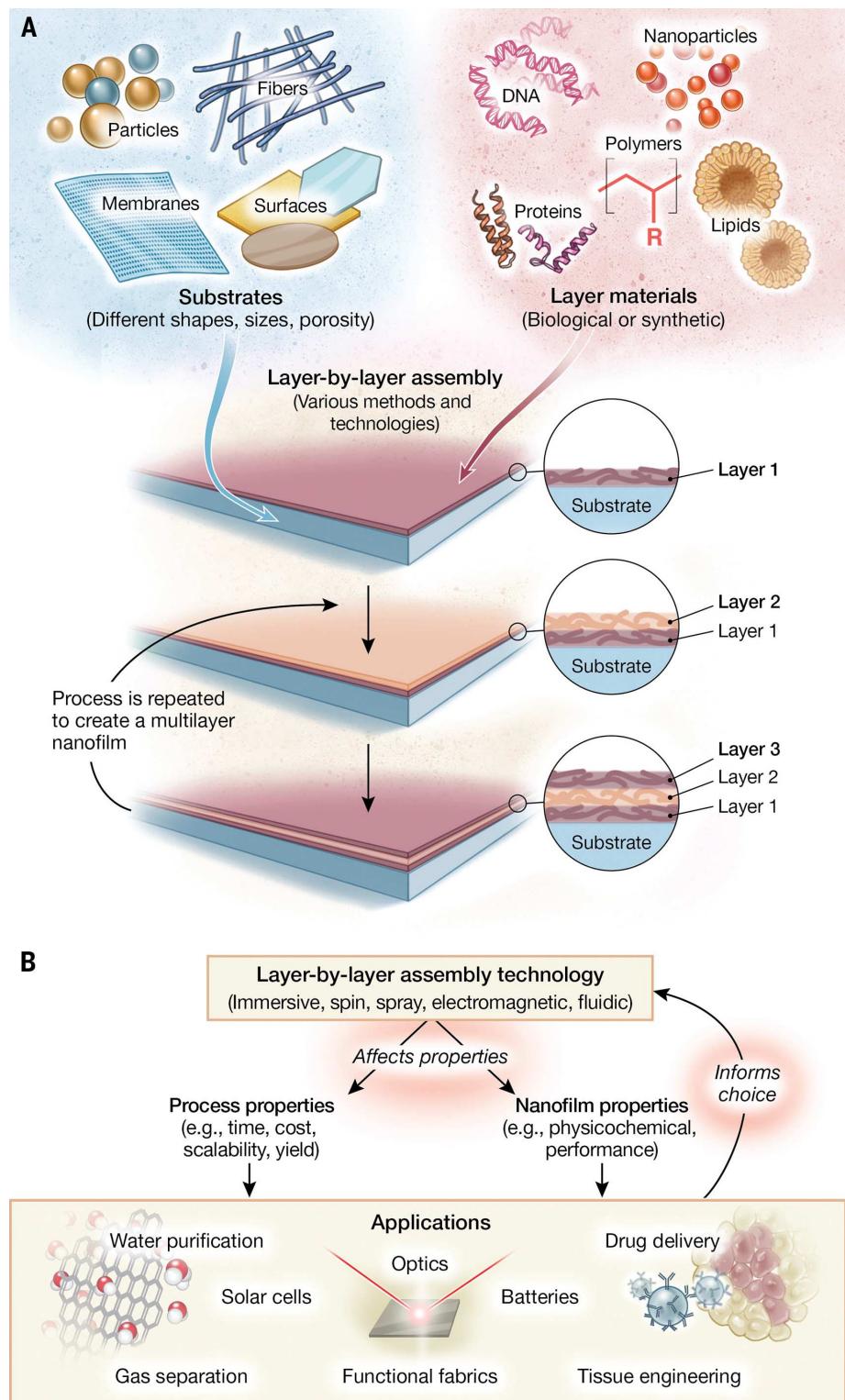


Fig. 1. Versatility of layer-by-layer assembly. (A) Schematic overview of LbL assembly and (B) an overview showing that the assembly technology influences film and process properties, as well as application areas. [Illustration credit: Alison E. Burke and Cassio Lynn]

particulate substrates is the attempt to avoid centrifugation, as it can lead to aggregation, is labor intensive, and is generally difficult to automate. A simple way to avoid centrifugation is to remove the need for washing steps. This can be achieved by adding exact amounts of polymer

calculated to saturate the surface of the particulate substrates (12, 37), rather than the high concentrations of excess polymer solution generally used (5, 6, 10). Initially, only two to three layers could be deposited before the particles start to aggregate (12), but more layers can be

deposited by incrementally measuring the zeta potential during assembly (37). Additionally, the use of constant mixing for soft particulate substrates such as emulsions (38) or sonication during layer deposition for hard particulate substrates like drug crystals (39) reduces aggregation. By optimizing the protocol, the saturation method gives a similar shell thickness to centrifugation-based assembly but is about three times faster (37, 39). This technology requires constant monitoring and surface area calculations to avoid adding excess polymer and therefore does not reduce manual involvement. A technology that focuses on decreasing manual involvement and reducing the need for centrifugation uses particulate substrates immobilized in agarose to convert collections of particulate substrates into a macroscopic substrate (33). This macroscopic collection of immobilized particles can be treated like a planar substrate and immersed in polymer solutions using a robotic dipping machine, allowing for full automation during the layering process. Although this technology generates films roughly half the thickness of those prepared by conventional centrifugation-based assembly, probably due to the impeded diffusion of polymers through the agarose hydrogel, ~80% of the particles can be recovered, which is an improvement over the ~90% loss that has been reported for centrifugation-based assembly at high layer numbers (21, 33).

Due to the ease of use and versatility of material and template choice, immersive assembly has been applied for numerous applications. For example, light-emitting diodes (LEDs) can be prepared from immersive assembly on planar substrates, with the polymer choice and multilayer thickness giving control over luminance and the turn-on voltage (29). Automated roll-to-roll immersive assembly can be used for depositing conductive and flame-retardant coatings (34). Planar substrates coated with particle multilayers can be used for the detection of small particles invisible to the naked eye through color shifts in the multilayer films (2). Glass slides can also be coated with particle multilayers for the preparation of antireflective, antifogging, and self-cleaning surfaces (24). Fusion microreactors coated with particles are more conducive toward reaction (27). Certain particulate substrates easily allow for the removal of the template particle, leaving behind hollow multilayer capsules. Similarly, drugs themselves can be used as the particulate templates, with both types suitable for drug delivery, (5, 6, 33, 38–40).

In summary, immersive assembly is the most commonly used LbL assembly technology and the de facto standard against which other technologies are compared. The simplicity of immersing substrates of almost any shape or size into containers with layering solution makes this technology easily accessible. The films produced have an interpenetrated structure and form “fuzzy” nanoassemblies that are almost synonymous with LbL assembly (3). Much recent work has been focused around shorter assembly times

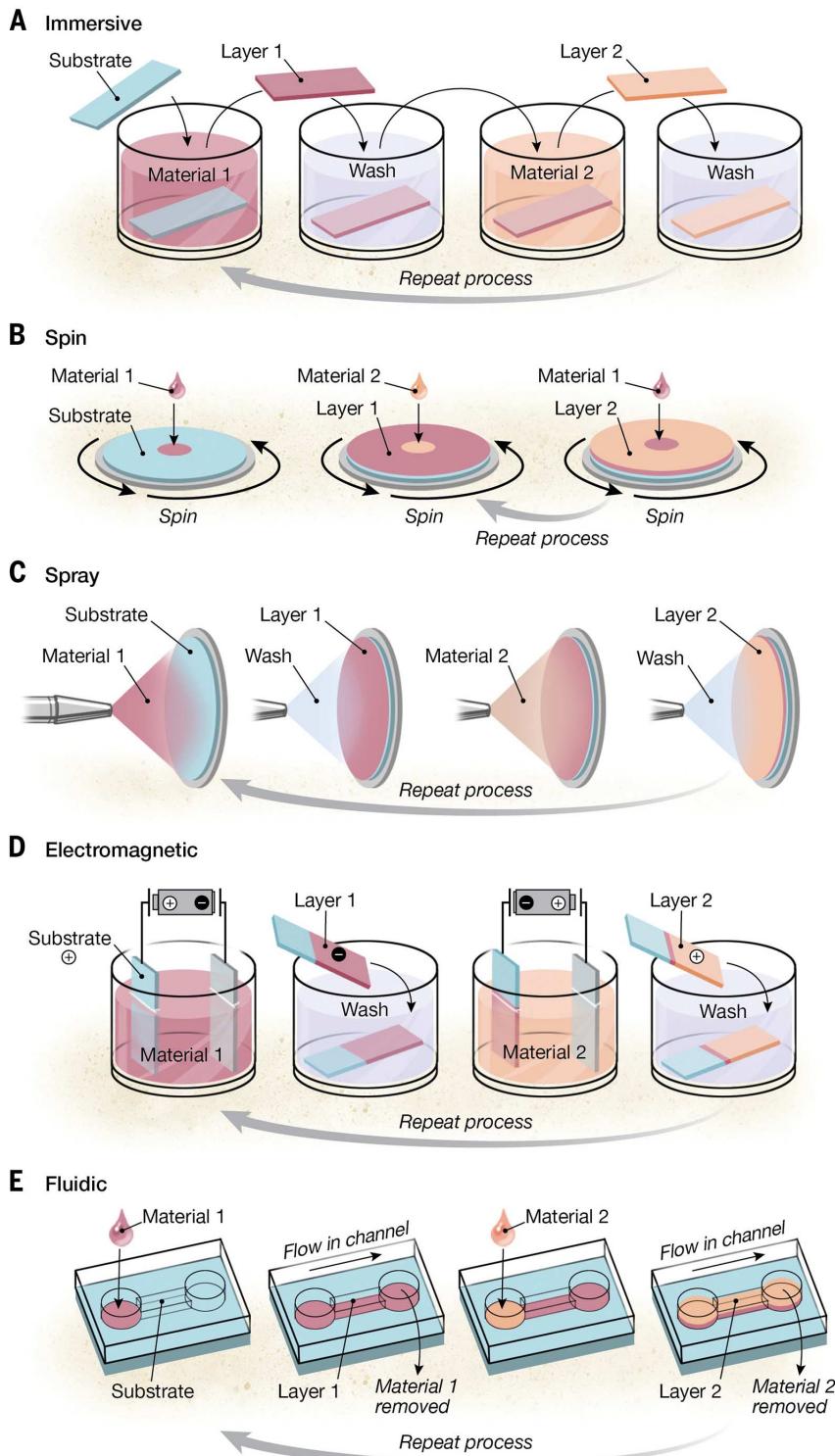


Fig. 2. Layer-by-layer assembly technologies. (A to E) Schematics of the five major technology categories for LbL assembly. [Illustration credit: Alison E. Burke and Cassio Lynn]

and automated systems with less manual intervention. For coating particulate substrates, there has also been considerable interest in technologies applicable to coating smaller, low-density particles (such as silica nanoparticles), which can be difficult to handle with the conventional centrifugation-based assembly. As immersive assembly typically requires more material than other technologies, especially to submerge large substrates on industrial scales, waste can be an

issue, although solutions can be reused as long as cross-contamination remains low. Immersive assembly has been the workhorse of LbL assembly and will undoubtedly continue to play an integral part in the development of new and improved thin films.

Spin assembly

Layer-by-layer assembly using spin coating (i.e., “spin assembly”) utilizes the common coating

technology of spinning a substrate to facilitate the deposition of materials (44). Although drying a substrate after immersive LbL assembly can be achieved through spinning (44), the majority of spin assembly is performed by either casting the solution onto a spinning substrate (42) or casting the solution onto a stationary substrate that is then spun (43). Spinning quickens the assembly process considerably, allowing for layers to be deposited in ~30 s due to the various

Table 1. Selected LbL assembly technologies and properties arising from using each technology. The table is intended to provide a general overview and is not exhaustive. NR indicates not reported in selected references. [Illustration credit: Alison E. Burke and Cassio Lynn]

LbL ASSEMBLY TECHNOLOGY	BUILDING BLOCKS		PROCESS PROPERTIES		NANOFILM PROPERTIES			
	Substrates	Substrate sizes*	Layer materials	Time per layer	Automated (current status)	Layer thickness, nm†	Roughness, nm	Layer structure
Immersive		10 nm-1 m		10 s-12 h		<1-15	1-20	Interpenetrated
Dipping (2, 25)	Planar	1-100 mm	Polymers, colloids	10-30 s (32) or 10-20 min‡ (26)	Yes	1-2 (25, 26)	1-10 (44, 49)	
Dewetting (8)	Planar	1-10 mm	Polymers, colloids§	30-60 s (8)	No	1-2 (8)	NR	
Roll-to-roll (9)	Flexible planar	100 mm-1 m	Polymers	2-5 min (9, 34)	Yes	1-15 (9, 34)	15-20 (9, 34)	
Centrifugation (5, 6)	Particulate	10 nm-10 µm	Polymers, colloids	20+ min (5, 6)	No	1-2 (5, 10)	3-10 (23, 94)	
Calculated saturation (12, 37)	Particulate	100 nm-1 µm	Charged polymers	5-10 min (39)	No	1-2 (39)	NR	
Immersive immobilization (33)	Particulate	100 nm-1 µm	Polymers	40-50 min (33)	Yes	<1 (33)	NR	
Creaming (11)	Emulsion	10 nm-1 µm	Polymers, colloids	0.5-12 h (11, 36)	No	1-7 (36, 38)	NR	
Spin		1-100 mm		10 s-5 min		<1-2	1-10	Stratified
Spin (14, 42, 43)	Planar	1-100 mm	Polymers, colloids	10-60 s (43)	Yes	<1-2 (42)	1-10 (44, 46)	
High gravity (15)	Planar	1-10 mm	Polymers, colloids	20 s-5 min (15, 49)	No	NR	1-2 (49)	
Spray		10 nm-10 m		<1 s-24 h		<1-15	1-10	Stratified
Spray (16, 52)	Planar	1 mm-10 m	Polymers	<1-30 s (59)	Yes	<1-5 (16, 51)	1-10 (51)	
Atomization (17)	None	10-100 nm	Charged polymers	12-24 h (17)	No	5-15 (17)	NR	
Spray immobilization (63)	Particulate	10-100 nm	Polymers	5-10 s (63)	Yes	2-4 (63)	NR	
Electromagnetic		10 nm-100 mm		1 s-20 min		1-20,000	10-30	Stratified
Electrodeposition (18, 20, 67, 73)	Planar	1-100 mm	Polymers, colloids	1 s-20 min¶ (65, 71)	No	2-20,000¶ (20, 67, 71)	10-30 (66, 70)	
Magnetic (19, 76)	Planar and Particulate	10 nm-100 mm	Polymers, colloids	15-20 min (19, 76)	No	1-2 (19, 75)	NR	
Electro-immobilization (13)	Particulate	10 nm-1 µm	Charged polymers	15-20 min (13)	No	2-3 (13)	NR	
Fluidic		100 nm-100 mm		10 s-45 min		<1-3	1-11	NR
Microfluidic planar (22)	Planar	10 µm-100 mm	Polymers	1-15 min (80, 81)	Yes	<1-3 (83, 84)	1-10 (80, 84)	
Microfluidic particulate (96-98)	Particulate	100 nm-10 µm	Polymers	10-60 s (96-98)	Yes	1-3 (96, 97)	NR	
Fluidized bed (23)	Particulate	1-10 µm	Polymers	3-5 min (23)	No	2-3 (23)	9-11 (23)	
Fluidic immobilization (86, 87)	Particulate	100 nm-1 µm	Polymers, colloids	5-45 min (86, 87)	No	1-2 (87)	NR	
Vacuum/filtration (21)	Particulate and fragile**	100 nm-1 µm	Polymers	10-20 min (95)	Yes	1-2 (94)	5-10 (94)	

*Typical order of magnitude substrate sizes are indicated. Larger or smaller substrate sizes are possible. †Typical thicknesses per layer for linearly growing films are indicated. Per-layer thicknesses for exponentially growing films vary widely, given the nonlinear growth profile. ‡Time with or without agitation, respectively. §Dewetting can make use of materials that are usually difficult to layer (e.g., materials with low charge or with low surface contact). ||Centrifugation processing time is highly variable due to manual pipetting and resuspension steps. ¶Thickness is dependent on time. **Fragile substrates, such as mammalian cells, can be layered using fluidic filtration.

forces governing the process (43). Furthermore, spin assembly allows for automation and the coating of substrates up to 10 cm in diameter using commercially available spin coaters (44, 45). However, standard spin coaters are generally designed for flat surfaces and are not amenable to the complex shapes accessible to immersive assembly.

Spin assembly typically results in more homogeneous films compared with immersive assembly. This is because assembly is driven by a collection of forces including electrostatic interactions, which cause the adsorption and rearrangement of polymers, and centrifugal, air shear, and viscous forces, which cause desorption of weakly bound polymers and dehydration of the films (43). These forces are also the reason why spin assembly can be orders of magnitude faster than immersive assembly. The salt concentration of the polymer solution has a larger effect at higher spin speeds, meaning that electrostatic forces play a greater role at low ionic strength, and shear forces dominate at high ionic strength (46). These shear

forces produce thinner, highly ordered films with specific layer interfaces when compared with immersive assembly, which produces thicker interpenetrated films (47). Specifically, the thickness for spin-assembled polymer films is generally linked to the spin speed, with higher speeds leading to thinner films (42). When depositing colloids, the forces experienced during spinning lead to a monolayer of colloids, whereas standard immersive assembly often leads to a pseudo-monolayer in which the substrate is not fully coated (43, 48). A comparison study of the differences between automated immersive assembly and automated spin assembly found that immersive-assembled prepared thicker, rougher films, whereas spinning resulted in thinner, smoother films (44). The films differed visually, as the spin-assembled films were transparent because of their distinct-layer stratification, and the immersive-assembled films were opaque due to their inhomogeneous, interpenetrated layers (Fig. 3). The contact angle and the relative

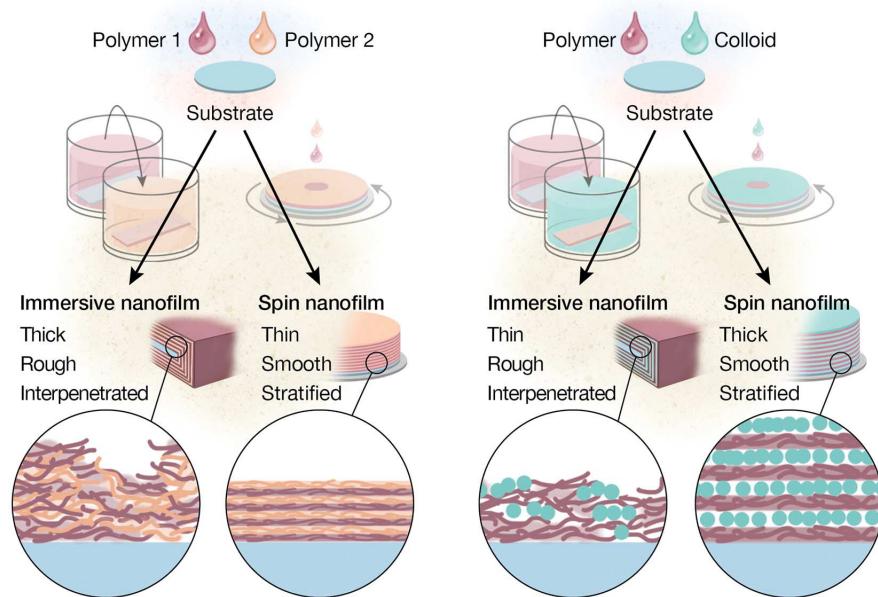
concentration of polymers were consistent across all bilayers for spin-assembled films, whereas immersive-assembled films became rougher with time, giving varying contact angle and relative concentration ratios between the two constituent layers (44). Another study, which compared spin assembly and immersive assembly, showed that clay nanocomposites in spin-assembled films have a higher degree of orientation (45). However, one issue that can result from spin assembly, which is not a concern for other assembly technologies, is that at higher ionic strengths of polymer solution, and also at lower spin speeds, the films can be thicker where the solution was cast when compared with the edges of the substrate (42, 46).

In a special case of spin assembly, the substrate can be placed in a closed container with a polymer solution or a colloidal dispersion parallel to the axis of rotation (rather than perpendicular). Upon spinning, centrifugal force pushes the layer material directly onto the substrate rather than across the substrate, hence the name “high-gravity assembly” (45). This allows for improved film deposition and uniformity, especially at low polymer concentrations, because the rotation and increased turbulence lower the thicknesses of both the laminar layer and the diffusing layer around the substrate. The adsorption equilibrium can be reached at least five times faster than immersive assembly and is controllable by the spin speed. Furthermore, polymer combinations that grow exponentially using immersive assembly also grow linearly using this technology. Similarly, the roughness is much lower (~2- to 10-fold) for LbL films assembled in this way (49).

Spin assembly typically produces substantially more organized films and multilayers than immersive assembly, which has made it a useful tool in preparing optical coatings with controllable and homogenous color (14) and for preparing transparent films (44). Similarly, spin assembly is useful for preparing LEDs with higher luminance than immersive assembly (41). A primary limitation for spin assembly in terms of application is that it is limited to coating small planar substrates, as increasing the substrate size requires higher spin speeds. Furthermore, spin coating of nonplanar surfaces is complicated.

In summary, spin assembly uses rotating substrates to deposit layers and remove excess coating material. Spin assembly typically produces thinner, more organized, and more stratified multilayers than does immersive assembly, and the process can be much faster. The spin coater needed for assembly is commonly accessible in many research environments and even some industrial settings, such as with the robotic wafer processing common in the semiconductor industry, which could facilitate translation from the laboratory to real-world applications. Furthermore, depositing multilayer films on nonflat surfaces, or even flat but rough surfaces, can be challenging due to the shear forces involved with film assembly. Nevertheless, the film and process properties arising from spin assembly, including smooth films assembled in a relatively short

A Creating different nanofilm properties using immersive vs spin LbL assembly



B Examples of polymer-polymer nanofilms using immersive vs spin LbL assembly

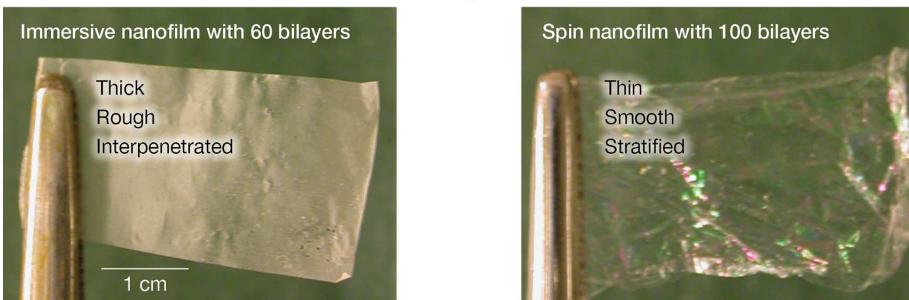


Fig. 3. Comparison between immersive and spin assembly. (A) Schematic comparison with layers of different materials. (B) Comparison between an immersive-assembled film (left) and a spin-assembled film (right). Films are made of hydrophobically modified poly(ethylene oxide) and poly(acrylic acid). [Adapted with permission from (44). Copyright 2008 American Chemical Society.] [Illustration credit: Alison E. Burke and Cassio Lynn]

time, continue to make this method an attractive choice.

Spray assembly

Spray LbL assembly is another assembly category, where films are assembled by aerosolizing polymer solutions and sequentially spraying them onto substrates (16). Although spraying air has been used to dry films during LbL assembly to reduce contamination (and align carbon nanotubes) (50), here we discuss spray assembly solely in the context of layer deposition. Standard spray assembly is much faster (as quick as ~6 s per layer) than immersive assembly (51) and approaches an industrial level far surpassing that of spin assembly (52, 53). Vacuum can also be used to further speed up the process by minimizing the lag time between spraying and washing, and vice versa, and to facilitate the spray coating of three-dimensional (3D) objects like membranes (54).

In spray assembly, the film properties—such as the morphology, uniformity, chemical composition, and selective membrane properties—can be tailored to be similar to those prepared by immersive assembly, with the film thickness influenced by suspension concentration, spray flow rate, spray duration, resting duration, whether or not the substrate is washed and for how long, and whether the solution is sprayed vertically or horizontally (16, 51, 55–57). This control arises from the two main forces governing the spray assembly process, namely, bulk movement in the actual spray and random movement in the liquid film (56). The random movement in the liquid film allows for polymer rearrangement and generates much higher convection close to the substrate, allowing for improved deposition. This is because of the submicron thickness of the liquid film at the substrate interface and because of the speed at which the spray contacts the substrate (16, 56). Washing the substrate generally produces thicker films than leaving the substrate unwashed, due to polymer rearrangements during washing (51). Like spin assembly, the films resulting from spray assembly have more distinct layers in comparison to immersive assembly (16, 53).

Spray assembly has also been combined with other technologies to leverage technology-specific advantages and automate the assembly process. For example, a disadvantage of spray LbL assembly is that the obtained films may not be homogeneous due to the effects of gravity draining, causing increased deposition in the vicinity of the solution drips, and because of irregular patterns caused by the spray nozzles at certain distances (51, 57). To address this problem, rotating the substrate during spray assembly allows for the preparation of more homogeneous films and subsecond spray times for each layer (56, 58, 59). By spraying rotating substrates, a majority of the polymer added to the substrate is adsorbed. In comparison, the vast majority of polymer remains in the coating solutions after immersive assembly. Therefore, applicable concentrations roughly 10 to 50 times less than those required for immersive assembly can be used for spray

assembly on rotating substrates (55, 59). Larger 3D substrates, such as tubular membranes, can also be coated by rotating the substrate during spraying (60). A further improvement has been the computer-aided automation of spray assembly to reduce manual processing (58, 60). Similar to the use of automated immersive assembly on QCM chips, the use of QCM chips for automated spray assembly enables feedback loop control and tracking of real-time film growth (61). Automated spray assembly has also been combined with roll-to-roll processing for coating industrial-size substrates (i.e., substrates that are tens of meters long) (62). Roll-to-roll spray assembly can also be used to coat particulate substrates with multilayer films by performing spray assembly on particulate substrates immobilized on top of a dissolvable surface (63).

A stand-alone spray assembly technology for coating particulate substrates uses surface acoustic waves of 1 to 10 nm in amplitude to atomize polymers and cargo (17). As the atomized droplets move through the air, the solvent evaporates and the polymer condenses into particle form, resulting in the first atomized solution becoming the template for subsequent coatings, with ~1000 carriers produced from each microliter of solution. The particles are dialyzed to remove excess polymer, added to a solution of oppositely charged polymer, and then re-atomized to coat the particles. This process can be repeated for multilayer assembly; however, the dialysis process increases the processing time of this technology to ~24 hours for each layer.

Spray assembly has found use for a wide variety of applications because it can be used to coat industrial-scale substrates with relative ease (62) and is not limited to planar substrates (54, 60, 63). Spray assembly has been used to prepare flame-retardant films over cotton cloth, where it was shown that spraying on vertically oriented substrates produced superior flame-retardant films compared with both spraying on horizontally oriented substrates or dipping (57). Clothing material was also coated with spray assembly to control air flux and provide chemical protection, potentially for use with military uniforms (54). Like other assembly technologies, spray assembly has been used to prepare antireflective coatings (61), and similarly, car tinting with structured coloring to reduce heating from infrared light (62). Membrane tubes could also be coated to improve the separation of organic dyes from water (60). Because the structure of the films can be controlled at the nanometer-level by the spray time, spray assembly can be used to control conductance in thin films in ways that are not available to other assembly technologies (59). Spray assembly has been used to prepare particles to examine cellular uptake of different coatings and aspect ratios of particles (63) and for gene delivery *in vitro* (17). Spray assembly has found use in diverse applications and industry because it offers rapid assembly times and is amenable to both automation and scale-up.

In summary, spray assembly produces multilayer films by aerosolizing coating solutions and

spraying them onto the substrate. The resulting films are typically well organized with distinct layers. Spray assembly is a quick and easy method to coat large or nonplanar substrates, although immersive assembly remains the method of choice for coating complex 3D substrates. Spray assembly is one of the most highly relevant technologies for industrial applications, as it is already widely used in industry.

Electromagnetic assembly

Electromagnetic assembly is based on the use of an applied electric or magnetic field to effect laying, such as by coating electrodes in polymer solutions or by moving magnetic particulate substrates in and out of coating solutions (18, 19). The former, commonly referred to as electrodeposition, is a well-established technology for coating materials using an applied voltage in electrolytic cells. In the standard electrodeposition setup, two electrodes are immersed in polymer solution, then an electric current is applied. The electrodes are then washed and placed into solution of an oppositely charged polymer; the polarities of the electrodes are reversed, and the process is repeated (64). Electrodeposition can be used to rapidly assemble ions, polymers, and colloids in much less time than in immersive assembly (18). For example, bimetallic mesoporous LbL films can be prepared by electrodeposition, with the electrodeposition time determining the layer thickness at ~1.5 nm/s (65). In another setup, the substrate can be placed between the two electrodes, allowing for planar substrates to be coated (66), or even immobilized particles (13). This technology results in films roughly twice the thickness of those resulting from centrifugation-based assembly. Electrodeposition can also use higher voltages, upwards of 30 V (13, 66); however, the assembly process for immobilized particles can take as long as 15 min per layer (13).

The thicknesses of the electrodeposited films are directly related to the voltage used during assembly, with the optimum voltage for achieving the thickest films dependent on the pH of the polymer solution (67). Higher voltages can cause desorption of the film as the electrode (i.e., the substrate) begins to repel the previously deposited layer. Generally, pH values lower than the pK_a (where K_a is the acid dissociation constant) of the polymers need lower voltages to reach peak thickness, and that peak thickness is also larger than the peak thickness at higher pH, closer to or above the pK_a of the polymers. However, if the voltage is raised high enough, a secondary peak thickness can be reached, allowing for the assembly of films at pH values otherwise difficult to grow using other technologies (67). The reason for this “valley” in thickness is that at high voltages, the electrolysis of water at the electrode plays a bigger role in hindering polymer adsorption; however, at even higher voltages (>3 V) the electrostatic interaction between the polymer and electrode exceeds any hindrance due to electrolysis (67). For example, polymer-enzyme films are roughly twice as thick when assembled at an optimal voltage of

1.2 V, compared with using lower or higher voltages (below 3 V) (68). These studies show that electrodeposition is similar to other LbL assembly technologies in the sense that the pH of the polymer solution, and therefore the configuration of the polymer itself, is crucial in controlling film thickness.

At higher voltages, electromagnetically assembled films are more interpenetrated than immersive-assembled films, which is in contrast to the highly stratified films prepared at lower voltages (66). When forming polymer-polymer films, the refractive index does not change substantially during film growth in a flow cell under an electric current, suggesting a more homogeneous deposition than in immersive assembly (69). Similarly, polymer-colloid films show high organization, as the refractive index decreases and transmittance increases when assembling films under higher voltages (70). Correspondingly, electrodeposited enzyme-polymer films are more uniform than immersive-assembled films, with 90% coverage of the substrate versus ~50% coverage, respectively (68). Because of this stratification and high surface coverage, electrodeposition allows for control over the spacing between layers (68).

Electrodeposition can also be achieved by using local effects at the electrodes, such as inducing redox reactions or changes in pH. The pH of the solution near the anode and cathode changes markedly from bulk solution to lower and higher pH values, respectively (71). The low pH near the anode can induce polymer deposition. However, this pH-induced electrodeposition is fairly limited, as only a few bilayers can be deposited (using materials such as alginate and chitosan) because the layers become too thick (tens of micrometers) for the electric current to penetrate, resulting in no pH change and therefore no deposition. Using a similar principle, covalently stabilized films can be prepared by generating copper(I) from copper(II) in situ at the electrode (i.e., the substrate) for cross-linking azide- and alkyne-containing polymers with copper-catalyzed “click” reactions (20). Polymers containing electrically sensitive click groups can also be electrocoupled, allowing for 500-nm-thick transparent, and therefore stratified and homogenous, films to be prepared in ~30 min (72). One-pot synthesis can be performed using the same basic principles by switching between oxidative and reductive reactions by alternating the voltage, allowing for wash-free assembly using electropolymerization (73).

Magnets, rather than electric currents, can be used to assemble LbL films on sensitive particulate templates, such as emulsions (74), or small templates difficult to pellet through centrifugation, such as sub-10-nm iron oxide nanoparticles (19). Template particles containing magnetic nanoparticles can be separated from the polymer solution using a magnet, which, similar to the filtration method, allows for nearly 100% of the particles to be recovered in a centrifugation-free LbL assembly process (75). Magnets or an external field can also be used to orient layered

magnetic nanoparticles on a planar substrate so that a subsequent layer of nanoparticles can deposit more rapidly and in an oriented fashion (76). This technology uses standard immersive assembly for the deposition of positively and negatively charged magnetic nanoparticles with application of a magnetic field between deposition steps. Therefore, the thickness does not increase in relation to standard immersive assembly; however, the absorbance of the film increases with application of the magnet, suggesting increased packing (76).

Electromagnetic assembly has found use in several different applications, as it can be used to form LbL films with compositions that are not readily assembled using other technologies. Bimetallic films of Pt and Pd layers have Brunauer-Emmett-Teller surface areas of ~40 m² g⁻¹ and therefore exhibit enhanced electrochemical activity in the methanol oxidation reaction, compared with single-layer films (65). Antireflective coatings can be prepared by adjusting the refractive index of the films by assembling the films at different voltages (70). Biological applications have also been explored, as biocompatible coatings can be formed using electromagnetic assembly, with *in vitro* tests confirming negligible cytotoxicity (71). Bienzyme films with bioelectric catalytic properties have higher surface coverage, and therefore activity, when compared with traditionally prepared films (68). The stratification of the assembled films is also conducive toward high-performance photoelectric devices (72) and separation membranes (66). Hollow polymer capsules (from micrometers to sub-100 nm in diameter) can also be prepared using electrodeposition on immobilized particles (13).

In summary, electromagnetic assembly uses electric or magnetic fields, typically in the form of electrodes in polymer solutions or magnetic particulate substrates, to deposit films. Electromagnetic assembly can exploit current-induced changes in pH or redox-reactions to effect film assembly, thus using a driving force substantially different from that of the other main assembly categories. Generally, electromagnetic-assembled films are thicker and more densely packed than films prepared using other LbL assembly methods (13, 68). Electromagnetic assembly is still not as common as some of the other technologies, and even though it requires special equipment and expertise, it does offer a different approach to multilayer film assembly (e.g., through magnetic handling of substrates and materials or through electrically induced assembly), thereby providing alternative opportunities for assembling films.

Fluidic assembly

Fluidic assembly can be used to deposit multilayers with fluidic channels, both by coating the channel walls and by coating a substrate placed or immobilized in a fluidic channel (77). The general method involves using pressure or vacuum to sequentially move polymer and washing solutions through the channels, which can be fluidic components, such as tubing or capillaries, or designed microfluidic networks (78, 79). Flow-

chamber-based QCM is a common fluidic assembly technology used for investigating thin-film properties and multilayer growth by providing crucial real-time information (22). Higher concentrations of polymer solution typically yield thicker films (79), with the contact time rather than the flow rate as the crucial factor determining the amount of adsorbed polymer under flow (80).

Fluidic assembly is typically implemented using a pump, capillary forces, or spinning to transport the liquid through the channels, although pipetting and static incubation can also be used. However, fluidic assembly strongly resembles immersive assembly when polymer solutions are allowed to remain in static contact with the substrate for more than 10 min (81, 82). Polymer and washing solutions loaded into channels with a pump or vacuum can deposit ~1.5-nm-thick layers in 5 to 10 min (83). Capillary forces can also be used to pull polymer solutions through microfluidic channels by placing droplets of solution at fluidic inlets, followed by spinning the substrate to remove the solution, allowing for ~1.2-nm-thick layers to be deposited in less than 2 min (84). Fluidic layering based on capillary forces is easy to implement, as capillary action does not require external active components, but it is not suitable for larger volumes or when dynamic control over the flow rate is needed.

Fluidic devices and perfusion chambers can also be used to achieve region-specific fluidic assembly or to perform fluidic assembly on more complicated 3D structures. For example, complex automated microfluidic devices can be used to assemble hundreds of layers in parallel using capillary flow and vacuum to fill and empty multiple channels (85). This enables the high-throughput screening of film libraries using small quantities of materials, as only a droplet is needed to fill a single microchannel. Region-specific films can be coated on substrates by affixing a geometric chamber over the substrate and then flowing the solution through the chamber and over the substrate (80). Perfusion chambers can be used for fluidic layering on complex 3D substrates such as sensitive biological substrates (like arteries), which must remain constantly hydrated during layering (86). Similarly, perfusion chambers can be used to hold agarose that contains immobilized particles for fluidic assembly (87). This technology not only allows for the deposition of polymers but also for the deposition of larger cargo, such as gold nanoparticles or liposomes, and produces films with nearly identical thickness to those prepared by standard centrifugation-based assembly (87).

Vacuum is typically used with other assembly technologies, such as spray assembly, or to remove the solution from channels in fluidic assembly, but it can also be used to form multilayers in a macrofluidic-type assembly, especially on unusual substrates like aerogels. Aerogels can be functionalized using vacuum assembly by pouring solutions of conducting polymers, biomolecules, or carbon nanotubes from the top and applying vacuum to pull these solutions down

through the aerogel (88). Vacuum assembly can also be used to deposit materials, such as reduced-graphene oxide, that would otherwise pose a challenge to the creation of uniform multilayers (89). For particulate substrates, vacuum assembly can be performed using separating filters, down to 200 nm in pore size, for centrifugation-free layering (21, 90). Vacuum is not applicable for all sensitive templates; however, for template particles including emulsions (91), cell islets (92), or calcium carbonate nanowires (93), a slight vacuum (~100 mbar) can facilitate the layering process (21). Less than 3% of the particles are lost during assembly using an optimized procedure, far less than with the calculated saturation-based method, where ~50% can be lost, or the centrifugation-based method, where more than 80% can be lost at high layer numbers (21). This combined filter-and-vacuum assembly technology yields a layer thickness of ~1.3 nm and a surface roughness of ~5 to 10 nm, which are both similar to those prepared via centrifugation-based assembly (94). A filtration setup has also been automated for coating cell islets, using a feedback loop for evacuating the fluid from the reaction chamber, thereby reducing the manual handling time by ~60% (95).

Like vacuum assembly, fluidic assembly is not restricted to planar substrates and is a viable alternative for centrifugation-free assembly on particulate substrates (96–98). Many fluidic assembly approaches coat emulsions or liquid crystals, as these materials are well studied in the fluidics field. Generally, the coating and washing solutions are deflected past the flow of particles by using physical gaps smaller than the particles so that the flow can enter perpendicular (96) or parallel to the particle flow stream (97). For fluidic assembly in parallel flow systems, larger template particles (~50 μm in diameter) are necessary, as they can be deflected in a zig-zag pattern using solid pillars at a ~45° angle to three parallel laminar flow streams: solution A, washing, and solution B. This gives a layer thickness of ~2 to 3 nm (97). A similar technology can be used to coat 15-μm beads with avidin and biotin, where higher deflection angles have a high correlation to failure rates, with angles of 1° optimal for a failure rate of virtually zero (99). Instead of pillars, specific geometries can be used to catch emulsions for the fluidic assembly of lipid layers (98). For coating lipid particles with polymers, tangential flow filtration can be used as a type of expedited dialysis for removing excess polymer solution (100). To coat sufficiently large and/or dense particles, a setup based on fluidized beds can be used. This allows for assembly that is ~eight times faster when compared with centrifugation-based assembly and produces films twice as thick (23). In this instance, the force of the washing or polymer solution lifting the particles is balanced against the force of gravity sedimenting the particles, resulting in a fluidized bed where washing and polymer solutions can be pushed past the particles. A similar setup can be used to coat larger (>100-μm) particles in packed columns, although these beds do

not need to be fluidized due to the large particle size, and gravity rather than a vacuum or pump-driven fluids can be used to pull the coating and washing solutions through the column (1).

Numerous applications have been introduced during the process of developing technologies for fluidic assembly. Many applications, such as improved capillary electrophoresis, are realized inside capillaries (81). Fluidic assembly can also be used to engineer complex flow patterns, such as having flow in opposite directions in the same capillary, simply by changing the outer coating of the capillary walls and generating flow with an electric current (82). Fluidic assembly is not limited to planar substrates: For example, multilayer coatings can be prepared on aerogels, resulting in improved compressive strength, wet-state super elasticity, fluorescence, and mechano responsive resistance, while also creating high charge-storage capacity (88). Damaged aortic porcine arteries can be repaired ex vivo with fluidic assembly, to protect the artery against unwanted blood coagulation, as well as to facilitate healing (86). Similarly, catheter tubing can be coated with antifungal multilayers to reduce fouling (78). Chromatography beads coated with multilayers of particles increase the surface area of the beads, thereby improving chromatography (1). Although fluidic assembly is typically performed on larger particles (tens or hundreds of micrometers in diameter), smaller particles (below ~5 μm in diameter) can be coated and loaded with functional cargo for potential drug delivery applications by combining microfluidics with immobilization (87). Fragile particulate substrates like emulsions can also be coated with lipids, using fluidic assembly for the generation of synthetic cells (98). Neuronal cells can be patterned with fluidic assembly (83), and cell islets can be coated to improve robustness, allowing for in vivo transplantation (92, 95). Fluidic assembly functions as a valuable tool for coating sensitive particulate substrates, like mammalian cells, that may be damaged using other technologies, such as during handling in centrifugation-based assembly.

In summary, fluidic assembly provides the means to assemble multilayers on surfaces not easily accessible to other methods (e.g., inside capillaries), provides new ways for region-specific patterning (e.g., by masking a surface with a fluidic channel), and increases the industrial capacity of multilayer assemblies (e.g., through parallelization of film assembly and decrease of reagent consumption). Although the specialized equipment and expertise required to set up (micro)fluidic systems can complicate the use of fluidic assembly, these advantages make it attractive for many applications.

Challenges both big and small

Over the past two decades, LbL assembly has undergone an explosive growth in usable materials and substrates. When taken together with all of the different assembly technologies available, it becomes obvious why LbL assembly is prevalent across a broad spectrum of disciplines. Despite

this extensive toolbox, relatively few multilayer films have had widespread impact outside of research environments. One focus for industrial applications is the identification of reliable, scalable, and resource-effective assembly processes, although this may require different approaches for macroscopic substrates and for microscopic particulate substrates.

For macroscopic substrates, improved high-throughput assembly methods for conformal coatings will play a key role. Immersive and spray roll-to-roll assembly are industrially relevant but only readily applicable to flexible planar substrates; therefore, innovation is needed in systems that can be easily scaled for coating large or numerous 3D macroscopic substrates. Similarly, reducing material waste during the coating process remains important, especially for valuable coating materials like biomolecules and custom polymers. Another challenge for films intended for in vivo biomedical application, such as drug delivery and tissue engineering, is ensuring sterility of the product. This is typically achieved through sterilization (heat, ultraviolet light, chemical treatment, etc.) just before use, which can affect film properties and performance. Finally, increasing the reliability and reproducibility of the films—for example, by increasing automation and reducing manual intervention—is crucial for extending knowledge about film properties and assembly technologies and also for applying the multilayer films in real-life applications.

Similar challenges exist for particulate substrates. One crucial difference is that several particulate assembly methods depend on centrifugation, which remains difficult to scale or combine with minimal-intervention high-throughput assembly. Furthermore, yield and size ranges need to be specified for the various technologies, as these details are often not determined. Detailed film properties (such as layer interpenetration, layer density, film stability or responsiveness, and permeability) that have primarily been studied on planar substrates also need to be investigated so that further comparisons between planar and particulate substrates can be drawn. Altogether, these challenges are not trivial and require focused efforts to overcome; they are also not unique to the field of LbL assembly. One way to address these challenges is to continue to be open and look for solutions in new and sometimes unexpected areas, both in neighboring and more distant fields; this has underpinned much of the technological innovation in LbL assembly.

Opportunities: Thinking outside the box

Layer-by-layer assembly is a firmly established technology and shows great promise in multiple, diverse fields. Much of the development up until now has focused on using new molecular driving forces for film assembly, thus enabling the use of a suite of substrates and layer materials. However, this enormous potential still remains largely limited to small-scale research settings and requires technological and methodological innovation. Despite a surge of new technologies,

many being recent developments, unmet challenges still remain, both for harnessing the specific strengths of different technologies for particular applications and also for developing new and improved technologies.

Although much work has been undertaken on establishing new assembly technologies, only a few studies have chosen a specific assembly technology for the material properties generated (e.g., stratification, density, roughness) rather than the processing properties used (e.g., ease-of-use, material and time savings, lowered involvement, larger batches). For example, for applications where electrical conductivity is important (such as fuel cells and batteries), the conductivity of an immersive-assembled film can be superior to that of spray-assembled films, which can be explained by differences in the interlayer organization of the constituent conductive layer materials (interpenetration versus stratification) (59). Conversely, for applications where optical clarity and/or wetting behavior is important, spin-assemble can allow for an optically transparent film with well-controlled water-contact angles to be assembled due to the smooth, stratified layers formed, whereas an immersive-assembled film can be translucent and with a contact angle that drastically changes depending on the number of layers deposited due to the rough, interpenetrated layers formed (Fig. 3) (44). However, layer structure is only one of the critical film properties to be taken into account when designing films for specific applications. For example, the higher surface coverage and layer density associated with electromagnetic assembly can allow for electrodeposited enzyme films to have higher enzymatic activity than comparable immersive-assembled films (68). Of course, the layer structure and density are not relevant if the desired film components cannot be layered, which can be an issue, for example, when using materials with low charge density (e.g., reduced-graphene oxide) or with a low surface area of contact (e.g., branched nanowires). In such cases, technologies such as dewetting and vacuum assembly enable film formation using constituents that cannot be easily layered using other technologies (8, 89). These examples demonstrate how the judicious choice of assembly technology can enable the assembly of new and improved thin films. As our understanding of the different technologies and how they compare to each other increases, so does the opportunity to let this insight help guide the development of the next generation of LbL assembled thin films.

It is noteworthy that the assembly technologies discussed herein were not originally developed for LbL assembly, and crossover technologies from other fields will continue to play an important part for new, and perhaps even revolutionary, developments. One interesting example involving industrial-scale layering was performed using a modified car wash for spray assembly on a full-sized car (101). Technologies long used in the pharmaceutical industry, such as methods used to treat, purify, and concentrate pharmaceuticals, may prove transformative for biomedical

applications. Similarly, using everyday objects like spray-paint cans could revolutionize assembly methods by essentially combining dewetting and spray assembly for rapid region-specific assembly with little to no material waste and no washing steps. Other combinations between existing assembly technologies should also help to expedite and automate the assembly process. Along these lines, technologies for assembly on particulate substrates are expected to continue to integrate immobilization methods, as they allow collections of particles to be treated like planar substrates, making accessible many of the planar assembly technologies discussed herein. Another promising approach for particulate substrates could be to use a type of “sponge” to absorb excess polymer from solution, thus removing the need to pellet the particles. In terms of future developments for applications, it will be important to understand the interaction between multilayer films and complex and natural environments, such as those found in the human body (40), outdoors, or in seawater. An important aspect of this could be the use of functional substrates capable of compounding the benefits of different multilayers in a synergistic fashion. Overall, the future of LbL assembly is bright, and as the black box of assembly technologies is slowly illuminated, great potential for innovation and application will be found.

REFERENCES AND NOTES

- J. J. Kirkland, Porous thin-layer modified glass bead supports for gas liquid chromatography. *Anal. Chem.* **37**, 1458–1461 (1965). doi: [10.1021/ac60231a004](https://doi.org/10.1021/ac60231a004)
- R. Iler, Multilayers of colloidal particles. *J. Colloid Interface Sci.* **21**, 569–594 (1966). doi: [10.1016/0095-8522\(66\)90018-3](https://doi.org/10.1016/0095-8522(66)90018-3)
- G. Decher, Fuzzy nanoassemblies: Toward layered polymeric multicomposites. *Science* **277**, 1232–1237 (1997). doi: [10.1126/science.277.5330.1232](https://doi.org/10.1126/science.277.5330.1232)
- P. Ott, K. Trenkenschuh, J. Gensel, A. Fery, A. Laschewsky, Free-standing membranes via covalent cross-linking of polyelectrolyte multilayers with complementary reactivity. *Langmuir* **26**, 18182–18188 (2010). doi: [10.1021/la103588z](https://doi.org/10.1021/la103588z); pmid: [21033763](#)
- F. Caruso, R. A. Caruso, H. Möhwald, Nanoengineering of inorganic and hybrid hollow spheres by colloidal templating. *Science* **282**, 1111–1114 (1998). doi: [10.1126/science.282.5391.1111](https://doi.org/10.1126/science.282.5391.1111); pmid: [9804547](#)
- E. Donath, G. B. Sukhorukov, F. Caruso, S. A. Davis, H. Möhwald, Novel hollow polymer shells by colloid-templated assembly of polyelectrolytes. *Angew. Chem. Int. Ed.* **37**, 2201–2205 (1998). doi: [10.1002/\(SICI\)1521-3773\(19980904\)37:16<2201::AID-ANIE2201>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1521-3773(19980904)37:16<2201::AID-ANIE2201>3.0.CO;2-E)
- X. Zhang, H. Chen, H. Zhang, Layer-by-layer assembly: From conventional to unconventional methods. *Chem. Commun.* **2007**, 1395–1405 (2007). doi: [10.1039/B615590A](https://doi.org/10.1039/B615590A); pmid: [17389973](#)
- B. S. Shim et al., Nanostructured thin films made by dewetting method of layer-by-layer assembly. *Nano Lett.* **7**, 3266–3273 (2007). doi: [10.1021/nl071245d](https://doi.org/10.1021/nl071245d); pmid: [1793537](#)
- K. Fujimoto, S. Fujita, B. Ding, S. Shiratori, Fabrication of layer-by-layer self-assembly films using roll-to-roll process. *Jpn. J. Appl. Phys.* **44**, L126–L128 (2005). doi: [10.1143/JJAP.44.L126](https://doi.org/10.1143/JJAP.44.L126)
- E. Donath et al., Nonlinear hairy layer theory of electrophoretic fingerprinting applied to consecutive layer by layer polyelectrolyte adsorption onto charged polystyrene latex particles. *Langmuir* **13**, 5294–5305 (1997). doi: [10.1021/la970090u](https://doi.org/10.1021/la970090u)
- D. O. Grigoriev, T. Bukreeva, H. Möhwald, D. G. Shchukin, New method for fabrication of loaded micro- and nanocontainers: Emulsion encapsulation by polyelectrolyte layer-by-layer deposition on the liquid core. *Langmuir* **24**, 999–1004 (2008). doi: [10.1021/la072873f](https://doi.org/10.1021/la072873f); pmid: [18163658](#)
- N. G. Hoogeveen, M. A. Cohen Stuart, G. J. Fleer, M. R. Böhmer, Formation and stability of multilayers of polyelectrolytes. *Langmuir* **12**, 3675–3681 (1996). doi: [10.1021/la951574](https://doi.org/10.1021/la951574)
- J. J. Richardson et al., Preparation of nano- and microcapsules by electrophoretic polymer assembly. *Angew. Chem. Int. Ed.* **52**, 6455–6458 (2013). doi: [10.1002/anie.201302092](https://doi.org/10.1002/anie.201302092); pmid: [23657949](#)
- I. M. Thomas, Single-layer TiO₂ and multilayer TiO₂-SiO₂ optical coatings prepared from colloidal suspensions. *Appl. Opt.* **26**, 4688–4691 (1987). doi: [10.1364/AO.26.004688](https://doi.org/10.1364/AO.26.004688); pmid: [20523426](#)
- L. Ma et al., Layer-by-layer self-assembly under high gravity field. *Langmuir* **28**, 9849–9856 (2012). doi: [10.1021/la30153w](https://doi.org/10.1021/la30153w); pmid: [22639916](#)
- J. B. Schlenoff, S. T. Dubas, T. Farhat, Sprayed polyelectrolyte multilayers. *Langmuir* **16**, 9968–9969 (2000). doi: [10.1021/la001312i](https://doi.org/10.1021/la001312i)
- A. Qi et al., Template-free synthesis and encapsulation technique for layer-by-layer polymer nanocarrier fabrication. *ACS Nano* **5**, 9583–9591 (2011). doi: [10.1021/nn20283n](https://doi.org/10.1021/nn20283n); pmid: [22059733](#)
- J. Sun, M. Gao, J. Feldmann, Electric field directed layer-by-layer assembly of highly fluorescent CdTe nanoparticles. *J. Nanosci. Nanotechnol.* **1**, 133–136 (2001). doi: [10.1166/jnn.2001.029](https://doi.org/10.1166/jnn.2001.029); pmid: [12914042](#)
- X. Hong et al., Fabrication of magnetic luminescent nanocomposites by a layer-by-layer self-assembly approach. *Chem. Mater.* **16**, 4022–4027 (2004). doi: [10.1021/cm049422o](https://doi.org/10.1021/cm049422o)
- G. Rydzek et al., Polymer multilayer films obtained by electrochemically catalyzed click chemistry. *Langmuir* **26**, 2816–2824 (2010). doi: [10.1021/la902874k](https://doi.org/10.1021/la902874k); pmid: [19950954](#)
- A. Voigt et al., Membrane filtration for microencapsulation and microcapsules fabrication by layer-by-layer polyelectrolyte adsorption. *Ind. Eng. Chem. Res.* **38**, 4037–4043 (1999). doi: [10.1021/ie9900925](https://doi.org/10.1021/ie9900925)
- C. Picart et al., Buildup mechanism for poly (l-lysine)/hyaluronic acid films onto a solid surface. *Langmuir* **17**, 7414–7424 (2001). doi: [10.1021/la010848g](https://doi.org/10.1021/la010848g)
- J. J. Richardson et al., Fluidized bed layer-by-layer microcapsule formation. *Langmuir* **30**, 10028–10034 (2014). doi: [10.1021/la502176g](https://doi.org/10.1021/la502176g); pmid: [25113552](#)
- D. Lee, M. F. Rubner, R. E. Cohen, All-nanoparticle thin-film coatings. *Nano Lett.* **6**, 2305–2312 (2006). doi: [10.1021/nl061776m](https://doi.org/10.1021/nl061776m); pmid: [17034102](#)
- G. Decher, J. D. Hong, J. Schmitt, Buildup of ultrathin multilayer films by a self-assembly process: III. Consecutively alternating adsorption of anionic and cationic polyelectrolytes on charged surfaces. *Thin Solid Films* **210–211**, 831–835 (1992). doi: [10.1016/0040-6090\(92\)90417-A](https://doi.org/10.1016/0040-6090(92)90417-A)
- S. T. Dubas, J. B. Schlenoff, Factors controlling the growth of polyelectrolyte multilayers. *Macromolecules* **32**, 8153–8160 (1999). doi: [10.1021/ma981927a](https://doi.org/10.1021/ma981927a)
- D. Peiffre, T. Corley, G. Halpern, B. Brinker, Utilization of polymeric materials in laser fusion target fabrication. *Polymer* **22**, 450–460 (1981). doi: [10.1016/0032-3861\(81\)90159-2](https://doi.org/10.1016/0032-3861(81)90159-2)
- G. L. Gaines Jr., Deposition of colloidal particles in monolayers and multilayers. *Thin Solid Films* **99**, 243–248 (1983). doi: [10.1016/0040-6090\(83\)90387-5](https://doi.org/10.1016/0040-6090(83)90387-5)
- A. Fou, O. Onitsuka, M. Ferreira, M. Rubner, B. Hsieh, Fabrication and properties of light-emitting diodes based on self-assembled multilayers of poly (phenylene vinylene). *J. Appl. Phys.* **79**, 7501–7509 (1996). doi: [10.1063/1.362421](https://doi.org/10.1063/1.362421)
- S. S. Shiratori, M. Yamada, Nano-scale control of composite polymer films by mass-controlled layer-by-layer sequential adsorption of polyelectrolytes. *Polym. Adv. Technol.* **11**, 810–814 (2000). doi: [10.1002/\(SICI\)1099-1581\(200008/12\)11:8<810::AID-PAT58>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1099-1581(200008/12)11:8<810::AID-PAT58>3.0.CO;2-4)
- C. G. Lölander, H. Arwin, J. C. Eriksson, I. Lundstrom, R. Larsson, Heparin surface film formation through adsorption of colloidal particles studied by ellipsometry and scanning electron microscopy. *Colloids Surf. B* **5**, 1–16 (1982). doi: [10.1016/0166-6622\(82\)80053-X](https://doi.org/10.1016/0166-6622(82)80053-X)
- Y. Fu et al., Facile and efficient approach to speed up layer-by-layer assembly: Dipping in agitated solutions. *Langmuir* **27**, 672–677 (2011). doi: [10.1021/la104524k](https://doi.org/10.1021/la104524k); pmid: [21166437](#)
- J. J. Richardson et al., Immersive polymer assembly on immobilized particles for automated capsule preparation. *Adv. Mater.* **25**, 6874–6878 (2013). doi: [10.1002/adma.201302696](https://doi.org/10.1002/adma.201302696); pmid: [24123174](#)
- A. J. Mateos, A. A. Cain, J. C. Grunlan, Large-scale continuous immersion system for layer-by-layer deposition of flame

- retardant and conductive nanocoatings on fabric. *Ind. Eng. Chem. Res.* **53**, 6409–6416 (2014). doi: [10.1021/ie500122u](https://doi.org/10.1021/ie500122u)
35. J. Li, H. D. H. Stöver, Pickering emulsion templated layer-by-layer assembly for making microcapsules. *Langmuir* **26**, 15554–15560 (2010). doi: [10.1021/la1020498](https://doi.org/10.1021/la1020498); pmid: [20812695](#)
36. H. Liu, X. Gu, M. Hu, Y. Hu, C. Wang, Facile fabrication of nanocomposite microcapsules by combining layer-by-layer self-assembly and Pickering emulsion templating. *RSC Adv.* **4**, 16751–16758 (2014). doi: [10.1039/c4ra00089g](https://doi.org/10.1039/c4ra00089g)
37. G. B. Sukhorukov *et al.*, Layer-by-layer self assembly of polyelectrolytes on colloidal particles. *Colloids Surf. A Physicochem. Eng. Asp.* **137**, 253–266 (1998). doi: [10.1016/S0927-7757\(98\)00213-1](https://doi.org/10.1016/S0927-7757(98)00213-1)
38. K. Szczepanowicz *et al.*, Formation of biocompatible nanocapsules with emulsion core and pegylated shell by polyelectrolyte multilayer adsorption. *Langmuir* **26**, 12592–12597 (2010). doi: [10.1021/la102061s](https://doi.org/10.1021/la102061s); pmid: [20604580](#)
39. T. G. Shutava, P. P. Pattekari, K. A. Arapov, V. P. Torchilin, Y. M. Lvov, Architectural layer-by-layer assembly of drug nanocapsules with PEGylated polyelectrolytes. *Soft Matter* **8**, 9418–9427 (2012). doi: [10.1039/c2sm25683e](https://doi.org/10.1039/c2sm25683e); pmid: [23144650](#)
40. Y. Yan, M. Björnalm, F. Caruso, Assembly of layer-by-layer particles and their interactions with biological systems. *Chem. Mater.* **26**, 452–460 (2014). doi: [10.1021/cm402126n](https://doi.org/10.1021/cm402126n)
41. H. Hong, R. Steitz, S. Kirstein, D. Davidov, Superlattice structures in poly(phenylenevinylene)-based self-assembled films. *Adv. Mater.* **10**, 1104–1108 (1998). doi: [10.1002/\(SICI\)1521-4095\(199810\)10:14<1104::AID-ADMA1104>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1521-4095(199810)10:14<1104::AID-ADMA1104>3.0.CO;2-F)
42. P. A. Chiarelli *et al.*, Controlled fabrication of polyelectrolyte multilayer thin films using spin-assembly. *Adv. Mater.* **13**, 1167–1171 (2001). doi: [10.1002/1521-4095\(200108\)13:15<1167::AID-ADMA1167>3.0.CO;2-A](https://doi.org/10.1002/1521-4095(200108)13:15<1167::AID-ADMA1167>3.0.CO;2-A)
43. J. Cho, K. Char, J. D. Hong, K. B. Lee, Fabrication of highly ordered multilayer films using a spin self-assembly method. *Adv. Mater.* **13**, 1076–1078 (2001). doi: [10.1002/1521-4095\(200107\)13:14<1076::AID-ADMA1076>3.0.CO;2-M](https://doi.org/10.1002/1521-4095(200107)13:14<1076::AID-ADMA1076>3.0.CO;2-M)
44. J. Seo, J. L. Lutkenhaus, J. Kim, P. T. Hammond, K. Char, Effect of the layer-by-layer (LbL) deposition method on the surface morphology and wetting behavior of hydrophobically modified PEO and PAA LbL films. *Langmuir* **24**, 7995–8000 (2008). doi: [10.1021/la800906x](https://doi.org/10.1021/la800906x); pmid: [18558781](#)
45. S. Vozar *et al.*, Automated spin-assisted layer-by-layer assembly of nanocomposites. *Rev. Sci. Instrum.* **80**, 023903 (2009). doi: [10.1063/1.3078009](https://doi.org/10.1063/1.3078009); pmid: [19256658](#)
46. P. A. Patel, A. V. Dobrynin, P. T. Mather, Combined effect of spin speed and ionic strength on polyelectrolyte spin assembly. *Langmuir* **23**, 12589–12597 (2007). doi: [10.1021/la7020676](https://doi.org/10.1021/la7020676); pmid: [17988161](#)
47. E. Kharlamieva, V. Kozlovskaya, J. Chan, J. F. Ankner, V. V. Tsukruk, Spin-assisted layer-by-layer assembly: Variation of stratification as studied with neutron reflectivity. *Langmuir* **25**, 14017–14024 (2009). doi: [10.1021/la9014042](https://doi.org/10.1021/la9014042); pmid: [19580306](#)
48. J. Schmitt *et al.*, Metal nanoparticle/polymer superlattice films: Fabrication and control of layer structure. *Adv. Mater.* **9**, 61–65 (1997). doi: [10.1002/adma.19970090114](https://doi.org/10.1002/adma.19970090114)
49. C. Jiang *et al.*, Controlled exponential growth in layer-by-layer multilayers using high gravity fields. *J. Mater. Chem. A* **2**, 14048–14053 (2014). doi: [10.1039/C4TA02437K](https://doi.org/10.1039/C4TA02437K)
50. B. S. Shim, N. A. Kotov, Single-walled carbon nanotube combing during layer-by-layer assembly: From random adsorption to aligned composites. *Langmuir* **21**, 9381–9385 (2005). doi: [10.1021/la050992s](https://doi.org/10.1021/la050992s); pmid: [16207007](#)
51. A. Izquierdo, S. S. Ono, J.-C. Voegel, P. Schaaf, G. Decher, Dipping versus spraying: Exploring the deposition conditions for speeding up layer-by-layer assembly. *Langmuir* **21**, 7558–7567 (2005). doi: [10.1021/la047407s](https://doi.org/10.1021/la047407s); pmid: [16042495](#)
52. M. Dierendonck, S. De Koker, R. De Rycke, B. G. De Geest, Just spray it—LbL assembly enters a new age. *Soft Matter* **10**, 804–807 (2014). doi: [10.1039/C3SM52202D](https://doi.org/10.1039/C3SM52202D); pmid: [24838052](#)
53. P. Schaaf, J. C. Voegel, L. Jierry, F. Boulmedais, Spray-assisted polyelectrolyte multilayer buildup: From step-by-step to single-step polyelectrolyte film constructions. *Adv. Mater.* **24**, 1001–1016 (2012). doi: [10.1002/adma.201104227](https://doi.org/10.1002/adma.201104227); pmid: [22278854](#)
54. K. C. Krogman, J. L. Lowery, N. S. Zacharia, G. C. Rutledge, P. T. Hammond, Spraying asymmetry into functional membranes layer-by-layer. *Nat. Mater.* **8**, 512–518 (2009). doi: [10.1038/nmat2430](https://doi.org/10.1038/nmat2430); pmid: [19377464](#)
55. M. H. Merrill, C. T. Sun, Fast, simple and efficient assembly of nanolayered materials and devices. *Nanotechnology* **20**, 075606 (2009). doi: [10.1088/0957-4484/20/7/075606](https://doi.org/10.1088/0957-4484/20/7/075606); pmid: [19417427](#)
56. W. D. Mulhearn, D. D. Kim, Y. Gu, D. Lee, Facilitated transport enhances spray layer-by-layer assembly of oppositely charged nanoparticles. *Soft Matter* **8**, 10419–10427 (2012). doi: [10.1039/c2sm26456k](https://doi.org/10.1039/c2sm26456k)
57. J. Alongi, F. Carosio, A. Frache, G. Malucelli, Layer by layer coatings assembled through dipping, vertical or horizontal spray for cotton flame retardancy. *Carbohydr. Polym.* **92**, 114–119 (2013). doi: [10.1016/j.carbpol.2012.08.086](https://doi.org/10.1016/j.carbpol.2012.08.086); pmid: [23218273](#)
58. K. C. Krogman, N. S. Zacharia, S. Schroeder, P. T. Hammond, Automated process for improved uniformity and versatility of layer-by-layer deposition. *Langmuir* **23**, 3137–3141 (2007). doi: [10.1021/la063085b](https://doi.org/10.1021/la063085b); pmid: [17288468](#)
59. F. S. Gittleson, D. J. Kohn, X. Li, A. D. Taylor, Improving the assembly speed, quality, and tunability of thin conductive multilayers. *ACS Nano* **6**, 3703–3711 (2012). doi: [10.1021/nn204384f](https://doi.org/10.1021/nn204384f); pmid: [22515634](#)
60. H. Q. Tang, S. L. Ji, L. L. Gong, H. X. Guo, G. J. Zhang, Tubular ceramic-based multilayer separation membranes using spray layer-by-layer assembly. *Polym. Chem.* **4**, 5621–5628 (2013). doi: [10.1039/c3py00617d](https://doi.org/10.1039/c3py00617d)
61. N. Fukao, K.-H. Kyung, K. Fujimoto, S. Shiratori, Automatic spray-LbL machine based on in-situ QCM monitoring. *Macromolecules* **44**, 2964–2969 (2011). doi: [10.1021/ma200024w](https://doi.org/10.1021/ma200024w)
62. K. C. Krogman, R. E. Cohen, P. T. Hammond, M. F. Rubner, B. N. Wang, Industrial-scale spray layer-by-layer assembly for production of biomimetic photonic systems. *Bioinspir. Biomim.* **8**, 045005 (2013). doi: [10.1088/1748-3182/8/4/045005](https://doi.org/10.1088/1748-3182/8/4/045005); pmid: [24263056](#)
63. S. W. Morton *et al.*, Scalable manufacture of built-to-order nanomedicine: Spray-assisted layer-by-layer functionalization of PRINT nanoparticles. *Adv. Mater.* **25**, 4707–4713 (2013). doi: [10.1002/adma.201302025](https://doi.org/10.1002/adma.201302025); pmid: [23813892](#)
64. P. R. Van Tassel, Polyelectrolyte adsorption and layer-by-layer assembly: Electrochemical control. *Curr. Opin. Colloid Interface Sci.* **17**, 106–113 (2012). doi: [10.1016/j.cocis.2011.08.008](https://doi.org/10.1016/j.cocis.2011.08.008)
65. H. Wang, S. Ishihara, K. Ariga, Y. Yamauchi, All-metal layer-by-layer films: Bimetallic alternate layers with accessible mesopores for enhanced electrocatalysis. *J. Am. Chem. Soc.* **134**, 10819–10821 (2012). doi: [10.1021/ja303773z](https://doi.org/10.1021/ja303773z); pmid: [22715877](#)
66. G. Zhang, L. Dai, L. Zhang, S. Ji, Effects of external electric field on film growth, morphology, and nanostructure of polyelectrolyte and nanohybrid multilayers onto insulating substrates. *Langmuir* **27**, 2093–2098 (2011). doi: [10.1021/la104120b](https://doi.org/10.1021/la104120b); pmid: [21280593](#)
67. Y. H. Ko *et al.*, Electric-field-assisted layer-by-layer assembly of weakly charged polyelectrolyte multilayers. *Macromolecules* **44**, 2866–2872 (2011). doi: [10.1021/ma102112a](https://doi.org/10.1021/ma102112a)
68. L. Shi *et al.*, Site-selective lateral multilayer assembly of bienzyme with polyelectrolyte on ITO electrode based on electric field-induced direct layer-by-layer deposition. *Biomacromolecules* **4**, 1161–1167 (2003). doi: [10.1021/bm030003e](https://doi.org/10.1021/bm030003e); pmid: [12959579](#)
69. A. P. Ngankam, P. R. Van Tassel, In situ layer-by-layer film formation kinetics under an applied voltage measured by optical waveguide lightmode spectroscopy. *Langmuir* **21**, 5865–5871 (2005). doi: [10.1021/la050066d](https://doi.org/10.1021/la050066d); pmid: [15952835](#)
70. Y. Omura, K. Kyung, S. Shiratori, S. Kim, Effects of applied voltage and solution pH in fabricating multilayers of weakly charged polyelectrolytes and nanoparticles. *Ind. Eng. Chem. Res.* **53**, 11727–11733 (2014). doi: [10.1021/je403736r](https://doi.org/10.1021/je403736r)
71. Z. Wang *et al.*, Electrodeposition of alginate/chitosan layer-by-layer composite coatings on titanium substrates. *Carbohydr. Polym.* **103**, 38–45 (2014). doi: [10.1016/j.carbpol.2013.12.007](https://doi.org/10.1016/j.carbpol.2013.12.007); pmid: [24528698](#)
72. M. Li *et al.*, Electrochemical-coupling layer-by-layer (ECC-LbL) assembly. *J. Am. Chem. Soc.* **133**, 7348–7351 (2011). doi: [10.1021/ja202768k](https://doi.org/10.1021/ja202768k); pmid: [21513274](#)
73. M. Li *et al.*, In situ switching layer-by-layer assembly: One-pot rapid layer assembly via alternation of reductive and oxidative electropolymerization. *Chem. Commun.* **49**, 6879–6881 (2013). doi: [10.1039/c3cc43629b](https://doi.org/10.1039/c3cc43629b); pmid: [23793107](#)
74. B. Mu, P. Liu, P. Du, Y. Dong, C. Lu, Magnetic-targeted pH-responsive drug delivery system via layer-by-layer self-assembly of polyelectrolytes onto drug-containing emulsion droplets and its controlled release. *J. Polym. Sci. A Polym. Chem.* **49**, 1969–1976 (2011). doi: [10.1002/pola.24623](https://doi.org/10.1002/pola.24623)
75. R. Wilson, D. G. Spiller, I. A. Prior, R. Bhatt, A. Hutchinson, Magnetic microspheres encoded with photoluminescent quantum dots for multiplexed detection. *J. Mater. Chem.* **17**, 4400–4406 (2007). doi: [10.1039/b708174j](https://doi.org/10.1039/b708174j)
76. S. Dey, K. Mohanta, A. J. Pal, Magnetic-field-assisted layer-by-layer electrostatic assembly of ferromagnetic nanoparticles. *Langmuir* **26**, 9627–9631 (2010). doi: [10.1021/la101132z](https://doi.org/10.1021/la101132z); pmid: [20469859](#)
77. Y. Wang *et al.*, Coupling electrodeposition with layer-by-layer assembly to address proteins within microfluidic channels. *Adv. Mater.* **23**, 5817–5821 (2011). doi: [10.1002/adma.201103726](https://doi.org/10.1002/adma.201103726); pmid: [22102376](#)
78. N. Raman, M.-R. Lee, S. P. Palecek, D. M. Lynn, Polymer multilayers loaded with antifungal β-peptides kill planktonic *Candida albicans* and reduce formation of fungal biofilms on the surfaces of flexible catheter tubes. *J. Control. Release* **191**, 54–62 (2014). doi: [10.1016/j.jconrel.2014.05.026](https://doi.org/10.1016/j.jconrel.2014.05.026); pmid: [24862322](#)
79. N. Madaboushi *et al.*, Microfluidics as a tool to understand the build-up mechanism of exponential-like growing films. *Macromol. Rapid Commun.* **33**, 1775–1779 (2012). doi: [10.1002/marc.201200353](https://doi.org/10.1002/marc.201200353); pmid: [22806940](#)
80. H.-J. Kim, K. Lee, S. Kumar, J. Kim, Dynamic sequential layer-by-layer deposition method for fast and region-selective multilayer thin film fabrication. *Langmuir* **21**, 8532–8538 (2005). doi: [10.1021/la0511182](https://doi.org/10.1021/la0511182); pmid: [16114968](#)
81. H. Katayama, Y. Ishihama, N. Asakawa, Stable cationic capillary coating with successive multiple ionic polymer layers for capillary electrophoresis. *Anal. Chem.* **70**, 5272–5277 (1998). doi: [10.1021/ac9805221](https://doi.org/10.1021/ac9805221); pmid: [9868917](#)
82. S. L. Barker, D. Ross, M. J. Tarlov, M. Gaitan, L. E. Locascio, Control of flow direction in microfluidic devices with polyelectrolyte multilayers. *Anal. Chem.* **72**, 5925–5929 (2000). doi: [10.1021/ac0008690](https://doi.org/10.1021/ac0008690); pmid: [11407588](#)
83. D. R. Reyes, E. M. Perruccio, S. P. Becerra, L. E. Locascio, M. Gaitan, Micropatterning neuronal cells on polyelectrolyte multilayers. *Langmuir* **20**, 8805–8811 (2004). doi: [10.1021/la042249a](https://doi.org/10.1021/la042249a); pmid: [15379510](#)
84. H. Jiang, S. Kim, K. Char, Multilayer line micropatterning using convective self-assembly in microfluidic channels. *Langmuir* **19**, 3094–3097 (2003). doi: [10.1021/la026788b](https://doi.org/10.1021/la026788b)
85. S. A. Castleberry, W. Li, D. Deng, S. Mayner, P. T. Hammond, Capillary flow layer-by-layer: A microfluidic platform for the high-throughput assembly and screening of polynayered film libraries. *ACS Nano* **8**, 6580–6589 (2014). doi: [10.1021/nn501963q](https://doi.org/10.1021/nn501963q); pmid: [24836460](#)
86. B. Thierry, F. M. Winnik, Y. Merhi, M. Tabrizian, Nanocoatings onto arteries via layer-by-layer deposition: Toward the in vivo repair of damaged blood vessels. *J. Am. Chem. Soc.* **125**, 7494–7495 (2003). doi: [10.1021/ja034321x](https://doi.org/10.1021/ja034321x); pmid: [12812471](#)
87. J. J. Richardson *et al.*, Convective polymer assembly for the deposition of nanostructures and polymer thin films on immobilized particles. *Nanoscale* **6**, 13416–13420 (2014). doi: [10.1039/C4NR04348K](https://doi.org/10.1039/C4NR04348K); pmid: [25273509](#)
88. M. Hamed *et al.*, Nanocellulose aerogels functionalized by rapid layer-by-layer assembly for high charge storage and beyond. *Angew. Chem. Int. Ed.* **52**, 12038–12042 (2013). doi: [10.1002/anie.201305137](https://doi.org/10.1002/anie.201305137); pmid: [24573788](#)
89. B.-S. Kong, J. Geng, H.-T. Jung, Layer-by-layer assembly of graphene and gold nanoparticles by vacuum filtration and spontaneous reduction of gold ions. *Chem. Commun.* **2009**, 2174–2176 (2009). doi: [10.1039/b821920f](https://doi.org/10.1039/b821920f); pmid: [19360184](#)
90. S. Hirjsäri, L. Peltonen, J. Hirvonen, Layer-by-layer polyelectrolyte coating of low molecular weight poly(lactic acid) nanoparticles. *Colloids Surf. B Biointerfaces* **49**, 93–99 (2006). doi: [10.1016/j.colsurfb.2006.03.009](https://doi.org/10.1016/j.colsurfb.2006.03.009); pmid: [16161460](#)
91. A. V. Sadovoy, M. V. Kiryukhin, G. B. Sukhorukov, M. N. Antipina, Kinetic stability of water-dispersed oil droplets encapsulated in a polyelectrolyte multilayer shell. *Phys. Chem. Chem. Phys.* **13**, 4005–4012 (2011). doi: [10.1039/c0cp01762k](https://doi.org/10.1039/c0cp01762k); pmid: [21240391](#)
92. J. T. Wilson *et al.*, Cell surface engineering with polyelectrolyte multilayer thin films. *J. Am. Chem. Soc.* **133**, 7054–7064 (2011). doi: [10.1021/ja110926z](https://doi.org/10.1021/ja110926z); pmid: [21491937](#)
93. T. Sasaki, M. Shimizu, Y. Wu, K. Sakurai, Chitosan derivatives/calcium carbonate composite capsules prepared by the layer-by-layer deposition method. *J. Nanomater.* **2008**, 1–8 (2008). doi: [10.1155/2008/185632](https://doi.org/10.1155/2008/185632)

94. S. Leporatti *et al.*, Scanning force microscopy investigation of polyelectrolyte nano- and microcapsule wall texture. *Langmuir* **16**, 4059–4063 (2000). doi: [10.1021/la991429a](https://doi.org/10.1021/la991429a)
95. J. M. Mets, J. T. Wilson, W. Cui, E. L. Chaikof, An automated process for layer-by-layer assembly of polyelectrolyte multilayer thin films on viable cell aggregates. *Adv. Healthcare Mater.* **2**, 266–270 (2013). doi: [10.1002/adhm.201200148](https://doi.org/10.1002/adhm.201200148); pmid: [23184761](#)
96. C. Priest *et al.*, Microfluidic polymer multilayer adsorption on liquid crystal droplets for microcapsule synthesis. *Lab Chip* **8**, 2182–2187 (2008). doi: [10.1039/b808826h](https://doi.org/10.1039/b808826h); pmid: [19023485](#)
97. C. Kantak, S. Beyer, L. Yobas, T. Bansal, D. Trau, A ‘microfluidic pinball’ for on-chip generation of layer-by-layer polyelectrolyte microcapsules. *Lab Chip* **11**, 1030–1035 (2011). doi: [10.1039/c0lc00381f](https://doi.org/10.1039/c0lc00381f); pmid: [21218225](#)
98. S. Matosevic, B. M. Paegel, Layer-by-layer cell membrane assembly. *Nat. Chem.* **5**, 958–963 (2013). doi: [10.1038/nchem.1765](https://doi.org/10.1038/nchem.1765); pmid: [24153375](#)
99. R. D. Sochol *et al.*, “Continuous flow layer-by-layer microbead functionalization via a micropost array railing system,” in *Proceedings of the 16th International Conference on Solid-State Sensors, Actuators and Microsystems (Transducers 2011)*, Beijing, 5 to 9 June 2011 (Institute of Electrical and Electronics Engineers, Piscataway, NJ, 2011), pp. 1761–1764.
100. S. Hirsjärvi, Y. Qiao, A. Royere, J. Bibette, J.-P. Benoit, Layer-by-layer surface modification of lipid nanocapsules. *Eur. J. Pharm. Biopharm.* **76**, 200–207 (2010). doi: [10.1016/j.ejpb.2010.07.010](https://doi.org/10.1016/j.ejpb.2010.07.010); pmid: [20656028](#)
101. K. Ariga *et al.*, Layer-by-layer nanorhitectonics: Invention, innovation, and evolution. *Chem. Lett.* **43**, 36–68 (2014). doi: [10.1246/cl.130987](https://doi.org/10.1246/cl.130987)

ACKNOWLEDGMENTS

This work was supported by the ARC under the Australian Laureate Fellowship (FL120100030 to F.C.) and the Australian government through an Australian Postgraduate Award (to M.B.). This work was also supported by the ARC Centre of Excellence in Convergent Bio-Nano Science and Technology (project number CE140100036).

10.1126/science.aaa2491



Technology-driven layer-by-layer assembly of nanofilms

Joseph J. Richardson *et al.*
Science **348**, (2015);
 DOI: 10.1126/science.aaa2491

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 23, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/aaa2491.full.html>

This article cites 100 articles, 2 of which can be accessed free:

<http://www.sciencemag.org/content/348/6233/aaa2491.full.html#ref-list-1>

This article appears in the following subject collections:

Materials Science

http://www.sciencemag.org/cgi/collection/mat_sci

RESEARCH ARTICLE SUMMARY

RNA IMAGING

Spatially resolved, highly multiplexed RNA profiling in single cells

Kok Hao Chen,^{1,*} Alistair N. Boettiger,^{1,*} Jeffrey R. Moffitt,^{1,*} Siyuan Wang,¹ Xiaowei Zhuang^{1,2†}

INTRODUCTION: The copy number and intracellular localization of RNA are important regulators of gene expression. Measurement of these properties at the transcriptome scale in single cells will give answers to many questions related to gene expression and regulation. Single-molecule RNA imaging approaches, such as single-molecule fluorescence *in situ* hybridization (smFISH), are powerful tools for counting and mapping RNA; however, the number of RNA species that can be simultaneously imaged in individual cells has been limited. This makes it challenging to perform transcriptomic analysis of single cells in a spatially resolved manner. Here, we report multiplexed error-robust FISH (MERFISH), a single-molecule imaging method that allows thousands of RNA

species to be imaged in single cells by using combinatorial FISH labeling with encoding schemes capable of detecting and/or correcting errors.

RATIONALE: We labeled each cellular RNA with a set of encoding probes, which contain targeting sequences that bind the RNA and readout sequences that bind fluorescently labeled readout probes. Each RNA species is encoded with a particular combination of readout sequences. We used successive rounds of hybridization and imaging, each with a different readout probe, to identify the readout sequences bound to each RNA and to decode the RNA. In principle, combinatorial labeling allows the number of detectable RNA species to

grow exponentially with the number of imaging rounds, but the detection errors also increase exponentially. To combat such accumulating errors, we exploited error-robust encoding schemes used in digital electronics, such as the extended Hamming code, in the design of

our encoding probes but modified these schemes in order to account for the error properties in FISH measurements. We assigned each RNA a binary word in our modified Hamming

code and encoded the RNA with a combination of readout sequences according to this binary word.

RESULTS: We first imaged 140 RNA species in human fibroblast cells using MERFISH with 16 rounds of hybridization and a modified Hamming code capable of both error detection and correction. We obtained ~80% detection efficiency and observed excellent correlation of RNA copy numbers determined with MERFISH with both bulk RNA sequencing data and conventional smFISH measurements of individual genes.

Next, we used an alternative MERFISH encoding scheme, which is capable of detecting but not correcting errors, to image 1001 RNA species in individual cells using only 14 rounds of hybridization. The observed RNA copy numbers again correlate well with bulk sequencing data. However, the detection efficiency is only one-third that of the error-correcting encoding scheme.

We performed correlation analysis of the 10^4 to 10^6 pairs of measured genes and identified many covarying gene groups that share common regulatory elements. Such grouping allowed us to hypothesize potential functions of ~100 unannotated or partially annotated genes of unknown functions. We further analyzed correlations in the spatial distributions of different RNA species and identified groups of RNAs with different distribution patterns in the cell.

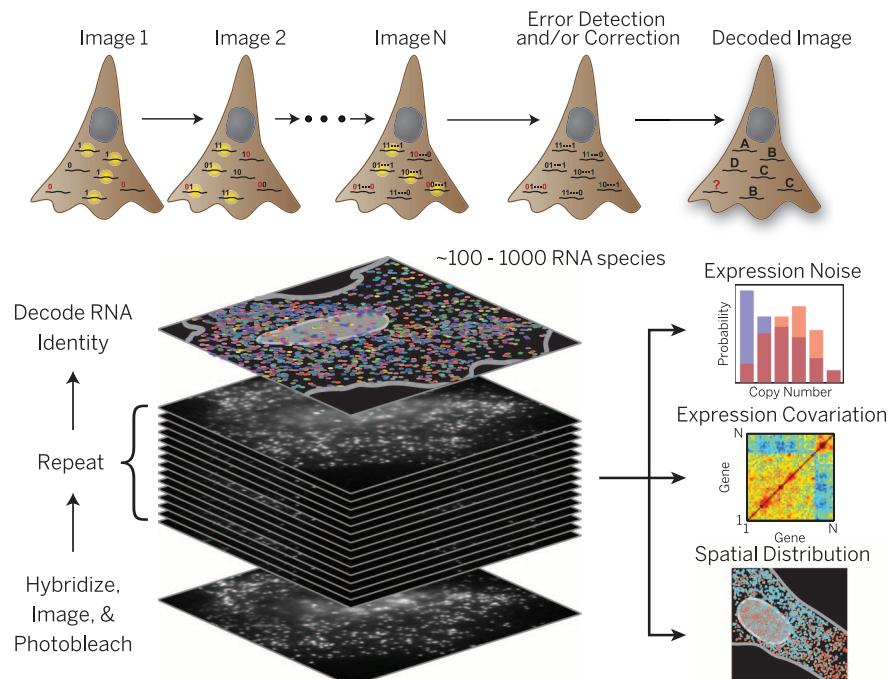
DISCUSSION: This highly multiplexed imaging approach enables analyses based on the variation and correlation of copy numbers and spatial distributions of a large number of RNA species within single cells. Such analyses should facilitate the delineation of regulatory networks and *in situ* identification of cell types. We envision that this approach will allow spatially resolved transcriptomes to be determined for single cells. ■

RELATED ITEMS IN SCIENCE

J. H. Lee et al., *Science* **343**, 1360–1363 (2014).

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

†Corresponding author. E-mail: zhuang@chemistry.harvard.edu
Cite this article as K. H. Chen et al., *Science* **348**, aaa6090 (2015). DOI:10.1126/science.aaa6090



MERFISH for transcriptome imaging. Numerous RNA species can be identified, counted, and localized in a single cell by using MERFISH, a single-molecule imaging approach that uses combinatorial labeling and sequential imaging with encoding schemes capable of detection and/or correction of errors. This highly multiplexed measurement of individual RNAs can be used to compute the gene expression profile and noise, covariation in expression among different genes, and spatial distribution of RNAs within single cells.

RESEARCH ARTICLE

RNA IMAGING

Spatially resolved, highly multiplexed RNA profiling in single cells

Kok Hao Chen,^{1,*} Alistair N. Boettiger,^{1,*} Jeffrey R. Moffitt,^{1*}
Siyuan Wang,¹ Xiaowei Zhuang^{1,2,†}

Knowledge of the expression profile and spatial landscape of the transcriptome in individual cells is essential for understanding the rich repertoire of cellular behaviors. Here, we report multiplexed error-robust fluorescence *in situ* hybridization (MERFISH), a single-molecule imaging approach that allows the copy numbers and spatial localizations of thousands of RNA species to be determined in single cells. Using error-robust encoding schemes to combat single-molecule labeling and detection errors, we demonstrated the imaging of 100 to 1000 distinct RNA species in hundreds of individual cells. Correlation analysis of the $\sim 10^4$ to 10^6 pairs of genes allowed us to constrain gene regulatory networks, predict novel functions for many unannotated genes, and identify distinct spatial distribution patterns of RNAs that correlate with properties of the encoded proteins.

System-wide analyses of the abundance and spatial organization of RNAs in single cells promise to transform our understanding in many areas of cell and developmental biology, such as the mechanisms of gene regulation, the heterogeneous behavior of cells, and the development and maintenance of cell fate (1). Single-molecule fluorescence *in situ* hybridization (smFISH) has emerged as a powerful tool for studying the copy number and spatial organization of RNAs in single cells either in isolation or in their native tissue context (2, 3). Taking advantage of its ability to map the spatial distributions of specific RNAs with high resolution, smFISH has revealed the importance of subcellular RNA localization in diverse processes such as cell migration, development, and polarization (4–8). In parallel, the ability of smFISH to precisely measure the copy numbers of specific RNAs without amplification bias has allowed quantitative measurement of the natural fluctuations in gene expression, which has in turn elucidated the regulatory mechanisms that shape such fluctuations and their role in a variety of biological processes (9–13).

Recent advances in imaging and analysis methods have allowed hundreds of smFISH measurements to be performed in an automated manner, substantially expanding our knowledge of the RNA expression profile and spatial organization in different organisms (14, 15). However, application of the smFISH approach to many systems-level questions remains limited by the number of RNA species that can be simultaneously measured in single cells. State-of-the-art efforts by

using combinatorial labeling with either color-based barcodes or sequential hybridization have enabled simultaneous measurements of 10 to 30 different RNA species in individual cells (16–19), yet many interesting biological questions would benefit from the measurement of hundreds to thousands of RNAs within a single cell. For example, analysis of how the expression profile of such a large number of RNAs vary from cell to cell and how these variations correlate among different genes could be used to systematically identify coregulated genes and map regulatory networks, knowledge of the subcellular organizations of numerous RNAs and their correlations could help elucidate molecular mechanisms underlying the establishment and maintenance of many local cellular structures, and RNA profiling of individual cells in native tissues could allow *in situ* identification of cell type.

Here, we report multiplexed error-robust FISH (MERFISH), a highly multiplexed smFISH imaging method that substantially increases the number of RNA species that can be simultaneously imaged in single cells by using combinatorial labeling and sequential imaging with error-robust encoding schemes. We demonstrated this transcriptome imaging approach by simultaneously measuring 140 RNA species with an encoding scheme that can both detect and correct errors and 1001 RNA species with an encoding scheme that can detect but not correct errors. Correlation analyses of the copy number variations and spatial distributions of these genes allowed us to identify groups of genes that are coregulated and groups of genes that share similar spatial distribution patterns inside the cell.

Combinatorial labeling with error-robust encoding schemes

Combinatorial labeling that identifies each RNA species by multiple (N) distinct signals offers a

route to rapidly increase the number of RNA species that can be probed simultaneously in individual cells (Fig. 1A). However, this approach to scaling up the throughput of smFISH to the systems scale faces a substantial challenge because not only does the number of addressable RNA species increases exponentially with N , but the detection error rates also grow exponentially with N (Fig. 1, B to D). Imagine a conceptually simple scheme to implement combinatorial labeling, in which each RNA species is encoded with a N -bit binary word, and the sample is probed with N corresponding rounds of hybridization, each round targeting only the subset of RNAs that should read “1” in the corresponding bit (fig. S1). N rounds of hybridization would allow $2^N - 1$ RNA species to be probed. With just 16 hybridizations, more than 64,000 RNA species—which should cover the entire human transcriptome, including both messenger RNAs (mRNAs) and noncoding RNAs (20)—could be identified (Fig. 1B, black symbols). However, as N increases, the fraction of RNAs properly detected (the calling rate) would rapidly decrease and, more troublingly, the fraction of RNAs that are identified as incorrect species (the misidentification rate) would rapidly increase (Fig. 1, C and D, black symbols). With realistic error rates per hybridization (measured below), the majority of RNA molecules would be misidentified after 16 rounds of hybridizations.

To address this challenge, we designed error-robust encoding schemes in which only a subset of the $2^N - 1$ words separated by a certain Hamming distance (21) were used to encode RNAs. In a codebook in which the minimum Hamming distance is 4 (HD4 code), at least four bits must be read incorrectly to change one valid code word into another (fig. S2A). As a result, every single-bit error produces a word that is exclusively close to a single RNA-encoding word, allowing such errors to be detected and corrected (fig. S2B). Double-bit errors produce words with an equal Hamming distance of 2 from multiple valid code words and, thus, can be detected but not corrected (fig. S2C). Such a code should substantially increase the calling rate and reduce the misidentification rate (Fig. 1, C and D, blue symbols). To further account for the fact that it is more likely to miss a hybridization event (an $1 \rightarrow 0$ error) than to misidentify a background spot as an RNA (an $0 \rightarrow 1$ error) in smFISH measurements, we designed a modified HD4 (MHD4) code, in which the number of 1 bits were kept both constant and relatively low—only four per word in this work—so as to reduce error and avoid biased detection. This MHD4 code should further increase the calling rate and reduce the misidentification rate (Fig. 1, C and D, purple symbols).

In addition to the error considerations, several practical challenges have also made it difficult to probe a large number of RNA species, such as the high cost of the massive number of distinct FISH probes needed and the long time required to complete many rounds of hybridization. An oligopaint approach has been previously developed to generate a large number of oligonucleotide probes to label chromosome DNA and to introduce nontargeting

¹Howard Hughes Medical Institute, Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. ²Department of Physics, Harvard University, Cambridge, MA 02138, USA.

*These authors contributed equally to this work. †Corresponding author. E-mail: zhuang@chemistry.harvard.edu

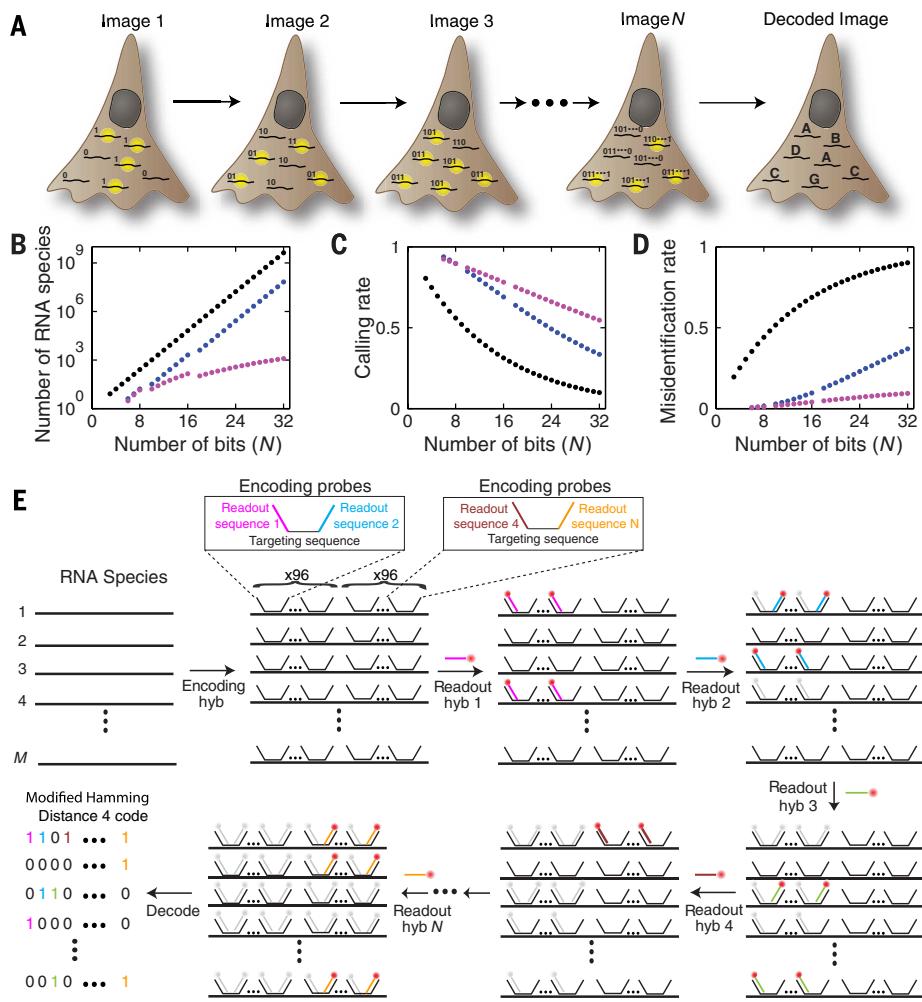


Fig. 1. MERFISH: A highly multiplexed smFISH approach enabled by combinatorial labeling and error-robust encoding. (A) Schematic depiction of the identification of multiple RNA species in N rounds of imaging. Each RNA species is encoded with a N -bit binary word, and during each round of imaging, only the subset of RNAs that should read 1 in the corresponding bit emit signal. (B to D) The number of addressable RNA species (B); the rate at which these RNAs are properly identified—the “calling rate” (C); and the rate at which RNAs are incorrectly identified as a different RNA species—the “misidentification rate” (D); plotted as a function of the number of bits (N) in the binary words encoding RNA. Black indicates a simple binary code that includes all $2^N - 1$ possible binary words. Blue indicates the HD4 code in which the Hamming distance separating words is 4. Purple indicates a modified HD4 (MHD4) code where the number of 1 bits are kept at four. The calling and misidentification rates are calculated with per-bit error rates of 10% for the 1→0 error and 4% for the 0→1 error. (E) Schematic diagram of the implementation of a MHD4 code for RNA identification. Each RNA species is first labeled with ~192 encoding probes that convert the RNA into a specific combination of readout sequences (Encoding hyb). These encoding probes each contain a central RNA-targeting region flanked by two readout sequences, drawn from a pool of N different sequences, each associated with a specific hybridization round. Encoding probes for a specific RNA species contain a particular combination of four of the N readout sequences, which correspond to the four hybridization rounds in which this RNA should read 1. N subsequent rounds of hybridization with the fluorescent readout probes are used to probe the readout sequences (hyb 1, hyb 2, ..., hyb N). The bound probes are inactivated by photobleaching between successive rounds of hybridization. For clarity, only one possible pairing of the readout sequences is depicted for the encoding probes; however, all possible pairs of the four readout sequences are used at the same frequency and distributed randomly along each cellular RNA in the actual experiments.

sites for secondary activities (22). Inspired by this approach, we designed a two-step labeling scheme to encode and read out cellular RNAs (Fig. 1E). First, we label cellular RNAs with a set of encoding probes, each probe comprising a

RNA targeting sequence and two flanking readout sequences. Four of the N distinct readout sequences were assigned to each RNA species based on the N -bit MHD4 code word of the RNA. Second, we identified these N readout sequences

with complementary FISH probes (the readout probes) via N rounds of hybridization and imaging, each round using a different readout probe. To increase the signal-to-background ratio, we labeled every cellular RNA with ~192 encoding probes. Because each encoding probe contained two of the four readout sequences associated with that RNA (Fig. 1E), a maximum of ~96 readout probes can bind to each cellular RNA per hybridization round. To generate the massive number of encoding probes required, we amplified them from array-derived oligonucleotide pools containing tens of thousands of custom sequences using a modified form of the oligopaint protocol comprising in vitro transcription followed by reverse transcription (fig. S3 and supplementary materials, materials and methods, “Probe Synthesis”) (22, 23). This two-step labeling approach dramatically diminished the total hybridization time for an experiment; we found that efficient hybridization to the readout sequences took only 15 min, whereas efficient direct hybridization to cellular RNA required more than 10 hours.

Measuring 140 genes with MERFISH by use of a 16-bit MHD4 code

To test the feasibility of this error-robust, multiplexed imaging approach, we performed a 140-gene measurement on human fibroblast cells (IMR90) using a 16-bit MHD4 code to encode 130 RNA species while leaving 10 code words as misidentification controls (table S1). After each round of hybridization with the fluorescent readout probes, cells were imaged by means of conventional wide-field imaging with an oblique-incidence illumination geometry. Fluorescent spots corresponding to individual RNAs were clearly detected and were then efficiently extinguished via a brief photo-bleaching step (Fig. 2A). The sample was stable throughout the 16 rounds of iterative labeling and imaging. The change in the number of fluorescent spots from round to round matched the expected change predicted on the basis of the relative abundances of RNA species targeted in each round derived from bulk sequencing, and we did not observe a systematic decreasing trend with increasing number of hybridization rounds (fig. S4A). The average brightness of the spots varied from round to round with a standard deviation of 40%, which is likely due to different binding efficiencies of the readout probes to the different readout sequences on the encoding probes (fig. S4B). We observed only a small, systematic decreasing trend in the spot brightness with increasing hybridization rounds, which was on average 4% per round (fig. S4B).

We then constructed binary words from the observed fluorescent spots based on their on-off patterns across the 16 hybridization rounds (Fig. 2, B to D). If the word exactly matched one of the 140 MHD4 code words (exact matches) or differed by only one bit (error-correctable matches), we assigned it to the corresponding RNA species (Fig. 2D). Within the single cell depicted in Fig. 2, A and B, more than 1500 RNA molecules corresponding to 87% of the 130 encoded RNA species were detected after error correction (Fig. 2E).

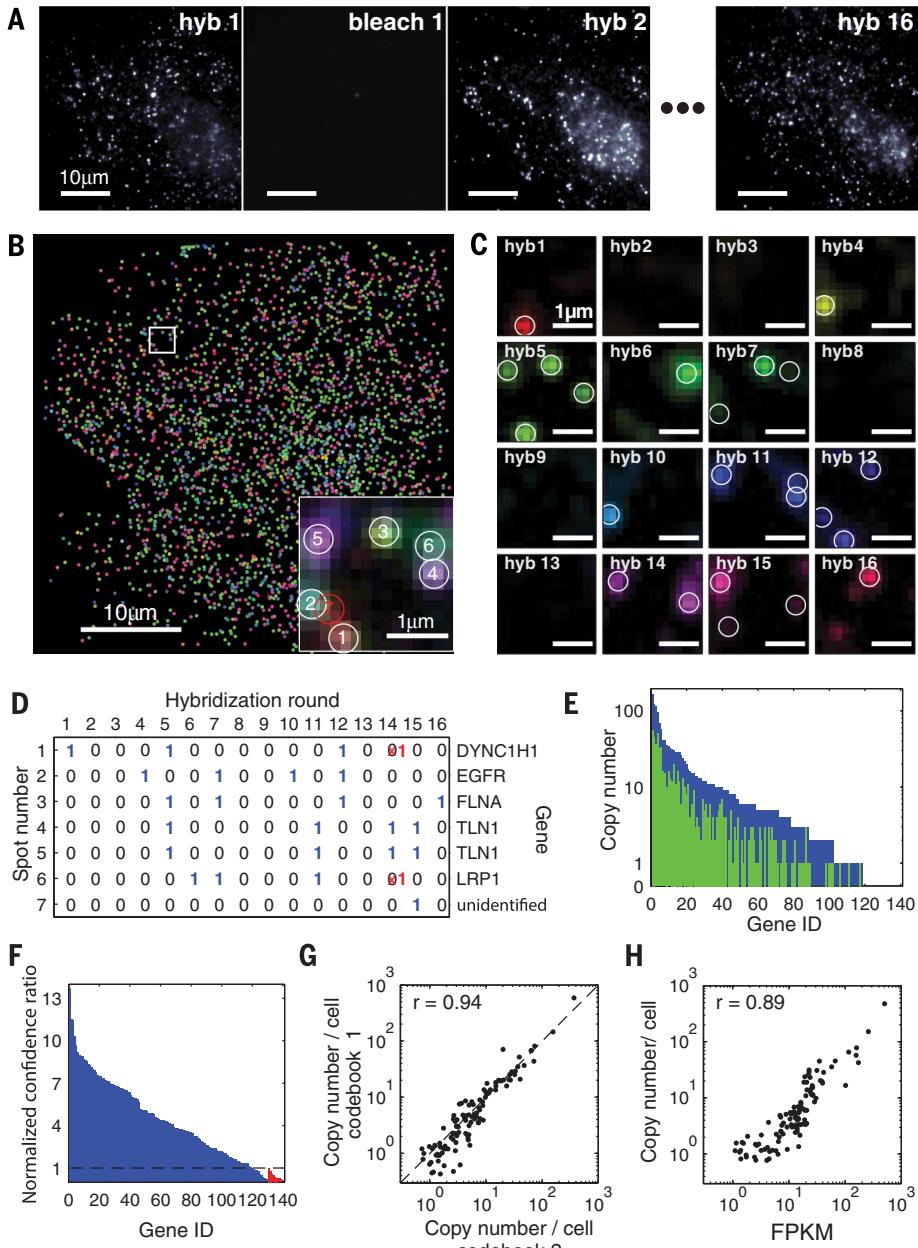


Fig. 2. Simultaneous measurement of 140 RNA species in single cells by use of MERFISH with a 16-bit MHD4 code. (A) Images of RNA molecules in an IMR90 cell after each hybridization round (hyb 1 to hyb 16). The images after photobleaching (for example, bleach 1) demonstrate efficient removal of fluorescent signals between hybridizations. (B) The localizations of all detected single molecules in this cell colored according to their measured binary words. (Inset) The composite, false-colored fluorescent image of the 16 hybridization rounds for the boxed subregion with numbered circles indicating potential RNA molecules. A red circle indicates an unidentifiable molecule, the binary word of which does not match any of the 16-bit MHD4 code words even after error correction. (C) Fluorescent images from each round of hybridization for the boxed subregion in (B), with circles indicating potential RNA molecules. (D) Corresponding words for the spots identified in (C). Red crosses represent the corrected bits. (E) The RNA copy number for each gene observed without (green) or with (blue) error correction in this cell. (F) The confidence ratio measured for the 130 RNA species (blue) and the 10 misidentification control words (red) normalized to the maximum value observed from the misidentification controls (dashed line). (G) Scatter plot of the average copy number of each RNA species per cell measured with two shuffled codebooks of the MHD4 code. The Pearson correlation coefficient is 0.94 with a P value of 1×10^{-53} . The dashed line corresponds to the $y = x$ line. (H) Scatter plot of the average copy number of each RNA species per cell versus the abundance determined by bulk sequencing in FPKM. The Pearson correlation coefficient between the logarithmic abundances of the two measurements was 0.89 with a P value of 3×10^{-39} .

Similar observations were made in ~400 cells from seven independent experiments. On average, ~4 times as many RNA molecules and ~2 times as many RNA species were detected per cell after error correction as compared with the values obtained before error correction (fig. S5).

Two types of errors can occur in the copy number measurement of each RNA species: (i) Some molecules of this RNA species are not detected, leading to a drop in calling rate, and (ii) some molecules from other RNA species are misidentified as this RNA species. To assess the extent of misidentification, we used the 10 misidentification control words—code words that were not associated with any cellular RNA. Although matches to these control words were observed, they occurred far less frequently than did the real RNA-encoding words: 95% of the 130 RNA-encoding words were counted more frequently than the median count for these control words. Moreover, we typically found the ratio of the number of exact matches to the number of matches with one-bit errors for a real RNA-encoding word to be substantially higher than the same ratios observed for the misidentification controls, as expected (fig. S6, A and B). Using this ratio as a measure of the confidence in RNA identification, we found that 91% of the 130 RNA species had a confidence ratio greater than the maximum confidence ratio observed for the misidentification controls (Fig. 2F), demonstrating a high accuracy of RNA identification. Subsequent analyses were conducted only on these 91% of genes.

To estimate the calling rate, we used the error-correction ability of the MHD4 code to determine the $1 \rightarrow 0$ error rates (10% on average) and $0 \rightarrow 1$ error rates (4% on average) for each hybridization round (fig. S6, C and D). Using these error rates, we estimated an ~80% calling rate for individual RNA species after error correction—~80% of the fluorescent spots corresponding to a RNA species were decoded correctly (fig. S6E). Although the remaining 20% of spots contributed to a loss in detection efficiency, most of them did not cause species misidentification because they were decoded as double-bit error words and discarded.

To test for potential technical bias in our measurements, we probed the same 130 RNAs species with a different MHD4 codebook by shuffling the code words among different RNA species (table S1) and changing the encoding probe sequences. Measurements with this alternative code gave similar misidentification and calling rates (fig. S7). The copy numbers of individual RNA species measured with these two codebooks showed excellent agreement with a Pearson correlation coefficient of 0.94 (Fig. 2G), indicating that the choice of encoding scheme did not bias the measured counts.

In order to validate the copy numbers derived from our MERFISH experiments, we performed conventional smFISH measurements on 15 of the 130 genes, spanning the full measured abundance range of three orders of magnitude. For each of these genes, both the average copy number and the copy number distribution across

many cells agreed quantitatively between our MERFISH and conventional smFISH measurements (fig. S8, A and B). The ratio of the copy numbers determined by these two approaches was 0.82 ± 0.06 (mean \pm SEM across the 15 measured RNA species) (fig. S8B), which agrees with the estimated 80% calling rate for our multiplexed imaging approach. The quantitative match between this ratio and our estimated calling rate over the full measured abundance range additionally supports our assessment that the misidentification error was low. Given that the agreement between the MERFISH and conventional smFISH results extended to the genes at the lowest measured abundance (<1 copy per cell) (fig. S8B), we estimate that our measurement sensitivity was better than 1 copy per cell.

As a final validation, we compared the abundance of each RNA species averaged over hundreds of cells to those obtained from a bulk RNA sequencing measurement that we performed on the same cell line. Our imaging results correlated remarkably well with bulk sequencing results, with a Pearson correlation coefficient of 0.89 (Fig. 2H).

High-throughput analysis of cell-to-cell variation in gene expression

The MERFISH approach allows parallelization of measurements of many individual RNA species and covariation analysis between different RNA species. We first illustrated the parallelization aspect by examining the cell-to-cell variation in the expression level of each of the measured genes (Fig. 3A). To quantify the measured variation, we computed the Fano factors, defined as the ratio of the variance to the mean RNA copy number, for all measured RNA species. The Fano factors substantially deviated from 1, the value expected for a simple Poisson process, for many genes and exhibited an increasing trend with the mean RNA abundance (Fig. 3B), which is consistent with a previous observation for other cell types (24). A simple model for promoter regulation—the promoter stochastically switches between on and off states with global constraints on the kinetic rates—has been previously suggested to rationalize such a trend (24, 25). According to this model, this trend of increasing Fano factors with mean RNA abundance can be explained by changes in the transcription rate and/or promoter off-switching rates but not by changes in the promoter on-switching rate.

Moreover, we identified several RNA species with substantially larger Fano factors than this average trend. For example, we found that *SLC5A3*, *CENPF*, *MKI67*, *TNC*, and *KIAA1199* displayed Fano factor values substantially higher than those of the other genes expressed at similar abundance levels. The high variability of some of these genes can be explained by their association with the cell cycle. For example, two of these particularly “noisy” genes, *MKI67* and *CENPF*, are both annotated as cell-cycle related genes (26), and based on their bimodal expression (Fig. 3C), we propose that their transcription is strongly regulated by the cell cycle. Other high-variability genes did not

show the same bimodal expression patterns and are not known to be associated with the cell cycle. Understanding the origin and implications of noisy gene expression is an active topic of current research (24).

Analysis of expression covariation among different genes

Analysis of covariations in the expression levels of different genes can reveal which genes are coregulated and elucidate gene regulatory pathways. At the population level, such analysis often requires the application of external stimuli to drive gene expression variation; hence, correlated expression changes can be observed among genes that share common regulatory elements influenced by the stimuli (27). At the single-cell level, one can take advantage of the natural stochastic fluctuations in gene expression for such analysis and can thus study multiple regulatory networks without having to stimulate each of them individually. Such covariation analysis can constrain regulatory networks, suggest new regulatory pathways, and predict function for unannotated genes based on associations with covarying genes (11, 28).

We applied this approach to the 140-gene measurements and examined the ~10,000 pairwise correlation coefficients that describe how the expression levels of each pair of genes covaried from cell to cell. Many of the highly variable genes showed tightly correlated or anticorrelated variations (Fig. 3C). To better understand the correlations for all gene pairs, we adopted a hierarchical clustering approach, commonly used in the analysis of both bulk and single-cell expression data (29, 30), to organize these genes on the basis of their correlation coefficients (Fig. 3D). From the cluster tree structure, we identified seven groups of genes with substantially correlated expression patterns (Fig. 3D and table S2). Within each of the seven groups, every gene showed significantly stronger average correlation with other members of the group than with genes outside the group (table S2). To further validate and understand these groups, we identified gene ontology (GO) terms (31) enriched in each of these seven groups. The enriched GO terms within each group shared similar functions and were largely specific to each group (Fig. 3E and table S2), validating the notion that the observed covariation in expression reflects some commonalities in the regulation of these genes.

Here, we describe two of these groups as illustrative examples. The predominant GO terms associated with group I were terms associated with the extracellular matrix (ECM) (Fig. 3, D and E, and table S2). Notable members of this group included ECM components—such as *FBNI*, *FBN2*, *COL5A*, *COL7A*, and *TNC*—and glycoproteins linking the ECM and cell membranes, such as *VCAN* and *THBS1*. The group also included an unannotated gene, *KIAA1199*, which we would predict to play a role in ECM metabolism on the basis of its association with this cluster. Indeed, this gene has recently been identified as an enzyme involved in the regulation of hyaluronan, which is a major sugar component of the ECM (32).

Group 6 contained many genes that encode vesicle transport proteins and proteins associated with cell motility (Fig. 3, D and E, and table S2). The vesicle transport genes included microtubule motors and related genes *DYNC1H1*, *CKAP1*, and factors associated with vesicle formation and trafficking, such as *DYNC1I3* and *RAB3B*. Again, we found an unannotated gene, *KIAA1462*, within this cluster. On the basis of its strong correlation with *DYNC1H1* and *DYNC1I3*, we predict that this gene may be involved in vesicle transport. The cell motility genes in this group included genes encoding actin-binding proteins such as *AFAPI*, *SPTAN1*, *SPTBNI*, and *MYH10*, and genes involved in the formation of adhesion complexes, such as *FLNA* and *FLNC*. Several guanosine triphosphatase (GTPase)-associated factors involved in the regulation of cell motility, attachment, and contraction also fell into this group, including *DOCK7*, *ROCK2*, *IQGAP1*, *PRKCA*, and *AMOTL1*. The observation that some cell motility genes correlated with vesicle transport genes is consistent with the role of vesicle transport in cell migration (33). An additional feature of group 6 is that a subset of these genes—in particular, those related to cell motility—were anticorrelated with members of the ECM group discussed above (Fig. 3D). This anticorrelation may reflect regulatory interactions that mediate the switching of cells between adherent and migratory states.

Mapping spatial distributions of RNAs

As an imaging-based approach, MERFISH also allowed us to investigate the spatial distributions of many RNA species simultaneously. Several patterns emerged from the visual inspection of individual genes, with some RNA transcripts enriched in the perinuclear region, some enriched in the cell periphery, and some scattered throughout the cell (Fig. 4A). To identify genes with similar spatial distributions, we determined the correlation coefficients for the spatial density profiles of all pairs of RNA species and organized these RNAs according to the pairwise correlations again using the hierarchical clustering approach. The correlation coefficient matrix showed groups of genes with correlated spatial organizations, and the two most notable groups with the strongest correlations are indicated in Fig. 4B. Group I RNAs appeared enriched in the perinuclear region, whereas group II RNAs appeared enriched near the cell periphery (Fig. 4C). Quantitative analysis of the distances between each RNA molecule and the cell nucleus or the cell periphery indeed confirmed this visual impression (Fig. 4D).

Group I contained genes encoding extracellular proteins such as *FBNI*, *FBN2*, and *THBS1*; secreted proteins such as *PAPPA*; and integral membrane proteins such as *LRP1* and *GPR107*. These proteins have no obvious commonalities in function. Rather, a GO analysis showed significant enrichment for location terms, such as extracellular region, basement membrane, or perivitelline space (Fig. 4E). To reach these locations, proteins must pass through the secretion pathway, which often requires translation of mRNA at the endoplasmic reticulum (ER) (34, 35). Thus, we propose

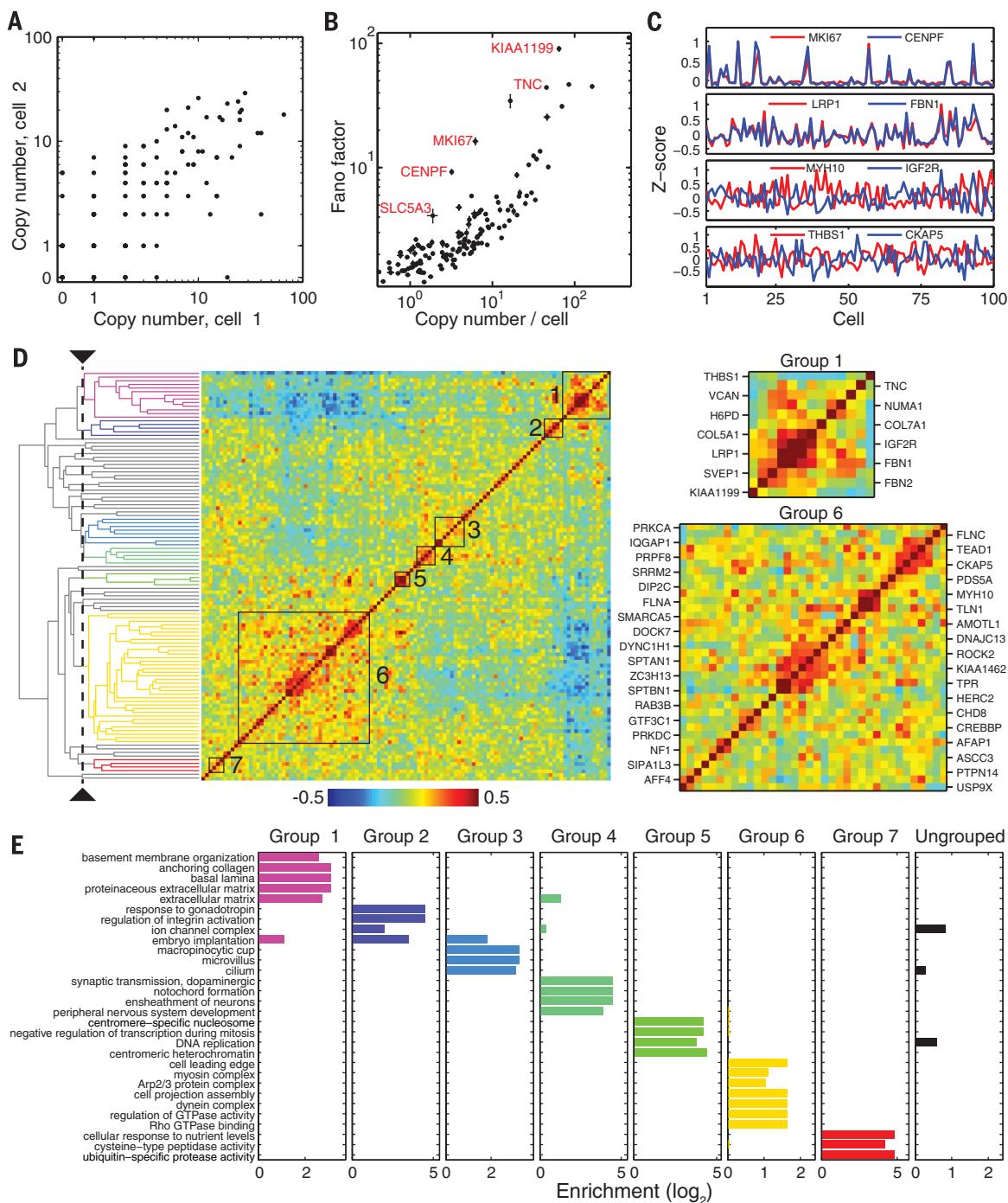


Fig. 3. Cell-to-cell variations and pairwise correlations for the RNA species determined from the 140-gene measurements. (A) Comparison of gene expression levels in two individual cells. (B) Fano factors for individual genes. Error bars represent standard error of the mean determined from seven independent data sets. (C) Z-scores of the expression variations of four example pairs of genes showing correlated (top two) or anticorrelated (bottom two) variation for 100 randomly selected cells. Z-score is defined as the difference from the mean normalized by the standard deviation. (D) Matrix of the pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes, shown together with the hierarchical clustering tree. The seven groups identified by a specific threshold on the cluster tree (dashed

line) are indicated by the black boxes in the matrix and colored lines on the tree, with gray lines on the tree indicating ungrouped genes. Different threshold choices on the cluster tree could be made to select either smaller subgroups with tighter correlations or larger super-groups containing more weakly coupled subgroups. Two of the seven groups are enlarged on the right. (E) Enrichment of 30 selected, statistically significantly enriched GO terms in the seven groups. Enrichment refers to the ratio of the fraction of genes within a group that have the specific GO term to the fraction of all measured genes having that term. Top 10 statistically significantly enriched GO terms for each of the seven groups are shown in table S2. Not all of the GO terms presented here are in the top 10 list.

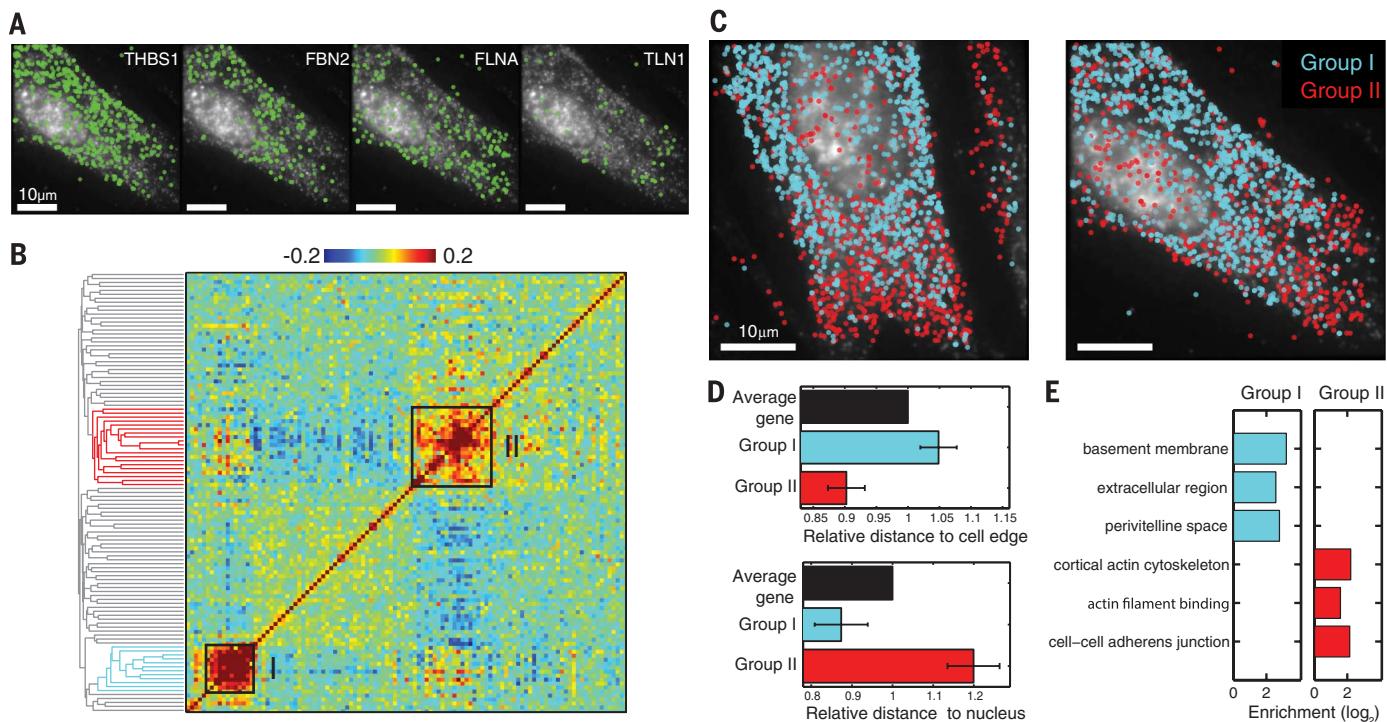


Fig. 4. Distinct spatial distributions of RNAs observed in the 140-gene measurements. (A) Examples of the spatial distributions observed for four different RNA species in a cell. (B) Matrix of the pairwise correlation coefficients describing the degree with which the spatial distributions of each gene pair is correlated, shown together with the hierarchical clustering tree. Two strongly correlating groups are indicated by the black boxes on the matrix

that the spatial pattern that we observed for these mRNAs reflects their cotranslational enrichment at the ER. The enrichment of these mRNAs in the perinuclear region (Fig. 4, C and D, light blue), where the rough ER resides, supports this conclusion.

Group II contained genes encoding the actin-binding proteins, including filamins FLNA and FLNC, talin TLN1, and spectrins SPTAN1 and SPTBN1; the microtubule-binding protein CKAP5; and the motor proteins MYH10 and DYNC1H1. This group was enriched with GO terms such as cortical actin cytoskeleton, actin filament binding, and cell-cell adherens junction (Fig. 4E). It has been shown previously that β -actin mRNA is enriched near the cell periphery in fibroblasts, as are mRNAs that encode members of the actin-binding Arp2/3 complex (36, 37). The enrichment of group II mRNAs in the peripheral region of the cells (Fig. 4, C and D) suggests that the spatial distribution of the group II genes might be related to the distribution of actin cytoskeleton mRNAs.

Measuring 1001 genes with MERFISH by use of a 14-bit MHD2 code

Last, we sought to further increase the throughput of our MERFISH measurement by simultaneously imaging ~1000 RNA species. This increase could be achieved with our MHD4 code by increasing the number of bits per code word to 32 while maintaining the number of 1 bits per word

at four (Fig. 1B). This could be implemented by either increasing the number of hybridization round to 32 or maintaining 16 rounds of hybridization, but using two-color imaging in each round. We pursued an alternative approach that did not require an increase in the number of hybridizations or color channels by relaxing the error correction requirement but keeping the error-detection capability. For example, by reducing the Hamming distance from 4 to 2, we could use all 14-bit words that contain four 1 bits to encode 1001 genes and probe these RNAs with only 14 rounds of hybridization. However, because a single error can produce a word equally close to two different code words, error correction is no longer possible for this modified Hamming-distance-2 (MHD2) code. Hence, we expect the calling rate to be lower and the misidentification rate to be higher with this encoding scheme.

To evaluate the performance of this 14-bit MHD2 code, we set aside 16 of the 1001 possible code words as misidentification controls and used the remaining 985 words to encode cellular RNAs (table S3). Among these 985 RNAs, we included 107 RNA species probed in the 140-gene experiments as an additional control. We performed the 1001-gene experiments in IMR90 cells by using a similar procedure as described above. To allow all encoding probes to be synthesized from a single 100,000-member oligopool, we reduced the number of encoding probes per RNA species to ~94. Fluorescent spots corresponding

and color on the tree. (C) The spatial distributions of all RNAs in the two groups in two example cells. Light blue symbols, group I genes; red symbols, group II genes. (D) Average distances for genes in group I and genes in group II to the cell edge or the nucleus normalized to the average distances for all genes. Error bars represent SEM across seven data sets. (E) Enrichment of GO terms in each of the two groups.

to individual RNA molecules were again clearly detected in each round of hybridization with the readout probes, and based on their on-off patterns, these spots were decoded into RNA (Fig. 5A and fig. S9, A and B). In the cell shown in Fig. 5A, 430 RNA species were detected, and similar results were obtained in ~200 imaged cells in three independent experiments.

As expected, the misidentification rate of this scheme was higher than that of the MHD4 code. Of all real RNA words, 77% were detected more frequently than the median count for the misidentification controls, instead of the 95% value observed in the MHD4 measurements. Using the same confidence ratio analysis as described above, we found that 73% (instead of 91% for the MHD4 measurements) of the 985 RNA species were measured with a confidence ratio larger than the maximum value observed for the misidentification controls (fig. S9C). RNA copy numbers measured from these 73% RNA species showed excellent correlation with our bulk RNA sequencing results (Pearson correlation coefficient $r = 0.76$) (Fig. 5B, black). The remaining 27% of the genes still exhibit good, albeit lower, correlation with the bulk RNA sequencing data ($r = 0.65$) (Fig. 5B, red), but we took the conservative measure of excluding them from further analysis.

The lack of an error correction capability also decreased the calling rate of each RNA species: When comparing the 107 RNA species common in both the 1001-gene and 140-gene measurements,

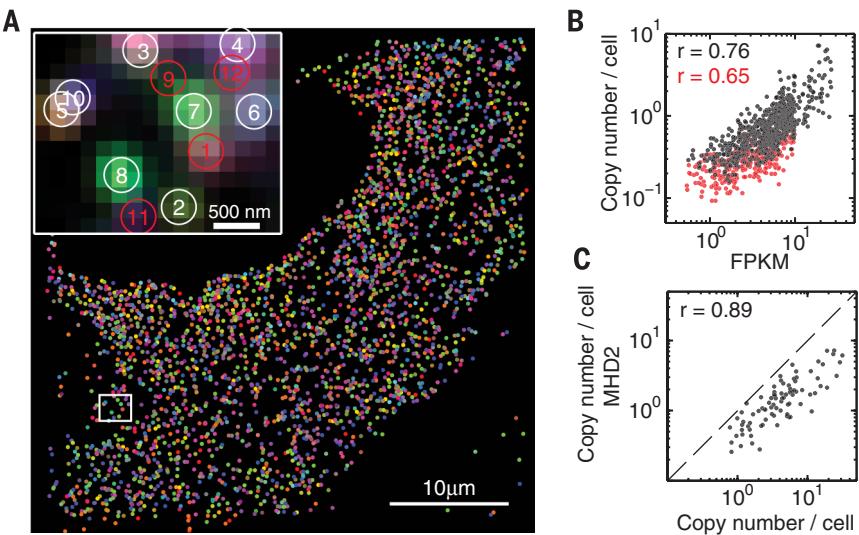


Fig. 5. Simultaneous measurements of 1001 RNA species in single cells by using MERFISH with a 14-bit MHD2 code. (A) The localizations of all detected single molecules in a cell colored based on their measured binary words. (Inset) The composite, false-colored fluorescent image of the 14 hybridization rounds for the boxed subregion with numbered circles indicating potential RNA molecules. Red circles indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Images of individual hybridization round are shown in fig. S9A. (B) Scatter plot of the average copy number per cell measured in the 1001-gene experiments versus the abundance measured via bulk sequencing. The black symbols are for the 73% of genes detected with confidence ratios higher than the maximum ratio observed for the misidentification controls. The Pearson correlation coefficient is 0.76 with a *P* value of 3×10^{-133} . The red symbols are for the remaining 27% of genes. The Pearson correlation coefficient is 0.65 with a *P* value of 3×10^{-33} . (C) Scatter plot of the average copy number for the 107 genes shared in both the 1001-gene measurement with the MHD2 code and the 140-gene measurement with the MHD4 code. The Pearson correlation coefficient is 0.89 with a *P* value of 9×10^{-30} . The dashed line corresponds to the $y = x$ line.

we found that the copy numbers per cell of these RNA species were lower in the 1001-gene measurements (Fig. 5C and fig. S9D). The total count of these RNAs per cell was $\sim 1/3$ of that observed in the 140-gene measurements. Thus, the lack of error correction in the MHD2 code reduced the calling rate to $\sim 30\%$ of that of the MHD4 code, which is consistent with the decrease in calling rate observed for the MHD4 code when error correction was not applied. As expected from the quantitative agreement between 140-gene measurements and conventional smFISH results, comparison of the 1001-gene measurements with conventional smFISH results for 10 RNA species also indicated a calling rate that is $\sim 1/3$ of that observed for the MHD4 code (fig. S8C). Despite the expected reduction in calling rate, the good correlations found between the copy numbers observed in the 1001-gene measurements and those observed in the 140-gene measurements, as well as in conventional smFISH and bulk RNA sequencing measurements, indicates that the relative abundance of these RNAs can be quantified with the MHD2 encoding scheme.

Simultaneously imaging ~ 1000 genes in individual cells substantially expanded our ability to detect coregulated genes. The matrix of pairwise correlation coefficients determined from the cell-to-cell variations in the expression levels of these genes is shown in Fig. 6A. Using the same hierarchical clustering analysis as described above,

we identified ~ 100 groups of genes with correlated expression (table S4). Nearly all of these ~ 100 groups showed statistically significant enrichment of functionally related GO terms (Fig. 6B and table S4). These included some of the groups identified in the 140-gene measurements, such as the group associated with cell-replication genes and the group associated with cell-motility genes (Fig. 6, A and B, groups 7 and 102), as well as many new groups. The groups identified here included 46 RNA species lacking any previous GO annotations, for which we can now hypothesize function on the basis of their group association (table S4). For example, *KIAA1462* is part of the cell motility group, as also shown in the 140-gene experiments, suggesting a potential role of this gene in cell motility (Fig. 6A, group 102). Likewise, *KIAA0355* is part of a new group enriched in genes associated with heart development (Fig. 6A, group 79), and *C17orf70* is part of a group associated with ribosomal RNA processing (Fig. 6A, group 22). Using these groupings, we can also hypothesize cellular functions for 61 transcription factors and other partially annotated proteins of unknown functions (table S4). For example, the transcription factors Z3CH13 and CHD8 are both members of the cell-motility group, suggesting their potential role in the transcriptional regulation of cell-motility genes. Although these predicted functions based on gene-association analysis require further validation, our covaria-

tion data provide a resource for generating hypotheses on gene function and regulation.

Discussion

We have developed a highly multiplexed detection scheme for transcriptomic-scale RNA imaging in single cells. Using combinatorial labeling, sequential hybridization and imaging, and two different error-robust encoding schemes, we simultaneously imaged either 140 or 1001 genes in hundreds of individual human fibroblast cells. Of the two encoding schemes presented here, the MHD4 code is capable of both error detection and error correction and hence can provide a higher calling rate and a lower misidentification rate than can the MHD2 code, which instead can only detect but cannot correct errors. MHD2, on the other hand, provides a faster scaling of the degree of multiplexing with the number of bits than can MHD4. Other error-robust encoding schemes can also be used for such multiplexed imaging, and experimenters can set the balance between detection accuracy and ease of multiplexing according to the specific requirements of the experiments.

By increasing the number of bits in the code words, it should be possible to further increase the number of detectable RNA species by using MERFISH with either MHD4 or MHD2 codes. Because of their much slower increase in error rates with the number of bits, we expect the error-correcting encoding schemes, such as MHD4, to be more favorable for scaling up the measurements. For example, using the MHD4 code with 32 total bits and four or six 1 bits would increase the number of addressable RNA species to 1240 or 27,776, respectively; the latter is the approximate scale of the human transcriptome. The predicted misidentification and calling rates are still reasonable for the 32-bit MHD4 code (shown in Fig. 1, C and D, purple for the MHD4 code with four 1 bits, and similar rates were calculated for the MHD4 code with six 1 bits). If more accurate measurements are desired, an additional increase in the number of bits would allow the use of encoding schemes with a Hamming distance greater than 4, further enhancing the error detection and correction capability. Although an increase in the number of bits by adding more hybridization rounds would increase the data collection time and potentially lead to sample degradation, these problems could be mitigated by using multiple colors to readout multiple bits in each round of hybridization.

As the degree of multiplexing is increased, it is important to consider the potential increase in the density of RNAs that need to be resolved in each round of imaging. On the basis of our imaging and sequencing results, we estimate that including the whole transcriptome of the IMR90 cells would lead to a total RNA density of ~ 200 molecules/ μm^3 . Using our current imaging and analysis methods, we could resolve 2 to 3 molecules/ μm^3 per hybridization round (38), which would reach a total RNA density of ~ 20 molecules/ μm^3 after 32 rounds of hybridization. This density should allow all but the top 10% most expressed genes to be imaged simultaneously or a subset of genes with even higher expression levels to be included.

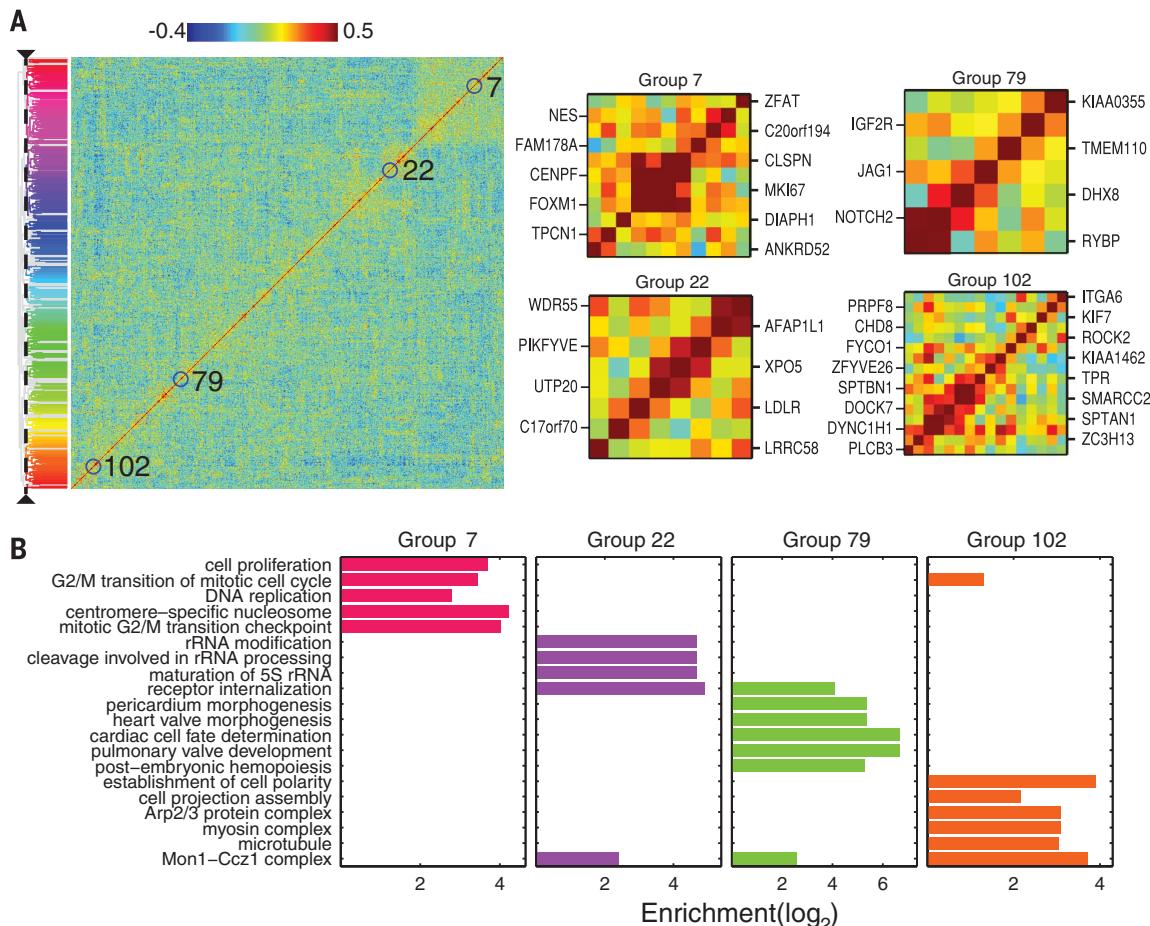


Fig. 6. Covariation analysis of the RNA species measured in the 1001-gene measurements. (A) Matrix of all pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes shown with the hierarchical clustering tree. The ~100 identified groups of correlated genes are indicated by color on the tree. Zoom-in of four of the groups described in the text are shown on the right. **(B)** Enrichment of 20 selected, statistically significantly enriched GO terms in the four groups. The statistically most significantly enriched GO terms (maximum 10) for each of the ~100 groups are shown in table S4.

By using more advanced image analysis algorithms to better resolve overlapping images of individual molecules, such as compressed sensing (39, 40), it would be possible to extend the resolvable density by approximately fourfold and thus allow nearly the entire transcriptome, except for the top 2% most expressed genes, to be imaged all together. Last, theoretical predictions (17) indicate that the use of superresolution imaging (41, 42) could increase the resolvable density to $\sim 10^5$ molecules/ μm^3 , which should be ample to address the entire transcriptome, even in cell types with RNA densities substantially higher than that of IMR90. However, RNAs in densely packed structures, such as p-bodies and stress granules, may still elude measurement.

We have illustrated the utility of the data derived from highly multiplexed RNA imaging by using covariation and correlation analysis to reveal distinct subcellular distribution patterns of RNAs, to constrain gene regulatory networks, and to predict functions for many previously unannotated or partially annotated genes with unknown functions. We anticipate that many more quantitative analyses could be applied to such data sets that include the spatial localization and copy number information of many RNA species in

individual cells. Given its ability to quantify RNAs across a wide range of abundances without amplification bias while preserving native context, we envision that MERFISH will enable many applications of *in situ* transcriptomic analyses of individual cells in culture or complex tissues.

Materials and Methods

Probe design

Each RNA species in our target set was randomly assigned a binary code word either from all 140 possible code words of the 16-bit MHD4 code or from all 1001 possible code words of the 14-bit MHD2 code, as we describe in the main text. The encoding schemes are provided in tables S1 and S3.

We used array-synthesized oligopools as templates to make the encoding probes (22, 23). The template molecule for each encoding probe contains three components: (i) a central targeting sequence for *in situ* hybridization to the target RNA, (ii) two flanking readout sequences designed to hybridize each of two distinct readout probes, and (iii) two flanking primer sequences to allow enzymatic amplification of the probes

(fig. S3). The readout sequences were taken from the 16 possible readout sequences, each corresponding to one hybridization round. The readout sequences were assigned to the encoding probes so that for any RNA species, each of the four readout sequences were distributed uniformly along the length of the target RNA and appeared at the same frequency. Template molecules for the 140-gene library also included a common 20-nucleotide (nt) priming region between the first polymerase chain reaction (PCR) primer and the first readout sequence. This priming sequence was used for the reverse transcription step described below. All template sequences are provided in table S5.

We embedded multiple experiments in a single array-synthesized oligopool and used PCR to selectively amplify only the oligos required for a specific experiment. Primer sequences for this indexed PCR reaction were generated from a set of orthogonal 25-nt sequences (43). These sequences were trimmed to 20 nt and selected for (i) a narrow melting temperature range (70 to 80°C), (ii) the absence of consecutive repeats of 3 or more identical nucleotides, and (iii) the presence of a GC clamp—one of the two 3' terminal bases must be G or C. To further improve specificity,

Table 1. Primer sequences used in the 140-gene and 1001-gene experiments.

Experiment name	Primer 1 sequence (index primer 1)	Primer 2 sequence (T7 promoter plus the reverse complement of index primer 2)
140-gene codebook 1	GTTGGTCGGCACTTGGGTGC	TAATACGACTCACTATAGGAAAGCCGGTTACCCGGTGG
140-gene codebook 2	CGATGCGCCAATTCCGGTTC	TAATACGACTCACTATAGGTGATCATCGCTCGGGGTTG
1001-gene	CGCGGGCTATATGCGAACCG	TAATACGACTCACTATAGGGCGTGGAGGGCATAAACGC

these sequences were then screened against the human transcriptome by using Basic Local Alignment Search Tool+ (BLAST+) (44), and primers with 14 or more contiguous bases of homology were eliminated. Last, BLAST+ was again used to identify and exclude primers that had an 11-nt homology region at the 3' end of any other primer or a 5-nt homology region at the 3' end of the T7 promoter. The forward primer sequences (primer 1) were determined as described above, whereas the reverse primers each contain a 20-nt sequence as described above plus a 20-nt T7 promoter sequence to facilitate amplification via in vitro transcription (primer 2). The primer sequences used in the 140-gene and 1001-gene experiments are listed in Table 1.

Thirty-nt-long readout sequences were created by concatenating fragments of the same orthogonal primer set generated above by combining one 20-nt primer with a 10-nt fragment of another. These readout sequences were then screened, by using BLAST+, for orthogonality with the index primer sequences and other readout sequences (no more than 11 nt of homology) and for potential off-target binding sites in the human genome (no more than 14 nt of homology). Fluorescently labeled readout probes with sequences complementary to the readout sequences were used to probe these readout sequences, one in each hybridization round. All used readout probes sequences are listed in Table 2.

The readout probes used for the 140-gene libraries were probes 1 through 16. The readout probes used for the 1001-gene experiment were probes 1 through 14. "/3Cy5Sp/" indicates a 3' Cy5 modification.

To design the central targeting sequences of the encoding probes, we first compiled the abundance of different transcripts in IMR90 cells using Cufflinks v2.1 (45), total RNA data from the Encyclopedia of DNA Elements (ENCODE) project (46), and human genome annotations from Gencode v18 (20). Probes were designed from gene models corresponding to the most abundant isoform by using OligoArray2.1 (47) with the following constraints: The target sequence region is 30-nt long; the melting temperatures of the hybridized region of the probe and cellular RNA target is greater than 70°C; there is no cross hybridization targets with melting temperatures greater than 72°C; there is no predicted internal secondary structures with melting temperatures greater than 76°C; and there is no contiguous repeats of six or more identical nucleotides. Melting temperatures were adjusted to optimize the specificity of these probes and minimize secondary structure while

Table 2. All used readout probes sequences.

Bit	Readout probes
1	CGCAACGCTTGGGACGGTTCAATCGGATC/3Cy5Sp/
2	CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/
3	ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/
4	CAAGTATGCAGCGCATTGACCGTCTCGTT/3Cy5Sp/
5	GCAGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/
6	AAGTCGTACGCCATGCGCAGCAATTCACT/3Cy5Sp/
7	CGAAACATCGGCCACGGTCCCGTTGAACCT/3Cy5Sp/
8	ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/
9	CGCGAAATCCCCTAACGAGCGTCCCTTG/3Cy5Sp/
10	GCATGAGTTGCCTGGCCTGCGACGACTAA/3Cy5Sp/
11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/
12	GGCCAATGGCCCAGGTCCGTACCGCAATT/3Cy5Sp/
13	TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/
14	CGCGCGGATCCGCTTGTGCGGAAACGGATAC/3Cy5Sp/
15	GCCTCGATTACGACGGATGTAATTGGCCG/3Cy5Sp/
16	GCCCGTATTCCCGCTTGCAGTAGGGCAAT/3Cy5Sp/

still producing sufficient numbers of probes for our libraries. To decrease computational cost, isoforms were divided into 1-kb regions for probe design. Using BLAST+, all potential probes that mapped to more than one cellular RNA species were rejected. Probes with multiple targets on the same RNA were kept.

For each gene in the 140-gene experiments, we generated 198 putative encoding probe sequences by concatenating the appropriate index primers, readout sequences, and targeting regions as shown in fig S3. To address the possibility that concatenation of these sequences introduced new regions of homology to off-target RNAs, we used BLAST+ to screen these putative sequences against all human ribosomal RNA (rRNA) and transfer RNA (tRNA) sequences as well as highly expressed genes [genes with fragments per kilobase per million reads (FPKM) > 10,000]. Probes with greater than 14 nt of homology to rRNAs or tRNAs or greater than 17 nt of homology to highly expressed genes were removed. After these cuts, we had ~192 (with a standard deviation of 2) probes per gene for both MHD4 codebooks used in the 140-gene experiments. We followed the same protocol for the 1001-gene experiments: Starting with 96 putative targeting sequences per gene, we obtained ~94 (with a standard deviation of 6) encoding probes per gene after these additional homology cuts. We decreased the number of encoding probes per RNA for the 1001-gene experiments so that these probes could be synthesized from a single 100,000-member oligopool as opposed to two separate pools. We designed each encoding probe to contain two of the four read-

out sequences associated with each code word; hence, only half of the bound encoding probes can bind readout probe during any given hybridization round. We used ~192 or ~94 encoding probes per RNA to obtain high signal-to-background ratios for individual RNA molecules. The number of encoding probes per RNA could be substantially reduced but still allow single RNA molecules to be identified (17, 48, 49). In addition, increasing the number of readout sequences per encoding probe or using optical sectioning methods to reduce the fluorescence background may allow further reduction in the number of the encoding probes per RNA.

We designed two types of misidentification controls. The first control—blank words—were not represented with encoding probes. The second type of control—no-target words—had encoding probes that were not targeting any cellular RNA. The targeting regions of these probes were composed of random nucleotide sequences subject to the same constraints used to design the RNA targeting sequences described above. Moreover, these random sequences were screened against the human transcriptome to ensure that they contain no substantial homology (>14-nt) to any human RNA. The 140-gene measurements contained five blank words and five no-target words. The 1001-gene measurements contained 11 blank words and five no-target words.

Probe synthesis

The encoding probes were synthesized by using the following four steps, and this synthesis protocol is illustrated in fig. S3.

Step 1: The template oligopool (CustomArray) was amplified via limited-cycle PCR on a Bio-Rad CFX96 by using primer sequences specific to the desired probe set. To facilitate subsequent amplification via *in vitro* transcription, the reverse primer contained the T7 promoter. All primers were synthesized by Integrated DNA Technologies (IDT). This reaction was column purified (Zymo DNA Clean and Concentrator, D4003).

Step 2: The purified PCR products were then further amplified ~200-fold and converted into RNA via a high yield *in vitro* transcription according to the manufacturer's instructions [New England Biolabs (NEB), E2040S]. Each 20 μL reaction contained ~1 μg of template DNA from above, 10 mM of each NTP, 1× reaction buffer, 1× RNase inhibitor (Promega RNasin, N2611) and 2 μL of the T7 polymerase. This reaction was incubated at 37°C for 4 hours to maximize yield. This reaction was not purified before the following steps.

Step 3: The RNA products from the above *in vitro* transcription reaction were then converted back into DNA via a reverse transcription reaction. Each 50-μL reaction contained the unpurified RNA produce from step 2 supplemented with 1.6 mM of each dNTP, 2 nmol of a reverse transcription primer, 300 units of Maxima H- reverse transcriptase (Thermo Scientific, EP0751), 60 units of RNasin, and a final 1× concentration of the Maxima RT buffer. This reaction was incubated at 50°C for 45 min, and the reverse transcriptase was inactivated at 85°C for 5 min. The templates for the 140-gene libraries contain a common priming region for this reverse transcription step; thus, a single primer was used for this step when creating these probes. Its sequence was CGGGTTAGCGCCGAAATG. A common priming region was not included for the 1001-gene library; thus, the reverse transcription was conducted with the forward primer: CGCGGGCTATATGCCAACCG.

Step 4: To remove the template RNA, 20 μL of 0.25 M EDTA and 0.5 N NaOH was added to the above reaction to selectively hydrolyze RNA, and the sample was incubated at 95°C for 10 min. This reaction was then immediately purified by means of column purification using a 100-μg-capacity column (Zymo Research, D4030) and the Zymo Oligo Clean and Concentrator protocol. The final probes were eluted in 100 μL of ribonuclease (RNase)-free deionized water, evaporated in a vacuum concentrator, and then resuspended in 10 μL of encoding hybridization buffer (recipe below). Probes were stored at -20°C. Denaturing polyacrylamide gel electrophoresis and absorption spectroscopy were used to confirm the quality of the probes and revealed that this probe synthesis protocol converts 90 to 100% of the reverse-transcription primer into full-length probe and of the probe that is constructed, 70 to 80% is recovered during the purification step. This protocol is similar to another recently published protocol (23) but provides a substantially larger yield.

Fluorescently labeled readout probes have sequences complementary to the readout sequences

described above and a Cy5 dye attached at the 3' end. These probes were synthesized and purified by means of high-performance liquid chromatography (HPLC) by IDT.

Sample preparation and labeling with encoding probes

Human primary fibroblasts (American Type Culture Collection, IMR90), a commonly used cell line with a previously determined transcriptome (46), were used in this work. These cells are relatively large and flat, facilitating wide-field imaging without the need for optical sectioning. Cells were cultured with Eagle's Minimum Essential Medium. Cells were plated on 22-mm, #1.5 coverslips (Bioptronic, 0420-0323-2) at 350,000 cells per coverslip and incubated at 37°C with 5% CO₂ for 48 to 96 hours within petri dishes. Cells were fixed for 20 min in 4% paraformaldehyde (Electron Microscopy Sciences, 15714) in 1× phosphate buffered saline (PBS; Ambion, AM9625) at room temperature, reduced for 5 min with 0.1% w/v sodium borohydride (Sigma, 480886) in water to reduce background fluorescence, washed three times with ice-cold 1× PBS, permeabilized for 2 min with 0.5% v/v Triton (Sigma, T8787) in 1× PBS at room temperature, and washed three times with ice cold 1× PBS.

Cells were incubated for 5 min in encoding wash buffer comprising 2× saline-sodium citrate buffer (SSC) (Ambion, AM9763), 30% v/v formamide (Ambion, AM9342), and 2 mM vanadyl ribonucleoside complex (NEB, S1402S). Ten microliters of 100 μM (140-gene experiments) or 200 μM (1001-gene experiments) encoding probes in encoding hybridization buffer was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37°C-hybridization oven for 18 to 36 hours. Encoding hybridization buffer is composed of encoding wash buffer supplemented with 1 mg/mL yeast tRNA (Life technologies, 15401-011) and 10% w/v dextran sulfate (Sigma, D8906-50G).

Cells were then washed with encoding wash buffer, incubated at 47°C for 10 min, and this wash was repeated for a total of three times. A 1:1000 dilution of 0.2-μm-diameter carboxylate-modified orange fluorescent beads (Life Technologies, F-8809) in 2×SSC was sonicated for 3 min and then incubated with the sample for 5 min. The beads were used as fiducial markers to align images obtained from multiple successive rounds of hybridization, as described below. The sample was washed once with 2×SSC, and then post-fixed with 4% v/v paraformaldehyde in 2×SSC at room temperature for 30 min. The sample was then washed three times with 2×SSC and either imaged immediately or stored for no longer than 12 hours at 4°C before imaging. All solutions were prepared as RNase-free.

MERFISH imaging

The sample coverslip was assembled into a Bioptronic's FCS2 flow chamber, and the flow through this chamber was controlled via a home-built fluidics system composed of three computer-

controlled eight-way valves (Hamilton, MVP and HVXM 8-5) and a computer-controlled peristaltic pump (Rainin, Dynamax RP-1). The sample was imaged on a home-built microscope constructed around an Olympus IX-71 body and a 1.45 NA, 100× oil immersion objective and configured for oblique incidence excitation. The objective was heated to 37°C with a Bioptronic objective heater. Constant focus was maintained throughout the imaging process with a home-built, autofocus system. Illumination was provided at 641, 561, and 405 nm by using solid-state lasers (MPB communications, VFL-P500-642; Coherent, 561-200CWCDRH; and Coherent, 1069413/AT) for excitation of our Cy5-labeled readout probes, the fiducial beads, and nuclear counterstains, respectively. These lines were combined with a custom dichroic (Chroma, zy405/488/561/647/752RP-UFI) and the emission was filtered with a custom dichroic (Chroma, ZET405/488/561/647-656/752m). Fluorescence was separated with a QuadView (Photometrics) by using the dichroics T560lpix, T650lpix, and 750dcxxr (Chroma) and the emission filters ET525/50m, WT59550m-2f, ET700/75m, and HQ770lp (Chroma) and imaged with an EMCCD camera (Andor, iXon-897). The camera was configured so that a pixel corresponds to 167 nm in the sample plane. The entire system was fully automated, so that imaging and fluid handling were performed for the entire experiment without user intervention.

Sequential hybridization, imaging, and bleaching proceeded as follows. One milliliter of 10 nM of the appropriate fluorescently labeled readout probe in readout hybridization buffer (2×SSC; 10% v/v formamide, 10% w/v dextran sulfate, and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 15 min. Then 2 mL of readout wash buffer (2×SSC, 20% v/v formamide, and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 3 min. Two milliliters of imaging buffer comprising 2×SSC, 50 mM TrisHCl pH 8, 10% w/v glucose, 2 mM Trolox (Sigma-Aldrich, 238813), 0.5 mg/mL glucose oxidase (Sigma-Aldrich, G2133), and 40 μg/mL catalase (Sigma-Aldrich, C30) was flown across the sample (50). Flow was then stopped, and then ~75 to 100 regions were exposed to ~25 mW 642-nm and 1 mW of 561-nm light and imaged. Each region was 40 by 40 μm. The laser powers were measured at the microscope back-port. Because the imaging buffer is sensitive to oxygen (51), the ~50 mL of imaging buffer used for a single experiment was made fresh at the beginning of the experiment and then stored under a layer of mineral oil throughout the measurement. Buffer stored in this fashion was stable for more than 24 hours.

After imaging, the fluorescence of the readout probes was extinguished via photobleaching. The sample was washed with 2 mL of photo-bleaching buffer (2×SSC and 2 mM vanadyl ribonucleoside complex), and each imaged region of the sample was exposed to 200 mW of 641-nm light for 3 s. To confirm the efficacy of

this photobleaching treatment, imaging buffer was reintroduced, and the sample was imaged as described above.

The above hybridization, imaging, and photo-bleaching process was repeated either 16 times for the 140-gene measurements by using the MHD4 code or 14 times for the 1001-gene measurements by using the MHD2 code. An entire experiment was typically completed in ~20 hours.

After completion of imaging, 2 mL of a 1:1000 dilution of Hoescht (ENZ-52401) in 2×SSC was flown through the chamber to label the nuclei of the cells. The sample was then washed immediately with 2 mL of 2×SSC followed by 2 mL of imaging buffer. Each region of the sample was then imaged once again with ~1 mW of 405-nm light.

Because we imaged cells using wide-field imaging with oblique-incidence illumination, without optical sectioning and *z*-scanning, we quantified the fraction of individual RNA species that was outside the axial range of our imaging geometry for six different RNA species using conventional smFISH. For this purpose, we optically sectioned these cells by collecting stacks of images at different focal depths through the entire depth of the cells. We aligned the images in consecutive focal planes and then computed for each cell the fraction of RNAs that were detected in the three-dimensional stack but not in the basal focal plane. We found that only a small fraction, $15 \pm 1\%$ (mean \pm SEM across six different RNA species) of RNA molecules were outside the imaging range of a fixed focal plane without *z*-scanning. These measurements also confirmed that our excitation geometry illuminated the full depth of our cells. From an imaging perspective, any optical sectioning technique could be used in MERFISH to allow the imaging of RNAs in thicker cells or tissues.

Construction of measured words

Fluorescent spots were identified and localized in each image by using a multi-Gaussian-fitting algorithm (38) assuming a Gaussian with a uniform width of 167 nm. This algorithm was used to allow partially overlapping spots to be distinguished and individually fit. RNA spots were distinguished from background signal—signal arising from probes bound nonspecifically, by setting the intensity threshold required to fit a spot with this software. Because of variation in the brightness of spots between rounds of hybridization, this threshold was adjusted appropriately for each hybridization round in order to minimize the combined average of the 1→0 and 0→1 error rates across all hybridization rounds (140-gene measurements) or to maximize the ratio of the number of measured words with four 1 bits to those with three or five 1 bits (1001-gene measurements). The location of the fiducial beads was identified in each frame by using a faster single-Gaussian fitting algorithm.

Images of the same sample region in different rounds of hybridization were registered by rotating and translating the image to align the two fiducial beads within the same image that were

most similar in location after a coarse initial alignment via image correlation. All images were aligned to a coordinate system established by the images collected in the first round of hybridization. The quality of this alignment was determined from the residual distance between five additional fiducial beads, and alignment error was typically ~20 nm.

Fluorescence spots in different hybridization rounds were connected into a single string, corresponding to a potential RNA molecule, if the distance between spots was smaller than 1 pixel (167 nm). For each string of spots, the on-off sequence of fluorescent signals in all hybridization rounds were used to assign a binary word to the potential RNA molecule, in which 1 was assigned to the hybridization rounds that contained a fluorescent signal above threshold and 0 was assigned to the other hybridization rounds. Measured words were then decoded into RNA species by using the 16-bit MHD4 code or the 14-bit MHD2 code discussed in the main text. In the case of the 16-bit MHD4 code, if the measured binary word matched the code word of a specific RNA perfectly or differed from the code word by one single bit, it was assigned to that RNA. In the case of the 14-bit MHD2 code, only if the measured binary word matched the code word of a specific RNA perfectly was it assigned to that RNA. To determine the copy number per cell, the number of each RNA species was counted in individual cells within each 40- by 40- μm imaging area. This number accounts for the majority but not all RNA molecules within a cell because a fraction of the cell could be outside the imaging area or focal depth. Tiling images of adjacent areas and adjacent focal planes could be used to improve the counting accuracy.

In the 140-gene experiments, some regions of the cell nucleus occasionally contained too much fluorescence signal to properly identify individual RNA spots. In the 1001-gene experiments, the cell nucleus generally contained too much fluorescent signal to allow identification of individual RNA molecules. These bright regions were excluded from all subsequent analysis. This work focuses on mRNAs, which are enriched in the cytoplasm. To estimate the fraction of mRNAs missed by excluding the nucleus region, we used conventional smFISH to quantify the fraction of molecules found inside the nucleus for six different mRNAs species. We found that only $5 \pm 2\%$ (mean \pm SEM across six RNA species) of these RNA molecules are found in the nucleus. Use of super-resolution imaging and/or optical sectioning could potentially allow individual molecules in these dense nucleus regions to be identified, which will be particularly useful for probing those noncoding RNAs that are enriched in the nucleus.

smFISH measurements of individual genes

Pools of 48 fluorescently labeled (Quasar 670) oligonucleotide probes per RNA were purchased from Biosearch Technologies. Thirty-nucleotide probe sequences were taken directly from a random subset of the targeting regions used for the

multiplexed measurements. Cells were fixed and permeabilized as described above. Ten microliters of 250 nM oligonucleotide probes in encoding hybridization buffer (described above) was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37°C-hybridization oven for 18 hours. Cells were then washed with encoding wash buffer (described above) at 37°C for 10 min, and this wash was repeated for a total of three times. The sample was then washed three times with 2×SSC and imaged in imaging buffer by using the same imaging geometry as described above for MERFISH imaging.

Bulk RNA sequencing

Total RNA was extracted from IMR90 cells cultured as above using the Zymo Quick RNA MiniPrep kit (R1054) according to the manufacturer's instructions. Polyadenylated [poly(A)] RNA was then selected (NEB, E7490), and a sequencing library was constructed by using the NEBNext Ultra RNA library preparation kit (NEB, E7530), amplified with custom oligonucleotides, and 150-base pair (bp) reads were obtained on a MiSeq. These sequences were aligned to the human genome (Gencode v18) and isoform abundance was computed with cufflinks (45).

Calculation of the predicted scaling and error properties of different encoding schemes

Analytic expressions were derived for the dependence of the number of possible code words, the calling rate, and the misidentification rate on N . The calling rate is defined as the fraction of RNA molecules that are properly identified. The misidentification rate is defined as the fraction of RNA molecules that are misidentified as a wrong RNA species. For encoding schemes with an error-detection capability, the calling rate and misidentification rate does not add up to 1 because a fraction of the molecules not called properly can be detected as errors and discarded and, hence, not misidentified as a wrong species. These calculations assume that the probability of misreading bits is constant for all hybridization rounds but differs for the 1→0 and 0→1 errors. Experimentally measured average 1→0 and 0→1 error rates (10 and 4%, respectively) were used for the estimates shown in Fig. 1, B to D. For simplicity, the word corresponding to all 0s was not removed from calculations.

For the simple binary encoding scheme in which all possible N -bit binary words are assigned to different RNA species, the number of possible code words is 2^N . The number of words that could be used to encode RNA is actually $2^N - 1$ because the code word "00...0" does not contain detectable fluorescence in any hybridization round, but for simplicity the word corresponding to all 0s was not removed from subsequent calculations. The error introduced by this approximation is negligible. For any given word with m 1s and $N - m$ 0s, the probability of measuring

that word without error—the fraction of RNAs that is properly called—is

$$(1 - p_1)^m (1 - p_0)^{N-m} \quad (1)$$

where p_1 is 1→0 error rate and p_0 is 0→1 error rate per bit. Because different words in this simple binary encoding scheme can have different numbers of 1 bits, the calling rate for different words will differ if $p_1 \neq p_0$. The average calling rate, reported in Fig. 1C, was determined from the weighted average of the value of Eq. 1 for all words. This weighted average is

$$\frac{1}{2^N} \sum_{m=0}^N \binom{N}{m} (1 - p_1)^m (1 - p_0)^{N-m} \quad (2)$$

where $\binom{N}{m}$ is the binomial coefficient and corresponds to the number of words with m 1 bits in this encoding scheme. Because in this encoding scheme every error produces a binary word that encodes a different RNA, the average misidentification rate for this encoding scheme, reported in Fig. 1D, follows directly from Eq. 2:

$$1 - \frac{1}{2^N} \sum_{m=0}^N \binom{N}{m} (1 - p_1)^m (1 - p_0)^{N-m} \quad (3)$$

To calculate the scaling and error properties of the extended Hamming distance 4 (HD4) code, we first created the generator matrix for the desired number of data bits using standard methods (21). The generator matrix determines the specific words that are present in a given encoding scheme and was used to directly determine the number of encoded words as a function of the number of bits. In this encoding scheme, the calling rate corresponds to the fraction of words measured without error as well as the fraction of words measured with a single-bit error. For code words with m 1 bits, this fraction is determined by the following expression:

$$(1 - p_1)^m (1 - p_0)^{N-m} + mp_1^1 (1 - p_1)^{m-1} (1 - p_0)^{N-m} + (N - m)p_0^1 (1 - p_1)^m (1 - p_0)^{N-m-1} \quad (4)$$

where the first term is the probability of not making any errors, the second term corresponds to the total probability of making one 1→0 error at any of the m 1 bits without making any other 0→1 errors, and the final term corresponds to the total probability of making one 0→1 error at any of the $N-m$ 0 bits without making any 1→0 errors. Because the number of 1 bits can differ between words in this encoding scheme, the average calling rate reported in Fig. 1C was computed from a weighted average over Eq. 4 for different values of m . The weight for each term was determined from the number of words that contain m 1 bits as determined from the generator matrix described above.

Because RNA-encoding words are separated by a minimum Hamming distance of 4, at least 4 errors are required to switch one word into another. If error correction is applied, then 3 or 5 errors could also convert one RNA into another. Thus, we estimate the misidentification rate from all possible combinations of 3-bit, 4-bit, and 5-bit errors for code words with m 1 bits. Technically, >5-bit errors could also convert one RNA into

another, but the probability of making such errors is negligible because of the small per-bit error rate. We approximate this expression with

$$\begin{aligned} & \sum_{i=0}^4 \binom{m}{i} \binom{N-m}{4-i} p_1^i p_0^{4-i} (1-p_1)^{m-i} (1-p_0)^{N-m-(4-i)} + \\ & \sum_{i=0}^3 \binom{m}{i} \binom{N-m}{3-i} p_1^i p_0^{3-i} (1-p_1)^{m-i} (1-p_0)^{N-m-(3-i)} + \\ & \sum_{i=0}^5 \binom{m}{i} \binom{N-m}{5-i} p_1^i p_0^{5-i} (1-p_1)^{m-i} (1-p_0)^{N-m-(5-i)} \end{aligned} \quad (5)$$

The first sum corresponds to all of the ways in which exactly four mistakes can be made. Similarly, the second and third sums correspond to all of the ways in which exactly three or five mistakes can be made. Equation 5 provides an upper bound for the misidentification rate because not all 3-, 4-, or 5-bit errors produce a word that matches or would be corrected to another legitimate word. Again because the number of 1 bits can differ between words, the average misidentification rate reported in Fig. 1D is calculated as a weighted average of Eq. 5 over the number of words that have m 1 bits.

To generate our MHD4 code in which the number of 1 bits for each code word is set to 4, we first generated the HD4 codes as described above, and then removed all code words that did not contain four 1s. The calling rate of this code, reported in Fig. 1C, was directly calculated from Eq. 4, but with $m = 4$ because all code words in this code have four 1 bits. The misidentification rate of this code, reported in Fig. 1D, was calculated by modifying Eq. 5 with the following considerations: (i) the number of 1 bits, m , was set to 4 and (ii) errors that produce words that do not contain three, four, or five 1 bits were excluded. Thus, the expression in Eq. 5 was simplified to

$$\begin{aligned} & \binom{4}{2} \binom{N-4}{2} p_1^2 p_0^2 (1-p_1)^2 (1-p_0)^{N-6} + \\ & \binom{4}{1} \binom{N-4}{2} p_1^1 p_0^2 (1-p_1)^3 (1-p_0)^{N-6} + \\ & \binom{4}{2} \binom{N-4}{1} p_1^2 p_0^1 (1-p_1)^2 (1-p_0)^{N-5} \\ & \binom{4}{2} \binom{N-4}{3} p_1^2 p_0^3 (1-p_1)^2 (1-p_0)^{N-7} + \\ & \binom{4}{3} \binom{N-4}{2} p_1^3 p_0^2 (1-p_1)^1 (1-p_0)^{N-6} \end{aligned} \quad (6)$$

Again, this expression is an upper bound on the actual misidentification rate because not all words with four 1s are valid code words.

Estimates of the 1→0 and 0→1 error rates for each hybridization round

To compute the probability of misreading a bit at a given hybridization round, we used the error-correcting properties of the MHD4 code. Briefly, the probabilities of 1→0 or 0→1 errors were derived in the following way. Let the probability of

making an error at the i th bit, $-i$ th hybridization round—be p_i and the actual number of RNA molecules of the given species be A , then the number of exact matches for this RNA will be $W_E = A \prod_{i=1}^{16} (1 - p_i)$, and the number of one-bit error corrected matches for this RNA corresponding to errors at the i th bit will be $W_i = A \frac{p_i}{(1-p_i)} \prod_{j=1}^{16} (1 - p_j)$. The p_i can be directly derived from the ratio: $W_i/W_E = \frac{p_i}{(1-p_i)}$. This ratio assumes that the 1-bit error-corrected counts were only generated from single-bit errors from the correct word and that multi-error contamination from other RNA words is negligible. Given that our error rate per hybridization round is small and that it takes at least three errors to convert one RNA-encoding word into a word that would be misidentified as another RNA, the above approximation should be a good one.

To compute the average 1→0 or 0→1 error probabilities for each of the 16 hybridization rounds, we use the above approach to calculate the per-bit error rates for each bit of every gene, sort these errors on the basis of whether they correspond to a 1→0 or a 0→1 error, and then take the average of these errors for each bit weighted by the number of counts observed for the corresponding gene.

Estimates of the calling rate for individual RNA species from actual imaging data

With the estimates of the 1→0 or 0→1 error probabilities for each round of hybridization as determined above, it is possible to estimate the calling rate for each RNA according to the specific word used to encode it. Specifically, the fraction of an RNA species that is called correctly is determined by

$$\sum_{i=1}^N \frac{p_i}{(1-p_i)} \prod_{j=1}^N (1 - p_j) \quad (7)$$

where the first term represents the probability of observing an exact match of the code word and the second term represents the probability of observing an error-corrected match (with 1-bit error). The values of the per-bit error rate p_i for each RNA species are determined by the specific code word for that RNA and the measured 1→0 or 0→1 error rates for each round of hybridization. If the code word of the RNA contains a 1 in the i th bit, then p_i is determined from the 1→0 error rate for the i th hybridization round; if the word contains a 0 in the i th bit, p_i is determined from the 0→1 error rate for the i th hybridization round.

Hierarchical clustering analysis of the co-variation in RNA abundance

Hierarchical clustering of the covariation in gene expression for both the 140- and 1001-gene experiments was conducted as follows. First, the distance between every pair of genes was determined as 1 minus the Pearson correlation coefficient of the cell-to-cell variation of the measured copy numbers of these two RNA species, both normalized by the total RNA counted in the cell. Thus, highly correlated genes are “closer” to

one another, and highly anticorrelated genes are “further” apart. An agglomerative hierarchical cluster tree was then constructed from these distances using the unweighted pair group method with arithmetic mean (UPGMA). Specifically, starting with individual genes, we constructed hierarchical clusters by identifying the two clusters (or individual genes) that are closest to one another according to the arithmetic mean of the distances between all intercluster gene pairs. The pairs of clusters (or individual genes) with the smallest distance are then grouped together, and the process is repeated. The matrix of pairwise correlations was then sorted according to the order of the genes within these trees.

Groups of genes with substantial covariations were identified by selecting a threshold on the hierarchical cluster tree (indicated by the dashed lines in Figs. 3D and 6A) that produced ~10 groups of genes, each of which contains at least four members for the 140-gene experiments or ~100 groups each of which contains at least three members for the 1001-gene experiments. One can change the threshold in order to identify either more tightly coupled smaller groups or larger groups with relatively loose coupling.

A probability value for the confidence that a gene belongs to a specific group was determined by computing the difference between the average correlation coefficient between that gene and all other members of that group and the average correlation coefficient between that gene and all other measured genes outside that group. The significance (*P* value) of this difference was determined with the student's *t* test and is provided in tables S2 and S4.

Because hierarchical clustering is inherently a one-dimensional analysis—any given genes can only be a member of a single group—this analysis does not allow all correlated gene groups to be identified. Higher-dimension analysis, such as principal component analysis or *k*-means clustering, could be used to identify more covarying gene clusters (30).

Analysis of RNA spatial distributions

To identify genes that have similar spatial distributions, we subdivided each of the measured cells into 2 by 2 regions and calculated the fraction of each RNA species present in each of these bins. To control for the fact that some regions of the cell naturally contain more RNA than others, we calculate the enrichment for each gene—the ratio of the observed fraction in a given region for a given RNA species to the average fraction observed for all genes in that same region. For each pair of RNA species, we then determined the Pearson correlation coefficient of the region-to-region variation in enrichment of these two RNA species for each cell and averaged the correlation coefficients over ~400 cells imaged in seven independent data sets. We then clustered RNA species on the basis of these average correlation coefficients using the same hierarchical clustering algorithm described above. Because of the large number of cells used for the analysis, we found that the coarse spatial binning (2 by 2

regions per cell) was sufficient to capture the spatial correlation between genes, and finer binning did not produce more significantly correlated groups.

To measure the distances of genes from the nuclei and from the cell edge, we first used brightness thresholds on our cell images to segment the nuclei and identify the cell edge. We then measured the distance from every RNA molecule to the nearest part of the nucleus and nearest part of the cell edge. For each data set, we computed the average distance for each RNA species averaged over all the cells measured. We then averaged these distances for the group I genes, group II genes, or all genes. Only those RNA species with at least 10 counts per cell were used in this analysis to minimize statistical error on the distance values.

GO analysis

Groups of genes were selected from the hierarchical trees as discussed above. A collection of GO terms (31) was determined for all measured RNA species as well as the RNA species associated with each group from the most recent human GO annotations (<http://geneontology.org/page/download-annotations>) by using both the annotated GO terms and terms immediately upstream or downstream of the found annotations. The enrichment of these annotations was calculated from the ratio of the fraction of genes within each group that have this term to the fraction of all measured genes that have this term and the *P* value for this enrichment was calculated via the hypergeometric function. Only statistically significantly enriched GO terms with a *P* value less than 0.05 were considered.

REFERENCES AND NOTES

- N. Crosetto, M. Bienko, A. van Oudenaarden, Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* **16**, 57–66 (2015). doi: [10.1038/ng3832](https://doi.org/10.1038/ng3832); pmid: [25446315](https://pubmed.ncbi.nlm.nih.gov/25446315/)
- A. M. Femino, F. S. Fay, K. Fogarty, R. H. Singer, Visualization of single RNA transcripts in situ. *Science* **280**, 585–590 (1998). doi: [10.1126/science.280.5363.585](https://doi.org/10.1126/science.280.5363.585); pmid: [9554849](https://pubmed.ncbi.nlm.nih.gov/9554849/)
- A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008). doi: [10.1038/nmeth.1253](https://doi.org/10.1038/nmeth.1253); pmid: [18806792](https://pubmed.ncbi.nlm.nih.gov/18806792/)
- A. J. Rodriguez, K. Czaplicki, J. S. Condeelis, R. H. Singer, Mechanisms and cellular roles of local protein synthesis in mammalian cells. *Curr. Opin. Cell Biol.* **20**, 144–149 (2008). doi: [10.1016/j.ceb.2008.02.004](https://doi.org/10.1016/j.ceb.2008.02.004); pmid: [18378131](https://pubmed.ncbi.nlm.nih.gov/18378131/)
- V. Balagopal, R. Parker, Polysomes, P bodies and stress granules: States and fates of eukaryotic mRNAs. *Curr. Opin. Cell Biol.* **21**, 403–408 (2009). doi: [10.1016/j.ceb.2009.03.005](https://doi.org/10.1016/j.ceb.2009.03.005); pmid: [19394210](https://pubmed.ncbi.nlm.nih.gov/19394210/)
- H. Jung, C. G. Gkogkas, N. Sonenberg, C. E. Holt, Remote control of gene function by local translation. *Cell* **157**, 26–40 (2014). doi: [10.1016/j.cell.2014.03.005](https://doi.org/10.1016/j.cell.2014.03.005); pmid: [24679524](https://pubmed.ncbi.nlm.nih.gov/24679524/)
- T. Gregor, H. G. Garcia, S. C. Little, The embryo as a laboratory: Quantifying transcription in Drosophila. *Trends Genet.* **30**, 364–375 (2014). doi: [10.1016/tig.2014.06.002](https://doi.org/10.1016/tig.2014.06.002); pmid: [25005921](https://pubmed.ncbi.nlm.nih.gov/25005921/)
- A. R. Buxbaum, G. Haimovich, R. H. Singer, In the right place at the right time: Visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.* **16**, 95–109 (2015). doi: [10.1038/nrm3918](https://doi.org/10.1038/nrm3918); pmid: [25549890](https://pubmed.ncbi.nlm.nih.gov/25549890/)
- D. R. Larson, R. H. Singer, D. Zenklusen, A single molecule view of gene expression. *Trends Cell Biol.* **19**, 630–637 (2009). doi: [10.1016/j.tcb.2009.08.008](https://doi.org/10.1016/j.tcb.2009.08.008); pmid: [19819144](https://pubmed.ncbi.nlm.nih.gov/19819144/)
- A. Raj, A. van Oudenaarden, Single-molecule approaches to stochastic gene expression. *Annu. Rev. Biophys.* **38**, 255–270 (2009). doi: [10.1146/annurev.biophys.37.032807.125928](https://doi.org/10.1146/annurev.biophys.37.032807.125928); pmid: [19416069](https://pubmed.ncbi.nlm.nih.gov/19416069/)
- B. Munsky, G. Neuert, A. van Oudenaarden, Using gene expression noise to understand gene regulation. *Science* **336**, 183–187 (2012). doi: [10.1126/science.1216379](https://doi.org/10.1126/science.1216379); pmid: [22499939](https://pubmed.ncbi.nlm.nih.gov/22499939/)
- M. Lagna, J. P. Bothma, M. Levine, Mechanisms of transcriptional precision in animal development. *Trends Genet.* **28**, 409–416 (2012). doi: [10.1016/tig.2012.03.006](https://doi.org/10.1016/tig.2012.03.006); pmid: [22513408](https://pubmed.ncbi.nlm.nih.gov/22513408/)
- T. Ha, Single-molecule methods: leap ahead. *Nat. Methods* **11**, 1015–1018 (2014). doi: [10.1038/nmeth.3107](https://doi.org/10.1038/nmeth.3107); pmid: [25264779](https://pubmed.ncbi.nlm.nih.gov/25264779/)
- Y. Taniguchi et al., Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010). doi: [10.1126/science.1188308](https://doi.org/10.1126/science.1188308); pmid: [20671182](https://pubmed.ncbi.nlm.nih.gov/20671182/)
- N. Battich, T. Stoeger, L. Pelkmans, Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nat. Methods* **10**, 1127–1133 (2013). doi: [10.1038/nmeth.2657](https://doi.org/10.1038/nmeth.2657); pmid: [24097269](https://pubmed.ncbi.nlm.nih.gov/24097269/)
- J. M. Levsy, S. M. Shenoy, R. C. Pezo, R. H. Singer, Single-cell gene expression profiling. *Science* **297**, 836–840 (2002). doi: [10.1126/science.1072241](https://doi.org/10.1126/science.1072241); pmid: [12161654](https://pubmed.ncbi.nlm.nih.gov/12161654/)
- E. Lubek, L. Cai, Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat. Methods* **9**, 743–748 (2012). doi: [10.1038/nmeth.2069](https://doi.org/10.1038/nmeth.2069); pmid: [22660740](https://pubmed.ncbi.nlm.nih.gov/22660740/)
- M. J. Levesque, A. Raj, Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. *Nat. Methods* **10**, 246–248 (2013). doi: [10.1038/nmeth.2372](https://doi.org/10.1038/nmeth.2372); pmid: [23416756](https://pubmed.ncbi.nlm.nih.gov/23416756/)
- E. Lubek, A. F. Coskun, T. Zhiyentayev, M. Ahmad, L. Cai, Single-cell *in situ* RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014). doi: [10.1038/nmeth.2892](https://doi.org/10.1038/nmeth.2892); pmid: [24681720](https://pubmed.ncbi.nlm.nih.gov/24681720/)
- J. Harrow et al., GENCODE: The reference human genome annotation for the ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012). doi: [10.1101/gr.135350.111](https://doi.org/10.1101/gr.135350.111); pmid: [22955987](https://pubmed.ncbi.nlm.nih.gov/22955987/)
- T. K. Moon, *Error Correction Coding: Mathematical Methods and Algorithms* (Wiley, New York, ed. 1, 2005).
- B. J. Beliveau et al., Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21301–21306 (2012). doi: [10.1073/pnas.1213181109](https://doi.org/10.1073/pnas.1213181109); pmid: [23236188](https://pubmed.ncbi.nlm.nih.gov/23236188/)
- Y. E. Murgha, J.-M. Rouillard, E. Gulari, Methods for the preparation of large quantities of complex single-stranded oligonucleotide libraries. *PLOS ONE* **9**, e94752 (2014). doi: [10.1371/journal.pone.0094752](https://doi.org/10.1371/journal.pone.0094752); pmid: [24733454](https://pubmed.ncbi.nlm.nih.gov/24733454/)
- A. Sanchez, I. Golding, Genetic determinants and cellular constraints in noisy gene expression. *Science* **342**, 1188–1193 (2013). doi: [10.1126/science.1242975](https://doi.org/10.1126/science.1242975); pmid: [24311680](https://pubmed.ncbi.nlm.nih.gov/24311680/)
- L. H. So et al., General properties of transcriptional time series in *Escherichia coli*. *Nat. Genet.* **43**, 554–560 (2011). doi: [10.1038/ng.821](https://doi.org/10.1038/ng.821); pmid: [21532574](https://pubmed.ncbi.nlm.nih.gov/21532574/)
- M. Safran et al., GeneCards Version 3: The human gene integrator. *Database (Oxford)* **2010**, baq020 (2010). doi: [10.1093/database/baq020](https://doi.org/10.1093/database/baq020); pmid: [20689021](https://pubmed.ncbi.nlm.nih.gov/20689021/)
- K. Dolinski, D. Botstein, Changing perspectives in yeast research nearly a decade after the genome sequence. *Genome Res.* **15**, 1611–1619 (2005). doi: [10.1101/gr.372750.105](https://doi.org/10.1101/gr.372750.105); pmid: [16339358](https://pubmed.ncbi.nlm.nih.gov/16339358/)
- O. Padovan-Merhar, A. Raj, Using variability in gene expression as a tool for studying gene regulation. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **5**, 751–759 (2013). doi: [10.1002/wsbm.1243](https://doi.org/10.1002/wsbm.1243); pmid: [23996796](https://pubmed.ncbi.nlm.nih.gov/23996796/)
- M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863–14868 (1998). doi: [10.1073/pnas.95.25.14863](https://doi.org/10.1073/pnas.95.25.14863); pmid: [9843981](https://pubmed.ncbi.nlm.nih.gov/9843981/)
- N. Gehlenborg et al., Visualization of omics data for systems biology. *Nat. Methods* **7** (suppl.), S56–S68 (2010). doi: [10.1038/nmeth.1436](https://doi.org/10.1038/nmeth.1436); pmid: [2095258](https://pubmed.ncbi.nlm.nih.gov/2095258/)
- M. Ashburner et al., Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000). doi: [10.1038/75556](https://doi.org/10.1038/75556); pmid: [10802651](https://pubmed.ncbi.nlm.nih.gov/10802651/)
- H. Yoshida et al., KIAA1199, a deafness gene of unknown function, is a new hyaluronan binding protein involved in hyaluronan depolymerization. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5612–5617 (2013). doi: [10.1073/pnas.1215432110](https://doi.org/10.1073/pnas.1215432110); pmid: [23509262](https://pubmed.ncbi.nlm.nih.gov/23509262/)
- D. A. Lauffenburger, A. F. Horwitz, Cell migration: A physically integrated molecular process. *Cell* **84**, 359–369 (1996). doi: [10.1016/S0092-8674\(00\)81280-5](https://doi.org/10.1016/S0092-8674(00)81280-5); pmid: [8608589](https://pubmed.ncbi.nlm.nih.gov/8608589/)
- T. A. Rapoport, Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**, 663–669 (2007). doi: [10.1038/nature06384](https://doi.org/10.1038/nature06384); pmid: [18046402](https://pubmed.ncbi.nlm.nih.gov/18046402/)
- C. H. Jan, C. C. Williams, J. S. Weissman, Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science* **346**, 1257521 (2014). doi: [10.1126/science.1257521](https://doi.org/10.1126/science.1257521); pmid: [25378630](https://pubmed.ncbi.nlm.nih.gov/25378630/)

36. J. B. Lawrence, R. H. Singer, Intracellular localization of messenger RNAs for cytoskeletal proteins. *Cell* **45**, 407–415 (1986). doi: [10.1016/0092-8674\(86\)90326-0](https://doi.org/10.1016/0092-8674(86)90326-0); pmid: [3698103](https://pubmed.ncbi.nlm.nih.gov/3698103/)
37. L. A. Mingle *et al.*, Localization of all seven messenger RNAs for the actin-polymerization nucleator Arp2/3 complex in the protrusions of fibroblasts. *J. Cell Sci.* **118**, 2425–2433 (2005). doi: [10.1242/jcs.02371](https://doi.org/10.1242/jcs.02371); pmid: [15923655](https://pubmed.ncbi.nlm.nih.gov/15923655/)
38. H. Babcock, Y. M. Sigal, X. Zhuang, A high-density 3D localization algorithm for stochastic optical reconstruction microscopy. *Opt. Nanoscopy* **1**, 6 (2012). doi: [10.1186/2192-2853-1-6](https://doi.org/10.1186/2192-2853-1-6); pmid: [25431749](https://pubmed.ncbi.nlm.nih.gov/25431749/)
39. L. Zhu, W. Zhang, D. Elnatan, B. Huang, Faster STORM using compressed sensing. *Nat. Methods* **9**, 721–723 (2012). doi: [10.1038/nmeth.1978](https://doi.org/10.1038/nmeth.1978); pmid: [22522657](https://pubmed.ncbi.nlm.nih.gov/22522657/)
40. H. P. Babcock, J. R. Moffitt, Y. Cao, X. Zhuang, Fast compressed sensing analysis for super-resolution imaging using L1-homotopy. *Opt. Express* **21**, 28583–28596 (2013). doi: [10.1364/OE.21.028583](https://doi.org/10.1364/OE.21.028583); pmid: [24514370](https://pubmed.ncbi.nlm.nih.gov/24514370/)
41. S. W. Hell, Microscopy and its focal switch. *Nat. Methods* **6**, 24–32 (2009). doi: [10.1038/nmeth.1291](https://doi.org/10.1038/nmeth.1291); pmid: [19116611](https://pubmed.ncbi.nlm.nih.gov/19116611/)
42. B. Huang, H. Babcock, X. Zhuang, Breaking the diffraction barrier: Super-resolution imaging of cells. *Cell* **143**, 1047–1058 (2010). doi: [10.1016/j.cell.2010.12.002](https://doi.org/10.1016/j.cell.2010.12.002); pmid: [21168201](https://pubmed.ncbi.nlm.nih.gov/21168201/)
43. Q. Xu, M. R. Schlabach, G. J. Hannon, S. J. Elledge, Design of 240,000 orthogonal 25mer DNA barcode probes. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2289–2294 (2009). doi: [10.1073/pnas.0812506106](https://doi.org/10.1073/pnas.0812506106); pmid: [19171886](https://pubmed.ncbi.nlm.nih.gov/19171886/)
44. C. Camacho *et al.*, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009). doi: [10.1186/1471-2105-10-421](https://doi.org/10.1186/1471-2105-10-421); pmid: [20003500](https://pubmed.ncbi.nlm.nih.gov/20003500/)
45. C. Trapnell *et al.*, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012). doi: [10.1038/nprot.2012.016](https://doi.org/10.1038/nprot.2012.016); pmid: [22383036](https://pubmed.ncbi.nlm.nih.gov/22383036/)
46. I. Dunham *et al.*, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). pmid: [22955616](https://pubmed.ncbi.nlm.nih.gov/22955616/)
47. J.-M. Rouillard, M. Zuker, E. Gulari, OligoArray 2.0: Design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res.* **31**, 3057–3062 (2003). doi: [10.1093/nar/gkg426](https://doi.org/10.1093/nar/gkg426); pmid: [12799432](https://pubmed.ncbi.nlm.nih.gov/12799432/)
48. M. Batish, A. Raj, S. Tyagi, in *Methods in Molecular Biology*, J. E. Gerst, Ed. (Humana Press, Totowa, NJ, 2011), vol. 714, pp. 3–13.
49. A. R. Buxbaum, B. Wu, R. H. Singer, Single β-actin mRNA detection in neurons reveals a mechanism for regulating its translatability. *Science* **343**, 419–422 (2014). doi: [10.1126/science.1242939](https://doi.org/10.1126/science.1242939); pmid: [24458642](https://pubmed.ncbi.nlm.nih.gov/24458642/)
50. I. Rasnik, S. A. McKinney, T. Ha, Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat. Methods* **3**, 891–893 (2006). doi: [10.1038/nmeth934](https://doi.org/10.1038/nmeth934); pmid: [17013382](https://pubmed.ncbi.nlm.nih.gov/17013382/)
51. X. Shi, J. Lim, T. Ha, Acidification of the oxygen scavenging system in single-molecule fluorescence studies: In situ sensing with a ratiometric dual-emission probe. *Anal. Chem.* **82**, 6132–6138 (2010). doi: [10.1021/ac1008749](https://doi.org/10.1021/ac1008749); pmid: [20583766](https://pubmed.ncbi.nlm.nih.gov/20583766/)

ACKNOWLEDGMENTS

We thank H. Babcock for technical advice and help with instrumentation and B. Birru for aid in error analysis. This work was in part supported by the National Institutes of Health. K.H.C. acknowledges a National Science Scholarship from the Agency for Science, Technology and Research of Singapore. A.N.B. acknowledges support by the Damon Runyon Foundation postdoctoral fellowship. J.R.M. acknowledges support from the Helen Hay Whitney Foundation postdoctoral fellowship. S.W. acknowledges support from Jane Coffins Child Foundation postdoctoral fellowship. X.Z. is a Howard Hughes Medical Institute investigator. RNA-seq data were deposited under the accession number GSE67685. MERFISH data are available at <http://zhuang.harvard.edu/merfish>. X.Z., K.H.C., A.N.B., J.R.M., and S.W. are inventors on a patent applied for by Harvard University that covers the MERFISH method described here; X.Z., J.R.M., and A.N.B. are inventors on a patent applied for by Harvard University that covers the probe synthesis method described here.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/aaa6090/suppl/DC1
Figs. S1 to S9
Tables S1 to S5

5 January 2015; accepted 19 March 2015
Published online 9 April 2015;
10.1126/science.aaa6090

RESEARCH ARTICLE

SUPERNOVAE

Old supernova dust factory revealed at the Galactic center

R. M. Lau,^{1*} T. L. Herter,¹ M. R. Morris,² Z. Li,³ J. D. Adams^{1,4}

Dust formation in supernova ejecta is currently the leading candidate to explain the large quantities of dust observed in the distant, early universe. However, it is unclear whether the ejecta-formed dust can survive the hot interior of the supernova remnant (SNR). We present infrared observations of ~0.02 solar masses of warm (~100 kelvin) dust seen near the center of the ~10,000-year-old Sagittarius A East SNR at the Galactic center. Our findings indicate the detection of dust within an older SNR that is expanding into a relatively dense surrounding medium (electron density $\sim 10^3$ centimeters $^{-3}$) and has survived the passage of the reverse shock. The results suggest that supernovae may be the dominant dust-production mechanism in the dense environment of galaxies of the early universe.

The search for the dominant formation mechanism of large quantities of dust detected in galaxies of the early universe remains an open investigation of profound importance. It both influences observed emission and greatly affects the formation of future generations of stars. Due to the short lifetimes of their progenitor stars and to their highly metal-enriched ejecta, supernovae (SNe) are believed to be efficient sources of dust production (1). However, the powerful supernova (SN) explosions and the resulting shocks are predicted to also be very effective at destroying and shattering dust: Depending on the energy of the explosion and the density of the surrounding medium, less than ~20% of the mass of the SN-condensed dust is expected to survive the passage of the reverse shock that is driven back into the ejecta due to the difference between the thermal pressure of the shocked circumstellar material and that of the expanding ejecta (2–4). In fact, recent studies have argued that SNe may be net destroyers of dust in present-day galaxies [e.g., in the Magellanic clouds (5, 6)] but net producers of dust in the earliest-forming galaxies in the universe (7). However, no direct observational evidence currently exists of the quantities of SN-condensed dust surviving the passage of the reverse shock through the ejecta.

Sagittarius A (Sgr A) is the well-studied remnant of a core-collapse SN, located 8 kpc away near the center of our Galaxy (8), that has

blown out a ~3' (7-pc)-diameter shell of non-thermal radio emission (Fig. 1) (9, 10). Based on the kinematics of maser spots immediately surrounding the supernova remnant (SNR), it is estimated to be ~5 pc from the Galactic center (11). An age of ~10⁴ years has been estimated for the SNR, based on predictions of its elongation from tidal forces (12) and, more recently, from the observed proper motion and displacement of the neutron star believed to be the remnant of the SN progenitor from the interior of the SNR (13). The ~10⁴-year time scale and the metal-rich hot ejecta from the center of the SNR imply that the reverse shock has reached the center of Sgr A East (14). Observations of dust associated with the Sgr A East SNR could therefore

provide estimates of the fraction of dust that survives the destructive passage of the reverse shock.

We use mid- and far-infrared (IR) images of Sgr A East from the Stratospheric Observatory for Infrared Astronomy (SOFIA) to argue the presence of warm ($T_d \sim 100$ K) dust near its center. The warm and dusty environment of the Galactic center (15) presents a challenge in definitively demonstrating that the IR emission is associated with the Sgr A East SNR ejecta. Our claim is substantiated by four major results from the observations: (i) The location of the IR-emitting region is consistent with the center of the SNR and is spatially anticorrelated with the hard x-ray emission, suggesting that dust is in a cooler region of the ejecta. (ii) Analysis of the dust temperatures and heating sources shows that the location constraints are consistent with Sgr A East because the IR-emitting region must be radiatively heated by the optical and ultraviolet photons from the central stellar cluster, which dominates the radiation field in the central parsecs of the Galaxy. (iii) The lack of cold dust emission at submillimeter wavelengths coincident with the IR-emitting region implies that it does not originate from a cloud along the line of sight. (iv) We show that the typical size of the emitting grains must be smaller than the dust in the surrounding interstellar medium (ISM), which is consistent with having been processed by the reverse shock of the SN.

We propose that the dust has survived the passage of the reverse shock due to the relatively large density of the surrounding medium, which slows the ejecta and places it into a temperature and density regime where it will undergo substantial radiative cooling on time scales much shorter than dust-destruction time scales. With an estimate of the surviving dust mass in Sgr A East, we discuss the viability of SNe as the

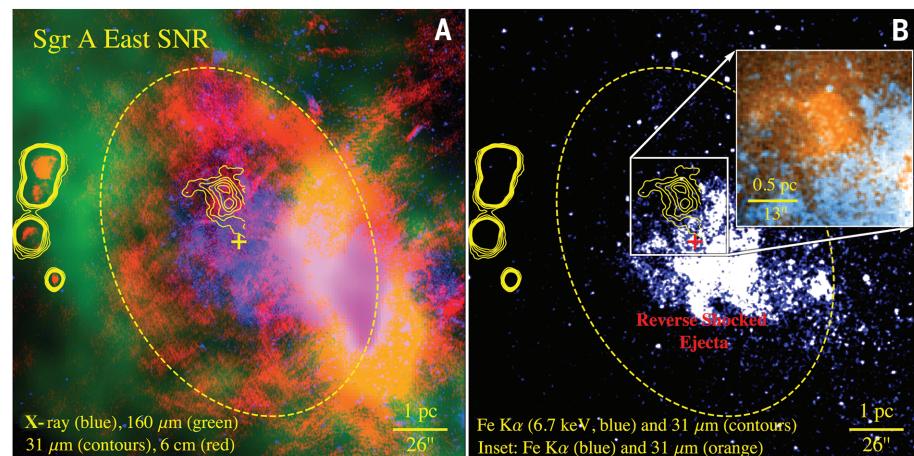


Fig. 1. Multiwavelength imaging of the Sgr A East supernova remnant. (A) Composite false-color image of the Sgr A East SNR overlaid with contours of the 31.5-μm emission east of the CND. North is up and east is left. The colors correspond to emission from 2 to 8 keV (blue; Chandra/ACIS-I), 160 μm (green; Herschel/PACS), and 6 cm (red; Very Large Array). The dotted ellipse delineates the boundary of the Sgr A East radio shell, with the cross indicating the location of its apparent center. (B) Fe K α (6.7 keV; blue) emission from the SNR overlaid with the 31.5-μm emission contours and a twice-magnified inset of the SNR's IR-emitting region (orange).

¹Department of Astronomy, Cornell University, Ithaca, NY 14853, USA. ²Department of Physics and Astronomy, University of California, Los Angeles, 430 Portola Plaza, Los Angeles, CA 90095, USA. ³School of Astronomy and Space Science, Nanjing University, 22 Hankou Road, Nanjing, Jiangsu 210093, China. ⁴Stratospheric Observatory for Infrared Astronomy (SOFIA) Science Center, Universities Space Research Association, NASA Ames Research Center, MS 232, Moffett Field, CA 94035, USA.
*Corresponding author. E-mail: ryan@astro.cornell.edu

dominant source of dust production in the early universe.

Observations and results

The Faint Object Infrared Camera for the SOFIA Telescope (FORCAST) (16) was used to obtain images of the Sgr A East SNR at 7.7, 19.7, 25.2, 31.5, and 37.1 μm . The spatially resolved hot and warm dust traced by these wavelength bands is shown in Fig. 2. We used the Chandra/Advanced CCD Imaging Spectrometer imaging array (ACIS-I) to obtain high-quality hard x-ray (2 to 8 keV) images of Sgr A East. Additionally, we incorporated archival mid-IR (5.8 and 8.0 μm) and submillimeter (70 and 160 μm) observations taken by the Spitzer Space Telescope's Infrared Array Camera (IRAC) (17, 18) and the Herschel Space Observatory's Photodetector Array Camera and Spectrometer (PACS) (19, 20), respectively (21).

Source morphology and location

The mid-IR (5.8 and 8 μm) and far-IR (19 to 37 μm) observations of the proposed SNR dust emission are cospatial and confined to the regions near the center of the Sgr A East radio shell (Fig. 1A). The central position of the dust emission is consistent with the expected location of dust having condensed within the ejecta. Furthermore, the dust is spatially anticorrelated with the hard x-ray continuum of the SNR ejecta (Fig. 1B). This anticorrelation suggests that the dust is located in a much cooler and less hostile region of the ejecta.

It is apparent in Fig. 1A that there is no major submillimeter emission coincident with the location of the IR-emitting region. This is in contrast to the Sgr A East HII regions that are located ~3 pc to the east in projection from the center

of the SNR and are the illuminated edges of the molecular cloud associated with the prominent ridge of submillimeter emission that extends along the eastern side of Sgr A East (22). The lack of submillimeter emission at the IR-emitting region therefore implies that it is not associated with a cold molecular cloud along the line of sight toward the center of Sgr A East.

Observed dust temperature

A key to understanding whether the dust is located interior to the Sgr A East SNR is its thermal structure. The closest known stellar heating source is the central cluster of massive young stars surrounding the Galactic black hole, located ~3 pc away in projection. We also consider heating via collisions with thermal electrons in the shocked ejecta of the SNR but conclude that it is negligible (23). The presence of the strong radiation field from the central cluster therefore presents a distinctive heating scenario for dust in Sgr A East because the radiative heating contribution, which typically arises from the interstellar radiation field in most SNRs, is usually negligible compared with collisional heating.

To investigate the heating source(s) and thermal structure of the possible SNR dust, we generate a color temperature map using the ratio of the deconvolved 31.5- and 37.1- μm (31/37) flux maps of Sgr A West and East observed by SOFIA/FORCAST (Fig. 3) (24, 25). We assume the dust emission is optically thin and takes the form $F_v \propto B_v(T_d)v^\beta$ (F_v , flux density; B_v , function for blackbody radiation; v , frequency at which the dust is emitting; β , frequency power-law index), where a value of 2 is adopted for β , which is typical for interstellar dust. The longer-wavelength

images are used to produce the color temperature map because the signal-to-noise ratio from the IR-emitting region at those wavelengths is higher than at 19.7 μm . The average color temperature derived from the ratio of the 19.7- and 37.1- μm (19/37) fluxes of the IR-emitting region is consistent with the average 31/37 color temperature.

It is apparent from the location of the temperature peak and the negative radial temperature gradient centered on Sagittarius A* (Sgr A*) that the luminous central cluster dominates the heating of the dust in the HII region and circumnuclear disk (CND) immediately surrounding Sgr A* (24). The proposed SNR dust exhibits a temperature of $\sim 100 \pm 8$ K (26), which is much greater than the ~ 75 K that would be predicted for 0.1- μm -sized grains from the radial temperature gradient centered approximately on Sgr A*, allowing for a $\sqrt{2}$ projection factor. We note that the ~ 75 -K temperature of the structure we refer to as the Northern Dust Cloud (see Fig. 3) is consistent with this gradient and is equidistant from Sgr A* in projection with the SNR dust. The four Sgr A East HII regions seen to the east of the SNR (Figs. 1 and 3) exhibit temperature maxima, but they are each heated locally by sources of 10^4 to $10^5 L_\odot$ (L_\odot , solar luminosity) (22, 25).

Heating source and dust size

Three different scenarios can be considered to explain why the proposed SNR dust is at a considerably higher temperature than expected from heating by the central stellar cluster: (i) The dust is locally heated by a luminous stellar source or sources, (ii) the dust is heated by energetic electron collisions in the shocked ejecta, or (iii) the dust is much smaller in size than the

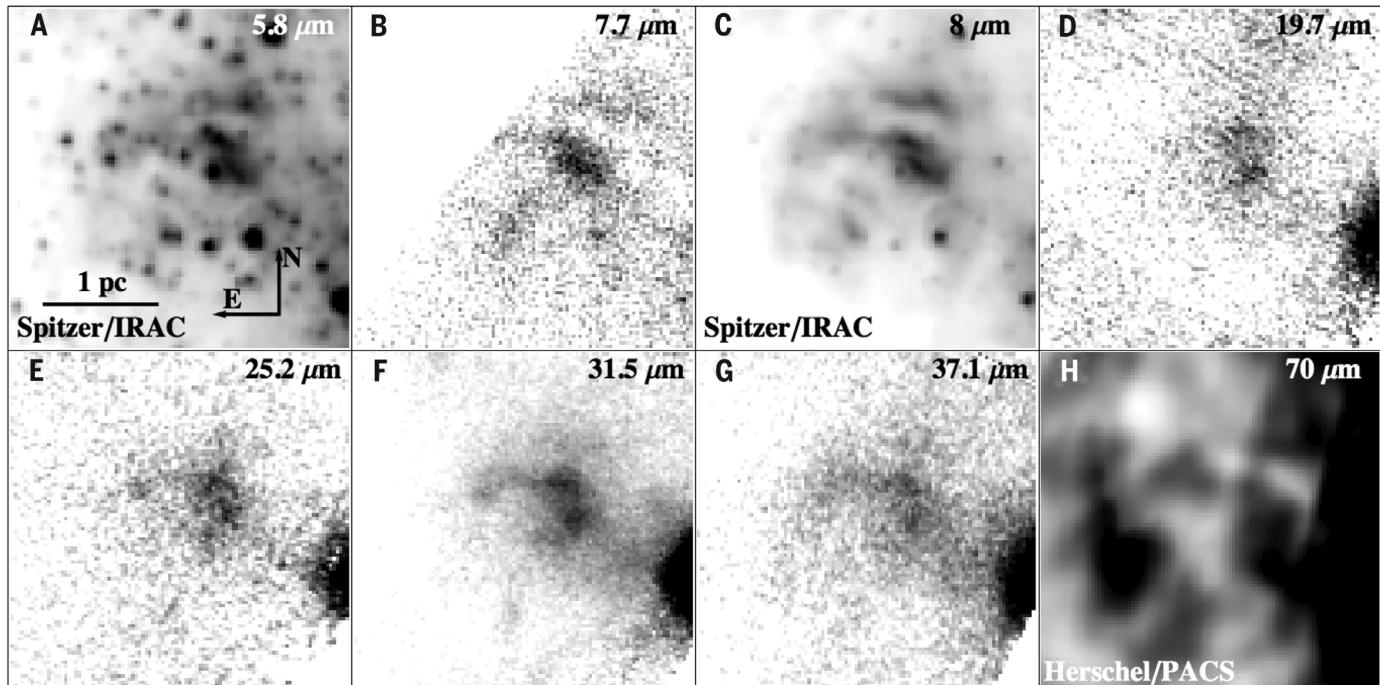


Fig. 2. Mid- and far-IR images of the IR-emitting region. (A to H) Spitzer/IRAC, SOFIA/FORCAST, and Herschel/PACS flux maps of the Sgr A East SNR dust.

dust associated with the CND and surrounding ISM. The first scenario would probably imply that the dust is associated with an illuminated molecular cloud along the line of sight toward Sgr A East, whereas the latter two scenarios would constrain the location of the dust to the interior of Sgr A East.

First, we consider dominant heating by a local stellar source, as is the case for the Sgr A East HII regions. Given the size of the SNR's IR-emitting region (~ 0.8 pc), we assume that any such local heating source(s) would be located at a distance of ~ 0.4 pc from the dust. For $0.1\text{-}\mu\text{m}$ -sized silicate grains that are in thermal equilibrium with the radiation field, the total stellar luminosity of the heating source would then have to be $\sim 2 \times 10^6 L_\odot$ to reproduce dust temperatures of ~ 100 K. A source with such a high luminosity, which is similar to that of luminous blue variables, would be easily detected in the near-IR (1.90 μm) (27, 28) and near-IR Spitzer/IRAC (18) observations of the region; however, no such heating source can be identified within several parsecs of the dust. Additionally, the total integrated infrared luminosity of the IR-emitting region is $\sim 7 \times 10^4 L_\odot$, which is too high if it were heated by a group of dusty asymptotic giant branch stars. Conversely, this total infrared luminosity is too low and the temperatures are too uniform for the IR emission to be associated with extremely massive, evolved dust-enshrouded sources, such as Wolf-Rayet stars ($L_* > 10^5 L_\odot$, where L_* is the stellar luminosity). We therefore rule out the possibility that a local heating source is responsible for the observed dust temperatures.

Dust heating due to collisions with electrons is commonly observed in SNRs, given the energetic conditions of the ejecta (29). Because of the apparent location of the dust outside the regions of hard x-ray emission (see Fig. 1B), the dust will be in a cooler and denser region

of the ejecta. Taking the electron density and temperature to be $\sim 100 \text{ cm}^{-3}$ and $\sim 10^5$ K, respectively, consistent with conditions that result in a short ejecta-cooling time scale ($\tau_{\text{cool}} << 10^4$ year), we find that the radiation by the central cluster will dominate the heating over electron collisions for all grain sizes.

In the final scenario, we propose that the grains composing the proposed SNR dust are smaller than the typical $\sim 0.1\text{-}\mu\text{m}$ size (24) and are therefore heated to higher temperatures due to lower heat capacities. The 31/37 dust temperature map in Fig. 3 is overlaid with several contours of the predicted dust equilibrium temperatures for $0.1\text{-}\mu\text{m}$ silicate and $0.01\text{-}\mu\text{m}$ amorphous carbon grains, where we have only considered radiative heating from the central cluster and assumed no blockage by the CND. The $0.1\text{-}\mu\text{m}$ dust temperature contours show clear agreement with the observed temperatures in the CND as well as the Northern Dust Cloud; however, the contours underpredict the temperatures observed for the proposed SNR dust. Shifting to the smaller $0.01\text{-}\mu\text{m}$ -sized grains shows much closer agreement between the predicted and observed dust temperatures (~ 100 K).

We conclude that the inconsistency between the observed dust temperature of the proposed SNR dust and the Northern Dust Cloud and CND is due to a composition of smaller grains, which is consistent with expectations for dust in a SNR in which fragmentation and thermal sputtering have taken place after grain formation.

Dust spectral energy distribution models

The observed spectral energy distribution (SED) provides important constraints on the size of the dust particles, their physical location and the heating source(s). With this in mind, we investigate six different regions in and around the SNR dust (Fig. 4) and use the DustEM code

(30) to produce dust models that fit the observed SEDs. Three of the regions cover sites of prominent far-IR emission: north clump, south clump, and east. The remaining three regions cover sites of strong emission observed at mid-IR and submillimeter wavelengths that may or may not be associated with the SNR: southeast, west (W), and north. For the models, we perform a linear least-squares fit to the dereddened fluxes observed at 5.8, 8.0, 19.7, 25.2, 31.5, and 37.1 μm . We assume that the dust is composed of two independent grain distributions: $\sim 0.04\text{-}\mu\text{m}$ -sized [large grain (LG)] and $\sim 0.001\text{-}\mu\text{m}$ -sized [very small grain (VSG)] amorphous carbon grains heated radiatively by the central cluster, which we modeled as a point source with a luminosity of $4 \times 10^7 L_\odot$ (31) and the spectrum of a stellar atmosphere with an effective temperature of 37,000 K (32), representative of the luminous stars that dominate the radiation from the cluster. Because the distance between the central cluster and the proposed SNR dust (d) is uncertain, we allow d to vary as a free parameter in our models. The best-fit models (Fig. 4 and Table I) show that d is consistent with the ~ 5 -pc separation distance between the SNR and Sgr A* estimated by kinematics of maser spots associated with the SNR (11).

The LG and VSG mass abundances are the other two free parameters for the models, in addition to the distance between the central cluster and the dust. Assuming a distinctly larger-sized ($0.1\text{-}\mu\text{m}$) dust distribution results in very poor SED fits for all of the regions except W. We therefore require an independent distribution of VSGs for which transient heating allows us to fit the 5.8- and 8.0- μm flux points. A modified black-body fit to the 5.8- and 8.0- μm points yields temperatures of ~ 350 K, which cannot be achieved if the dust is equilibrium-heated. The results from our best-fit models show that the VSG-to-LG

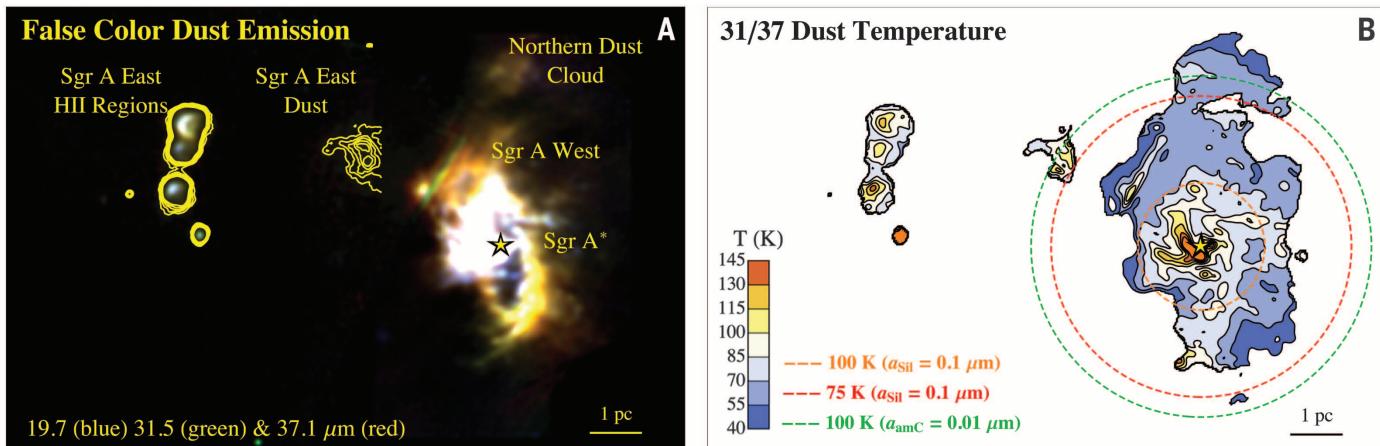


Fig. 3. SOFIA/FORCAST flux and temperature maps of Sgr A East and West.

(A) False-color image of dust emission from the Sgr A West and East regions. The colors correspond to 19.7 μm (blue), 31.5 μm (green), and 37.1 μm (red), and the 31.5- μm flux contours east of the CND are overlaid. The central cluster is located in the subparsec vicinity of Sgr A*. N, north; E, east; W, west; SE, southeast; NC, north clump; SC, south clump. (B) 31/37 dust temperature map of the Sgr A West and East regions. Overlaid are the theoretical dust temperature contours at 75 K (red)

and 100 K (orange) for $0.1\text{-}\mu\text{m}$ -sized silicate grains and at 100 K (green) for $0.01\text{-}\mu\text{m}$ -sized amorphous carbon grains, assuming equilibrium radiative heating by the central cluster. The apparent linear structure northeast of Sgr A* seen in the temperature contours is an artifact from combining the images of Sgr A East and Sgr A West, and the temperature peak at the south of Sgr A West is probably associated with an embedded source that appears more prominently in the mid-IR Spitzer/IRAC images (18). In both panels, north is up and east is left.

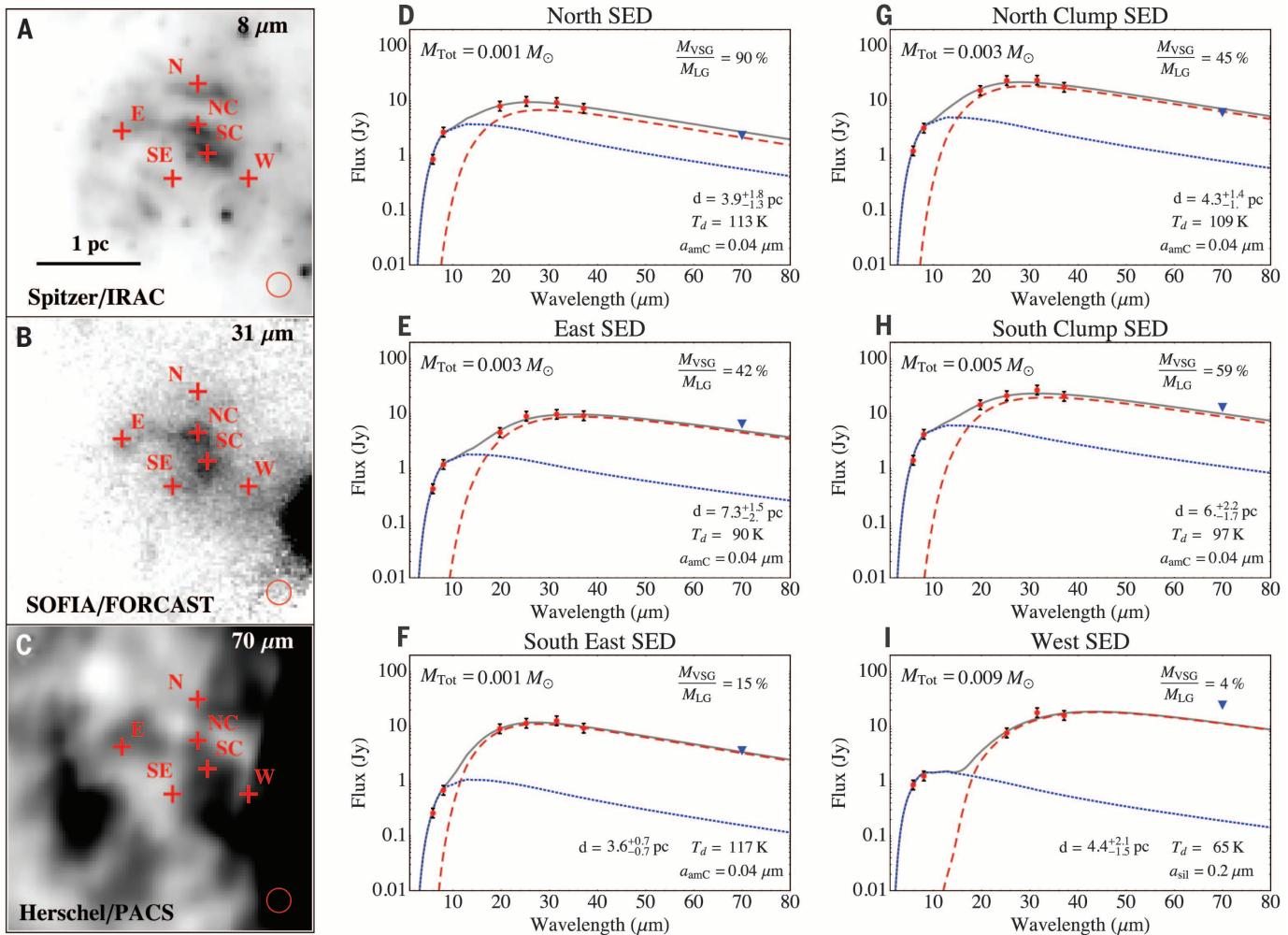


Fig. 4. Spectral energy distribution models fit to the measured flux throughout the IR-emitting region. (A to C) Sgr A East dust emission region. North is up and east is left. Overlaid are the six locations across the region where SEDs were extracted. The aperture size used to extract the fluxes is shown in the lower right corner. Fluctuations in the amplitudes of the background emission in the 70-μm Herschel/PACS image can vary on the order of the

source flux. Fluxes extracted from the 70-μm image are therefore treated as upper limits. (D to I) Best-fit DustEM models of the six regions. The blue dotted line and the red dashed line correspond to emission from VSG and LG distributions, respectively. Blue inverted triangles denote upper limits of the 70-μm flux measured by Herschel/PACS. Error bars indicate $\pm 20\%$ of the flux measured from Spitzer/IRAC and SOFIA/FORCAST images. Jy, jansky.

mass abundance ratio for all regions except W ranges from 15 to 90% (Fig. 4), which suggests a VSG enhancement when compared with the typical ratio of the Milky Way ISM (~13%) (33), as well as that of the compact Sgr A East HII regions in the 50 km s⁻¹ cloud ~3 pc east of the proposed SNR dust (~2 to 4%) (25).

To produce adequate fits to the SED of region W, we are required to either increase d to ~10 pc or increase the larger grain size to 0.2 μm. Because $d \sim 10$ pc would suggest that the region W dust is located twice as far away as the dust at the other regions, we adopt the more likely interpretation that the dust grains are larger. Regardless of interpretation, the resulting inconsistent model fits to the region W SED strongly suggest that region W is not associated with the other regions and might thus be an extension of the dust distribution of the CND, as its proximity to the CND might suggest.

Table 1. Assumed and derived dust properties of the dust SED model fit to the Sgr A East SNR dust. $a_{\text{amC}}^{\text{VSG}}$ and $a_{\text{amC}}^{\text{LG}}$ are the amorphous carbon grain sizes assumed for the VSG and LG components, respectively, of the model, given in units of micrometers. The LG component for the model fit to the West region is composed of silicates, like the dust in the CND (24). T_d is the temperature of the LG distribution, d is the distance between the dust in the region and the central stellar cluster, M_{Tot} is the total dust mass fit to the region, and $M_{\text{VSG}}/M_{\text{LG}}$ is the mass ratio of the VSGs to the LGs. Errors provided are 1σ , as determined from the weighted linear least-squares fits.

Region	$a_{\text{amC}}^{\text{VSG}}$	$a_{\text{amC}}^{\text{LG}}$	T_d (K)	d (pc)	$M_{\text{Tot}} (\times 10^{-3} M_{\odot})$	$M_{\text{VSG}}/M_{\text{LG}}$
North	0.001	0.04	113^{+19}_{-15}	$3.9^{+1.8}_{-1.3}$	$1.1^{+1.1}_{-0.6}$	$0.90^{+0.13}_{-0.14}$
N. clump	0.001	0.04	109^{+11}_{-11}	$4.3^{+1.4}_{-1.0}$	$2.6^{+1.9}_{-1.0}$	$0.45^{+0.06}_{-0.06}$
S. clump	0.001	0.04	97^{+12}_{-11}	$6.0^{+2.2}_{-1.7}$	$5.3^{+4.7}_{-2.5}$	$0.59^{+0.08}_{-0.09}$
East	0.001	0.04	90^{+8}_{-8}	$7.3^{+2.0}_{-1.5}$	$2.9^{+1.9}_{-1.1}$	$0.42^{+0.06}_{-0.06}$
Southeast	0.001	0.04	117^{+9}_{-8}	$3.6^{+0.7}_{-0.7}$	$0.9^{+0.4}_{-0.3}$	$0.15^{+0.02}_{-0.01}$
West	0.008	0.2	65^{+10}_{-8}	$4.4^{+14.3}_{-1.5}$	$9.2^{+14.3}_{-5.7}$	$0.04^{+0.01}_{-0.01}$

Our model fits do not constrain the composition of the LG distribution, which can also be modeled with silicate grains. Substituting am-

orphous carbon for silicates in the models decreases the predicted total dust mass by only ~10%. However, silicates for the VSG distribution are ruled

out, as they yield poor fits due to the steepness of the shorter-wavelength side of the 9.7- μm feature. Altering the size of the LG distribution by $\pm 0.02 \mu\text{m}$ changes the total mass by only $\sim 20\%$. Polycyclic aromatic hydrocarbons are also ruled out for the VSG composition, based on the ratio of the 7.7- and 8- μm fluxes (23). The resulting SED dust models and the fitted parameters are shown in Fig. 4 and Table 1, respectively.

Submillimeter excess and dust mass

In our model fitting, we omit the 70- μm flux to determine whether this flux is associated with the warm dust emission probed by the far-IR or is enhanced relative to the model due to the presence of a cooler dust distribution. Owing to the high background fluctuations at 70- μm images throughout the Galactic center, we assume that the extracted 70- μm fluxes shown in the SEDs (Fig. 4) are upper limits. The absence of a 70- μm excess indicates that it is unlikely to be associated with a molecular cloud along the line of sight toward the SNR.

The dust model fit to the SED of the full region covering the proposed SNR dust yields a total dust mass of $\sim 0.02^{+0.008}_{-0.006} M_{\odot}$ (M_{\odot} , solar mass), where the LG component composes $\sim 45\%$ of the total mass. This mass estimate is consistent with the summed dust mass of the five sampled regions, excluding W ($\sim 0.013 M_{\odot}$), as they cover approximately three-quarters of the angular size of the SNR dust emission.

Sgr A East expansion and dust survival in a nonuniform-density medium

The location, thermal structure, and SED of the dust indicate that it is very likely associated with the Sgr A East SNR. We describe a theoretical framework of the evolution of the SN ejecta to explain the morphology, size composition, and apparent location of the surviving dust within the SNR. Our interpretation is based on comparing the ejecta-cooling time scales to the dust-destruction time scales due to thermal sputtering during the expansion of the ejecta into an asymmetric surrounding medium.

Molecular line observations toward Sgr A East reveal the presence of dense molecular clouds that appear to be interacting with the SNR, especially at the northern and eastern edges of the remnant (9, 34). Newly condensed dust from the initial explosion is assumed to be present uniformly throughout the SN ejecta, and the densities of the surrounding medium to the northeast (NE) and southwest (SW) are calculated as $\sim 10^3 \text{ cm}^{-3}$ (14) and $\sim 100 \text{ cm}^{-3}$, respectively. This ambient density gradient implies that there will be an asymmetry in the ejecta densities, as well as in the strength and speed of the forward and reverse shock, which will influence the survival and destruction time scales of dust within the ejecta. After passage of the reverse shock, the density of the ejecta to the NE will therefore be ~ 10 times greater than the ejecta to the SW, and the shocked gas temperatures to the NE will be ~ 5 times lower than that to the SW. Hard x-ray observations of the Sgr A East ejecta re-

veal an electron temperature of $T_e \sim 2 \times 10^7 \text{ K}$ and density of $n_e \sim 10 \text{ cm}^{-3}$ (14, 35). This hot ejecta can be associated with the diffuse SW expansion (Fig. 1B), whereas the NE ejecta, which is located at the projected position of the observed dust, will be unobservable in hard x-ray emission, given an initial temperature of $T_e \sim 4 \times 10^6 \text{ K}$ and density of $n_e \sim 100 \text{ cm}^{-3}$.

The time scale for complete destruction due to thermal sputtering for $a = 0.04\text{-}\mu\text{m}$ -sized grains is ~ 3000 years in the hot, diffuse region of the ejecta and ~ 1000 years in the cooler, dense region (23). These time scales do not include the erosion effects due to the kinetic sputtering that occurs when the SN-condensed dust initially encounters the reverse shock (3). These effects include calculations beyond the scope of this paper; therefore, our estimates apply to dust that has survived the initial kinetic sputtering and can be treated as an upper limit on the dust lifetimes. Given the estimated age of the SNR, this destruction time scale implies that a large fraction of the LGs ($a \sim 0.04 \mu\text{m}$) and VSGs ($a \sim 0.001 \mu\text{m}$) we adopt for our SED models will be destroyed. However, unlike the diffuse, hot SW ejecta, the dense, cooler ejecta at the NE undergoes substantial radiative cooling that occurs on time scales much shorter than the dust-sputtering lifetimes.

In the metal-enriched environment of SN ejecta, iron will dominate the radiative cooling at temperatures greater than 10^6 K (36). Assuming that there is $\sim 0.15 M_{\odot}$ of iron within the ejecta (35), with density in the NE 10 times that in the SW, and that the ejecta is in collisional ionization equilibrium in each region, we estimate the cooling time scale of the NE ejecta to be $\tau_{\text{cool}} \sim 400$ years, which is much shorter than the estimated sputtering time scale. Assuming the NE ejecta cools to as low as $T_e \sim 10^5 \text{ K}$, the destruction time scales are $\sim 10^6$ years for $a = 0.04\text{-}\mu\text{m}$ -sized grains, which is sufficiently long—even for VSGs—to survive within the ejecta. As the ejecta cools to even lower temperatures, the thermal sputtering becomes negligible. Toward the SW, we find that the cooling time scale for the ejecta is $\tau_{\text{cool}} \sim 3 \times 10^4$ years, consistent with the detection of hard x-rays there, given the estimated age of the SNR. Unfortunately, the high levels of extinction toward Sgr A East does not currently allow for observations of optical cooling lines or soft x-ray emission to directly confirm our hypothesis. However, the prominent spatial anticorrelation of the dust emission and the hard x-ray emission in Sgr A East strongly suggests that dust-survival conditions are much more favorable in cooler regions of the ejecta (see Fig. 1B).

The SN-condensed dust is then injected into the surrounding ISM, with minimal erosion from thermal sputtering and nonthermal kinetic sputtering. Assuming that the progenitor of Sgr A East was comoving with the surrounding medium, the relative velocity between the dust and the ISM can be estimated to be $\sim 100 \text{ km s}^{-1}$, based on the age of the SNR ($\sim 10^4$ years) and the approximate distance traveled by the SN-condensed dust from the apparent center of the SNR ($\sim 1 \text{ pc}$). After the dust has been slowed from 100 km s^{-1}

to 10 km s^{-1} by collisions with the surrounding ISM, the decrease in grain radius due to kinetic sputtering will be less than 10% (3, 37).

Grain-grain collisions and the VSG-LG mass ratio

The temperatures of the NE ejecta fall in a regime where grain-grain collisions become important, if we assume that the thermal velocities of the dust grains are closely coupled to those of the gas. Unlike sputtering by ions, which erodes the grain atom by atom, grain-grain collisions are efficient at redistributing mass from larger ($a \sim 0.1 \mu\text{m}$) to smaller ($a \lesssim 0.005 \mu\text{m}$) grains by shattering or fragmentation (38). We attribute the enhanced VSG-LG mass ratio of the SNR dust derived from the SED models (Fig. 4 and Table 1) to grain-grain collisions, which occur between 0.005- and 0.1- μm grains over ~ 60 years (23). Therefore, grain-grain collisions and the fragmentation of larger grains into smaller grains occur on time scales shorter than the age of the SNR and the dust-sputtering lifetimes, which should lead to a substantial enhancement in the small-grain mass abundance relative to that of the interstellar medium, consistent with our models.

SN dust mass–survival fraction and implications for galaxies of the early universe

Our results show that, given a dense surrounding environment, SN-condensed dust can survive the destructive passage of the reverse shock to be injected into the ISM. The mass of the Sgr A East SNR dust provides an estimate of the mass-survival fraction of dust initially condensed in the ejecta. SNe dust-production models for a progenitor with a mass of 13 to $20 M_{\odot}$ predict that $\sim 0.3 M_{\odot}$ of dust forms in the ejecta (2, 3); therefore, our derived SNR dust mass of $\sim 0.2 M_{\odot}$ implies that 7% of the total initial dust mass survived the passage of the reverse shock. This number is very uncertain, however, because estimates of dust masses produced in SNe are not well constrained, given the uncertainties in the microphysics that dictate grain formation.

Excluding SN1987A, infrared observations of SNRs much younger than Sgr A East reveal that far smaller quantities of dust are formed in the ejecta than predicted by models ($\lesssim 0.1 M_{\odot}$) [(39) and references therein]. If we assume that $0.1 M_{\odot}$ of dust is initially formed in SNe, which is the quantity of dust detected in Cassiopeia A and the Crab Nebula (39), our results imply that 20% of the initial dust mass survives the reverse shock to be injected into the ISM.

We apply our results to the dust-formation and -survival rates in the ISM of the galaxies in the early universe. In a scenario where the galaxy's star-formation history undergoes a single short and intense burst followed by a calm period with a much lower rate of star formation, a dust yield of $\sim 0.15 M_{\odot}$ per SN is required to produce the large inferred quantities of dust (40). Our current results suggest that it is difficult to produce the observed dust mass in such galaxies if only ~ 10 to 20% of the dust survives

the reverse shock. However, because stars in early-universe galaxies form in substantially denser regions than those in local galaxies [(41) and references therein], the dust mass–survival rate is probably greater than that which we infer. Additionally, if each SNe produced as much dust as observed in the ejecta of SN1987A ($\sim 0.5 M_{\odot}$) (42, 43), SNe could reasonably account for the dust production. These findings are consistent with SNe being a dominant dust-production mechanism in galaxies of the early universe (7).

REFERENCES AND NOTES

- E. Dwek, *Astrophys. J.* **501**, 643–665 (1998).
- S. Bianchi, R. Schneider, *Mon. Not. R. Astron. Soc.* **378**, 973–982 (2007).
- T. Nozawa et al., *Astrophys. J.* **666**, 955–966 (2007).
- D. W. Silvia, B. D. Smith, J. M. Shull, *Astrophys. J.* **748**, 12 (2012).
- T. Temini et al., *Astrophys. J.*, in press; preprint available at <http://arxiv.org/abs/1411.4574> (2014).
- M. Lakićević et al., *Astrophys. J.*, in press; preprint available at <http://arxiv.org/abs/1410.5709> (2014).
- E. Dwek et al., *Astrophys. J.* **788**, L30 (2014).
- M. J. Reid et al., *Annu. Rev. Astron. Astrophys.* **31**, 345–372 (1993).
- R. D. Ekers, J. H. van Gorkom, U. H. Schwarz, W. M. Goss, *Astron. Astrophys.* **122**, 143–150 (1983).
- F. Yusef-Zadeh, M. R. Morris, *Astrophys. J.* **320**, 545–561 (1987).
- F. Yusef-Zadeh, D. A. Roberts, W. M. Goss, D. A. Frail, A. J. Green, *Astrophys. J.* **512**, 230–236 (1999).
- K. I. Uchida, M. R. Morris, E. Serabyn, D. Fong, T. Meseroll, in *The Central Regions of the Galaxy and Galaxies*, Y. Sofue, Ed. (International Astronomical Union Symposia series no. 184, Springer, Dordrecht, Netherlands, 1998), p. 317.
- J.-H. Zhao, M. R. Morris, W. M. Goss, *Astrophys. J.* **146**, 8 (2013).
- Y. Maeda et al., *Astrophys. J.* **570**, 671–687 (2002).
- M. R. Morris, E. Serabyn, *Annu. Rev. Astron. Astrophys.* **34**, 645–701 (1996).
- T. L. Herter et al., *Publ. Astron. Soc. Pac.* **125**, 1393–1404 (2013).
- G. G. Fazio et al., *Astrophys. J. Suppl. Ser.* **154**, 10–17 (2004).
- S. Stolovy et al., *J. Phys. Conf. Ser.* **54**, 176–182 (2006).
- G. L. Pilbratt et al., *Astron. Astrophys.* **518**, L1 (2010).
- A. Poglitsch et al., *Astron. Astrophys.* **518**, L2 (2010).
- See the supplementary materials for more details on the SOFIA/FORCAST and Chandra/ACIS-I observations.
- E. Mills et al., *Astrophys. J.* **735**, 84 (2011).
- See the supplementary materials on Science Online.
- R. M. Lau, T. L. Herter, M. R. Morris, E. E. Becklin, J. D. Adams, *Astrophys. J.* **775**, 37 (2013).
- R. M. Lau, T. L. Herter, M. R. Morris, J. D. Adams, *Astrophys. J.* **794**, 108 (2014).
- The 1σ errors are derived from the ±20% flux measurement uncertainty (see supplementary materials).
- Q. D. Wang et al., *Mon. Not. R. Astron. Soc.* **402**, 895–902 (2010).
- H. Dong et al., *Mon. Not. R. Astron. Soc.* **417**, 114–135 (2011).
- E. Dwek et al., *Astrophys. J.* **722**, 425–434 (2010).
- M. Compiègne et al., *Astron. Astrophys.* **525**, A103 (2011).
- R. Schödel et al., *Astron. Astrophys.* **566**, A47 (2014).
- F. Castelli, R. L. Kurucz, <http://arxiv.org/abs/astro-ph/0405087> (2004).
- J.-P. Bernard et al., *Astron. J.* **136**, 919–945 (2008).
- E. Serabyn, J. H. Lacy, J. M. Achternmann, *Astrophys. J.* **395**, 166–173 (1992).
- S. Park et al., *Astrophys. J.* **631**, 964–975 (2005).
- O. Gnat, G. J. Ferland, *Astrophys. J. Suppl. Ser.* **199**, 20 (2012).
- SN-condensed dust is also predicted to survive in diffuse environments ($n_{\text{e}} \sim 1 \text{ cm}^{-3}$) (3) such as the ISM of the Magellanic clouds (5, 6). However, the SN ejecta and dust would be much more diffuse in a $\sim 10^4$ -year-old SNR in such an environment and would therefore be difficult to distinguish from dust in the surrounding ISM.
- A. P. Jones, A. G. G. M. Tielens, D. J. Hollenbach, *Astrophys. J.* **469**, 740 (1996).
- C. Gall, J. Hjorth, A. C. Andersen, *Astron. Astrophys. Rev.* **19**, 43 (2011).
- E. Dwek, I. Cherdneff, *Astrophys. J.* **727**, 63 (2011).
- M. Shirazi, J. Brinchmann, A. Rahmati, *Astrophys. J.* **787**, 120 (2014).
- M. Matsuura et al., *Science* **333**, 1258–1261 (2011).
- R. Indebetouw et al., *Astrophys. J.* **782**, L2 (2014).

ACKNOWLEDGMENTS

We thank the rest of the FORCAST team, M. Hankins, G. Gull, J. Schoenwald, and C. Henderson, the Universities Space Research Association (USRA) Science and Mission Ops teams, and the entire SOFIA staff. Additionally, we thank E. Dwek and the anonymous referees for their insightful comments. Z.L. acknowledges support from the Recruitment Program of Global Youth Experts. This work is based on observations made with the NASA/German Aerospace Center (DLR) SOFIA. SOFIA science mission operations are conducted jointly by the USRA (under NASA contract NAS2-97001) and the Deutsches SOFIA Institut (under DLR contract 50 OK 0901). Financial support for FORCAST was provided by NASA through award 8500-98-014 issued by USRA. Data presented in this paper can be accessed from supplementary materials data S1 (23).

SUPPLEMENTARY MATERIALS

- www.sciencemag.org/content/348/6233/suppl/DC1
- Materials and Methods
- Supplementary Text
- Fig. S1
- Table S1
- References (44–48)
- Data S1

5 November 2014; accepted 6 March 2015
Published online 19 March 2015;
10.1126/science.aaa2208

REPORTS

GALAXY EVOLUTION

Isolated compact elliptical galaxies: Stellar systems that ran away

Igor Chilingarian^{1,2,*} and Ivan Zolotukhin^{2,3}

Compact elliptical galaxies form a rare class of stellar system (~30 presently known) characterized by high stellar densities and small sizes and often harboring metal-rich stars. They were thought to form through tidal stripping of massive progenitors, until two isolated objects were discovered where massive galaxies performing the stripping could not be identified. By mining astronomical survey data, we have now found 195 compact elliptical galaxies in all types of environment. They all share similar dynamical and stellar population properties. Dynamical analysis for nonisolated galaxies demonstrates the feasibility of their ejection from host clusters and groups by three-body encounters, which is in agreement with numerical simulations. Hence, isolated compact elliptical and isolated quiescent dwarf galaxies are tidally stripped systems that ran away from their hosts.

Compact elliptical galaxies are thought to form through the hierarchical merging of smaller building blocks into larger systems (1, 2), and the history of these interactions is imprinted in their observable properties. Some galaxies, such as ultra-compact dwarfs (3) and compact ellipticals (cEs) (4–7), show evidence of strong tidal interactions with massive neighboring galaxies (8) that stripped most of the stars from the compact galaxies' progenitors. cEs are rare galaxies with high stellar densities that resemble centers of giant ellipticals but have masses that are about two orders of magnitude smaller [$M \sim 10^9$ solar mass (M_{\odot})]. They are found mostly in the cores of galaxy clusters next to massive central galaxies, which is in alignment with the above hypothesis for their evolution.

The recent discoveries of isolated cE galaxies (9, 10) that do not belong to any galaxy cluster or group raised another round of debate about cE formation: whether they all formed through the tidal stripping, or through a different mechanism

of formation, such as mergers of dwarf galaxies with specific morphologies and configurations (10). Dwarf-dwarf galaxy mergers do happen in vicinities of massive galaxies (11, 12). However, neither have they been observed in low-density environments, nor do any of the remnants resemble properties of cE galaxies. The existence of a substantial number of isolated cEs will hence imply notably higher dwarf-dwarf merger rates than predicted by numerical simulations (13) and challenge the currently accepted hierarchical structure formation paradigm.

We demonstrated that all known cE galaxies are outliers from the universal optical-ultraviolet color-color-magnitude relation of galaxies (14). We could therefore perform a search for cE galaxies not only in the centers of rich clusters and groups as has been done before (5), but across all environments using data from wide-field imaging surveys, the optical ground-based Sloan Digital Sky Survey [SDSS, (15)] and the ultraviolet all-sky survey carried out by the GALaxy Evolution eXplorer [GALEX, (16)] spacecraft, which are all publicly available in the Virtual Observatory.

First, we created an initial list of candidates (supplementary materials) from the sample of galaxies having spectra in the SDSS and, hence, known distances by selecting outliers above +0.035 mag in the optical ($g - r$) color from the

¹Smithsonian Astrophysical Observatory, 60 Garden Street MS09, Cambridge, MA 02138, USA. ²Sternberg Astronomical Institute, Moscow State University, 13 Universitetsky prospect, Moscow 119992, Russia. ³L'Institut de Recherche en Astrophysique et Planétologie, 9 Avenue du Colonel Roche BP 44346, 31028, Toulouse Cedex 4, France.

*Corresponding author. E-mail: igor.chilingarian@cfa.harvard.edu



Old supernova dust factory revealed at the Galactic center

R. M. Lau *et al.*

Science **348**, 413 (2015);

DOI: 10.1126/science.aaa2208

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 24, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/413.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2015/03/18/science.aaa2208.DC1.html>

This article cites 40 articles, 4 of which can be accessed free:

<http://www.sciencemag.org/content/348/6233/413.full.html#ref-list-1>

This article appears in the following subject collections:

Astronomy

<http://www.sciencemag.org/cgi/collection/astronomy>

the reverse shock. However, because stars in early-universe galaxies form in substantially denser regions than those in local galaxies [(41) and references therein], the dust mass–survival rate is probably greater than that which we infer. Additionally, if each SNe produced as much dust as observed in the ejecta of SN1987A ($\sim 0.5 M_{\odot}$) (42, 43), SNe could reasonably account for the dust production. These findings are consistent with SNe being a dominant dust-production mechanism in galaxies of the early universe (7).

REFERENCES AND NOTES

- E. Dwek, *Astrophys. J.* **501**, 643–665 (1998).
- S. Bianchi, R. Schneider, *Mon. Not. R. Astron. Soc.* **378**, 973–982 (2007).
- T. Nozawa et al., *Astrophys. J.* **666**, 955–966 (2007).
- D. W. Silvia, B. D. Smith, J. M. Shull, *Astrophys. J.* **748**, 12 (2012).
- T. Temini et al., *Astrophys. J.*, in press; preprint available at <http://arxiv.org/abs/1411.4574> (2014).
- M. Lakićević et al., *Astrophys. J.*, in press; preprint available at <http://arxiv.org/abs/1410.5709> (2014).
- E. Dwek et al., *Astrophys. J.* **788**, L30 (2014).
- M. J. Reid et al., *Annu. Rev. Astron. Astrophys.* **31**, 345–372 (1993).
- R. D. Ekers, J. H. van Gorkom, U. H. Schwarz, W. M. Goss, *Astron. Astrophys.* **122**, 143–150 (1983).
- F. Yusef-Zadeh, M. R. Morris, *Astrophys. J.* **320**, 545–561 (1987).
- F. Yusef-Zadeh, D. A. Roberts, W. M. Goss, D. A. Frail, A. J. Green, *Astrophys. J.* **512**, 230–236 (1999).
- K. I. Uchida, M. R. Morris, E. Serabyn, D. Fong, T. Meseroll, in *The Central Regions of the Galaxy and Galaxies*, Y. Sofue, Ed. (International Astronomical Union Symposia series no. 184, Springer, Dordrecht, Netherlands, 1998), p. 317.
- J.-H. Zhao, M. R. Morris, W. M. Goss, *Astrophys. J.* **146**, 8 (2013).
- Y. Maeda et al., *Astrophys. J.* **570**, 671–687 (2002).
- M. R. Morris, E. Serabyn, *Annu. Rev. Astron. Astrophys.* **34**, 645–701 (1996).
- T. L. Herter et al., *Publ. Astron. Soc. Pac.* **125**, 1393–1404 (2013).
- G. G. Fazio et al., *Astrophys. J. Suppl. Ser.* **154**, 10–17 (2004).
- S. Stolovy et al., *J. Phys. Conf. Ser.* **54**, 176–182 (2006).
- G. L. Pilbratt et al., *Astron. Astrophys.* **518**, L1 (2010).
- A. Poglitsch et al., *Astron. Astrophys.* **518**, L2 (2010).
- See the supplementary materials for more details on the SOFIA/FORCAST and Chandra/ACIS-I observations.
- E. Mills et al., *Astrophys. J.* **735**, 84 (2011).
- See the supplementary materials on Science Online.
- R. M. Lau, T. L. Herter, M. R. Morris, E. E. Becklin, J. D. Adams, *Astrophys. J.* **775**, 37 (2013).
- R. M. Lau, T. L. Herter, M. R. Morris, J. D. Adams, *Astrophys. J.* **794**, 108 (2014).
- The 1σ errors are derived from the ±20% flux measurement uncertainty (see supplementary materials).
- Q. D. Wang et al., *Mon. Not. R. Astron. Soc.* **402**, 895–902 (2010).
- H. Dong et al., *Mon. Not. R. Astron. Soc.* **417**, 114–135 (2011).
- E. Dwek et al., *Astrophys. J.* **722**, 425–434 (2010).
- M. Compiègne et al., *Astron. Astrophys.* **525**, A103 (2011).
- R. Schödel et al., *Astron. Astrophys.* **566**, A47 (2014).
- F. Castelli, R. L. Kurucz, <http://arxiv.org/abs/astro-ph/0405087> (2004).
- J.-P. Bernard et al., *Astron. J.* **136**, 919–945 (2008).
- E. Serabyn, J. H. Lacy, J. M. Achternmann, *Astrophys. J.* **395**, 166–173 (1992).
- S. Park et al., *Astrophys. J.* **631**, 964–975 (2005).
- O. Gnat, G. J. Ferland, *Astrophys. J. Suppl. Ser.* **199**, 20 (2012).
- SN-condensed dust is also predicted to survive in diffuse environments ($n_{\text{e}} \sim 1 \text{ cm}^{-3}$) (3) such as the ISM of the Magellanic clouds (5, 6). However, the SN ejecta and dust would be much more diffuse in a $\sim 10^4$ -year-old SNR in such an environment and would therefore be difficult to distinguish from dust in the surrounding ISM.
- A. P. Jones, A. G. G. M. Tielens, D. J. Hollenbach, *Astrophys. J.* **469**, 740 (1996).
- C. Gall, J. Hjorth, A. C. Andersen, *Astron. Astrophys. Rev.* **19**, 43 (2011).
- E. Dwek, I. Cherdneff, *Astrophys. J.* **727**, 63 (2011).
- M. Shirazi, J. Brinchmann, A. Rahmati, *Astrophys. J.* **787**, 120 (2014).
- M. Matsuura et al., *Science* **333**, 1258–1261 (2011).
- R. Indebetouw et al., *Astrophys. J.* **782**, L2 (2014).

ACKNOWLEDGMENTS

We thank the rest of the FORCAST team, M. Hankins, G. Gull, J. Schoenwald, and C. Henderson, the Universities Space Research Association (USRA) Science and Mission Ops teams, and the entire SOFIA staff. Additionally, we thank E. Dwek and the anonymous referees for their insightful comments. Z.L. acknowledges support from the Recruitment Program of Global Youth Experts. This work is based on observations made with the NASA/German Aerospace Center (DLR) SOFIA. SOFIA science mission operations are conducted jointly by the USRA (under NASA contract NAS2-97001) and the Deutsches SOFIA Institut (under DLR contract 50 OK 0901). Financial support for FORCAST was provided by NASA through award 8500-98-014 issued by USRA. Data presented in this paper can be accessed from supplementary materials data S1 (23).

SUPPLEMENTARY MATERIALS

- www.sciencemag.org/content/348/6233/413/suppl/DC1
- Materials and Methods
- Supplementary Text
- Fig. S1
- Table S1
- References (44–48)
- Data S1

5 November 2014; accepted 6 March 2015
Published online 19 March 2015;
10.1126/science.aaa2208

REPORTS

GALAXY EVOLUTION

Isolated compact elliptical galaxies: Stellar systems that ran away

Igor Chilingarian^{1,2,*} and Ivan Zolotukhin^{2,3}

Compact elliptical galaxies form a rare class of stellar system (~30 presently known) characterized by high stellar densities and small sizes and often harboring metal-rich stars. They were thought to form through tidal stripping of massive progenitors, until two isolated objects were discovered where massive galaxies performing the stripping could not be identified. By mining astronomical survey data, we have now found 195 compact elliptical galaxies in all types of environment. They all share similar dynamical and stellar population properties. Dynamical analysis for nonisolated galaxies demonstrates the feasibility of their ejection from host clusters and groups by three-body encounters, which is in agreement with numerical simulations. Hence, isolated compact elliptical and isolated quiescent dwarf galaxies are tidally stripped systems that ran away from their hosts.

Compact elliptical galaxies are thought to form through the hierarchical merging of smaller building blocks into larger systems (1, 2), and the history of these interactions is imprinted in their observable properties. Some galaxies, such as ultra-compact dwarfs (3) and compact ellipticals (cEs) (4–7), show evidence of strong tidal interactions with massive neighboring galaxies (8) that stripped most of the stars from the compact galaxies' progenitors. cEs are rare galaxies with high stellar densities that resemble centers of giant ellipticals but have masses that are about two orders of magnitude smaller [$M \sim 10^9$ solar mass (M_{\odot})] . They are found mostly in the cores of galaxy clusters next to massive central galaxies, which is in alignment with the above hypothesis for their evolution.

The recent discoveries of isolated cE galaxies (9, 10) that do not belong to any galaxy cluster or group raised another round of debate about cE formation: whether they all formed through the tidal stripping, or through a different mechanism

of formation, such as mergers of dwarf galaxies with specific morphologies and configurations (10). Dwarf-dwarf galaxy mergers do happen in vicinities of massive galaxies (11, 12). However, neither have they been observed in low-density environments, nor do any of the remnants resemble properties of cE galaxies. The existence of a substantial number of isolated cEs will hence imply notably higher dwarf-dwarf merger rates than predicted by numerical simulations (13) and challenge the currently accepted hierarchical structure formation paradigm.

We demonstrated that all known cE galaxies are outliers from the universal optical-ultraviolet color-color-magnitude relation of galaxies (14). We could therefore perform a search for cE galaxies not only in the centers of rich clusters and groups as has been done before (5), but across all environments using data from wide-field imaging surveys, the optical ground-based Sloan Digital Sky Survey [SDSS, (15)] and the ultraviolet all-sky survey carried out by the GALaxy Evolution eXplorer [GALEX, (16)] spacecraft, which are all publicly available in the Virtual Observatory.

First, we created an initial list of candidates (supplementary materials) from the sample of galaxies having spectra in the SDSS and, hence, known distances by selecting outliers above +0.035 mag in the optical ($g - r$) color from the

¹Smithsonian Astrophysical Observatory, 60 Garden Street MS09, Cambridge, MA 02138, USA. ²Sternberg Astronomical Institute, Moscow State University, 13 Universitetsky prospect, Moscow 119992, Russia. ³L'Institut de Recherche en Astrophysique et Planétologie, 9 Avenue du Colonel Roche BP 44346, 31028, Toulouse Cedex 4, France.
*Corresponding author. E-mail: igor.chilingarian@cfa.harvard.edu

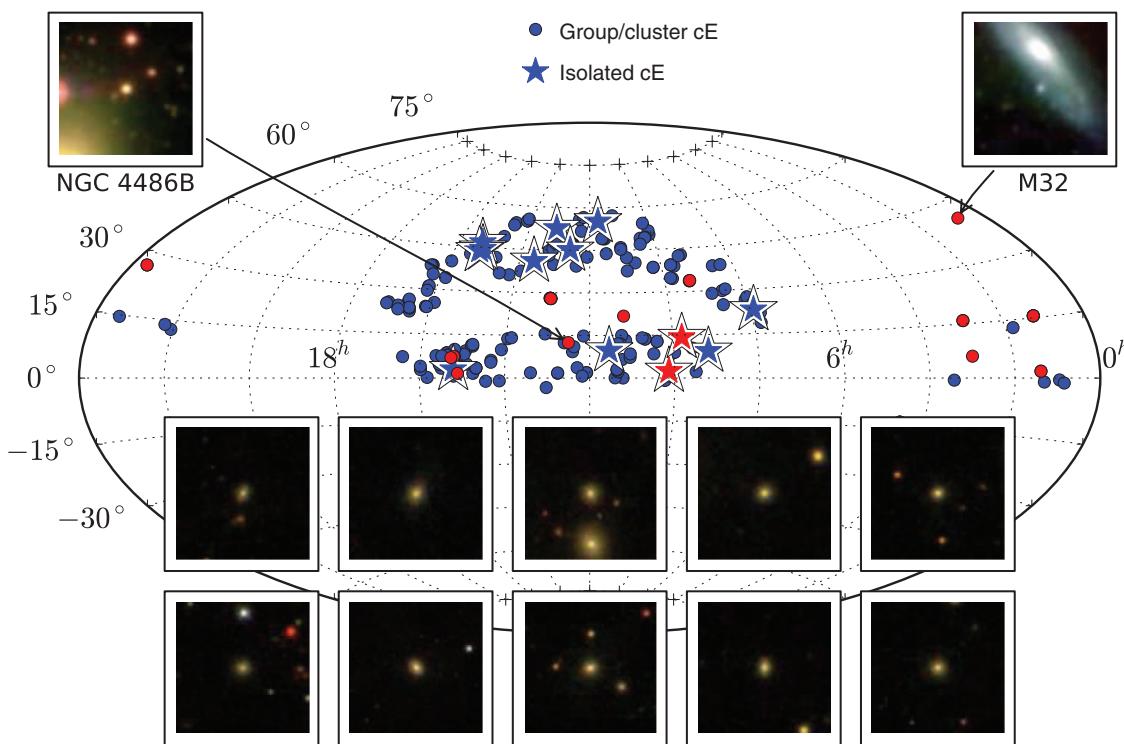


Fig. 1. A sample of compact elliptical galaxies in all types of environment. Our current sample of compact elliptical galaxies (blue symbols) is compared with a data set compiled from the literature (red symbols). Dots and stars denote group/cluster and isolated compact elliptical galaxies, respectively. Square panels in the bottom part of the figure show representatives of the current sample, and top corner insets display Messier 32 and NGC 4486, prototypical cEs in the local universe, as they would look with the SDSS telescope at a 130 Mpc distance ($z = 0.03$). Each inset panel covers a 20 by 20 kpc region centered on a cE.

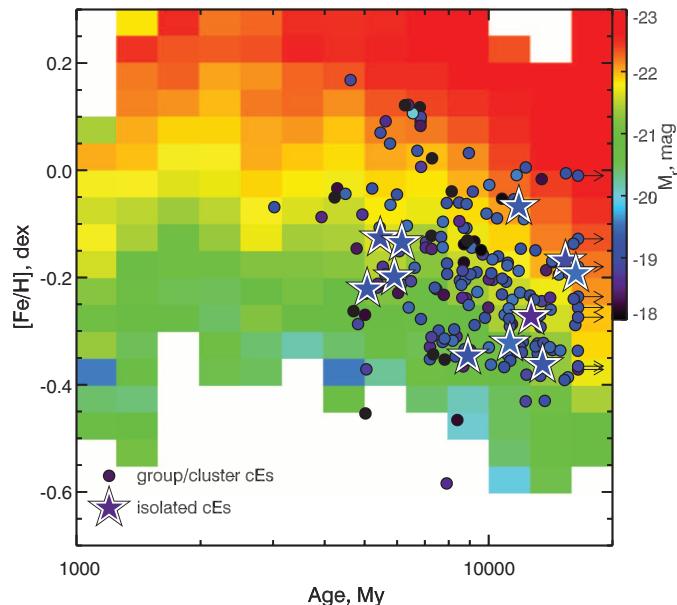
universal relation (14). We chose low-luminosity galaxies [$L < 4 \times 10^9$ solar luminosity (L_\odot) or absolute magnitude ($M_g > -18.7$ mag)] that had small half-light radii ($R_e < 0.6$ kpc) or were spatially unresolved in SDSS images; did not show substantial ellipticity, which was essential for removing edge-on spiral galaxies; had the redshifts in the range of $0.007 < z < 0.08$ (distances between 30 and 340 Mpc); and either possessed red near-ultraviolet colors [$(NUV - r) > 4$ mag] or remained undetected in the NUV band. We constrained by color and also removed objects that have emission lines in their spectra in order to exclude any objects with recent or ongoing star formation.

We then fitted their SDSS spectra against a grid of stellar population models using the “NBursts” code (17) and obtained mean ages, metallicities, and velocity dispersions of their stars. We rejected candidates with stellar ages younger than 4 billion years and introduced an additional constraint based on stellar velocity dispersions ($\sigma > 60$ km/s). Stellar systems in equilibrium that are dynamically supported by random motions of stars, as most elliptical galaxies are, have their dynamical masses (M_{vir}), half-mass radii (R_e), and global velocity dispersions (σ_v) connected by the simple virial relation $M_{\text{vir}} = 9.96 R_e \sigma_v^2/G$ (18, 19). Therefore, for a galaxy with known velocity dispersion and a stellar mass (M_*) derived from its luminosity and stellar population parameters, we can estimate the lower limit for the half-light radius (if a galaxy contains dark matter, its real half-light radius will be larger because $M_{\text{vir}} > M_*$). Hence, we can firmly reject physically extended objects such as “normal” dwarf elliptical galaxies that are unresolved in SDSS images

Fig. 2. Comparison of stellar population properties of isolated and nonisolated compact elliptical galaxies to a reference sample of elliptical galaxies from the SDSS. Mean stellar ages and metallicities of cE galaxies hosted by groups and clusters (round symbols) and isolated cEs (stars) were determined from the modeling of their SDSS spectra by using the NBursts full-spectrum fitting. The plotting symbols are color-coded according to the r -band absolute magnitudes derived from the SDSS photometry. We also computed ages and metallicities for a reference sample of 82,500 elliptical galaxies from SDSS DR7 in the same fashion, along with the median r -band absolute magnitude for every bin of age-metallicity parameter space. These magnitudes are shown as a background map, with the colors on the same scale as cE galaxies. This figure demonstrates that (i) stellar populations of isolated and group/cluster cEs do not differ statistically and (ii) cEs are on average much fainter than are normal elliptical galaxies of the same stellar age/metallicity.

because of their large distances by selecting only objects with high velocity dispersions.

In this fashion, we constructed a sample of 195 galaxies (Fig. 1 and supplementary materials). We then cross-matched this list against the SDSS Galaxy Groups catalog (20) and established their



group/cluster membership. For seven objects without counterparts in the Galaxy Groups catalog, we identified possible host galaxies (in most cases, group centers) located between 750 kpc and 3.3 Mpc in projection. Because some bright and extended galaxies were missing from the SDSS

spectroscopic sample, and therefore also from the Galaxy Groups catalog, we used NASA/Infrared Processing and Analysis Center (IPAC) Extragalactic Database (NED) for the identification of host galaxies for 45 cEs. Our final sample contains 56 galaxies in clusters, 128 in groups, and 11 isolated or field cE galaxies. Eight galaxies (supplementary materials) exhibit prominent tidal streams similar to the two known cEs (8).

Ages and metallicities (Fig. 2 and supplementary materials), colors, and luminosities of 11 isolated cE galaxies do not show a statistically significant difference from those of galaxies being members of clusters and groups. The Kolmogorov-Smirnov (KS) probabilities of these properties for isolated and nonisolated subsamples to be derived from the same underlying distribution range from 30 (for luminosities) to 80% (metallicities). Our cE sample does not show any correlation between the metallicity and the stellar mass, conversely to normal elliptical galaxies, which exhibit a rather tight mass-metallicity relation (Fig. 2).

For rich groups and clusters hosting cE galaxies, we built caustic diagrams (21–23) that present differences of radial velocities of cluster members from the cluster center versus projected distances. A galaxy position on such a diagram reflects its dynamical status: objects sitting deep inside the cluster potential well are located inside the distribution, whereas galaxies near the edges are barely gravitationally bound to the host cluster or infalling onto it for the first time.

We constructed an ensemble cluster by normalizing individual cluster and group data by their velocity dispersions and sizes for 33 structures from our sample, each of which included over 20 member galaxies (24, 25). Then, we computed its caustic diagram in order to visualize the phase space pattern of the infalling galaxy population and overplotted our cE galaxies on it (Fig. 3). The cE population strongly differs from other cluster members. The KS tests for projected distance and radial velocity distributions reject the hypothesis of cE and cluster member samples being derived from the same parent population at the 97 and 98% levels. Numerical simulations of tidal stripping (5, 26) suggest that a progenitor galaxy, even if it approaches a cluster center on a very extended radial orbit, will lose a major fraction of its orbital energy because of dynamical friction, become gravitationally locked in the inner region of a cluster on a tightly bound orbit, and will finally be accreted by the host galaxy after a few billion years. Many cEs from our sample exhibit this behavior (Fig. 3 and supplementary materials). However, we see a number of cE galaxies close to the edges of the caustics, suggesting that they are barely gravitationally bound to the cluster potential because they do not belong to the infalling population, as we demonstrated. This looks completely unrealistic in the case of a one-to-one galaxy encounter resulting in tidal stripping, but in the case of a three- or multiple-body encounter, this situation becomes significantly more likely.

An interaction of binary stars with the central supermassive black hole is one accepted scenario for the creation of hypervelocity stars (27) in our

Galaxy: One of the binary components is ejected, whereas the other one falls onto the black hole. Numerical simulations suggest (28) that three-body encounters are responsible for putting Milky Way satellites on extreme orbits, going as far as 3 Mpc away. Even though typical galaxy clusters have much wider and deeper potential wells than that of the Local Group, three- and even multiple-body encounters must happen much more frequently in those dense environments. Therefore, a certain probability exists that some of them will lead to the gravitational ejection of galaxies participating in the interaction to extreme radial orbits with the apocentric distances of a few megaparsecs (29). A three-body encounter that might eject a cE galaxy from its host cluster or group does not have to happen during the cE formation through tidal stripping; that is, two galaxies do not have to fall onto the cluster/group center at the same time. When a cE progenitor is tidally stripped, it quickly

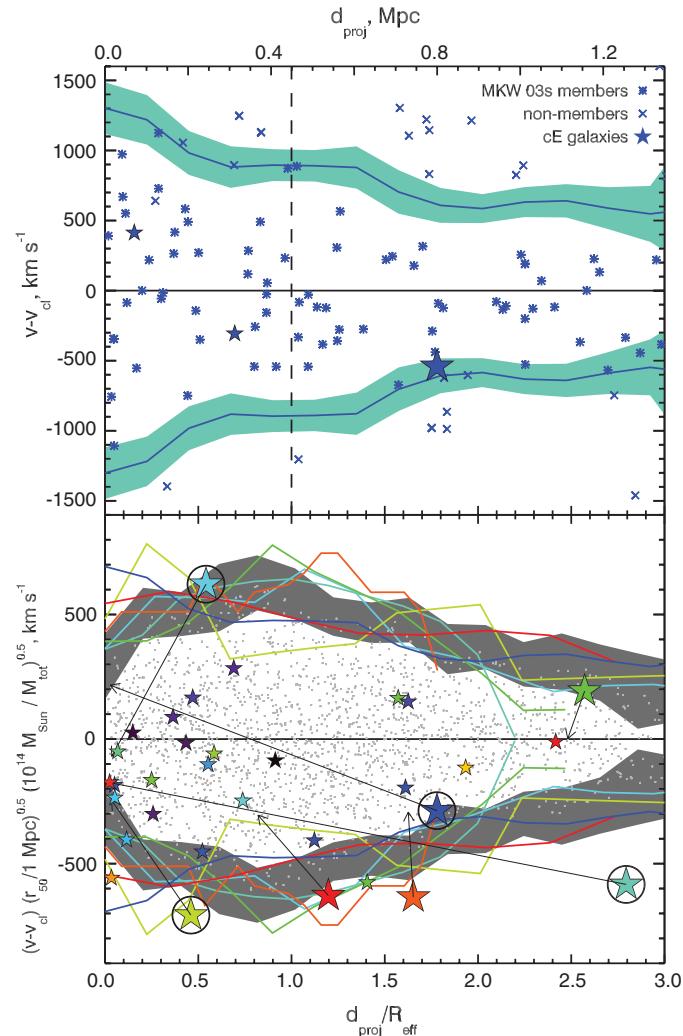
settles on a tightly bound, rapidly decaying orbit (5), and if another galaxy infalls later, but before a newly formed cE has been accreted (hundreds of millions to a couple of billion years), the three-body encounter becomes possible.

We estimate the probability of a close three-body encounter geometrically. Numerical simulations suggest (30) that over a typical cE lifetime of 2 billion years (5), an average brightest cluster galaxy (BCG) must have experienced three or four mergers with massive ($M_* > \sim 10^{10} M_\odot$) galaxies. We assume that (i) a typical cE resides on a quasicircular orbit within $r_{\text{cE}} \sim 120$ kpc from a host BCG galaxy after correction for projection effects (fig. S4), (ii) galaxies infall on a BCG on radial orbits from random directions, and (iii) a three-body encounter will be sufficiently close if a cE passes within $r_{3b} \approx 20$ kpc from a massive infalling galaxy. Hence, the probability is as a volume ratio of a cylinder of radius r_{3b} , height r_{cE} , and a

Fig. 3. Positions of cE galaxies identified on

caustic diagrams in galaxy clusters and rich groups with more than 20 members in the SDSS. (Top) An example of a caustic diagram showing projected distances and radial velocity differences of the members of the cluster MKW 03s that includes cE galaxies shown as stars. The derived caustics shown as solid blue lines roughly correspond to the escape velocity for a galaxy at a given distance from the cluster center. Shaded areas show statistical 1σ uncertainties of the caustic line computation.

(Bottom) The caustic lines for an ensemble cluster of 2592 individual members are shown as gray shaded areas and light gray dots, respectively. The caustic lines normalized by corresponding velocity dispersions and half-mass radii (R_{eff}) are shown for six individual clusters as colored solid lines.



Small stars denote cEs located deeply inside the potential wells of these six clusters, and large stars (with same colors as caustic lines) indicate cEs that have projected radial velocities of at least 85% of the caustic amplitude. Vectors point to the location of identified host galaxies for corresponding cEs. Three of them are different than the cluster of central galaxies, indicating that these cEs belong to subgroups inside clusters and that the caustic diagnostic is irrelevant for them. However, the four circled cEs are examples of galaxies barely gravitationally bound to their host clusters.

sphere of radius r_{CE} , $P_{3\text{b}} = 3/4(r_{3\text{b}}/r_{\text{CE}})^2 \approx 0.02$, or ~6 to 8% for three or four merger events.

In our sample of cluster and group cE galaxies, we indeed see numerous examples in which a cE resides only 20 to 80 kpc in projection from an ongoing major merger scene or several other massive cluster members are visible in the cE vicinity apart from the massive central cluster/group galaxy. Also, there is a known example of a globular cluster in the Virgo cluster (31) that was likely ejected at the speed of 2500 km/s and became gravitationally unbound to the cluster and its central galaxy, Messier 87.

We conclude that the tidal stripping process can explain all observational manifestations of compact elliptical galaxies, including the formation of isolated cEs whose existence was suggested as a strong counter-argument for tidal stripping (9). The ejection of cEs from central regions of galaxy clusters by three-body encounters is a channel for these galaxies to survive for an extended period of time in the violent cluster environment, where they would otherwise be accreted by massive hosts on a time scale of 2 billion to 3 billion years. The 11 isolated cEs probably represent a population of runaway galaxies that received sufficient kick velocities to leave their host clusters or groups forever.

The gravitational ejection mechanism may also explain the very existance of extremely rare isolated quiescent dwarf galaxies (32), where the star formation quenching is usually explained by environmental effects. These systems are more spatially extended than cEs and do not exhibit substantial tidal stripping footprints. This suggests that they never came very close to cluster/group centers, and therefore, the three-body encounter probability for them should be lower than that for cEs, although still nonnegligible.

REFERENCES AND NOTES

- S. D. M. White, C. S. Frenk, *Astrophys. J.* **379**, 52 (1991).
- S. Cole, C. G. Lacey, C. M. Baugh, C. S. Frenk, *Mon. Not. R. Astron. Soc.* **319**, 168–204 (2000).
- M. J. Drinkwater et al., *Nature* **423**, 519–521 (2003).
- J. Price et al., *Mon. Not. R. Astron. Soc.* **397**, 1816–1835 (2009).
- I. Chilingarian et al., *Science* **326**, 1379–1382 (2009).
- I. V. Chilingarian, G. Bergond, *Mon. Not. R. Astron. Soc.* **405**, L11–L15 (2010).
- M. A. Norris et al., *Mon. Not. R. Astron. Soc.* **443**, 1151–1172 (2014).
- A. P. Huxor, S. Phillips, J. Price, R. Harniman, *Mon. Not. R. Astron. Soc.* **414**, 3557–3565 (2011).
- A. P. Huxor, S. Phillips, J. Price, *Mon. Not. R. Astron. Soc.* **430**, 1956–1960 (2013).
- S. Paudel, T. Lisker, K. S. A. Hansson, A. P. Huxor, *Mon. Not. R. Astron. Soc.* **443**, 446–453 (2014).
- R. M. Rich et al., *Nature* **482**, 192–194 (2012).
- N. C. Amorisco, N. W. Evans, G. van de Ven, *Nature* **507**, 335–337 (2014).
- O. Fakhouri, C.-P. Ma, M. Boylan-Kolchin, *Mon. Not. R. Astron. Soc.* **406**, 2267–2278 (2010).
- I. V. Chilingarian, I. Y. Zolotukhin, *Mon. Not. R. Astron. Soc.* **419**, 1727–1739 (2012).
- K. N. Abazajian et al., *J. Suppl. Ser.* **182**, 543–558 (2009).
- D. C. Martin et al., *Astrophys. J.* **619**, L1–L6 (2005).
- I. Chilingarian, P. Prugniel, O. Sil'chenko, M. Koleva, *Stellar Populations as Building Blocks of Galaxies*, A. Vazdekis, R. R. Peletier, Eds. (Cambridge Univ. Press, Cambridge, UK, 2007), vol. 241 of IAU Symposium, pp. 175–176, arXiv:0709.3047.
- L. Spitzer, *Astrophys. J.* **158**, L139 (1969).
- L. Hernquist, *Astrophys. J.* **356**, 359 (1990).
- E. Tempel, E. Tago, L. J. Liivamägi, *Astron. Astrophys.* **540**, A106 (2012).
- N. Kaiser, *Mon. Not. R. Astron. Soc.* **227**, 1–21 (1987).

- A. Diaferio, M. J. Geller, *Astrophys. J.* **481**, 633–643 (1997).
- A. Diaferio, *Mon. Not. R. Astron. Soc.* **309**, 610–622 (1999).
- R. G. Carlberg, H. K. C. Yee, E. Ellingson, *Astrophys. J.* **478**, 462–475 (1997).
- A. Biviano, M. Girardi, *Astrophys. J.* **585**, 205–214 (2003).
- J. Pfeffer, H. Baumgardt, *Mon. Not. R. Astron. Soc.* **433**, 1997–2005 (2013).
- W. R. Brown, M. J. Geller, S. J. Kenyon, M. J. Kurtz, *Astrophys. J.* **622**, L33–L36 (2005).
- L. V. Sales, J. F. Navarro, M. G. Abadi, M. Steinmetz, *Mon. Not. R. Astron. Soc.* **379**, 1475–1483 (2007).
- A. R. Wetzel, J. L. Tinker, C. Conroy, F. C. Bosch, *Mon. Not. R. Astron. Soc.* **439**, 2687–2700 (2014).
- G. De Lucia, J. Blaizot, *Mon. Not. R. Astron. Soc.* **375**, 2–14 (2007).
- N. Caldwell et al., *Astrophys. J.* **787**, L11 (2014).
- M. Geha, M. R. Blanton, R. Yan, J. L. Tinker, *Astrophys. J.* **757**, 85 (2012).

ACKNOWLEDGMENTS

The authors are grateful to F. Combes (Observatoire de Paris), I. Katkov (Sternberg Astronomical Institute), and M. Kurtz (Smithsonian Astrophysical Observatory) for useful discussions and critical reading of the manuscript. This result emerged from the tutorial run by the authors at the Astronomical Data Analysis Software and Systems conference in 2012. The authors acknowledge support by the Russian Science Foundation project 14-22-00041 "VOLGA—A View On the

Life of Galaxies." The project used computational resources funded by the M. V. Lomonosov Moscow State University Program of Development. This research has made use of Aladin developed by the Centre de Données Astronomiques de Strasbourg; TOPCAT and STILTS software packages developed by M. Taylor; "exploreedsds" script by G. Mamou; the VizieR catalog access tool, CDS, Strasbourg, France; and the NASA/IPAC NED, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, NASA, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is www.sdss.org. GALEX and SDSS databases used in our study are available via the CasJobs web-site http://skyserver.sdss.org/CasJobs.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/418/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S4
References (33–46)
Table S1

19 November 2014; accepted 20 March 2015
10.1126/science.aaa3344

ORGANIC CHEMISTRY

A rhodium catalyst for single-step styrene production from benzene and ethylene

Benjamin A. Vaughan,¹ Michael S. Webster-Gardiner,¹ Thomas R. Cundari,^{2*} T. Brent Gunnoe^{1*}

Rising global demand for fossil resources has prompted a renewed interest in catalyst technologies that increase the efficiency of conversion of hydrocarbons from petroleum and natural gas to higher-value materials. Styrene is currently produced from benzene and ethylene through the intermediacy of ethylbenzene, which must be dehydrogenated in a separate step. The direct oxidative conversion of benzene and ethylene to styrene could provide a more efficient route, but achieving high selectivity and yield for this reaction has been challenging. Here, we report that the Rh catalyst (^{F1}DAB)Rh(TFA)(η²-C₂H₄) [^{F1}DAB is *N,N'*-bis(pentafluorophenyl)-2,3-dimethyl-1,4-diaza-1,3-butadiene; TFA is trifluoroacetate] converts benzene, ethylene, and Cu(II) acetate to styrene, Cu(I) acetate, and acetic acid with 100% selectivity and yields ≥95%. Turnover numbers >800 have been demonstrated, with catalyst stability up to 96 hours.

Vinyl arenes are important precursors for fine chemical synthesis, as well as for the preparation of plastics and elastomers (1–5). For example, styrene is produced globally on a scale of ~18.5 million tons (2). Current methods for the large-scale production of vinyl arenes involve multiple steps, typically beginning with arene alkylation using a Friedel-Crafts (e.g., AlCl₃ with HF) or zeolite catalyst followed by energy-intensive dehydrogenation of the alkyl group (Fig. 1) (1–6). Friedel-Crafts catalysis suffers from the use of harsh acids, including HF, low selectivity for the monoalkylated

product (polyalkylation is inherent to the mechanism), and the generation of stoichiometric waste (2). Zeolite catalysts have improved the process for benzene alkylation, yet these catalysts still require high temperatures (generally 350° to 450°C) and give polyalkylated products (2, 7–10).

An alternative method for the production of vinyl arenes is a direct and single-step oxidative arene vinylation (Fig. 1). If the terminal oxidant is oxygen from air (either introduced in situ or used to recycle a different in situ oxidant), the net reaction is the conversion of benzene, ethylene, and oxidant to styrene and water (11). Acid-based (i.e., Friedel-Crafts or zeolite catalysts) catalysis occurs by electrophilic aromatic substitution and does not offer a viable pathway to directly generate vinyl arenes. Transition metal complexes that catalyze ethylene hydrophenylation by benzene

¹Department of Chemistry, University of Virginia, Charlottesville, VA 22903, USA. ²Center for Advanced Scientific Computing and Modeling, Department of Chemistry, University of North Texas, Denton, TX 76203, USA.

*Corresponding author. E-mail: tbg7h@virginia.edu (T.B.G.), t@unt.edu (T.R.C.)



Isolated compact elliptical galaxies: Stellar systems that ran away
 Igor Chilingarian and Ivan Zolotukhin
Science **348**, 418 (2015);
 DOI: 10.1126/science.aaa3344

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 24, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/418.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2015/04/22/348.6233.418.DC1.html>

This article cites 44 articles, 22 of which can be accessed free:

<http://www.sciencemag.org/content/348/6233/418.full.html#ref-list-1>

This article appears in the following subject collections:

Astronomy

<http://www.sciencemag.org/cgi/collection/astronomy>

sphere of radius r_{CE} , $P_{3\text{b}} = 3/4(r_{3\text{b}}/r_{\text{CE}})^2 \approx 0.02$, or ~6 to 8% for three or four merger events.

In our sample of cluster and group cE galaxies, we indeed see numerous examples in which a cE resides only 20 to 80 kpc in projection from an ongoing major merger scene or several other massive cluster members are visible in the cE vicinity apart from the massive central cluster/group galaxy. Also, there is a known example of a globular cluster in the Virgo cluster (31) that was likely ejected at the speed of 2500 km/s and became gravitationally unbound to the cluster and its central galaxy, Messier 87.

We conclude that the tidal stripping process can explain all observational manifestations of compact elliptical galaxies, including the formation of isolated cEs whose existence was suggested as a strong counter-argument for tidal stripping (9). The ejection of cEs from central regions of galaxy clusters by three-body encounters is a channel for these galaxies to survive for an extended period of time in the violent cluster environment, where they would otherwise be accreted by massive hosts on a time scale of 2 billion to 3 billion years. The 11 isolated cEs probably represent a population of runaway galaxies that received sufficient kick velocities to leave their host clusters or groups forever.

The gravitational ejection mechanism may also explain the very existance of extremely rare isolated quiescent dwarf galaxies (32), where the star formation quenching is usually explained by environmental effects. These systems are more spatially extended than cEs and do not exhibit substantial tidal stripping footprints. This suggests that they never came very close to cluster/group centers, and therefore, the three-body encounter probability for them should be lower than that for cEs, although still nonnegligible.

REFERENCES AND NOTES

- S. D. M. White, C. S. Frenk, *Astrophys. J.* **379**, 52 (1991).
- S. Cole, C. G. Lacey, C. M. Baugh, C. S. Frenk, *Mon. Not. R. Astron. Soc.* **319**, 168–204 (2000).
- M. J. Drinkwater et al., *Nature* **423**, 519–521 (2003).
- J. Price et al., *Mon. Not. R. Astron. Soc.* **397**, 1816–1835 (2009).
- I. Chilingarian et al., *Science* **326**, 1379–1382 (2009).
- I. V. Chilingarian, G. Bergond, *Mon. Not. R. Astron. Soc.* **405**, L11–L15 (2010).
- M. A. Norris et al., *Mon. Not. R. Astron. Soc.* **443**, 1151–1172 (2014).
- A. P. Huxor, S. Phillips, J. Price, R. Harniman, *Mon. Not. R. Astron. Soc.* **414**, 3557–3565 (2011).
- A. P. Huxor, S. Phillips, J. Price, *Mon. Not. R. Astron. Soc.* **430**, 1956–1960 (2013).
- S. Paudel, T. Lisker, K. S. A. Hansson, A. P. Huxor, *Mon. Not. R. Astron. Soc.* **443**, 446–453 (2014).
- R. M. Rich et al., *Nature* **482**, 192–194 (2012).
- N. C. Amorisco, N. W. Evans, G. van de Ven, *Nature* **507**, 335–337 (2014).
- O. Fakhouri, C.-P. Ma, M. Boylan-Kolchin, *Mon. Not. R. Astron. Soc.* **406**, 2267–2278 (2010).
- I. V. Chilingarian, I. Y. Zolotukhin, *Mon. Not. R. Astron. Soc.* **419**, 1727–1739 (2012).
- K. N. Abazajian et al., *J. Suppl. Ser.* **182**, 543–558 (2009).
- D. C. Martin et al., *Astrophys. J.* **619**, L1–L6 (2005).
- I. Chilingarian, P. Prugniel, O. Sil'chenko, M. Koleva, *Stellar Populations as Building Blocks of Galaxies*, A. Vazdekis, R. R. Peletier, Eds. (Cambridge Univ. Press, Cambridge, UK, 2007), vol. 241 of IAU Symposium, pp. 175–176, arXiv:0709.3047.
- L. Spitzer, *Astrophys. J.* **158**, L139 (1969).
- L. Hernquist, *Astrophys. J.* **356**, 359 (1990).
- E. Tempel, E. Tago, L. J. Liivamägi, *Astron. Astrophys.* **540**, A106 (2012).
- N. Kaiser, *Mon. Not. R. Astron. Soc.* **227**, 1–21 (1987).

- A. Diaferio, M. J. Geller, *Astrophys. J.* **481**, 633–643 (1997).
- A. Diaferio, *Mon. Not. R. Astron. Soc.* **309**, 610–622 (1999).
- R. G. Carlberg, H. K. C. Yee, E. Ellingson, *Astrophys. J.* **478**, 462–475 (1997).
- A. Biviano, M. Girardi, *Astrophys. J.* **585**, 205–214 (2003).
- J. Pfeffer, H. Baumgardt, *Mon. Not. R. Astron. Soc.* **433**, 1997–2005 (2013).
- W. R. Brown, M. J. Geller, S. J. Kenyon, M. J. Kurtz, *Astrophys. J.* **622**, L33–L36 (2005).
- L. V. Sales, J. F. Navarro, M. G. Abadi, M. Steinmetz, *Mon. Not. R. Astron. Soc.* **379**, 1475–1483 (2007).
- A. R. Wetzel, J. L. Tinker, C. Conroy, F. C. Bosch, *Mon. Not. R. Astron. Soc.* **439**, 2687–2700 (2014).
- G. De Lucia, J. Blaizot, *Mon. Not. R. Astron. Soc.* **375**, 2–14 (2007).
- N. Caldwell et al., *Astrophys. J.* **787**, L11 (2014).
- M. Geha, M. R. Blanton, R. Yan, J. L. Tinker, *Astrophys. J.* **757**, 85 (2012).

ACKNOWLEDGMENTS

The authors are grateful to F. Combes (Observatoire de Paris), I. Katkov (Sternberg Astronomical Institute), and M. Kurtz (Smithsonian Astrophysical Observatory) for useful discussions and critical reading of the manuscript. This result emerged from the tutorial run by the authors at the Astronomical Data Analysis Software and Systems conference in 2012. The authors acknowledge support by the Russian Science Foundation project 14-22-00041 “VOLGA—A View On the

Life of Galaxies.” The project used computational resources funded by the M. V. Lomonosov Moscow State University Program of Development. This research has made use of Aladin developed by the Centre de Données Astronomiques de Strasbourg; TOPCAT and STILTS software packages developed by M. Taylor; “exploreedsds” script by G. Mamou; the VizieR catalog access tool, CDS, Strasbourg, France; and the NASA/IPAC NED, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, NASA, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is www.sdss.org. GALEX and SDSS databases used in our study are available via the CasJobs web-site http://skyserver.sdss.org/CasJobs.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/418/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S4
References (33–46)
Table S1

19 November 2014; accepted 20 March 2015
10.1126/science.aaa3344

ORGANIC CHEMISTRY

A rhodium catalyst for single-step styrene production from benzene and ethylene

Benjamin A. Vaughan,¹ Michael S. Webster-Gardiner,¹ Thomas R. Cundari,^{2*} T. Brent Gunnoe^{1*}

Rising global demand for fossil resources has prompted a renewed interest in catalyst technologies that increase the efficiency of conversion of hydrocarbons from petroleum and natural gas to higher-value materials. Styrene is currently produced from benzene and ethylene through the intermediacy of ethylbenzene, which must be dehydrogenated in a separate step. The direct oxidative conversion of benzene and ethylene to styrene could provide a more efficient route, but achieving high selectivity and yield for this reaction has been challenging. Here, we report that the Rh catalyst (^{F1}DAB)Rh(TFA)(η²-C₂H₄) [^{F1}DAB is *N,N'*-bis(pentafluorophenyl)-2,3-dimethyl-1,4-diaza-1,3-butadiene; TFA is trifluoroacetate] converts benzene, ethylene, and Cu(II) acetate to styrene, Cu(I) acetate, and acetic acid with 100% selectivity and yields ≥95%. Turnover numbers >800 have been demonstrated, with catalyst stability up to 96 hours.

Vinyl arenes are important precursors for fine chemical synthesis, as well as for the preparation of plastics and elastomers (1–5). For example, styrene is produced globally on a scale of ~18.5 million tons (2). Current methods for the large-scale production of vinyl arenes involve multiple steps, typically beginning with arene alkylation using a Friedel-Crafts (e.g., AlCl₃ with HF) or zeolite catalyst followed by energy-intensive dehydrogenation of the alkyl group (Fig. 1) (1–6). Friedel-Crafts catalysis suffers from the use of harsh acids, including HF, low selectivity for the monoalkylated

product (polyalkylation is inherent to the mechanism), and the generation of stoichiometric waste (2). Zeolite catalysts have improved the process for benzene alkylation, yet these catalysts still require high temperatures (generally 350° to 450°C) and give polyalkylated products (2, 7–10).

An alternative method for the production of vinyl arenes is a direct and single-step oxidative arene vinylation (Fig. 1). If the terminal oxidant is oxygen from air (either introduced in situ or used to recycle a different in situ oxidant), the net reaction is the conversion of benzene, ethylene, and oxidant to styrene and water (11). Acid-based (i.e., Friedel-Crafts or zeolite catalysts) catalysis occurs by electrophilic aromatic substitution and does not offer a viable pathway to directly generate vinyl arenes. Transition metal complexes that catalyze ethylene hydrophenylation by benzene

¹Department of Chemistry, University of Virginia, Charlottesville, VA 22903, USA. ²Center for Advanced Scientific Computing and Modeling, Department of Chemistry, University of North Texas, Denton, TX 76203, USA.

*Corresponding author. E-mail: tbg7h@virginia.edu (T.B.G.), t@unt.edu (T.R.C.)

C–H activation followed by ethylene insertion into a metal–phenyl bond have been reported as alternatives to acid-based catalysts (Fig. 2) (12–25). For these catalysts, β -hydride elimination from a M–CH₂CH₂Ph intermediate and dissociation of styrene provides a route for the direct oxidative vinylylation of benzene (Fig. 2).

Previously, our groups have studied the use of platinum(II) catalysts for the hydrophenylation of ethylene to produce ethylbenzene (16–19, 26–28). Through a combination of experimental and computational mechanistic studies, we discerned a competing β -hydride elimination pathway from Pt–CH₂CH₂Ph intermediates to form a Pt–styrene hydride complex, which can lead to the formation of free styrene (28). Unfortunately, the formation

of styrene leads to catalyst decomposition (27). We proposed that this catalyst decomposition is the result of unstable Pt(II)–hydride complexes, which are formed from β -hydride elimination, that react to release H₂ and produce metallic Pt. The thermodynamic driving force for the formation of Pt(s) presents a substantial challenge to achieving long-lived vinyl arene production with these catalysts (Fig. 2, inset) (27). Given that the formation and decomposition of Pt(II)–H species is problematic, we sought to design catalysts using isoelectronic Rh(I) in anticipation that Rh(I)–H would exhibit greater stability compared with related Pt species (Fig. 2, inset) (27).

Figure 2 shows a targeted catalytic cycle for the direct oxidative vinylylation of benzene to pro-

duce styrene. Despite precedent for the key steps in this catalytic cycle, designing a selective catalyst represents a substantial challenge because many competing side reactions (shown in red) are likely to have activation barriers that are similar to or lower than those of the reactions along the desired catalytic cycle. In addition to these possible side reactions, designing a molecular catalyst that achieves high turnover numbers (TONs) is difficult because the oxidative conditions and the presence of potentially reactive metal–hydride intermediates could be anticipated to result in catalyst decomposition.

Table 1 compares previously reported homogeneous catalysts for direct oxidative styrene synthesis from ethylene and benzene (29–34). Generally, all suffer from one or more of the following drawbacks: low selectivity, low yield, low TON, and/or use of oxidants that cannot be regenerated using oxygen. Notable results include the work of Hong and co-workers, who reported a Rh₄(CO)₁₂ catalyst that gave, to our knowledge, the highest TON of styrene (472). In tandem with this process, liberated dihydrogen is consumed by two equivalents of ethylene and one equivalent of CO to produce 3-pentanone with 809 turnovers (TOs) (29). Sanford and co-workers reported that (3,5-dichloropyridyl)Pd(OAc)₂ catalyzes styrene production with 100% selectivity and 6.6 TOs for styrene (33% overall yield) using PhCO₂tBu, an oxidant that cannot be recycled with oxygen (33). Here, we report a rhodium catalyst for the selective one-step production of styrene from benzene, ethylene, and Cu(II) salts. We chose a Cu(II) salt as the in situ oxidant because of industrial precedent for recycling reduced Cu(I) using oxygen. In the commercial Wacker–Hoechst process for

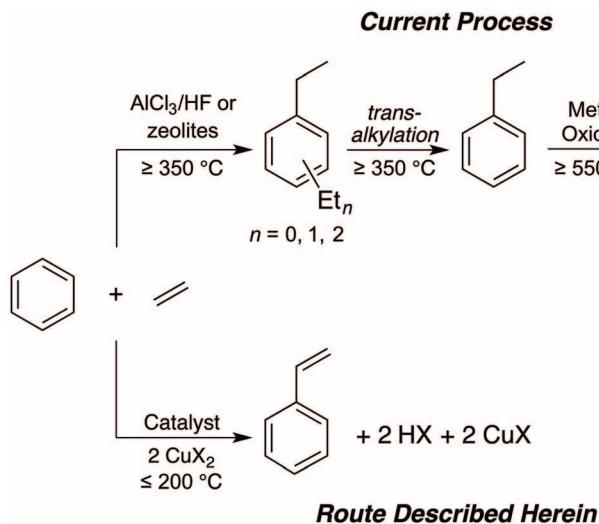


Fig. 1. Comparison of the current route to styrene production and the single-step route described herein.

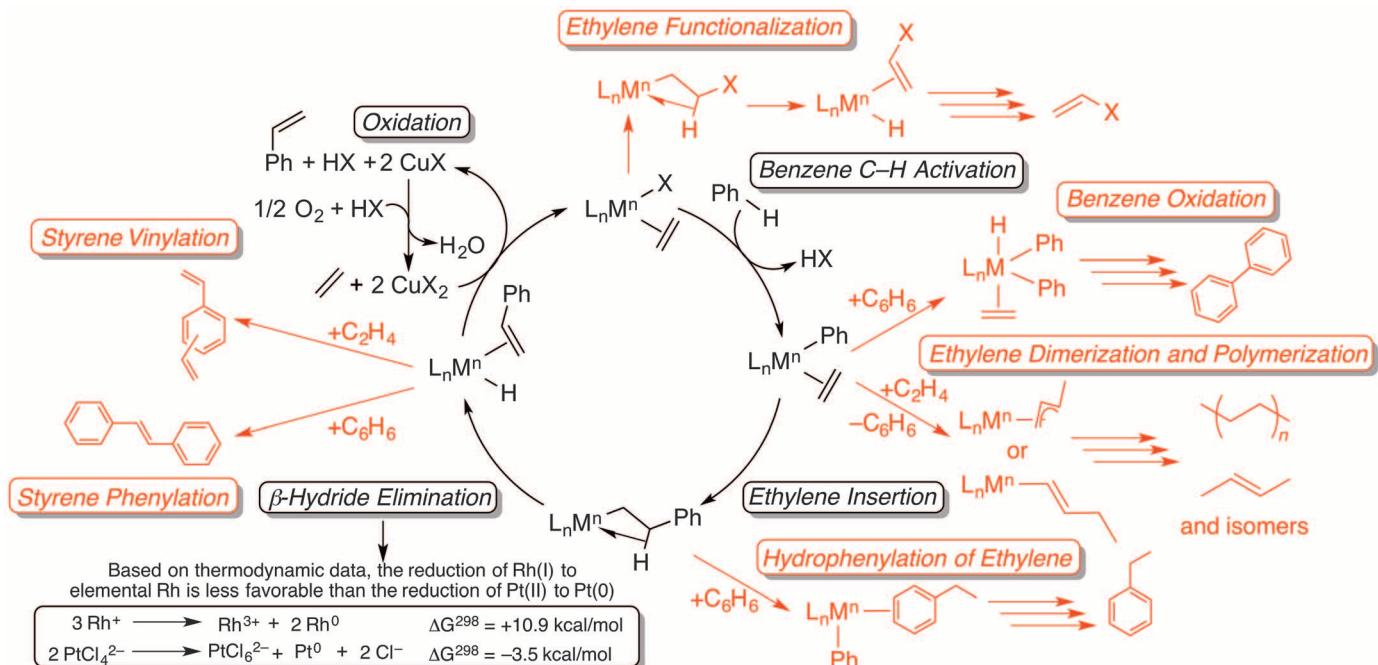


Fig. 2. Proposed cycle for transition metal–catalyzed styrene production from benzene and ethylene using CuX₂ as an oxidant. The cuprous (CuX) product could be recycled back to the cupric state using air, as shown at the upper left. Potential side reactions that a selective catalyst must avoid are shown in red.

ethylene oxidation (35, 36), use of oxygen to re-oxidize Cu(I) to Cu(II) has proven viable both in situ and in a second step (37).

We recently reported the synthesis of an electron-deficient Rh(I) complex (^{19}F DAB)Rh(TFA)(η^2 -COE) [^{19}F DAB is *N,N*-bis(pentafluorophenyl)-2,3-dimethyl-1,4-diaza-1,3-butadiene; TFA is trifluoroacetate; COE is cyclooctene] and demonstrated that this complex is an active catalyst for arene H/D exchange in trifluoroacetic acid (38). Given that arene

C-H activation is a key step in transition metal-catalyzed oxidative arene vinylation, we hypothesized that this Rh(I) complex might be an effective catalyst precursor for styrene production. Because the COE ligand would likely exchange for ethylene, the ethylene analog (^{19}F DAB)Rh(TFA)(η^2 -C₂H₄) (**1**) was independently synthesized as our catalyst precursor (Fig. 3).

Heating a 20-mL benzene solution of **1** [0.001 mole percent (mol %) relative to benzene] with

Table 1. Comparison of previously reported catalysts for styrene production. Selectivity is defined as the ratio of turnovers of styrene to total turnovers (all products) and is given as a percentage. Yield of styrene is reported relative to the limiting reagent. acac, acetylacetone; DBM, dibenzoylmethane; DCP, 3,5-dichloropyridine; HPA, H₃PMo₁₂O₄₀ · 30H₂O; TFA, trifluoroacetate.

Catalyst	Oxidant	TON	Selectivity (%)	Yield (%)
Rh ₄ (CO) ₁₂ (29)	C ₂ H ₄ /CO	472	37	19
(acac) ₂ Rh(Cl)(H ₂ O) (30)	Cu(OAc) ₂	24	89	36
Rh(PMe ₃) ₂ (CO)(Cl) (34)	hv	3	38	18
Pd(OAc) ₂ (31)	AgOAc	0.59	44	12
Pd(OAc) ₂ (30)	Cu(OAc) ₂ /O ₂	19	29	5
(DBM)Pd(OAc) ₂ (32)	HPA	100	58	2
(3,5-DCP)Pd(OAc) ₂ (33)	PhCO ₃ ^t Bu	6.6	100	33
(^{19}F DAB)Rh(TFA)(C ₂ H ₄)	Cu(OAc) ₂	115	100	96
(^{19}F DAB)Rh(TFA)(C ₂ H ₄)	Cu(OAc) ₂	835	100	70

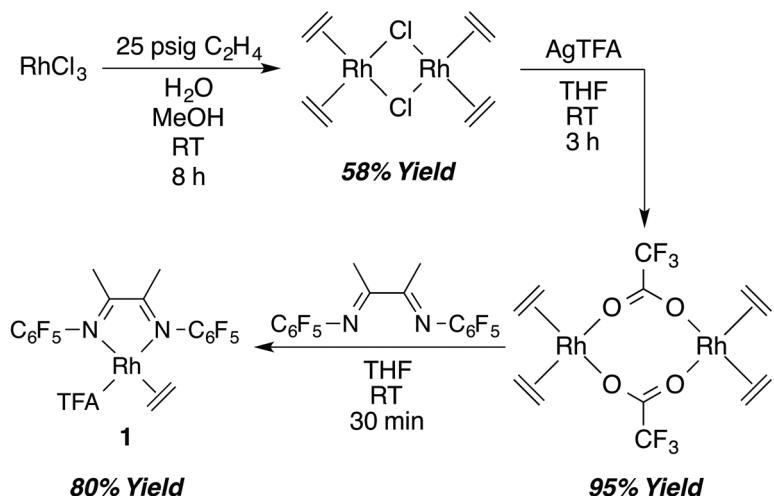
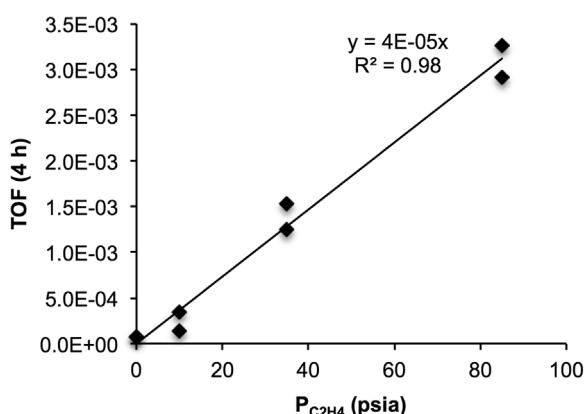


Fig. 3. Synthesis of (^{19}F DAB)Rh(TFA)(η^2 -C₂H₄) (**1**) (RT, room temperature).

Fig. 4. Effect of ethylene pressure on catalysis with (^{19}F DAB)Rh(TFA)(η^2 -C₂H₄) (**1**). Reaction conditions: 0.001 mol % **1**, 120 equivalents Cu(OAc)₂, 150°C, 4 hours. Data for two independent reactions are shown.



ethylene and Cu(OAc)₂ (120 equivalents relative to **1**) to 150°C affords 58 to 62 TOs of styrene after 24 hours (for all TOs reported, two runs were performed, and both results are given). Samples of the reaction mixture were analyzed by gas chromatography-flame ionization detector (GC/FID) using relative peak areas versus an internal standard (decane). This corresponds to quantitative yield based on the Cu(II) limiting reagent. The calculated yield here assumes that two equivalents of Cu(II) are consumed to produce two equivalents of Cu(I) per equivalent of styrene. No other products were observed upon analysis of the reaction mixture by GC-mass spectrometry or GC/FID, indicating high selectivity for styrene production. Detection limits for the instruments were equivalent to ~1 TO of product. Specifically, we looked for evidence of stilbene, biphenyl, and vinyl acetate production, because these are the most commonly observed by-products in previously reported catalysis (Table 1). Control reactions with [Rh(μ -TFA)(η^2 -C₂H₄)₂]₂, a precursor to complex **1**, afforded <5 TOs of styrene after 24 hours, with or without Cu(OAc)₂, highlighting the importance of the ^{19}F DAB ligand. Control reactions with Cu(OAc)₂ alone also afforded no styrene formation.

With a competent catalyst in hand, we next sought to optimize reaction conditions. The effect of oxidant identity on catalysis with **1** was the first parameter investigated. Both soluble {copper 2-ethylhexanoate [Cu(OHex)₂] and copper pivalate [Cu(OPiv)₂] and insoluble {copper acetate [Cu(OAc)₂] and copper trifluoroacetate hydrate [Cu(TFA)₂] Cu(II) salts were screened. Figure S2 shows a plot of turnovers versus time for the various Cu(II) oxidants. Using a turnover frequency (TOF) calculated after 4 hours of reaction, soluble Cu(OHex)₂ gives the fastest initial rate with a TOF of $2.8 \times 10^{-3} \text{ s}^{-1}$, but the reaction does not reach 100% yield relative to oxidant until 28 hours, which may indicate that catalyst deactivation occurs. Cu(OAc)₂ affords a slower initial rate than Cu(OHex)₂, with a TOF of $2.8 \times 10^{-4} \text{ s}^{-1}$ after 4 hours, but this oxidant provides a more stable catalytic process. Both Cu(TFA)₂ and Cu(OPiv)₂ afford slower initial rates; reactions with Cu(OPiv)₂ reach 92% yield after 28 hours, whereas reactions with Cu(TFA)₂ produce only 19 TOs of styrene (32% yield) after 20 hours.

To study catalyst longevity, we varied the amount of Cu(OAc)₂. Between 60 and 240 equivalents (relative to **1**), the yield of styrene relative to oxidant was always >95% (fig. S3). These near-quantitative yields demonstrate that the catalytic process using **1** as a precursor is stable and long-lived. For a reaction using 0.0001 mol % **1** and 2400 equivalents of Cu(OAc)₂, the catalyst remained active over a period of 96 hours and afforded a TON of 817 to 852. A plot of TOF versus time shows that the Rh catalyst is stable through at least 96 hours (fig. S1). The tolerance of **1** to a large excess of oxidant without any decrease in activity is promising. The effect of temperature on catalysis was also examined (fig. S4). Generally, the rate of reaction increased with temperature; however, at 180°C, rapid catalyst deactivation led to a low TON (<10 TOs) after 12 hours. Minimal activity (<1 TO) was

also observed at temperatures <100°C. The optimal temperature proved to be 150°C.

We also observed that the reaction rate increased with increasing ethylene pressure. To determine the TOF, we measured TO after 4 hours of reaction. Figure 4 shows a plot of TOF versus ethylene pressure. A linear correlation is observed. Thus, the reaction rate appears to have a first-order dependence on ethylene concentration. This is in contrast to previously reported Pt(II) and Ru(II) catalysts for the hydrophenylation of ethylene, which show an inverse dependence on ethylene pressure (14, 17). For the Pt and Ru catalysts, $M(CH_2CH_2Ph)(\eta^2-C_2H_4)$ complexes were identified as the likely catalyst resting states. The opposite dependence on ethylene pressure using **1** as catalyst precursor signals a likely change in the catalyst resting state.

To gain further insight into the reaction mechanism, we ran the reaction in a 1:1 molar mixture of C_6H_6 and C_6D_6 . After 1 hour, a k_H/k_D (ratio of rate of reaction of protio-benzene and perdeutero-benzene) of 3.1(2) was determined by examining the ratio of undeuterated styrene [mass/charge ratio (m/z) = 104] to styrene- d_5 (m/z = 109) in the mass spectra from three independent experiments (fig. S5). After 2 hours, the observed isotope effect was 3.0(2), statistically equivalent to the data at 1 hour (fig. S5). Thus, the observed k_H/k_D of ~3.1 likely reflects a kinetic isotope effect (KIE) for the catalytic cycle. The KIE is consistent with other transition metal-mediated C–H activation reactions. (39, 40) The primary KIE supports the hypothesis that a Rh catalyst is facilitating a metal-mediated C–H activation process, which occurs before or during the turnover-limiting step. No change in the isotopic distribution for benzene was observed over the course of the reaction, and no styrene- d_{6-8} products were observed except those predicted by the natural abundance of deuterium in ethylene.

Although more detailed studies are required to understand the reactivity profile of **1**, we believe that the highly electron-withdrawing perfluorophenyl groups on the ^{Fl}DAB ligand help suppress irreversible oxidation to inactive Rh(III) in the presence of Cu(II), possibly facilitate associative ligand exchange between free ethylene and coordinated styrene, and facilitate rapid ethylene insertion into Rh–Ph bonds. Challenges that remain for the continued development of this class of catalyst include increasing activity with the aim of achieving higher conversions of benzene.

REFERENCES AND NOTES

- G. A. Olah, Á. Molnár, *Hydrocarbon Chemistry* (Wiley, Hoboken, NJ, ed. 2, 2003).
- C. Perego, P. Pollesel, in *Advances in Nanoporous Materials*, E. Stefan, Ed. (Elsevier, Oxford, 2010), vol. 1, pp. 97–149.
- S.-S. Chen, in *Kirk-Othmer Encyclopedia of Chemical Technology*, A. Seidel, M. Bickford, Eds. (Wiley, Hoboken, NJ, 2000), pp. 325–357.
- H. A. Wittcoff, B. G. Reuben, J. S. Plotkin, in *Industrial Organic Chemicals* (Wiley, Hoboken, NJ, 2004), pp. 100–166.
- Process Evaluation/Research Planning (PERP) Program Report, Styrene/Ethylbenzene, (PERP Report 91-9, Chem Systems, Inc., New York, 1992).
- M. Lucchini, A. Galeotti, “Improved process for the dehydrogenation of alkyl-aromatic hydrocarbons for the production of vinyl-aromatic monomers” (International Patent WO2007073918A1, 2007).
- C. Perego, P. Ingallina, *Catal. Today* **73**, 3–22 (2002).
- C. Perego, P. Ingallina, *Green Chem.* **6**, 274 (2004).
- J. Čejka, B. Wichterlová, *Catal. Rev.* **44**, 375–421 (2002).
- I. M. Gerzeliev, S. N. Khadzhiev, I. E. Sakharova, *Petrol. Chem.* **51**, 39–48 (2011).
- CRC Handbook of Chemistry and Physics* (CRC Press, Boca Raton, FL, 1977).
- M. Lail, B. N. Arrowood, T. B. Gunnoe, *J. Am. Chem. Soc.* **125**, 7506–7507 (2003).
- M. Lail et al., *Organometallics* **23**, 5007–5020 (2004).
- N. A. Foley, J. P. Lee, Z. Ke, T. B. Gunnoe, T. R. Cundari, *Acc. Chem. Res.* **42**, 585–597 (2009).
- E. E. Joslin et al., *Organometallics* **31**, 6851–6860 (2012).
- J. R. Andreatta, B. A. McKeown, T. B. Gunnoe, *J. Organomet. Chem.* **696**, 305–315 (2011).
- B. A. McKeown et al., *J. Am. Chem. Soc.* **133**, 19131–19152 (2011).
- B. A. McKeown, H. E. Gonzalez, T. B. Gunnoe, T. R. Cundari, M. Sabat, *ACS Catal.* **3**, 1165–1171 (2013).
- B. A. McKeown, B. M. Prince, Z. Ramiro, T. B. Gunnoe, T. R. Cundari, *ACS Catal.* **4**, 1607–1615 (2014).
- S. A. Burgess et al., *Chem. Sci.* **5**, 4355–4366 (2014).
- W. D. Jones, J. A. Maguire, G. P. Rosini, *Inorg. Chim. Acta* **270**, 77–86 (1998).
- A. T. Luedtke, K. I. Goldberg, *Angew. Chem. Int. Ed.* **47**, 7694–7696 (2008).
- T. Matsumoto, D. J. Taube, R. A. Periana, H. Taube, H. Yoshida, *J. Am. Chem. Soc.* **122**, 7414–7415 (2000).
- T. Matsumoto, R. A. Periana, D. J. Taube, H. Yoshida, *J. Mol. Catal. A* **180**, 1–18 (2002).
- J. Oxgaard, R. P. Muller, W. A. Goddard 3rd, R. A. Periana, *J. Am. Chem. Soc.* **126**, 352–363 (2004).
- B. A. McKeown, N. A. Foley, J. P. Lee, T. B. Gunnoe, *Organometallics* **27**, 4031–4033 (2008).
- B. A. McKeown et al., *Organometallics* **32**, 3903–3913 (2013).
- B. A. McKeown et al., *Organometallics* **32**, 2857–2865 (2013).
- P. Hong, H. Yamazaki, *J. Mol. Catal.* **26**, 297–311 (1984).
- D. Taube, R. Periana, T. Matsumoto, “Oxidative coupling of olefins and aromatics using a rhodium catalyst and a copper(II) redox agent” (U.S. Patent 6127590A, 2000).
- Y. Fujiwara, I. Noritani, S. Danno, R. Asano, S. Teranishi, *J. Am. Chem. Soc.* **91**, 7166–7169 (1969).
- T. Yamada, A. Sakakura, S. Sakaguchi, Y. Obora, Y. Ishii, *New J. Chem.* **32**, 738 (2008).
- A. Kubota, M. H. Emmert, M. S. Sanford, *Org. Lett.* **14**, 1760–1763 (2012).
- K. Sasaki, T. Sakakura, Y. Tokunaga, K. Wada, M. Tanaka, *Chem. Lett.* **17**, 685–688 (1988).
- American Chemical Society, *Chem. Eng. News Archive* **39** (16), 52–55 (1961).
- W. A. Herrmann, in *Catalysis from A to Z*, B. Cornils, W. A. Herrmann, M. Muhrer, C.-H. Wong, Eds. (Wiley, Weinheim, Germany, 2007), pp. 1512–1524.
- M. Eckert, G. Fleischmann, R. Jira, H. M. Bolt, K. Golka, in *Ullmann's Encyclopedia of Industrial Chemistry* (Wiley-VCH Verlag, Weinheim, Germany, 2000), pp. 1–17.
- M. S. Webster-Gardiner et al., *Cat. Sci. Tech.* **5**, 96–100 (2015).
- M. Gómez-Gállego, M. A. Sierra, *Chem. Rev.* **111**, 4857–4963 (2011).
- W. D. Jones, *Acc. Chem. Res.* **36**, 140–146 (2003).

ACKNOWLEDGMENTS

The authors acknowledge support from the U.S. Department of Energy, Office of Basic Energy Sciences [DE-SC0000776 (T.B.G.) and DE-FG02-03ER15387 (T.R.C.)] for studies of styrene catalysis; the Center for Catalytic Hydrocarbon Functionalization, an Energy Frontier Research Center (award DE-SC0001298), which funded the initial catalyst discovery; and an AES Corporation Graduate Fellowship in Energy Research (M.S.W.-G). The authors also thank B. McKeown, G. Fortman, S. Kalman (University of Virginia), and R. Nielsen (California Institute of Technology) for helpful discussions.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/421/suppl/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References (41–43)

3 November 2014; accepted 12 March 2015
10.1126/science.aaa2260

SELF-ASSEMBLY

Selective assemblies of giant tetrahedra via precisely controlled positional interactions

Mingjun Huang,¹ Chih-Hao Hsu,¹ Jing Wang,¹ Shan Mei,¹ Xuehui Dong,¹ Yiwen Li,¹ Mingxuan Li,¹ Hao Liu,¹ Wei Zhang,¹ Takuzo Aida,² Wen-Bin Zhang,^{3*} Kan Yue,^{1*} Stephen Z. D. Cheng^{1*}

Self-assembly of rigid building blocks with explicit shape and symmetry is substantially influenced by the geometric factors and remains largely unexplored. We report the selective assembly behaviors of a class of precisely defined, nanosized giant tetrahedra constructed by placing different polyhedral oligomeric silsesquioxane (POSS) molecular nanoparticles at the vertices of a rigid tetrahedral framework. Designed symmetry breaking of these giant tetrahedra introduces precise positional interactions and results in diverse selectively assembled, highly ordered supramolecular lattices including a Frank-Kasper A15 phase, which resembles the essential structural features of certain metal alloys but at a larger length scale. These results demonstrate the power of persistent molecular geometry with balanced enthalpy and entropy in creating thermodynamically stable supramolecular lattices with properties distinct from those of other self-assembling soft materials.

Self-assembled hierarchical structures in soft materials have been intensely studied. Among them, assemblies of building blocks with specific geometric shapes and symmetry are of particular interest. As the sim-

plest case, ordered structures constructed from packing of spherical motifs have been a classic yet dynamic research field that can be traced back to the study of metals and metal alloys. Most metal atoms, viewed as congruent spheres, typically tend

**A rhodium catalyst for single-step styrene production from benzene and ethylene**Benjamin A. Vaughan *et al.**Science* **348**, 421 (2015);

DOI: 10.1126/science.aaa2260

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 23, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/421.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2015/04/22/348.6233.421.DC1.html>

This article appears in the following subject collections:

Chemistry

<http://www.sciencemag.org/cgi/collection/chemistry>

also observed at temperatures <100°C. The optimal temperature proved to be 150°C.

We also observed that the reaction rate increased with increasing ethylene pressure. To determine the TOF, we measured TO after 4 hours of reaction. Figure 4 shows a plot of TOF versus ethylene pressure. A linear correlation is observed. Thus, the reaction rate appears to have a first-order dependence on ethylene concentration. This is in contrast to previously reported Pt(II) and Ru(II) catalysts for the hydrophenylation of ethylene, which show an inverse dependence on ethylene pressure (14, 17). For the Pt and Ru catalysts, $M(CH_2CH_2Ph)(\eta^2-C_2H_4)$ complexes were identified as the likely catalyst resting states. The opposite dependence on ethylene pressure using **1** as catalyst precursor signals a likely change in the catalyst resting state.

To gain further insight into the reaction mechanism, we ran the reaction in a 1:1 molar mixture of C_6H_6 and C_6D_6 . After 1 hour, a k_H/k_D (ratio of rate of reaction of protio-benzene and perdeutero-benzene) of 3.1(2) was determined by examining the ratio of undeuterated styrene [mass/charge ratio (m/z) = 104] to styrene- d_5 (m/z = 109) in the mass spectra from three independent experiments (fig. S5). After 2 hours, the observed isotope effect was 3.0(2), statistically equivalent to the data at 1 hour (fig. S5). Thus, the observed k_H/k_D of ~3.1 likely reflects a kinetic isotope effect (KIE) for the catalytic cycle. The KIE is consistent with other transition metal-mediated C–H activation reactions. (39, 40) The primary KIE supports the hypothesis that a Rh catalyst is facilitating a metal-mediated C–H activation process, which occurs before or during the turnover-limiting step. No change in the isotopic distribution for benzene was observed over the course of the reaction, and no styrene- d_{6-8} products were observed except those predicted by the natural abundance of deuterium in ethylene.

Although more detailed studies are required to understand the reactivity profile of **1**, we believe that the highly electron-withdrawing perfluorophenyl groups on the ^{Fl}DAB ligand help suppress irreversible oxidation to inactive Rh(III) in the presence of Cu(II), possibly facilitate associative ligand exchange between free ethylene and coordinated styrene, and facilitate rapid ethylene insertion into Rh–Ph bonds. Challenges that remain for the continued development of this class of catalyst include increasing activity with the aim of achieving higher conversions of benzene.

REFERENCES AND NOTES

- G. A. Olah, Á. Molnár, *Hydrocarbon Chemistry* (Wiley, Hoboken, NJ, ed. 2, 2003).
- C. Perego, P. Pollesel, in *Advances in Nanoporous Materials*, E. Stefan, Ed. (Elsevier, Oxford, 2010), vol. 1, pp. 97–149.
- S.-S. Chen, in *Kirk-Othmer Encyclopedia of Chemical Technology*, A. Seidel, M. Bickford, Eds. (Wiley, Hoboken, NJ, 2000), pp. 325–357.
- H. A. Wittcoff, B. G. Reuben, J. S. Plotkin, in *Industrial Organic Chemicals* (Wiley, Hoboken, NJ, 2004), pp. 100–166.
- Process Evaluation/Research Planning (PERP) Program Report, Styrene/Ethylbenzene, (PERP Report 91-9, Chem Systems, Inc., New York, 1992).
- M. Lucchini, A. Galeotti, “Improved process for the dehydrogenation of alkyl-aromatic hydrocarbons for the production of vinyl-aromatic monomers” (International Patent WO2007073918A1, 2007).
- C. Perego, P. Ingallina, *Catal. Today* **73**, 3–22 (2002).
- C. Perego, P. Ingallina, *Green Chem.* **6**, 274 (2004).
- J. Čejka, B. Wichterlová, *Catal. Rev.* **44**, 375–421 (2002).
- I. M. Gerzeliev, S. N. Khadzhiev, I. E. Sakharova, *Petrol. Chem.* **51**, 39–48 (2011).
- CRC Handbook of Chemistry and Physics* (CRC Press, Boca Raton, FL, 1977).
- M. Lail, B. N. Arrowood, T. B. Gunnoe, *J. Am. Chem. Soc.* **125**, 7506–7507 (2003).
- M. Lail et al., *Organometallics* **23**, 5007–5020 (2004).
- N. A. Foley, J. P. Lee, Z. Ke, T. B. Gunnoe, T. R. Cundari, *Acc. Chem. Res.* **42**, 585–597 (2009).
- E. E. Joslin et al., *Organometallics* **31**, 6851–6860 (2012).
- J. R. Andreatta, B. A. McKeown, T. B. Gunnoe, *J. Organomet. Chem.* **696**, 305–315 (2011).
- B. A. McKeown et al., *J. Am. Chem. Soc.* **133**, 19131–19152 (2011).
- B. A. McKeown, H. E. Gonzalez, T. B. Gunnoe, T. R. Cundari, M. Sabat, *ACS Catal.* **3**, 1165–1171 (2013).
- B. A. McKeown, B. M. Prince, Z. Ramiro, T. B. Gunnoe, T. R. Cundari, *ACS Catal.* **4**, 1607–1615 (2014).
- S. A. Burgess et al., *Chem. Sci.* **5**, 4355–4366 (2014).
- W. D. Jones, J. A. Maguire, G. P. Rosini, *Inorg. Chim. Acta* **270**, 77–86 (1998).
- A. T. Luedtke, K. I. Goldberg, *Angew. Chem. Int. Ed.* **47**, 7694–7696 (2008).
- T. Matsumoto, D. J. Taube, R. A. Periana, H. Taube, H. Yoshida, *J. Am. Chem. Soc.* **122**, 7414–7415 (2000).
- T. Matsumoto, R. A. Periana, D. J. Taube, H. Yoshida, *J. Mol. Catal. A* **180**, 1–18 (2002).
- J. Oxgaard, R. P. Muller, W. A. Goddard 3rd, R. A. Periana, *J. Am. Chem. Soc.* **126**, 352–363 (2004).
- B. A. McKeown, N. A. Foley, J. P. Lee, T. B. Gunnoe, *Organometallics* **27**, 4031–4033 (2008).
- B. A. McKeown et al., *Organometallics* **32**, 3903–3913 (2013).
- B. A. McKeown et al., *Organometallics* **32**, 2857–2865 (2013).
- P. Hong, H. Yamazaki, *J. Mol. Catal.* **26**, 297–311 (1984).
- D. Taube, R. Periana, T. Matsumoto, “Oxidative coupling of olefins and aromatics using a rhodium catalyst and a copper(II) redox agent” (U.S. Patent 6127590A, 2000).
- Y. Fujiwara, I. Noritani, S. Danno, R. Asano, S. Teranishi, *J. Am. Chem. Soc.* **91**, 7166–7169 (1969).
- T. Yamada, A. Sakakura, S. Sakaguchi, Y. Obora, Y. Ishii, *New J. Chem.* **32**, 738 (2008).
- A. Kubota, M. H. Emmert, M. S. Sanford, *Org. Lett.* **14**, 1760–1763 (2012).
- K. Sasaki, T. Sakakura, Y. Tokunaga, K. Wada, M. Tanaka, *Chem. Lett.* **17**, 685–688 (1988).
- American Chemical Society, *Chem. Eng. News Archive* **39** (16), 52–55 (1961).
- W. A. Herrmann, in *Catalysis from A to Z*, B. Cornils, W. A. Herrmann, M. Muhrer, C.-H. Wong, Eds. (Wiley, Weinheim, Germany, 2007), pp. 1512–1524.
- M. Eckert, G. Fleischmann, R. Jira, H. M. Bolt, K. Golka, in *Ullmann's Encyclopedia of Industrial Chemistry* (Wiley-VCH Verlag, Weinheim, Germany, 2000), pp. 1–17.
- M. S. Webster-Gardiner et al., *Cat. Sci. Tech.* **5**, 96–100 (2015).
- M. Gómez-Gállego, M. A. Sierra, *Chem. Rev.* **111**, 4857–4963 (2011).
- W. D. Jones, *Acc. Chem. Res.* **36**, 140–146 (2003).

ACKNOWLEDGMENTS

The authors acknowledge support from the U.S. Department of Energy, Office of Basic Energy Sciences [DE-SC0000776 (T.B.G.) and DE-FG02-03ER15387 (T.R.C.)] for studies of styrene catalysis; the Center for Catalytic Hydrocarbon Functionalization, an Energy Frontier Research Center (award DE-SC0001298), which funded the initial catalyst discovery; and an AES Corporation Graduate Fellowship in Energy Research (M.S.W.-G.). The authors also thank B. McKeown, G. Fortman, S. Kalman (University of Virginia), and R. Nielsen (California Institute of Technology) for helpful discussions.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/421/suppl/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References (41–43)
3 November 2014; accepted 12 March 2015
10.1126/science.aaa2260

SELF-ASSEMBLY

Selective assemblies of giant tetrahedra via precisely controlled positional interactions

Mingjun Huang,¹ Chih-Hao Hsu,¹ Jing Wang,¹ Shan Mei,¹ Xuehui Dong,¹ Yiwen Li,¹ Mingxuan Li,¹ Hao Liu,¹ Wei Zhang,¹ Takuzo Aida,² Wen-Bin Zhang,^{3*} Kan Yue,^{1*} Stephen Z. D. Cheng^{1*}

Self-assembly of rigid building blocks with explicit shape and symmetry is substantially influenced by the geometric factors and remains largely unexplored. We report the selective assembly behaviors of a class of precisely defined, nanosized giant tetrahedra constructed by placing different polyhedral oligomeric silsesquioxane (POSS) molecular nanoparticles at the vertices of a rigid tetrahedral framework. Designed symmetry breaking of these giant tetrahedra introduces precise positional interactions and results in diverse selectively assembled, highly ordered supramolecular lattices including a Frank-Kasper A15 phase, which resembles the essential structural features of certain metal alloys but at a larger length scale. These results demonstrate the power of persistent molecular geometry with balanced enthalpy and entropy in creating thermodynamically stable supramolecular lattices with properties distinct from those of other self-assembling soft materials.

Self-assembled hierarchical structures in soft materials have been intensely studied. Among them, assemblies of building blocks with specific geometric shapes and symmetry are of particular interest. As the sim-

plest case, ordered structures constructed from packing of spherical motifs have been a classic yet dynamic research field that can be traced back to the study of metals and metal alloys. Most metal atoms, viewed as congruent spheres, typically tend

to hold 12 neighbors (the coordination number, CN, is thus 12) in local environments, forming the most efficient packing scheme with tetrahedral interstices (1). This type of structure allows three possible variations: face-centered cubic (the cuboctahedron), hexagonal close-packed (the twinned cuboctahedron), and the topologically close-packed icosahedron (or “icosahedral coordination”).

In metal alloys, different metal atoms with various sizes and electronic states are involved. Frank and Kasper (2) studied the stability of icosahedral lattices and proved that distorted icosahedra could be accommodated with topologically close-packed Kasper polyhedra, which allow even higher coordination numbers (CN = 14, 15, and 16) in metal alloy crystals. This class of metal alloy crystal structures is referred as the “Frank-Kasper” phases, including the A15 phase (with an A_3B stoichiometry such as Cr₃Si), the Friauf-Laves phase (with an A_2B stoichiometry such as Zn₂Mg), the σ phase (with an AB stoichiometry such as CrFe), and others (3). Some Frank-Kasper phases are viewed as periodic approximates of aperiodic “quasicrystals.” Therefore, they provide a platform to understand how to fill in space with different spherical motifs and how to achieve properties related to their characteristic structural features of low lattice symmetry and high coordination numbers.

A typical cubic unit cell of the A15 phase (Fig. 1A) consists of six A units (pale red spheres) in 14-fold Kasper polyhedra and two B units (dark red spheres) in 12-fold icosahedral coordination (Fig. 1B) with a space group of $Pm\bar{3}n$ (O_h^3). The projection view along the $\langle 001 \rangle$ direction (Fig. 1C) displays a regular two-dimensional (2D) 4⁴ tiling pattern (4). Recently, examples of the A15, σ , and quasicrystalline phases constructed by nano- and micrometer-sized “deformable” spheres, micelles, and colloids were reported in many systems, including spherical dendrimers (5–8), ABC star-triblock copolymers (9), micelles of linear diblock or tetrablock copolymers in the bulk (10–12) or in solution (13), binary nanoparticle lattices (14), and mesoporous silica produced from surfactant micelles (15). In particular, formation of the A15 phase in dendrimers has been attributed to the presence of soft “squishy surface layers” composed of alkyl chains, which can deform to minimize steric interactions (5, 16) and surface contact area among the spheres (the Weaire-Phelan structure) (17, 18).

Constructing ordered phases with the use of shaped building blocks other than spheres has yet to be demonstrated. Recent computer simulation results revealed possible crystalline and liquid crystalline structures from the packing

of polyhedra (19). Among all the polyhedra, the tetrahedron is the simplest. Rigid tetrahedron building blocks have been shown to form quasi-crystalline and crystalline phases with high packing fractions (20). However, related experimental investigation remains largely unexplored in terms of both observations of ordered structures and their formation mechanisms. Shape-persistent molecular nanoparticles, such as derivatives of POSS (21), fullerenes (22), polyoxometalates (23), and proteins (24), offer great opportunities to

construct nanosized giant tetrahedra with atomic precision (25, 26).

Here, we present an experimental study of giant tetrahedra constructed by attaching four POSS cages with different functional groups to a rigid tetrahedral core (Fig. 2). They are distinguished from the reported dendrimer and block copolymer systems (12) by the absence of any flexible alkyl or polymeric chains. Self-assembly of these giant tetrahedra is mediated by interactions among the POSS nanoclusters and the

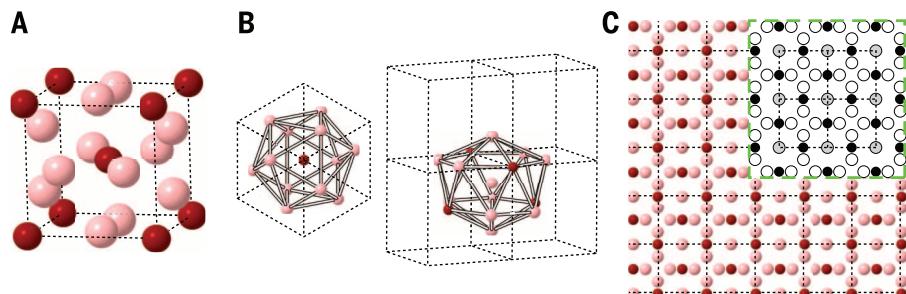


Fig. 1. Schematic illustration of the A15 phase. (A) In an A15 cubic unit cell, the dark red and pale red colors represent different coordination environments. (B) Schemes of CN = 12 and CN = 14 coordination environments in the A15 lattice. (C) 2D-projected view of the A15 lattice along the $\langle 001 \rangle$ direction. The inset shows a 2D 4⁴ tiling pattern along the z axis. The spheres at the sparse layers ($z/4$ and $3z/4$) are represented by gray circles; the spheres at the dense layers are shown by black and white circles ($z/2$ and z).

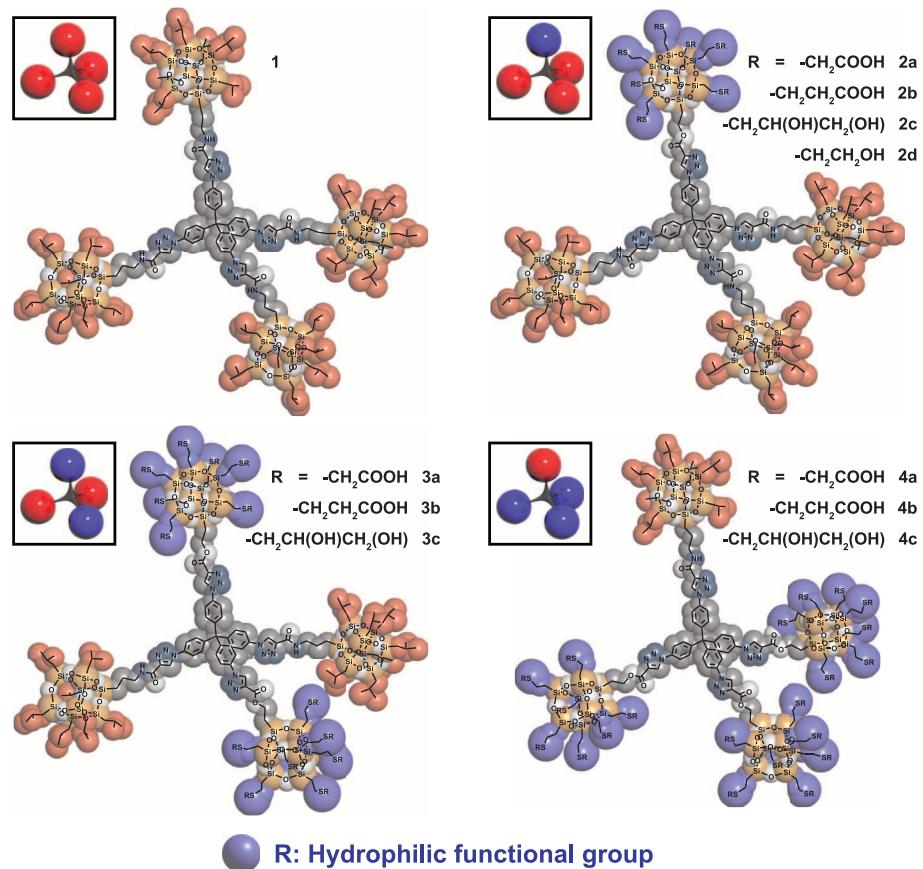


Fig. 2. Chemical structures and molecular models (shown in shadow) of the four categories of giant tetrahedra. Cartoons in the boxes are corresponding simplifications of the giant tetrahedra, in which blue spheres represent hydrophilic POSS cages and red spheres represent hydrophobic BPOSS cages.

¹Department of Polymer Science, College of Polymer Science and Polymer Engineering, University of Akron, Akron, OH 44325, USA. ²Department of Chemistry and Biotechnology, School of Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. ³Key Laboratory of Polymer Chemistry and Physics of Ministry of Education, Center for Soft Matter Science and Engineering, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, P.R. China.

*Corresponding author. E-mail: scheng@uakron.edu (S.Z.D.C.); ky13@zips.uakron.edu (K.Y.); webin@pku.edu.cn (W.-B.Z.)

overall molecular symmetry. Various ordered supramolecular lattices, including the Frank-Kasper A15 phase, are observed in this system by tuning the numbers of hydrophilic or hydrophobic POSS cages in each molecule and the functional groups on the hydrophilic POSS cages.

Giant tetrahedra **1** to **4** with different partitions of hydrophobic and hydrophilic POSS cages were synthesized by sequentially applying two “click” reactions (fig. S1): the copper-catalyzed azide-alkyne [3+2] cycloaddition reaction and the thiol-ene reaction (27, 28). The hydrophobic POSS cages have seven isobutyl groups (BPOSS) and the hydrophilic POSS cages have either hydroxyl or carboxylic acid groups (Fig. 2). Incorporation of different POSS cages results in competing interactions (i.e., collective hydrogen-bonding interactions among the hydrophilic POSS cages and the crystallization of BPOSS cages) to drive self-assembly; tuning the number of hydrophobic or hydrophilic POSS cages systematically varies molecular symmetry of the giant tetrahedra. Nuclear magnetic resonance and mass spectroscopy results (figs. S2 and S3) confirmed their structural precision and high purity. We expect that geometric and interactional factors jointly determine their self-assembly behaviors.

Giant tetrahedron **1** contains four identical BPOSS cages. A crystalline structure with a triclinic unit cell and a space group of *P*1 has been determined (fig. S4 and table S1), based on the combination of selected-area electron diffraction (SAED; fig. S4A) data from its single crystals and wide-angle x-ray diffraction (WAXD; fig. S4B) data from the bulk sample (28). In the simulated molecular packing, the tetrahedral cores adopt an interpenetrated stacking manner to form geometrically locked columns, which are surrounded by a shell of crystalline BPOSS cages (fig. S4, D and E). To maximize the contacts among the crystalline BPOSS cages, the lattice is distorted from hexagonally packed cylinders toward lower symmetry.

Replacing one BPOSS cage with a hydrophilic POSS cage in **1** lowers the molecular symmetry to *C*3v and results in giant tetrahedra **2a** to **2c**. At 25°C, density-frustrated lamellar supramolecular structures with a three-layer packing periodicity are observed in **2a** to **2c**, as supported by the combined small-angle x-ray scattering (SAXS) and WAXD results (Fig. 3A) with a scattering vector (*q*) value ratio of 1:2:3. Besides, the strongest diffraction peak at 1.09 nm in the WAXD pattern is attributed to the characteristic diffraction of crystalline BPOSS domains (29). A bright-field (BF) transmission electron microscope (TEM) image of microtomed thin-sectioned **2a** samples (Fig. 3B) and its fast Fourier transform (FFT) pattern (Fig. 3B, inset) also confirm the lamellar structure. The measured periodicities of 4.3 to 4.7 nm (Table 1) can only accommodate two layers of BPOSS and one interdigitated layer of the hydrophilic POSS cages (Fig. 3G) (estimated ~4.5 nm). Despite the unmatched numbers of hydrophobic and hydrophilic POSS cages, crystallization of BPOSS cages dominates and preferentially creates flat interfaces (30), leading to the formation of frustrated supramolecular lamellae.

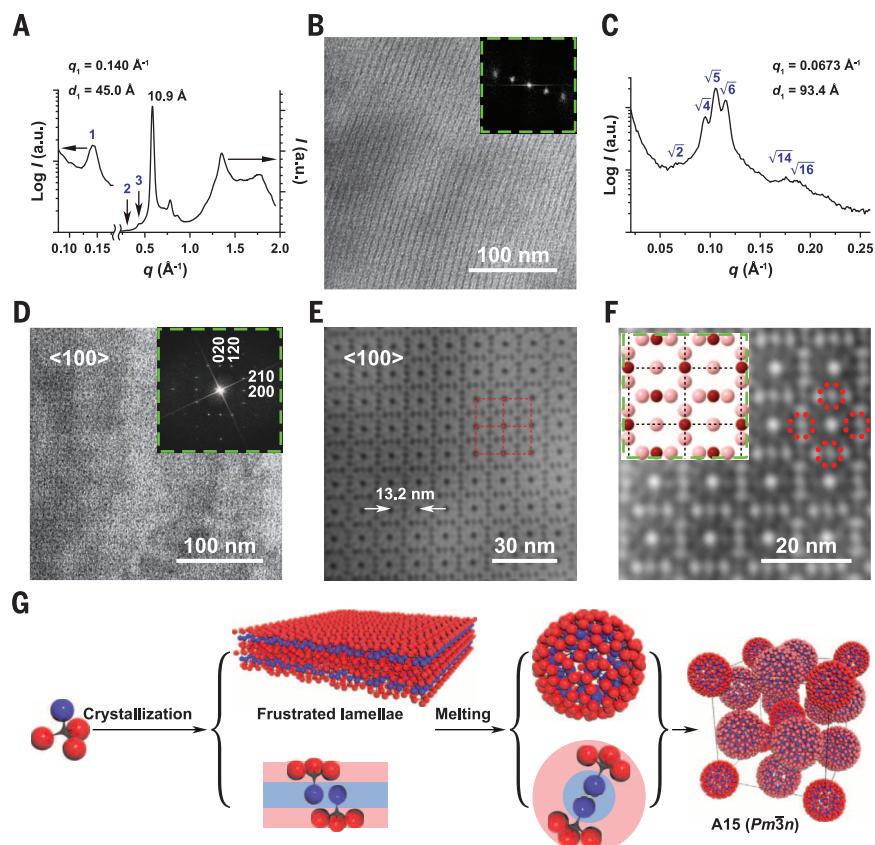


Fig. 3. Selectively assembled structures of **2a.** (A) Combined SAXS and WAXD profiles of **2a** evaporated from tetrahydrofuran-acetonitrile (THF/MeCN) mixed solvents at 25°C. (B) BF TEM image and corresponding FFT pattern (inset) of a microtomed thin-sectioned **2a** sample. (C) SAXS pattern of **2a** after the sample was heated to above its *T_m* and annealed at 140°C for 12 hours. (D) A {100} plane of an A15 supramolecular lattice was identified by the BF TEM image after the thin-sectioned sample was stained by RuO₄. The inset is the FFT pattern of this image. (E) Fourier filtering of the image shown in (D) revealed a clear view of the 2D 4⁴ tiling along the (100) direction. (F) Inverse colored and magnified image of (E). White spheres represent the hydrophilic POSS domains with different sizes. The inset shows a simulated projection view along the <100> direction. Spheres in the red-dot circles correspond to the dark red ones shown in Fig. 1A. (G) Schematic illustrations of the selective assembly mechanism and molecular packing in the A15 lattice.

Table 1. Supramolecular lattice analysis of the giant tetrahedra with different symmetry. Lattice I structures were formed by slow evaporation of the sample solutions in THF/MeCN mixed solvents at 25°C; lattice II structures were formed after annealing treatment. *d_I* is the determined periodicity of the lamellar structures.

Molecule	Lattice I	<i>T_m</i> (°C)	<i>d_I</i> (nm)	Lattice II	<i>d_{II}</i> (nm)	M
1	Triclinic crystal	138	—	Disordered	—	—
2a	Lamellae	145	4.50	A15	13.2*	38/50§
2b	Lamellae	144	4.60	A15	14.2*	46/61§
2c	Lamellae	146	4.67	A15	14.6*	49/65§
2d	Lamellae	126	4.50	bcc	8.2*	44§
3a	Lamellae	177	5.40	Double gyroid	6.35	—
3b	Lamellae	172	5.90	Lamellae	6.40†	—
3c	Lamellae	180	6.00	Lamellae	7.30†	—
4a	Disordered	—	—	Cylinder	6.22‡	6.7#
4b	Disordered	—	—	Cylinder	6.28‡	6.6#
4c	Disordered	—	—	Cylinder	6.70‡	7.4#

*Dimensions of the A15 or bcc unit cells. †Lamellar periodicities in supramolecular lattice II. ‡The corresponding principal (01) spacing between the cylinders in the honeycomb-like hexagonal lattices. §Calculated numbers of giant tetrahedra in two types of spheres in the A15 lattice (the size ratios of two types of spheres in **2b** and **2c** are assumed to be 1.1, the same as in **2a**) or in each sphere in the bcc lattice. #Average number of giant tetrahedra within 1-nm-thick cross section of the cylinders in the honeycomb-like hexagonal supramolecular lattices.

After **2a** was heated to 180°C (above its melting point T_m ; Table 1) and immediately cooled to and annealed at 140°C for 12 hours, an entirely different SAXS pattern was observed (Fig. 3C). The WAXD pattern indicates that BPOSS cages were amorphous in this structure (fig. S5A). Both **2b** and **2c** exhibited virtually identical SAXS patterns upon the same thermal treatment (fig. S5, B and C). The observed q value ratios (Fig. 2C and fig. S5, B and C) are $\sqrt{2}:\sqrt{4}:\sqrt{5}:\sqrt{6}$, which is characteristic of the A15 phase (5). A cubic unit cell with $a = 13.2$ nm can be deduced for **2a**. The lattice assignment is further validated by TEM images of the microtomed, RuO₄-stained thin-sectioned samples of **2a** (~80 nm thick). The BF TEM image in Fig. 3D exhibits the arrangement of spheres along the <100> direction of the A15 phase in real space. Its FFT pattern is shown in the inset of Fig. 3D (also in fig. S5D) with major diffractions assigned. Fourier filtering treatment provides a clear view of the regular 2D 4⁴ tiling pattern along the <100> direction (Fig. 3E). From this image, the measured distance between two closest neighboring squares is 13.2 nm, which is consistent with the value calculated from the

SAXS result. Setting Fig. 3E in inverse contrast makes it easier to identify the fine features of the spherical packing (Fig. 3F). It is observed that spheres in the red-dot circles (Fig. 3F), which correspond to the dark red spheres in Fig. 1A with CN = 12, are smaller relative to their neighbors (pale red spheres in Fig. 1A with CN = 14). On the basis of the average size ratio between these two types of spheres (1.1 ± 0.06), we estimate that these two types of spheres contain 38 and 50 giant tetrahedra, respectively (28). These results support the existence of two types of spheres with different coordination environments in a single-component system, in contrast to metal alloys with different types of atoms. Moreover, the number of giant tetrahedra in each sphere is found to increase with increasing strength of the collective hydrogen-bonding interactions and the molecular masses from **2a** to **2c** (Table 1). The formation mechanism of the A15 phase is illustrated in Fig. 3G. When the frustrated lamellar crystals melt, the hydrophilic POSS cages form spherical aggregates via collective hydrogen bonding, while BPOSS cages originally located in the neighboring top and bottom lamellar layers undergo a 2D scrolling to form the shell. The self-assembled spheres finally pack into the A15 supramolecular lattice.

Dendrimers with a poly(benzyl ether) core and a dodecyl corona are known to form spheres that further pack into A15 lattices (5, 7), which can be explained by the soft “squishy surface layers” that promote deformation of the spheres to maximize entropy and minimize interfaces (17, 18, 31, 32). The molecular geometry of giant tetrahedra **2a** to **2c** also prefers the formation of spheres in the first step. Without any flexible chains, it is proposed that extra degrees of freedom (such as the excluded volume of BPOSS cages and the nonclose packing of the hydrophilic POSS cages via hydrogen bonding) contribute to the size differentiations of the assembled spherical motifs, which entropically favor more space and looser packing to form the A15 phase. Furthermore, it is believed that the deformability is associated with the size of the spheres, because the interstitial gaps become larger as the size of spheres increases (16).

To prove this assumption, we synthesized **2d** (Fig. 2 and fig. S6A) containing a hydrophilic POSS cage with the weakest hydrogen-bonding interaction and the smallest molar mass among **2a** to **2d**. After similar thermal treatment, a body-centered cubic (bcc) lattice composed of only one type of spheres was found (fig. S6). Each sphere contains 44 giant tetrahedra **2d**. This number provides a reasonable estimation of the upper size limit of nondeformable spheres assembled from this series of giant tetrahedra, because a small fraction of the A15 phase can also be identified from the TEM image of thin-sectioned **2d** samples (fig. S6F). Any spheres larger than this size would deform as the result of nonclose packing of the hydrophilic POSS cages at the spherical center (33) to better fit into the supramolecular lattice-packing requirements with lower symmetry.

Giant tetrahedra **3a** to **3c** are more symmetric in terms of both volume fractions and interactions. “Double-layered” lamellar supramolecular lattices (30, 34) are observed for **3a** to **3c** at 25°C, due to the dominating BPOSS crystallization (fig. S7A). Layer thicknesses of these lamellar structures were determined from SAXS results (Table 1 and fig. S7B), and they match the estimated values from molecular packing models (28). Their high-temperature structures were obtained by annealing above their T_m at 180°C for 3 hours and subsequent quenching into liquid nitrogen to suppress crystallization of BPOSS cages (fig. S7C). SAXS and TEM results (Fig. 4, A and D) indicate that the lamellar structures of **3b** and **3c** remain but have increased lamellae d -spacings relative to their room-temperature structures; this is mainly attributed to the disordered BPOSS packing and thermal expansion. On the other hand, a highly ordered double-gyroid supramolecular lattice (space group $Ia\bar{3}d$) forms in **3a** after similar treatment (Fig. 4, B and E, and fig. S7D). In the TEM image (Fig. 4E), the darker regions are hydrophilic POSS domains embedded in the hydrophobic matrix composed of BPOSS cages and the tetrahedral cores. Formation of

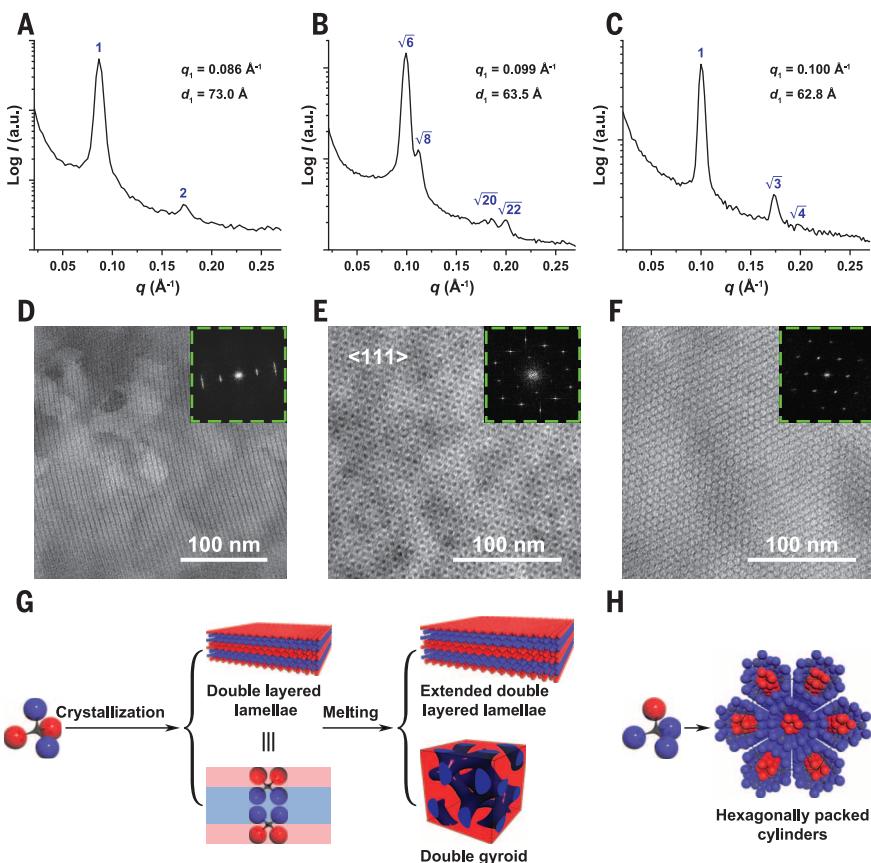


Fig. 4. Selectively assembled structures from giant tetrahedra **3 and **4**.** (A to C) SAXS patterns of **3c** (A), **3a** (B), and **4b** (C) were taken at 25°C after corresponding thermal treatments. (D) BF TEM image of thin-sectioned **3c** confirms the lamellar lattice deduced from the SAXS result shown in (A). (E) BF TEM image of thin-sectioned and RuO₄-stained **3a** confirms the double-gyroid lattice deduced from the SAXS result shown in (B). (F) BF TEM image of thin-sectioned **4b** confirms the honeycomb-like hexagonal lattice deduced from the SAXS result shown in (C). In (D) to (F), the insets are the FFT patterns of the TEM images. (G) Schematic illustration of the selective assembly mechanisms and packing models of **3a** to **3c**. (H) Schematic packing models of **4a** to **4c**.

such a double-gyroid phase from the rigid and symmetric giant tetrahedron **3a** reflects the ubiquity of the gyroid structure, implying the subtle influence of the slightly different volume fractions and interactions on the selective assembly of these giant tetrahedra (Fig. 4G).

Giant tetrahedra **4a** to **4c** failed to crystallize in similar solvent evaporation processes because of the low volume fraction of BPOSS cages that does not favor the formation of continuous 2D flat crystals (fig. S8). At such a volume fraction, an inverse spherical phase such as bcc or Al15 was expected. However, after thermal annealing at 130°C, only ordered hexagonal cylinder phases were observed in **4a** to **4c**, as revealed by the *q* value ratio of $1:\sqrt{3}:\sqrt{4}$ in their SAXS patterns (fig. S8B, Fig. 4C, and fig. S8C for **4a**, **4b**, and **4c**, respectively) and the honeycomb-like hexagonal structure observed in BF TEM images (Fig. 4F). In the proposed schematic packing model of **4a** to **4c** (Fig. 4H), BPOSS cages are wrapped into centers of the columns while hydrophilic POSS cages with strong collective hydrogen bonding form the continuous matrix. In sharp contrast to the packing of **2a** to **2c** at higher temperatures, **4a** to **4c** tend to maximize the contacts of hydrophilic POSS cages (and thus the extent of collective hydrogen-bonding formation), which substantially minimizes the overall free energy of the system.

Symmetry breaking on accurately controlled positional interactions of nanosized giant tetrahedra has been used to construct the Frank-Kasper Al15 phase and other ordered supramolecular lattices. The diverse self-assembly behaviors of these giant tetrahedra reveal that rigid, single-component soft-matter systems offer potential for building supramolecular “metal alloy analogs.” The subtle competition between the persistent molecular geometry and the deformability driven by interaction terms dictates the selective assembly of the giant tetrahedra. Because of the “click” synthesis, this system is highly tunable in terms of core structure, nanoparticle functionality, and feature size. The concepts and formation mechanisms of these supramolecular structures could be extended to other giant-polyhedra molecules with different topologies and chemical compositions.

REFERENCES AND NOTES

- M. D. Graef, M. E. McHenry, *Structure of Materials: An Introduction to Crystallography, Diffraction and Symmetry* (Cambridge Univ. Press, Cambridge, ed. 2, 2012).
- F. C. Frank, J. S. Kasper, *Acta Crystallogr.* **11**, 184–190 (1958).
- F. C. Frank, J. S. Kasper, *Acta Crystallogr.* **12**, 483–499 (1959).
- G. Ungar, X. Zeng, *Soft Matter* **1**, 95–106 (2005).
- V. S. K. Balagurusamy, G. Ungar, V. Percec, G. Johansson, *J. Am. Chem. Soc.* **119**, 1539–1555 (1997).
- X. Zeng et al., *Nature* **428**, 157–160 (2004).
- S. D. Hudson et al., *Science* **278**, 449–452 (1997).
- G. Ungar, Y. Liu, X. Zeng, V. Percec, W. D. Cho, *Science* **299**, 1208–1211 (2003).
- K. Hayashida, T. Dotera, A. Takano, Y. Matsushita, *Phys. Rev. Lett.* **98**, 195502 (2007).
- J. Zhang, F. S. Bates, *J. Am. Chem. Soc.* **134**, 7636–7639 (2012).
- S. Lee, M. J. Blumle, F. S. Bates, *Science* **330**, 349–353 (2010).
- S. Lee, C. Leighton, F. S. Bates, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 17723–17731 (2014).
- S. Fischer et al., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1810–1814 (2011).
- D. V. Talapin et al., *Nature* **461**, 964–967 (2009).
- C. Xiao, N. Fujita, K. Miyasaka, Y. Sakamoto, O. Terasaki, *Nature* **487**, 349–353 (2012).
- C. R. Iacovella, A. S. Keys, S. C. Glotzer, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20935–20940 (2011).
- P. Ziherl, R. D. Kamien, *J. Phys. Chem. B* **105**, 10147–10158 (2001).
- G. M. Grason, B. A. DiDonna, R. D. Kamien, *Phys. Rev. Lett.* **91**, 058304 (2003).
- P. F. Damasceno, M. Engel, S. C. Glotzer, *Science* **337**, 453–457 (2012).
- A. Haji-Akbari et al., *Nature* **462**, 773–777 (2009).
- D. B. Cordes, P. D. Lickiss, F. Rataboul, *Chem. Rev.* **110**, 2081–2173 (2010).
- X. Yu et al., *J. Am. Chem. Soc.* **134**, 7780–7787 (2012).
- A. Dolbecq, E. Dumas, C. R. Mayer, P. Mialane, *Chem. Rev.* **110**, 6009–6048 (2010).
- Y.-T. Lai, D. Cascio, T. O. Yeates, *Science* **336**, 1129 (2012).
- W.-B. Zhang et al., *Macromolecules* **47**, 1221–1239 (2014).
- X. Yu et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10078–10083 (2013).
- K. Yue et al., *Macromolecules* **45**, 8126–8134 (2012).
- See supplementary materials on Science Online.
- A. R. Bassindale et al., *Dalton Trans.* **2003**, 2945–2949 (2003).
- H. Liu et al., *J. Am. Chem. Soc.* **136**, 10691–10699 (2014).
- P. Ziherl, R. D. Kamien, *Phys. Rev. Lett.* **85**, 3528–3531 (2000).
- D. Weaire, R. Phelan, *Philos. Mag. Lett.* **69**, 107–110 (1994).
- V. Percec et al., *J. Am. Chem. Soc.* **130**, 13079–13094 (2008).
- Y. Li et al., *J. Am. Chem. Soc.* **133**, 10712–10715 (2011).

ACKNOWLEDGMENTS

Supported by NSF grant DMR-1408872. We thank B. Lotz for helpful discussion. The MALDI-TOF MS analysis was assisted by K. Guo, C. Shi, and C. Wesdemiotis. Use of the Advanced Photon Source at Argonne National Laboratory was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract DE-AC02-06CH11357.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/424/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
Table S1
References (35–38)

5 November 2014; accepted 2 March 2015
10.1126/science.aaa2421

METHANE CYCLING

Nonequilibrium clumped isotope signals in microbial methane

David T. Wang,^{1,2} Danielle S. Gruen,^{1,2} Barbara Sherwood Lollar,³ Kai-Uwe Hinrichs,⁴ Lucy C. Stewart,⁵ James F. Holden,⁵ Alexander N. Hristov,⁶ John W. Pohlman,⁷ Penny L. Morrill,⁸ Martin Könneke,⁴ Kyle B. Delwiche,⁹ Eoghan P. Reeves,¹ Chelsea N. Sutcliffe,³ Daniel J. Ritter,¹⁰ Jeffrey S. Seewald,² Jennifer C. McIntosh,¹⁰ Harold F. Hemond,⁹ Michael D. Kubo,¹¹ Dawn Cardace,¹² Tori M. Hoehler,¹¹ Shuhei Ono^{1*}

Methane is a key component in the global carbon cycle, with a wide range of anthropogenic and natural sources. Although isotopic compositions of methane have traditionally aided source identification, the abundance of its multiply substituted “clumped” isotopologues (for example, $^{13}\text{CH}_3\text{D}$) has recently emerged as a proxy for determining methane-formation temperatures. However, the effect of biological processes on methane’s clumped isotopologue signature is poorly constrained. We show that methanogenesis proceeding at relatively high rates in cattle, surface environments, and laboratory cultures exerts kinetic control on $^{13}\text{CH}_3\text{D}$ abundances and results in anomalously elevated formation-temperature estimates. We demonstrate quantitatively that H_2 availability accounts for this effect. Clumped methane thermometry can therefore provide constraints on the generation of methane in diverse settings, including continental serpentinization sites and ancient, deep groundwaters.

Carbon ($^{13}\text{C}/^{12}\text{C}$) and hydrogen (D/H) isotope ratios of methane are widely applied for distinguishing microbial from thermogenic methane in the environment (1–7), as well as for apportioning pathways of microbial methane production (8–10). This bulk isotope approach, however, is largely based on empirical observations, and different origins of methane often yield overlapping characteristic isotope signals (3, 7, 11–13). Beyond conventional bulk isotope ratios, it has become possible to precisely measure the abundance of multiply substituted “clumped” isotopologues (e.g., $^{13}\text{CH}_3\text{D}$) (14, 15). In particular, the abundance of clumped isotopes makes it possible to obtain information

about the temperature at which C–H bonds were formed or last equilibrated (14) (fig. S1). Formation temperatures of both thermogenic and microbial methane in natural gas reservoirs can be estimated on the basis of clumped isotopologues (16). The mechanisms by which isotopologues attain distributions consistent with thermodynamic equilibrium, however, remain unclear because bulk methane isotopes ($\delta^{13}\text{C}$ and δD) often reflect kinetic isotope fractionations (13, 17), and H isotope exchange between methane and water is sluggish (18).

To test whether clumped methane thermometry can be widely applied for methane sources beyond natural gas reservoirs, we examined



Selective assemblies of giant tetrahedra via precisely controlled positional interactions

Mingjun Huang *et al.*

Science **348**, 424 (2015);

DOI: 10.1126/science.aaa2421

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 23, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/424.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2015/04/22/348.6233.424.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/348/6233/424.full.html#related>

This article **cites 35 articles**, 9 of which can be accessed free:

<http://www.sciencemag.org/content/348/6233/424.full.html#ref-list-1>

This article has been **cited by 1** articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/348/6233/424.full.html#related-urls>

This article appears in the following **subject collections**:

Materials Science

http://www.sciencemag.org/cgi/collection/mat_sci

such a double-gyroid phase from the rigid and symmetric giant tetrahedron **3a** reflects the ubiquity of the gyroid structure, implying the subtle influence of the slightly different volume fractions and interactions on the selective assembly of these giant tetrahedra (Fig. 4G).

Giant tetrahedra **4a** to **4c** failed to crystallize in similar solvent evaporation processes because of the low volume fraction of BPOSS cages that does not favor the formation of continuous 2D flat crystals (fig. S8). At such a volume fraction, an inverse spherical phase such as bcc or Al5 was expected. However, after thermal annealing at 130°C, only ordered hexagonal cylinder phases were observed in **4a** to **4c**, as revealed by the *q* value ratio of $1:\sqrt{3}:\sqrt{4}$ in their SAXS patterns (fig. S8B, Fig. 4C, and fig. S8C for **4a**, **4b**, and **4c**, respectively) and the honeycomb-like hexagonal structure observed in BF TEM images (Fig. 4F). In the proposed schematic packing model of **4a** to **4c** (Fig. 4H), BPOSS cages are wrapped into centers of the columns while hydrophilic POSS cages with strong collective hydrogen bonding form the continuous matrix. In sharp contrast to the packing of **2a** to **2c** at higher temperatures, **4a** to **4c** tend to maximize the contacts of hydrophilic POSS cages (and thus the extent of collective hydrogen-bonding formation), which substantially minimizes the overall free energy of the system.

Symmetry breaking on accurately controlled positional interactions of nanosized giant tetrahedra has been used to construct the Frank-Kasper Al5 phase and other ordered supramolecular lattices. The diverse self-assembly behaviors of these giant tetrahedra reveal that rigid, single-component soft-matter systems offer potential for building supramolecular “metal alloy analogs.” The subtle competition between the persistent molecular geometry and the deformability driven by interaction terms dictates the selective assembly of the giant tetrahedra. Because of the “click” synthesis, this system is highly tunable in terms of core structure, nanoparticle functionality, and feature size. The concepts and formation mechanisms of these supramolecular structures could be extended to other giant-polyhedra molecules with different topologies and chemical compositions.

REFERENCES AND NOTES

- M. D. Graef, M. E. McHenry, *Structure of Materials: An Introduction to Crystallography, Diffraction and Symmetry* (Cambridge Univ. Press, Cambridge, ed. 2, 2012).
- F. C. Frank, J. S. Kasper, *Acta Crystallogr.* **11**, 184–190 (1958).
- F. C. Frank, J. S. Kasper, *Acta Crystallogr.* **12**, 483–499 (1959).
- G. Ungar, X. Zeng, *Soft Matter* **1**, 95–106 (2005).
- V. S. K. Balagurusamy, G. Ungar, V. Percec, G. Johansson, *J. Am. Chem. Soc.* **119**, 1539–1555 (1997).
- X. Zeng et al., *Nature* **428**, 157–160 (2004).
- S. D. Hudson et al., *Science* **278**, 449–452 (1997).
- G. Ungar, Y. Liu, X. Zeng, V. Percec, W. D. Cho, *Science* **299**, 1208–1211 (2003).
- K. Hayashida, T. Dotera, A. Takano, Y. Matsushita, *Phys. Rev. Lett.* **98**, 195502 (2007).
- J. Zhang, F. S. Bates, *J. Am. Chem. Soc.* **134**, 7636–7639 (2012).
- S. Lee, M. J. Blumle, F. S. Bates, *Science* **330**, 349–353 (2010).
- S. Lee, C. Leighton, F. S. Bates, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 17723–17731 (2014).
- S. Fischer et al., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1810–1814 (2011).
- D. V. Talapin et al., *Nature* **461**, 964–967 (2009).
- C. Xiao, N. Fujita, K. Miyasaka, Y. Sakamoto, O. Terasaki, *Nature* **487**, 349–353 (2012).
- C. R. Iacovella, A. S. Keys, S. C. Glotzer, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20935–20940 (2011).
- P. Ziherl, R. D. Kamien, *J. Phys. Chem. B* **105**, 10147–10158 (2001).
- G. M. Grason, B. A. DiDonna, R. D. Kamien, *Phys. Rev. Lett.* **91**, 058304 (2003).
- P. F. Damasceno, M. Engel, S. C. Glotzer, *Science* **337**, 453–457 (2012).
- A. Hajj-Akbari et al., *Nature* **462**, 773–777 (2009).
- D. B. Cordes, P. D. Lickiss, F. Rataboul, *Chem. Rev.* **110**, 2081–2173 (2010).
- X. Yu et al., *J. Am. Chem. Soc.* **134**, 7780–7787 (2012).
- A. Dolbecq, E. Dumas, C. R. Mayer, P. Mialane, *Chem. Rev.* **110**, 6009–6048 (2010).
- Y.-T. Lai, D. Cascio, T. O. Yeates, *Science* **336**, 1129 (2012).
- W.-B. Zhang et al., *Macromolecules* **47**, 1221–1239 (2014).
- X. Yu et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10078–10083 (2013).
- K. Yue et al., *Macromolecules* **45**, 8126–8134 (2012).
- See supplementary materials on Science Online.
- A. R. Bassindale et al., *Dalton Trans.* **2003**, 2945–2949 (2003).
- H. Liu et al., *J. Am. Chem. Soc.* **136**, 10691–10699 (2014).
- P. Ziherl, R. D. Kamien, *Phys. Rev. Lett.* **85**, 3528–3531 (2000).
- D. Weaire, R. Phelan, *Philos. Mag. Lett.* **69**, 107–110 (1994).
- V. Percec et al., *J. Am. Chem. Soc.* **130**, 13079–13094 (2008).
- Y. Li et al., *J. Am. Chem. Soc.* **133**, 10712–10715 (2011).

ACKNOWLEDGMENTS

Supported by NSF grant DMR-1408872. We thank B. Lotz for helpful discussion. The MALDI-TOF MS analysis was assisted by K. Guo, C. Shi, and C. Wesdemiotis. Use of the Advanced Photon Source at Argonne National Laboratory was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract DE-AC02-06CH11357.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/424/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
Table S1
References (35–38)

5 November 2014; accepted 2 March 2015
10.1126/science.aaa2421

METHANE CYCLING

Nonequilibrium clumped isotope signals in microbial methane

David T. Wang,^{1,2} Danielle S. Gruen,^{1,2} Barbara Sherwood Lollar,³ Kai-Uwe Hinrichs,⁴ Lucy C. Stewart,⁵ James F. Holden,⁵ Alexander N. Hristov,⁶ John W. Pohlman,⁷ Penny L. Morrill,⁸ Martin Könneke,⁴ Kyle B. Delwiche,⁹ Eoghan P. Reeves,¹ Chelsea N. Sutcliffe,³ Daniel J. Ritter,¹⁰ Jeffrey S. Seewald,² Jennifer C. McIntosh,¹⁰ Harold F. Hemond,⁹ Michael D. Kubo,¹¹ Dawn Cardace,¹² Tori M. Hoehler,¹¹ Shuhei Ono^{1*}

Methane is a key component in the global carbon cycle, with a wide range of anthropogenic and natural sources. Although isotopic compositions of methane have traditionally aided source identification, the abundance of its multiply substituted “clumped” isotopologues (for example, $^{13}\text{CH}_3\text{D}$) has recently emerged as a proxy for determining methane-formation temperatures. However, the effect of biological processes on methane’s clumped isotopologue signature is poorly constrained. We show that methanogenesis proceeding at relatively high rates in cattle, surface environments, and laboratory cultures exerts kinetic control on $^{13}\text{CH}_3\text{D}$ abundances and results in anomalously elevated formation-temperature estimates. We demonstrate quantitatively that H_2 availability accounts for this effect. Clumped methane thermometry can therefore provide constraints on the generation of methane in diverse settings, including continental serpentinization sites and ancient, deep groundwaters.

Carbon ($^{13}\text{C}/^{12}\text{C}$) and hydrogen (D/H) isotope ratios of methane are widely applied for distinguishing microbial from thermogenic methane in the environment (1–7), as well as for apportioning pathways of microbial methane production (8–10). This bulk isotope approach, however, is largely based on empirical observations, and different origins of methane often yield overlapping characteristic isotope signals (3, 7, 11–13). Beyond conventional bulk isotope ratios, it has become possible to precisely measure the abundance of multiply substituted “clumped” isotopologues (e.g., $^{13}\text{CH}_3\text{D}$) (14, 15). In particular, the abundance of clumped isotopes makes it possible to obtain information

about the temperature at which C–H bonds were formed or last equilibrated (14) (fig. S1). Formation temperatures of both thermogenic and microbial methane in natural gas reservoirs can be estimated on the basis of clumped isotopologues (16). The mechanisms by which isotopologues attain distributions consistent with thermodynamic equilibrium, however, remain unclear because bulk methane isotopes ($\delta^{13}\text{C}$ and δD) often reflect kinetic isotope fractionations (13, 17), and H isotope exchange between methane and water is sluggish (18).

To test whether clumped methane thermometry can be widely applied for methane sources beyond natural gas reservoirs, we examined

methane samples from diverse systems, including lakes, wetlands, cow rumen, laboratory cultures of methanogenic microbes, and geological settings that may support abiogenic methane production. We used a recently developed tunable laser spectroscopy technique (14, 19) to measure the relative abundances of four methane isotopologues ($^{12}\text{CH}_4$, $^{13}\text{CH}_4$, $^{12}\text{CH}_3\text{D}$, and $^{13}\text{CH}_3\text{D}$).

Our measurements for dominantly thermogenic gases from the Marcellus and Utica shales (1, 20) yielded $\Delta^{13}\text{CH}_3\text{D}$ -based temperatures of $147^{+25}_{-22}^\circ\text{C}$ and $160^{+29}_{-25}^\circ\text{C}$, respectively. The clumped isotope temperature for the Marcellus Shale sample is comparable to, although slightly lower than, estimates by Stolper *et al.* (16) of 179° to 207°C (Fig. 1). In addition, microbial methane in pore waters and gas hydrates from northern Cascadia margin sediments (3) and from wells producing from coal seams in the Powder River Basin (2, 21) yielded $\Delta^{13}\text{CH}_3\text{D}$ temperatures of 12° to 42°C and 35° to 52°C , respectively. These are consistent with their expected low formation temperatures. Furthermore, thermogenic methane sampled from a hydrothermal vent in the Guaymas Basin, Gulf of California (6), yielded a $\Delta^{13}\text{CH}_3\text{D}$ temperature of $326^{+170}_{-95}^\circ\text{C}$, within error of the measured vent temperature (299°C) (22). Therefore, our data provide independent support of the hypothesis that $^{13}\text{CH}_3\text{D}$ abundance reflects the temperature at which methane is generated in these sedimentary basins (16).

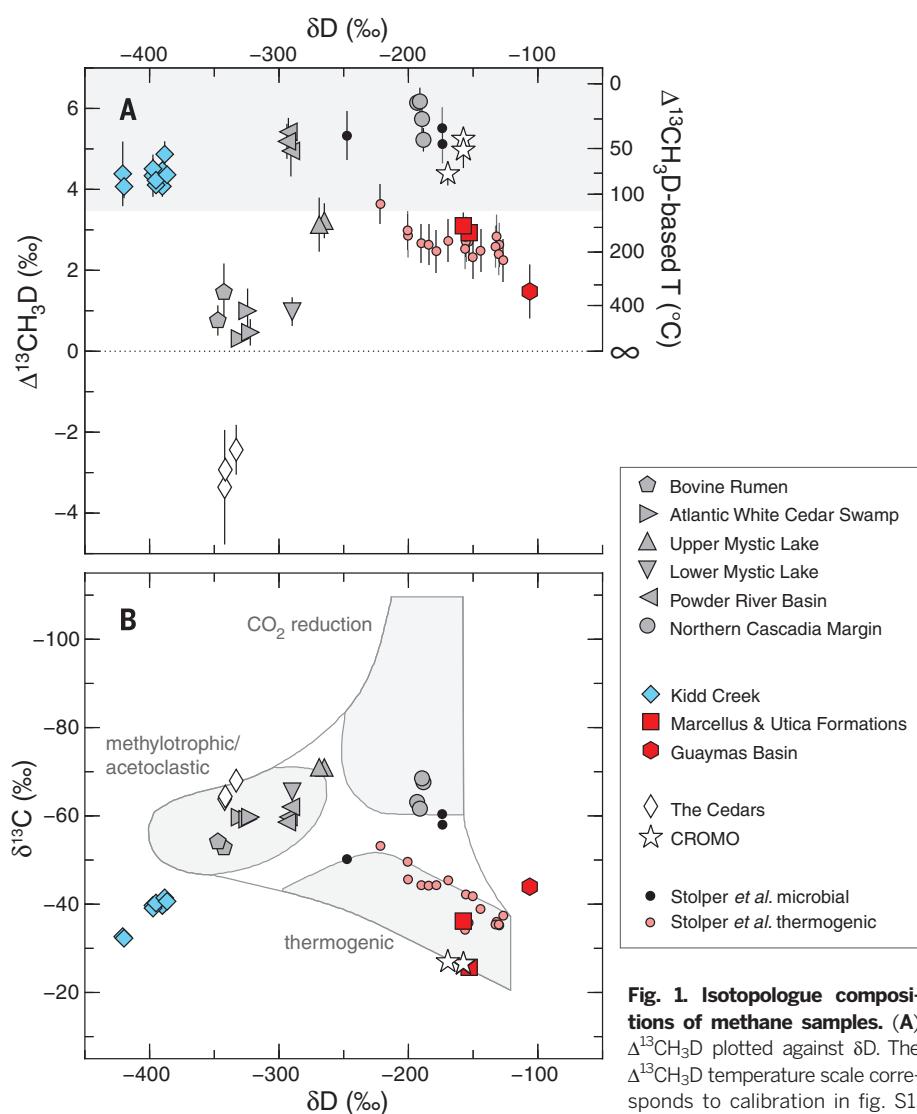
In contrast, we found that methane sampled from lakes, a swamp, and the rumen of a cow carries $^{13}\text{CH}_3\text{D}$ signals that correspond to anomalously high $\Delta^{13}\text{CH}_3\text{D}$ temperatures (139° to 775°C) (Fig. 1A) that are well above the environmental temperatures ($<40^\circ\text{C}$). Such signals are clearly not controlled by equilibrium. Notably, a positive correlation between $\Delta^{13}\text{CH}_3\text{D}$ and the extent of D/H fractionation between methane and environmental water [$\varepsilon_{\text{methane/water}}$ (23)] suggests a strong link between isotopologue (i.e., $^{13}\text{CH}_3\text{D}$) and isotope (D/H) disequilibria. In contrast, the above-mentioned methane samples from sedimentary basins appear to have attained hydrogen isotope equilibrium with associated waters

at or near the temperatures indicated by the $\Delta^{13}\text{CH}_3\text{D}$ data (Fig. 2).

To confirm these observations from the natural environment, we demonstrated that strong disequilibrium $^{13}\text{CH}_3\text{D}$ signals are also produced by cultures of methanogenic archaea in the laboratory (Fig. 3). Thermophilic methanogens cultured at 40° to 85°C produced methane with $\Delta^{13}\text{CH}_3\text{D}$ values from $+0.5$ to $+2.3$ per mil (‰) (corresponding to $\Delta^{13}\text{CH}_3\text{D}$ temperatures of 216° to 620°C), and mesophilic methanogens cultured at ambient temperature produced methane with conspicuously “anticlumped” signatures (i.e., values of $\Delta^{13}\text{CH}_3\text{D} < 0\text{‰}$, for which no apparent temperature can be expressed) as low as -1.3‰ (Fig. 3). Methane from cultures is also characterized by large kinetic D/H fractionation with respect to water (17, 24). Because laboratory cul-

tures are grown under optimal conditions (high H_2 and high CO_2), these anticlumped $\Delta^{13}\text{CH}_3\text{D}$ and low $\varepsilon_{\text{methane/water}}$ values are primarily expressions of kinetic isotope effects. Consequently, the distribution of samples with $\Delta^{13}\text{CH}_3\text{D}$ and $\varepsilon_{\text{methane/water}}$ values in Fig. 2 can be explained by microbial methanogenesis operating on a spectrum between fully kinetic (low $\Delta^{13}\text{CH}_3\text{D}$ and low $\varepsilon_{\text{methane/water}}$) and equilibrium (high $\Delta^{13}\text{CH}_3\text{D}$ and high $\varepsilon_{\text{methane/water}}$) end members.

We constructed a mathematical framework to describe the controls on the correlation of $\Delta^{13}\text{CH}_3\text{D}$ and $\varepsilon_{\text{methane/water}}$ signals from hydrogenotrophic methanogenesis. The model largely follows those developed for microbial sulfate reduction (25, 26) and predicts the isotopologue compositions of product methane as a result of a series of enzymatic reactions (fig. S4) (19). Using isotope



¹Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ²Marine Chemistry and Geochemistry Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA. ³Department of Earth Sciences, University of Toronto, Toronto, Ontario M5S 3B1, Canada. ⁴MARUM Center for Marine Environmental Sciences and Department of Geosciences, University of Bremen, Bremen D-28359, Germany. ⁵Department of Microbiology, University of Massachusetts, Amherst, MA 01003, USA. ⁶Department of Animal Science, Pennsylvania State University, University Park, PA 16802, USA. ⁷U.S. Geological Survey (USGS), Woods Hole Coastal and Marine Science Center, Woods Hole, MA 02543, USA. ⁸Department of Earth Sciences, Memorial University of Newfoundland, St John's, Newfoundland and Labrador A1B 3X5, Canada. ⁹Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ¹⁰Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ 85721, USA. ¹¹NASA Ames Research Center, Moffett Field, CA 94035, USA. ¹²Department of Geosciences, University of Rhode Island, Kingston, RI 02881, USA.

*Corresponding author. E-mail: sono@mit.edu

fractionation factors estimated from theory, experiments, and observations as input parameters (table S3) (19), our model reproduces the observed correlation between $\Delta^{13}\text{CH}_3\text{D}$ and $\epsilon_{\text{methane/water}}$ of natural samples (Fig. 2). The isotopologue compositions of product methane reflect the degree of metabolic reversibility. Fully reversible reactions yield equilibrium end members (27), whereas irreversible reactions result in kinetic (disequilibrium) end-member signals. In this model, the reversibility is linked to available free energy (26, 27), in this case expressed as H_2 concentration ($[\text{H}_2]$). The model can explain the relationship among $[\text{H}_2]$, $\epsilon_{\text{methane/water}}$ (28), and $\Delta^{13}\text{CH}_3\text{D}$ via Michaelis-Menten kinetics and can predict the observed patterns in diverse settings, ranging from marine sediments (low $[\text{H}_2]$, high $\Delta^{13}\text{CH}_3\text{D}$ and $\epsilon_{\text{methane/water}}$) to bovine rumen (high $[\text{H}_2]$, low $\Delta^{13}\text{CH}_3\text{D}$ and $\epsilon_{\text{methane/water}}$) (Fig. 4). We note that mixing of methane sources with different $\delta^{13}\text{C}$ and δD values or oxidation of methane could also alter the relationships over the primary signal of microbial methanogenesis (19). Likewise, inheritance of clumping signals from precursor organic substrates (e.g., via acetoclastic or methylotrophic methanogenesis) cannot be ruled out entirely and awaits experimental validation.

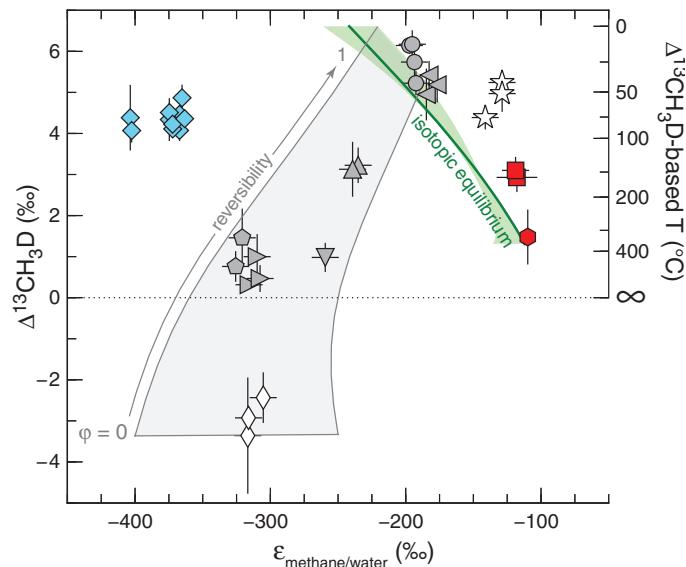
We showed above that the combination of $\Delta^{13}\text{CH}_3\text{D}$ and $\epsilon_{\text{methane/water}}$ values provides mechanistic constraints on whether methane was formed under kinetic versus near-equilibrium conditions. Next, we used this framework to place constraints on the origins of methane at two sites of present-day serpentinization in Phanerozoic ophiolites [The Cedars (29) and Coast Range Ophiolite Microbial Observatory (CROMO) (30)] in northern California, as well as in deep (>2 km below surface) fracture fluids with billion-year residence times in the Kidd Creek mine, Canada (5, 31).

Methane collected from groundwater springs associated with serpentinization at The Cedars yielded antoclumped $\Delta^{13}\text{CH}_3\text{D}$ signals (-3‰) with low $\epsilon_{\text{methane/water}}$ values (Figs. 1A and 2). The data plot along the microbial (kinetic) trend defined in Fig. 2, supporting a previous hypothesis that methane at The Cedars is being produced by active microbial methanogenesis (29). The exceptionally high H_2 concentration (up to 50% by volume in bubbles) at The Cedars indicate the massive excess of electron donors. This, along with severe inorganic carbon limitation [due to high pH (>11) and precipitation of carbonate minerals (29)], drives the formation of methane carrying strong kinetic imprints, consistent with the observed antoclumped $\Delta^{13}\text{CH}_3\text{D}$ signals (Fig. 4).

Despite the similarity in geologic setting, methane associated with serpentinization at CROMO (30) revealed very different $\Delta^{13}\text{CH}_3\text{D}$ values, which correspond to low apparent temperatures (42° to 76°C) and plot close to the equilibrium line (Fig. 2). Although the conventional $\delta^{13}\text{C}$ and δD values of methane from CROMO are nearly identical to those of the Utica Shale sample (Fig. 1B), methane at CROMO carries much higher $\Delta^{13}\text{CH}_3\text{D}$ values (Fig. 1A). The origin of methane at the CROMO site remains unresolved (30), but the comparably

Fig. 2. Extent of clumped and hydrogen isotopic disequilibrium in methane.

Symbols and vertical error bars are the same as those in Fig. 1. Horizontal error bars represent uncertainties on estimates of $\epsilon_{\text{methane/water}}$ (23) (table S4). The solid green curve represents isotopic equilibrium, with the $\epsilon_{\text{methane/water}}$ calibration given by (36). Green shading represents ranges of $\epsilon_{\text{methane/water}}$ calibrations from published reports (fig. S3). Gray shading represents



model predictions from this study, for microbial methane formed between 0° and 40°C. Metabolic reversibility (φ) increases from bottom ($\varphi = 0$, fully kinetic) to top ($\varphi \rightarrow 1$, equilibrium) within this field (19).

Fig. 3. $\Delta^{13}\text{CH}_3\text{D}$ values of methane produced by hydrogenotrophic methanogens in batch cultures reflect kinetic effects.

Data and error bars are from table S2. The green line represents clumped isotopologue equilibrium (i.e., samples for which $\Delta^{13}\text{CH}_3\text{D}$ temperature is equal to growth temperature) (fig. S1).

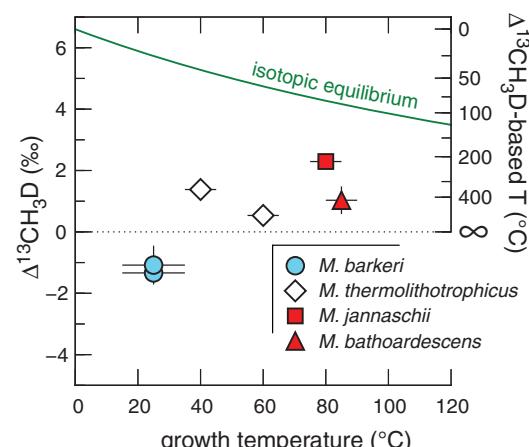
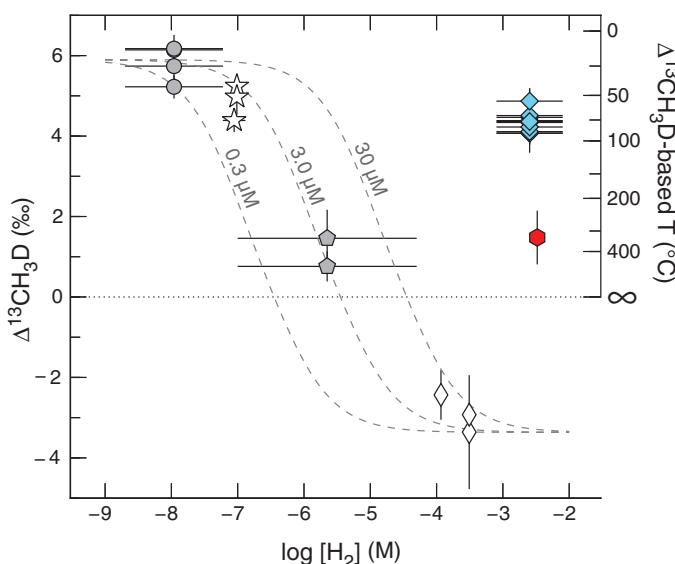


Fig. 4. Relationships between $\Delta^{13}\text{CH}_3\text{D}$ and H_2 concentration for microbial methane.

Symbols and vertical error bars are the same as in Fig. 1. The H_2 data are from table S4; when a range of $[\text{H}_2]$ values is given, points are plotted at the geometric mean of the maximum and minimum values. Dashed lines represent model predictions for microbial methane produced at 20°C, calculated using Michaelis-Menten constants (K_M) of 0.3, 3.0, and 30 μM H_2 . Data for samples of dominantly nonmicrobial methane from Guaymas Basin and Kidd Creek are plotted for comparison.



high $\Delta^{13}\text{CH}_3\text{D}$ values at CROMO suggest that methane here could be sourced from a mixture of thermogenic and microbial methane. Alternatively, lower H_2 availability at CROMO, compared with The Cedars (table S4), may support microbial methanogenesis under near-equilibrium conditions (Fig. 4). Regardless, the different isotopologue signatures in methane from CROMO versus The Cedars demonstrate that distinct processes contribute to methane formation in these two serpentinization systems.

Deep, ancient fracture fluids in the Kidd Creek mine in the Canadian Shield (31) contain copious quantities of both dissolved methane and hydrogen (5). The Kidd Creek methane occupies a distinct region in the diagram of $\Delta^{13}\text{CH}_3\text{D}$ versus $\varepsilon_{\text{methane/water}}$ (Fig. 2), due to strong D/H disequilibria between methane and water (4) and low- $\Delta^{13}\text{CH}_3\text{D}$ temperature signals of 56° to 90°C that are consistent with other temperature estimates for these groundwaters (4). Although the specific mechanisms by which the proposed abiotic hydrocarbons at Kidd Creek are generated remain under investigation (5, 32), the distinct isotopologue signals provide further support for the hypothesis that methane here is neither microbial nor thermogenic.

Our results demonstrate that measurements of $\Delta^{13}\text{CH}_3\text{D}$ provide information beyond the simple formation temperature of methane. The combination of methane and water hydrogen-isotope fractionation and $\Delta^{13}\text{CH}_3\text{D}$ abundance enables the differentiation of methane that has been formed at extremely low rates in the subsurface (3, 21, 27) from methane formed in cattle and surface environments in which methanogenesis proceeds at comparatively high rates (33, 34).

REFERENCES AND NOTES

- F. J. Baldassare, M. A. McCaffrey, J. A. Harper, *Am. Assoc. Pet. Geol. Bull.* **98**, 341–372 (2014).
- R. M. Flores, C. A. Rice, G. D. Stricker, A. Warden, M. S. Ellis, *Int. J. Coal Geol.* **76**, 52–75 (2008).
- J. Pohlman, M. Kaneko, V. Heuer, R. Coffin, M. Whiticar, *Earth Planet. Sci. Lett.* **287**, 504–512 (2009).
- B. Sherwood Lollar et al., *Geochim. Cosmochim. Acta* **72**, 4778–4795 (2008).
- B. Sherwood Lollar, T. D. Westgate, J. A. Ward, G. F. Slater, G. Lacrampe-Couloume, *Nature* **416**, 522–524 (2002).
- J. Welhan, J. Lupton, *Am. Assoc. Pet. Geol. Bull.* **71**, 215–223 (1987).
- M. J. Whiticar, *Org. Geochem.* **16**, 531–547 (1990).
- R. A. Burke Jr., C. S. Martens, W. M. Sackett, *Nature* **332**, 829–831 (1988).
- C. K. McCalley et al., *Nature* **514**, 478–481 (2014).
- M. J. Whiticar, E. Faber, M. Schoell, *Geochim. Cosmochim. Acta* **50**, 693–709 (1986).
- G. Etiope, B. Sherwood Lollar, *Rev. Geophys.* **51**, 276–299 (2013).
- M. Schoell, *Chem. Geol.* **71**, 1–10 (1988).
- M. J. Whiticar, *Chem. Geol.* **161**, 291–314 (1999).
- S. Ono et al., *Anal. Chem.* **86**, 6487–6494 (2014).
- D. A. Stolper et al., *Geochim. Cosmochim. Acta* **126**, 169–191 (2014).
- D. A. Stolper et al., *Science* **344**, 1500–1503 (2014).
- D. L. Valentine, A. Chidthaisong, A. Rice, W. S. Reeburgh, S. C. Tyler, *Geochim. Cosmochim. Acta* **68**, 1571–1590 (2004).
- E. P. Reeves, J. S. Seewald, S. P. Sylva, *Geochim. Cosmochim. Acta* **77**, 582–599 (2012).
- Materials and methods are available as supplementary materials on Science Online.
- R. Burruss, C. Laughrey, *Org. Geochem.* **41**, 1285–1296 (2010).
- B. L. Bates, J. C. McIntosh, K. A. Lohse, P. D. Brooks, *Chem. Geol.* **284**, 45–61 (2011).
- E. P. Reeves, J. M. McDermott, J. S. Seewald, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5474–5479 (2014).
- The abundance of $^{13}\text{CH}_3\text{D}$ is captured by a metric, $\Delta^{13}\text{CH}_3\text{D}$, that quantifies its deviation from a random distribution of isotopic substitutions among all isotopologues in a sample of methane: $\Delta^{13}\text{CH}_3\text{D} = \ln Q$, where Q is the reaction quotient of the isotopic exchange reaction $^{13}\text{CH}_4 + ^{12}\text{CH}_3\text{D} \rightleftharpoons ^{13}\text{CH}_3\text{D} + ^{12}\text{CH}_4$. The reported δ values are conventional isotopic notation, e.g., $\delta\text{D} = (\text{D}/\text{H})_{\text{sample}}/(\text{D}/\text{H})_{\text{reference}} - 1$. Mass spectrometric measurements yield $\Delta_{18}\text{D}$, a parameter that quantifies the combined abundance of $^{13}\text{CH}_3\text{D}$ and $^{12}\text{CH}_2\text{D}_2$. For most natural samples of methane, $\Delta_{18}\text{D}$ temperature is expected to be directly relatable to $\Delta^{13}\text{CH}_3\text{D}$ temperature, as measured by laser spectroscopy. The D/H fractionation between methane and environmental water is defined as $\varepsilon_{\text{methane/water}} = (\text{D}/\text{H})_{\text{methane}}/(\text{D}/\text{H})_{\text{water}} - 1$.
- M. Balabane, E. Galimov, M. Hermann, R. Letolle, *Org. Geochem.* **11**, 115–119 (1987).
- C. Rees, *Geochim. Cosmochim. Acta* **37**, 1141–1162 (1973).
- B. A. Wing, I. Halevy, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18116–18125 (2014).
- T. Holler et al., *Proc. Natl. Acad. Sci. U.S.A.* **108**, E1484–E1490 (2011).
- R. A. Burke Jr., *Chemosphere* **26**, 55–67 (1993).
- P. L. Morrill et al., *Geochim. Cosmochim. Acta* **109**, 222–240 (2013).
- D. Cardace et al., *Sci. Drill.* **16**, 45–55 (2013).
- G. Holland et al., *Nature* **497**, 357–360 (2013).
- B. Sherwood Lollar, T. C. Onstott, G. Lacrampe-Couloume, C. J. Ballentine, *Nature* **516**, 379–382 (2014).
- K. A. Johnson, D. E. Johnson, *J. Anim. Sci.* **73**, 2483–2492 (1995).
- C. Varadarajan, H. F. Hemond, *J. Geophys. Res.* **117**, G02004 (2012).
- K. Takai et al., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10949–10954 (2008).
- Y. Horibe, H. Craig, *Geochim. Cosmochim. Acta* **59**, 5209–5217 (1995).

ACKNOWLEDGMENTS

We thank J. Hayes, R. Summons, A. Whitehill, S. Zaarur, C. Ruppel, L. T. Bryndzia, N. Blair, D. Vinson, K. Nealson, and M. Schrenk for

discussions; W. Olszewski, D. Nelson, G. Lacrampe-Couloume, and B. Topçuoğlu for technical assistance; A. Whitehill, G. Luo, A. Apprill, K. Twing, W. Brazelton, A. Wray, J. Oh, A. Rowe, G. Chadwick, and A. Rietze for assistance in the field; R. Michener for the $\delta\text{D}_{\text{water}}$ analyses; L. T. Bryndzia (Shell) for providing the shale gas samples; R. Dias (USGS) for sharing the NGS samples; and R. Raiche, D. McCrary, S. Moore (Homestake Mining Co.), the staff of the McLaughlin Natural Reserve, and the well operators for access to wells. Grants from the NSF (EAR-1250394 to S.O. and EAR-1322805 to J.C.M.), N. R. Braunsdorf and D. J. H. Smit of Shell PTI/EG (to S.O.), the Deep Carbon Observatory (to S.O., B.S.L., M.K., and K.-U.H.), the Natural Sciences and Engineering Research Council of Canada (to B.S.L.), and the Gottfried Wilhelm Leibniz Program of the Deutsche Forschungsgemeinschaft (HI 616-14-1 to K.-U.H. and M.K.) supported this study. D.T.W. was supported by a National Defense Science and Engineering Graduate Fellowship. D.S.G. was supported by the Neil and Anna Rasmussen Foundation Fund, the Grayce B. Kerr Fellowship, and a Shell-MITEI Graduate Fellowship. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. government. All data used to support the conclusions in this manuscript are provided in the supplementary materials. Author contributions: D.T.W. and S.O. developed the methods, analyzed data, and performed modeling. D.T.W. and D.S.G. performed isotopic analyses. D.S.G., L.C.S., J.F.H., M.K., K.-U.H., and S.O. designed and/or conducted microbiological experiments. D.T.W., D.S.G., B.S.L., P.L.M., K.B.D., A.N.H., C.N.S., M.D.K., D.J.R., J.C.M., D.C., and S.O. designed and/or executed the field-sampling campaigns. D.T.W. and S.O. wrote the manuscript with input from all authors.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/428/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 to S6
References (37–87)

5 December 2014; accepted 18 February 2015

Published online 5 March 2015:

10.1126/science.aaa4326

ISOTOPE GEOCHEMISTRY

Biological signatures in clumped isotopes of O_2

Laurence Y. Yeung,^{1,2*} Jeanine L. Ash,^{1,*†} Edward D. Young¹

The abundances of molecules containing more than one rare isotope have been applied broadly to determine formation temperatures of natural materials. These applications of “clumped” isotopes rely on the assumption that isotope-exchange equilibrium is reached, or at least approached, during the formation of those materials. In a closed-system terrarium experiment, we demonstrate that biological oxygen (O_2) cycling drives the clumped-isotope composition of O_2 away from isotopic equilibrium. Our model of the system suggests that unique biological signatures are present in clumped isotopes of O_2 —and not formation temperatures. Photosynthetic O_2 is depleted in $^{18}\text{O}^{18}\text{O}$ and $^{17}\text{O}^{18}\text{O}$ relative to a stochastic distribution of isotopes, unlike at equilibrium, where heavy-isotope pairs are enriched. Similar signatures may be widespread in nature, offering new tracers of biological and geochemical cycling.

Statistical thermodynamics predicts that heavy isotopes will be bound together in a molecule more often than predicted by chance alone, provided the system is at isotopic equilibrium (1, 2). This preference for heavy-isotope pairing and its variation with temperature forms the basis of clumped-isotope thermometry (3–5), a class of approaches based on precise measurements of molecules containing

more than one rare isotope. When isotope-exchange reactions facilitate the equilibration of heavy-isotope pairs, the resulting isotopic distribution

*Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, CA 90095, USA.

†Department of Earth Science, Rice University, Houston, TX 77005, USA.

*Corresponding author. E-mail: lyeung@rice.edu (L.Y.Y.); jash@ucla.edu (J.L.A.) †These authors contributed equally to this work.



Nonequilibrium clumped isotope signals in microbial methane

David T. Wang *et al.*
Science **348**, 428 (2015);
 DOI: 10.1126/science.aaa4326

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 23, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/428.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2015/03/04/science.aaa4326.DC1.html>

A list of selected additional articles on the Science Web sites related to this article can be found at:

<http://www.sciencemag.org/content/348/6233/428.full.html#related>

This article cites 78 articles, 15 of which can be accessed free:
<http://www.sciencemag.org/content/348/6233/428.full.html#ref-list-1>

This article appears in the following subject collections:
 Geochemistry, Geophysics
http://www.sciencemag.org/cgi/collection/geochem_phys

high $\Delta^{13}\text{CH}_3\text{D}$ values at CROMO suggest that methane here could be sourced from a mixture of thermogenic and microbial methane. Alternatively, lower H_2 availability at CROMO, compared with The Cedars (table S4), may support microbial methanogenesis under near-equilibrium conditions (Fig. 4). Regardless, the different isotopologue signatures in methane from CROMO versus The Cedars demonstrate that distinct processes contribute to methane formation in these two serpentinization systems.

Deep, ancient fracture fluids in the Kidd Creek mine in the Canadian Shield (31) contain copious quantities of both dissolved methane and hydrogen (5). The Kidd Creek methane occupies a distinct region in the diagram of $\Delta^{13}\text{CH}_3\text{D}$ versus $\varepsilon_{\text{methane/water}}$ (Fig. 2), due to strong D/H disequilibria between methane and water (4) and low- $\Delta^{13}\text{CH}_3\text{D}$ temperature signals of 56° to 90°C that are consistent with other temperature estimates for these groundwaters (4). Although the specific mechanisms by which the proposed abiotic hydrocarbons at Kidd Creek are generated remain under investigation (5, 32), the distinct isotopologue signals provide further support for the hypothesis that methane here is neither microbial nor thermogenic.

Our results demonstrate that measurements of $\Delta^{13}\text{CH}_3\text{D}$ provide information beyond the simple formation temperature of methane. The combination of methane and water hydrogen-isotope fractionation and $\Delta^{13}\text{CH}_3\text{D}$ abundance enables the differentiation of methane that has been formed at extremely low rates in the subsurface (3, 21, 27) from methane formed in cattle and surface environments in which methanogenesis proceeds at comparatively high rates (33, 34).

REFERENCES AND NOTES

- F. J. Baldassare, M. A. McCaffrey, J. A. Harper, *Am. Assoc. Pet. Geol. Bull.* **98**, 341–372 (2014).
- R. M. Flores, C. A. Rice, G. D. Stricker, A. Warden, M. S. Ellis, *Int. J. Coal Geol.* **76**, 52–75 (2008).
- J. Pohlman, M. Kaneko, V. Heuer, R. Coffin, M. Whiticar, *Earth Planet. Sci. Lett.* **287**, 504–512 (2009).
- B. Sherwood Lollar et al., *Geochim. Cosmochim. Acta* **72**, 4778–4795 (2008).
- B. Sherwood Lollar, T. D. Westgate, J. A. Ward, G. F. Slater, G. Lacrampe-Couloume, *Nature* **416**, 522–524 (2002).
- J. Welhan, J. Lupton, *Am. Assoc. Pet. Geol. Bull.* **71**, 215–223 (1987).
- M. J. Whiticar, *Org. Geochem.* **16**, 531–547 (1990).
- R. A. Burke Jr., C. S. Martens, W. M. Sackett, *Nature* **332**, 829–831 (1988).
- C. K. McCalley et al., *Nature* **514**, 478–481 (2014).
- M. J. Whiticar, E. Faber, M. Schoell, *Geochim. Cosmochim. Acta* **50**, 693–709 (1986).
- G. Etiope, B. Sherwood Lollar, *Rev. Geophys.* **51**, 276–299 (2013).
- M. Schoell, *Chem. Geol.* **71**, 1–10 (1988).
- M. J. Whiticar, *Chem. Geol.* **161**, 291–314 (1999).
- S. Ono et al., *Anal. Chem.* **86**, 6487–6494 (2014).
- D. A. Stolper et al., *Geochim. Cosmochim. Acta* **126**, 169–191 (2014).
- D. A. Stolper et al., *Science* **344**, 1500–1503 (2014).
- D. L. Valentine, A. Chidthaisong, A. Rice, W. S. Reeburgh, S. C. Tyler, *Geochim. Cosmochim. Acta* **68**, 1571–1590 (2004).
- E. P. Reeves, J. S. Seewald, S. P. Sylva, *Geochim. Cosmochim. Acta* **77**, 582–599 (2012).
- Materials and methods are available as supplementary materials on *Science Online*.
- R. Burruss, C. Laughrey, *Org. Geochem.* **41**, 1285–1296 (2010).
- B. L. Bates, J. C. McIntosh, K. A. Lohse, P. D. Brooks, *Chem. Geol.* **284**, 45–61 (2011).
- E. P. Reeves, J. M. McDermott, J. S. Seewald, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5474–5479 (2014).
- The abundance of $^{13}\text{CH}_3\text{D}$ is captured by a metric, $\Delta^{13}\text{CH}_3\text{D}$, that quantifies its deviation from a random distribution of isotopic substitutions among all isotopologues in a sample of methane: $\Delta^{13}\text{CH}_3\text{D} = \ln Q$, where Q is the reaction quotient of the isotopic exchange reaction $^{13}\text{CH}_4 + ^{12}\text{CH}_2\text{D} \rightleftharpoons ^{13}\text{CH}_3\text{D} + ^{12}\text{CH}_2\text{D}$. The reported δ values are conventional isotopic notation, e.g., $\delta\text{D} = (\text{D}/\text{H})_{\text{sample}}/(\text{D}/\text{H})_{\text{reference}} - 1$. Mass spectrometric measurements yield $\Delta_{18}\text{D}$, a parameter that quantifies the combined abundance of $^{13}\text{CH}_3\text{D}$ and $^{12}\text{CH}_2\text{D}_2$. For most natural samples of methane, $\Delta_{18}\text{D}$ temperature is expected to be directly relatable to $\Delta^{13}\text{CH}_3\text{D}$ temperature, as measured by laser spectroscopy. The D/H fractionation between methane and environmental water is defined as $\varepsilon_{\text{methane/water}} = (\text{D}/\text{H})_{\text{methane}}/(\text{D}/\text{H})_{\text{water}} - 1$.
- M. Balabane, E. Galimov, M. Hermann, R. Letolle, *Org. Geochem.* **11**, 115–119 (1987).
- C. Rees, *Geochim. Cosmochim. Acta* **37**, 1141–1162 (1973).
- B. A. Wing, I. Halevy, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18116–18125 (2014).
- T. Holler et al., *Proc. Natl. Acad. Sci. U.S.A.* **108**, E1484–E1490 (2011).
- R. A. Burke Jr., *Chemosphere* **26**, 55–67 (1993).
- P. L. Morrill et al., *Geochim. Cosmochim. Acta* **109**, 222–240 (2013).
- D. Cardace et al., *Sci. Drill.* **16**, 45–55 (2013).
- G. Holland et al., *Nature* **497**, 357–360 (2013).
- B. Sherwood Lollar, T. C. Onstott, G. Lacrampe-Couloume, C. J. Ballentine, *Nature* **516**, 379–382 (2014).
- K. A. Johnson, D. E. Johnson, *J. Anim. Sci.* **73**, 2483–2492 (1995).
- C. Varadharajan, H. F. Hemond, *J. Geophys. Res.* **117**, G02004 (2012).
- K. Takai et al., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10949–10954 (2008).
- Y. Horibe, H. Craig, *Geochim. Cosmochim. Acta* **59**, 5209–5217 (1995).

ACKNOWLEDGMENTS

We thank J. Hayes, R. Summons, A. Whitehill, S. Zaarur, C. Ruppel, L. T. Bryndzia, N. Blair, D. Vinson, K. Nealson, and M. Schrenk for

discussions; W. Olszewski, D. Nelson, G. Lacrampe-Couloume, and B. Topçuoğlu for technical assistance; A. Whitehill, G. Luo, A. Apprill, K. Twing, W. Brazelton, A. Wray, J. Oh, A. Rowe, G. Chadwick, and A. Rietze for assistance in the field; R. Michener for the $\delta\text{D}_{\text{water}}$ analyses; L. T. Bryndzia (Shell) for providing the shale gas samples; R. Dias (USGS) for sharing the NGS samples; and R. Raiche, D. McCrary, S. Moore (Homestake Mining Co.), the staff of the McLaughlin Natural Reserve, and the well operators for access to wells. Grants from the NSF (EAR-1250394 to S.O. and EAR-1322805 to J.C.M.), N. R. Braunsdorf and D. J. H. Smit of Shell PTI/EG (to S.O.), the Deep Carbon Observatory (to S.O., B.S.L., M.K., and K.-U.H.), the Natural Sciences and Engineering Research Council of Canada (to B.S.L.), and the Gottfried Wilhelm Leibniz Program of the Deutsche Forschungsgemeinschaft (HI 616-14-1 to K.-U.H. and M.K.) supported this study. D.T.W. was supported by a National Defense Science and Engineering Graduate Fellowship. D.S.G. was supported by the Neil and Anna Rasmussen Foundation Fund, the Grayce B. Kerr Fellowship, and a Shell-MITEI Graduate Fellowship. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. government. All data used to support the conclusions in this manuscript are provided in the supplementary materials. Author contributions: D.T.W. and S.O. developed the methods, analyzed data, and performed modeling. D.T.W. and D.S.G. performed isotopic analyses. D.S.G., L.C.S., J.F.H., M.K., K.-U.H., and S.O. designed and/or conducted microbiological experiments. D.T.W., D.S.G., B.S.L., P.L.M., K.B.D., A.N.H., C.N.S., M.D.K., D.J.R., J.C.M., D.C., and S.O. designed and/or executed the field-sampling campaigns. D.T.W. and S.O. wrote the manuscript with input from all authors.

SUPPLEMENTARY MATERIALS

www.science.org/content/348/6233/428/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 to S6
References (37–87)

5 December 2014; accepted 18 February 2015

Published online 5 March 2015:

10.1126/science.aaa4326

ISOTOPE GEOCHEMISTRY

Biological signatures in clumped isotopes of O_2

Laurence Y. Yeung,^{1,2*} Jeanine L. Ash,^{1,*†} Edward D. Young¹

The abundances of molecules containing more than one rare isotope have been applied broadly to determine formation temperatures of natural materials. These applications of “clumped” isotopes rely on the assumption that isotope-exchange equilibrium is reached, or at least approached, during the formation of those materials. In a closed-system terrarium experiment, we demonstrate that biological oxygen (O_2) cycling drives the clumped-isotope composition of O_2 away from isotopic equilibrium. Our model of the system suggests that unique biological signatures are present in clumped isotopes of O_2 —and not formation temperatures. Photosynthetic O_2 is depleted in $^{18}\text{O}^{18}\text{O}$ and $^{17}\text{O}^{18}\text{O}$ relative to a stochastic distribution of isotopes, unlike at equilibrium, where heavy-isotope pairs are enriched. Similar signatures may be widespread in nature, offering new tracers of biological and geochemical cycling.

Statistical thermodynamics predicts that heavy isotopes will be bound together in a molecule more often than predicted by chance alone, provided the system is at isotopic equilibrium (1, 2). This preference for heavy-isotope pairing and its variation with temperature forms the basis of clumped-isotope thermometry (3–5), a class of approaches based on precise measurements of molecules containing

more than one rare isotope. When isotope-exchange reactions facilitate the equilibration of heavy-isotope pairs, the resulting isotopic distribution

*Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, CA 90095, USA.

†Department of Earth Science, Rice University, Houston, TX 77005, USA.

*Corresponding author. E-mail: lyeung@rice.edu (L.Y.Y.); jash@ucla.edu (J.L.A.) †These authors contributed equally to this work.

has indeed been shown to achieve equilibrium across a wide range of temperatures (4, 6–8); however, isotopic equilibrium is the exception rather than the rule in nature. Biogenic substances, for example, are often formed through irreversible enzymatic reactions for which isotope-exchange equilibrium cannot be expected a priori. Yet, many natural materials with kinetically constrained and/or biological origins (e.g., carbonate shells) show only minor departures from equilibrium isotope fractionation (9–11). Large biological and physical effects on heavy-isotope pairing could complicate the interpretation of emerging clumped-isotope thermometers in methane, O₂, and other candidate systems (4, 5, 12).

Here, we consider photosynthetic O₂ formation from water at the oxygen-evolving complex of Photosystem II (OEC). In the OEC, O–O bond formation occurs at the end of a five-step light-dependent sequence (Fig. 1). This reaction most likely does not equilibrate O–O isotope pairs given the lack of isotopic equilibration between water and the O₂ produced (13–16). We argue that the tendency for two heavy oxygen isotopes to be bound together during oxygenic photosynthesis reflects primarily the isotopic preferences of water molecules binding to the OEC. These patterns of heavy-isotope pairing should be apparent in clumped isotopes of O₂. Measurements of the ¹⁸O¹⁸O (mass 36) and ¹⁷O¹⁸O (mass 35) isotopologues of O₂, together with bulk isotopic ratios (¹⁸O/¹⁶O and ¹⁷O/¹⁶O), characterize the number of heavy-isotope pairs in a sample relative to the number expected by chance alone (i.e., the stochastic distribution). These deviations are quantified as Δ₃₆ and Δ₃₅ values: Excesses of ¹⁸O¹⁸O and ¹⁷O¹⁸O relative to the stochastic distribution of isotopes in the sample results in Δ₃₆ > 0 and Δ₃₅ > 0, respectively. A deficit in ¹⁸O¹⁸O and ¹⁷O¹⁸O results in Δ₃₆ < 0 and Δ₃₅ < 0.

The Δ₃₆ and Δ₃₅ signatures of oxygenic photosynthesis can thus be estimated by assigning each water-binding site its own isotopic fractionation factor α = ¹⁸R_{bound}/¹⁸R_{water}, where ¹⁸R is the ratio of ¹⁸O to ¹⁶O atoms in each reservoir. At natural isotopic abundances, the bulk isotopic composition of photosynthetic O₂ is the weighted sum of those contributions—i.e., ¹⁸R_p ≈ 12; [(¹⁸R_{water} × α_A) + (¹⁸R_{water} × α_B)], with binding sites A and B each contributing one of two oxygen atoms in each O₂ molecule. The probability of generating ¹⁸O–¹⁸O bonds is therefore ³⁶R_p = (¹⁸R_{water} × α_A) (¹⁸R_{water} × α_B). The stochastic distribution of ¹⁸O atoms is calculated from the bulk ¹⁸O/¹⁶O ratio as ³⁶R_{stochastic} = (¹⁸R_p)². The expression for Δ_{36,p} then reduces to (17)

$$\Delta_{36,p} = \left[\frac{\alpha_A \alpha_B}{\frac{1}{4}(\alpha_A + \alpha_B)^2} - 1 \right] \quad (1)$$

Equation 1 reveals that, in all cases, Δ_{36,p} ≤ 0; contrary to the enhanced isotope pairing that would be expected at isotopic equilibrium, there is an apparent aversion to heavy-isotope pairing associated with photosynthetic O₂ production. If the isotopic preferences at each water-binding site are equal (α_A = α_B), then Δ_{36,p} = 0. If the

binding sites are not equivalent (α_A ≠ α_B), as isotope-labeling studies indicate (18, 19), then 0 ≥ Δ_{36,p} > −0.9 per mil (‰) for plausible α-values between 0.97 and 1.03 (20, 21). A similar expression can be derived for Δ_{35,p} values, which are predicted to be about half those of Δ_{36,p} (see the supplementary text). These values cannot be interpreted as formation temperatures because all equilibrated samples have Δ_n ≥ 0 (2). Photosynthesis should therefore impart a distinct non-equilibrium clumped-isotope signature on O₂.

We conducted a closed-system terrarium experiment with six water hyacinths (*Eichhornia crassipes*) to explore the effects of biological oxygen cycling on five isotopologues of O₂ (17). The terrarium was illuminated with fluorescent lights on a 12-hour/12-hour light-dark cycle. Headspace samples were purified and analyzed over a 1-year period for both the bulk and clumped isotopic composition of O₂. We found that biological oxygen cycling altered isotopic ordering in the headspace O₂, yielding apparent steady-state Δ₃₆ and Δ₃₅ values that are inconsistent with O₂ formation temperatures and more consistent with the predicted photosynthetic endmembers (Fig. 2 and table S3). The Δ₃₆ and Δ₃₅ values of O₂ were driven down from atmospheric values [2‰ and 1‰, respectively (4)] and down past equilibrium values at 25°C (1.5‰ and 0.8‰, respectively), finally approaching an apparent isotopic steady state at the stochastic distribution of isotopes (Δ₃₆ = −0.01 ± 0.08‰, and Δ₃₅ = 0.0 ± 0.1‰; 1 SEM, n = 4). The plant community shifted to an algae-dominated ecosystem during the first 6 months, altering the isotopic, chemical, and physical properties of the terrarium (fig. S1). However, the clumped-isotope composition of the headspace O₂ evolved steadily toward its apparent steady state, similar to the evolution of the oxygen triple-isotope composition. Steady-state Δ¹⁷O values were 165 parts per million (ppm), consistent with those reported in similar experiments (22, 23).

Dark incubations of the terrarium, which consumed up to 35% of the headspace O₂, caused Δ₃₆ values to increase linearly with time up to ~1‰ (Fig. 2). The Δ₃₅ values, in contrast, remained generally constant (means of Δ₃₅ = 0.1 ± 0.1‰ and 0.1 ± 0.05‰; 1 SD). Returning to light-dark cycles restored the clumped-isotope composition to its apparent steady-state value after 6 months (Δ₃₆ = −0.09 ± 0.06‰, and Δ₃₅ = 0.0 ± 0.1‰; 1 SEM, n = 3). To test the veracity of these measurements, headspace O₂ samples drawn from both light and dark incubations were photolytically equilibrated at known temperatures (4). The equilibrations yielded Δ₃₆ and Δ₃₅ values of O₂ consistent with isotope-exchange equilibrium (table S3), suggesting that our observations are unlikely to be analytical artifacts. Atmospheric O₂ leaking into the terrarium would increase δ¹⁸O far too rapidly relative to Δ₃₆ to explain these observations. The observed clumped-isotope variations therefore most likely arise from biological and physical processes inside the terrarium.

We constructed a two-reservoir model of O₂ (i.e., in headspace and water) in the terrarium

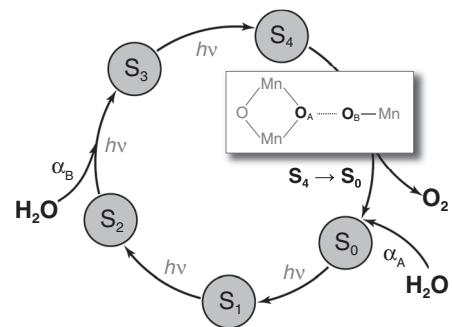


Fig. 1. Conceptual diagram of O₂ formation at the OEC. The five-step Kok cycle for the water-splitting reaction 2H₂O + 4hv → O₂ + 4H⁺ + 4e[−] is shown without electron flow (32). Transitions between intermediate oxidation states of the OEC (S₀ to S₄) occur upon absorption of visible light. The water-binding sequence is based on experimental results (19, 33, 34), which also indicate that water substrates are exchangeable at least up to state S₃ on chemically distinct binding sites (18, 19). The O–O bond is formed during the S₄-to-S₀ transition, expressing the isotopic fractionations α_A and α_B from water substrate binding.

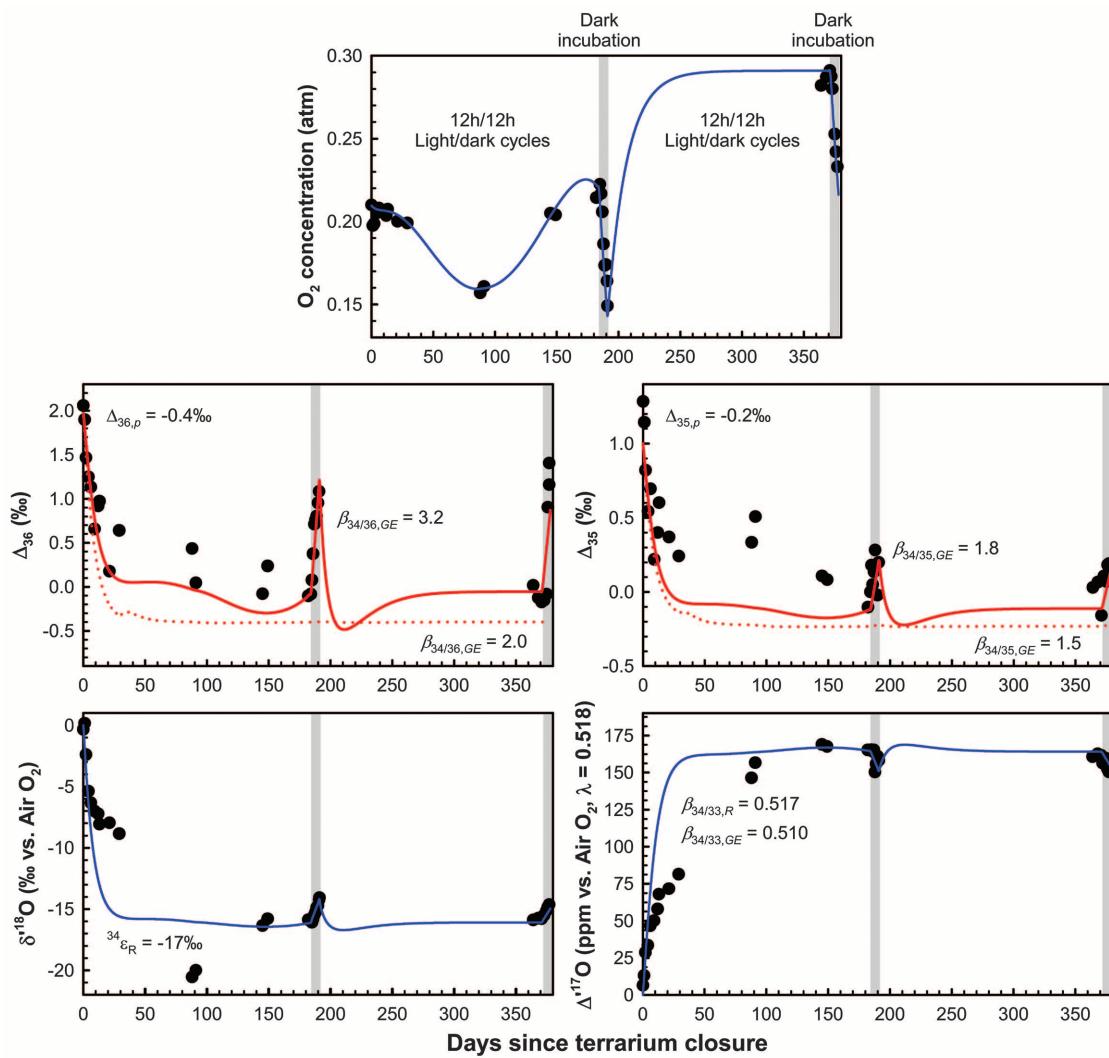
that accounts for photosynthetic O₂ formation, fractionation of O₂ due to respiration, and air-water gas exchange (17). We included kinetic isotope fractionation for gas transfer into and out of solution [³⁴α_{GE,kinetic} = 0.9972 for ¹⁸O/¹⁶O (24)]. The model was run with a range of plausible isotope fractionation factors for respiration [³⁴α_R = 0.97–0.99 (25, 26)] and gas-exchange rates (24, 27) to examine the sensitivity of the terrarium headspace to changes in those quantities. The oxygen triple-isotope composition of the terrarium water was measured and used as the bulk isotopic composition of photosynthetic O₂ (13, 15, 17). No single set of parameters explained all of the isotopic variations during the entire experiment, likely due to the evolving biological community, so we focus on isotopic variations at steady state and during dark incubations.

The increase of headspace Δ₃₆ and Δ₃₅ values in the dark implies that the apparent steady-state values near zero can only be reached if light-dependent processes drive Δ₃₆ and Δ₃₅ values below zero. Equation 1 suggests that photosynthesis could be the relevant mechanism, because the O₂ generated is likely to have Δ_{36,p} and Δ_{35,p} values less than zero. To estimate the composition of this source, we note that kinetic and equilibrium isotope effects for relevant photosynthetic fractionations are probably in the range 0.96 > δ¹⁸α > 1.04 (20, 21), which we broaden to a more conservative plausible range of 0.9 > δ¹⁸α > 1.1. This range of isotope effects gives lower limits on Δ_{36,p} and Δ_{35,p} of −10‰ and −5‰, respectively.

If the Δ₃₆ increase during dark incubations were solely caused by fractionation in respiration, then large isotope effects in water-enzyme binding would be required: Δ_{36,p} < −10‰ is needed to achieve steady-state values of Δ₃₆ near zero (17). In addition, the associated Δ_{35,p} < −5‰

Fig. 2. Evolution of concentration and O₂ isotope isopologue composition in the terrarium. Observations (data points) are compared with model results (curves). Uncertainties are not shown for clarity, but long-term analytical uncertainties in O₂ concentration, $\delta^{18}\text{O}$, $\Delta^{17}\text{O}$, Δ_{36} , and Δ_{35} are 1%, 0.04‰, 5 ppm, 0.17‰, and 0.3‰, respectively.

A single isotopologue discrimination factor (${}^{34}\epsilon_R = -17\text{\textperthousand}$) is used here to illustrate steady-state behavior in $\delta^{18}\text{O}$ and $\Delta^{17}\text{O}$; a more detailed model run yields better agreement for $\delta^{18}\text{O}$ and $\Delta^{17}\text{O}$ but similar results for Δ_{36} and Δ_{35} . Mass-dependent exponents used in the model, $\beta_{34/n}$, are labeled, with subscripts *R* and *GE* denoting values for respiration and gas exchange, respectively. For $\beta_{34/35,GE}$ and $\beta_{34/36,GE}$, two model runs are shown to illustrate their effects on the Δ_{36} and Δ_{35} time traces (17).



endmember composition causes poor agreement between measured and modeled Δ_{35} values (fig. S4C). Furthermore, an increase in respiration rates would drive Δ_{36} and Δ_{35} values higher, whereas a decrease in respiration rates would drive the O₂ toward its $\Delta_{36,p}$ and $\Delta_{35,p}$ photosynthetic values (17). Therefore, when the O₂ cycle was out of balance in the first 6 months, Δ_{36} would have fluctuated inversely with O₂ concentration (fig. S4, B and C). Instead, both Δ_{36} and Δ_{35} decreased nearly monotonically.

Isotopologue fractionation during nonequilibrium O₂ gas exchange could explain the increases of headspace Δ_{36} and Δ_{35} values during dark incubations. The fractionation in headspace ${}^{16}\text{O}^{18}\text{O}/{}^{16}\text{O}_2$ is closer to that for gas exchange than that for respiration (${}^{34}\alpha_{\text{observed}} = 0.995$ versus ${}^{34}\alpha_{\text{GE,kinetic}} = 0.9972$ versus ${}^{34}\alpha_R \sim 0.98$), suggesting that the Δ_{36} and Δ_{35} increases are similarly dominated by gas exchange. Modeling the mass dependence of gas exchange using the dark incubation data yields $\Delta_{36,p}$ and $\Delta_{35,p}$ values within a plausible range (i.e., $\Delta_{36,p} = -0.4\text{\textperthousand}$, $\Delta_{35,p} = -0.2\text{\textperthousand}$) (Fig. 2). The evolution of Δ_{36} and Δ_{35} is also more robust to imbalances in the O₂ cycle (17). Other

oxygen-consumption mechanisms, such as sulfide oxidation, could impart additional isotopologue signatures (28), so attributing isotopologue discrimination in the dark to a single process is necessarily a simplification. Indeed, the implied mass dependence of O₂ consumption in the dark terrarium is unusual, and it merits further investigation (17). A detailed understanding of isotopologue fractionation factors will require more controlled experiments of isolated biological and physical processes. Yet, the specific isotopologue discrimination during dark incubations does not affect the conclusion that photosynthesis generates O₂ with an “anticlumped” isotopologue distribution (i.e., $\Delta_{36} \leq 0$ and $\Delta_{35} \leq 0$). This biological signature in O₂ may be readily observed in the surface ocean, where it could be used to constrain gross primary productivity by exploiting the contrast between biological and atmospheric O₂ clumped-isotope signatures (29). Isotopic ordering in atmospheric O₂ is relatively unaffected by biological O₂ cycling because photochemical equilibration of O₂ exceeds rates of biological cycling by at least a factor of 100 (4, 30). Using a biological endmember composition of

$\Delta_{36} = 0$, we calculate that biological effects on the tropospheric Δ_{36} budget are therefore most likely on the order of 0.01‰.

Our observations indicate that variations in the isotopologue abundance of even simple molecules like O₂ capture the chemistry of complex natural systems. Broader application of these techniques could yield insights into the mechanisms of biomolecule synthesis, e.g., methanogenesis, nitrogen reduction during denitrification, and molecular hydrogen release during nitrogen fixation (31). Moreover, because clumped-isotope signatures can depend only on isotope fractionation factors and not on the isotopic composition of substrates, a new class of reservoir-insensitive approaches for tracing biogeochemical cycling could emerge from these molecular-scale insights.

REFERENCES AND NOTES

- P. Richet, Y. Bottinga, M. Javoy, *Annu. Rev. Earth Planet. Sci.* **5**, 65–110 (1977).
- Z. Wang, E. A. Schauble, J. M. Eiler, *Geochim. Cosmochim. Acta* **68**, 4779–4797 (2004).
- P. Ghosh et al., *Geochim. Cosmochim. Acta* **70**, 1439–1456 (2006).

4. L. Y. Yeung, J. L. Ash, E. D. Young, *J. Geophys. Res.* **119**, 10 (2014).
5. D. A. Stolper et al., *Science* **344**, 1500–1503 (2014).
6. H. P. Affek, *Am. J. Sci.* **313**, 309–325 (2013).
7. B. H. Passy, G. A. Henkes, *Earth Planet. Sci. Lett.* **351**–352, 223–236 (2012).
8. D. A. Stolper et al., *Geochim. Cosmochim. Acta* **126**, 169–191 (2014).
9. W. Guo, J. L. Mosenfelder, W. A. Goddard III, J. M. Eiler, *Geochim. Cosmochim. Acta* **73**, 7203–7225 (2009).
10. J. Tang, M. Dietzel, A. Fernandez, A. K. Tripathi, B. E. Rosenheim, *Geochim. Cosmochim. Acta* **134**, 120–136 (2014).
11. H. P. Affek, S. Zarrar, *Geochim. Cosmochim. Acta* **143**, 319–330 (2014).
12. S. Ono et al., *Anal. Chem.* **86**, 6487–6494 (2014).
13. R. D. Guy, M. L. Fogel, J. A. Berry, *Plant Physiol.* **101**, 37–47 (1993).
14. C. L. R. Stevens, D. Schultz, C. Van Baalen, P. L. Parker, *Plant Physiol.* **56**, 126–129 (1975).
15. Y. Helman, E. Barkan, D. Eisenstadt, B. Luz, A. Kaplan, *Plant Physiol.* **138**, 2292–2298 (2005).
16. H. C. Urey, L. J. Grief, *J. Am. Chem. Soc.* **57**, 321–327 (1935).
17. Materials and methods are available as supplementary materials on *Science Online*.
18. W. Hillier, T. Wydrzynski, *Coord. Chem. Rev.* **252**, 306–317 (2008).
19. L. Rapatskiy et al., *J. Am. Chem. Soc.* **134**, 16619–16634 (2012).
20. A. M. Angeles-Boza et al., *Chem. Sci.* **5**, 1141 (2014).
21. A. M. Angeles-Boza, J. P. Roth, *Inorg. Chem.* **51**, 4722–4729 (2012).
22. B. Luz, E. Barkan, M. L. Bender, M. H. Thiemens, K. A. Boering, *Nature* **400**, 547–550 (1999).
23. A. Angert, S. Rachmilevitch, E. Barkan, B. Luz, *Global Biogeochem. Cycles* **17**, 1030 (2003).
24. M. Knox, P. D. Quay, D. Wilbur, *J. Geophys. Res.* **97** (C12), 20335–20343 (1992).
25. B. Luz, E. Barkan, *Geochim. Cosmochim. Acta* **69**, 1099–1110 (2005).
26. M. H. Cheah et al., *Anal. Chem.* **86**, 5171–5178 (2014).
27. K. E. Tempest, S. Emerson, *Mar. Chem.* **153**, 39–47 (2013).
28. R. S. Thurston, K. W. Mandernack, W. C. Shanks III, *Chem. Geol.* **269**, 252–261 (2010).
29. L. W. Juraneck, P. D. Quay, *Annu. Rev. Mar. Sci.* **5**, 503–524 (2013).
30. L. Y. Yeung, E. D. Young, E. A. Schauble, *J. Geophys. Res.* **117**, D18306 (2012).
31. B. M. Hoffman, D. Lukyanov, D. R. Dean, L. C. Seefeldt, *Acc. Chem. Res.* **46**, 587–595 (2013).
32. B. Kok, B. Forbush, M. McGloin, *Photochem. Photobiol.* **11**, 457–475 (1970).
33. T. Noguchi, *Phil. Trans. R. Soc. B* **363**, 1189–1195 (2008).
34. N. Cox et al., *Science* **345**, 804–808 (2014).

ACKNOWLEDGMENTS

We thank H. Hu and N. Levin for performing oxygen triple-isotope analyses of the terrarium water at Johns Hopkins University, and E. Schauble for helpful discussions during the course of this work. This research was supported in part by the National Science Foundation (EAR-1049655 and DGE-1144087), the National Aeronautics and Space Administration Cosmochemistry program, and the Deep Carbon Observatory. The data and model parameters used in this study are available in the supplementary materials (tables S1 to S3).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/431/suppl/DC1
 Materials and Methods
 Supplementary Text
 Figs. S1 to S5
 Tables S1 to S3
 References (35–50)

7 January 2015; accepted 13 March 2015
 10.1126/science.aaa6284

RESEARCH FUNDING

Big names or big ideas: Do peer-review panels select the best science proposals?

Danielle Li^{1*}† and Leila Agha^{2,3*}†

This paper examines the success of peer-review panels in predicting the future quality of proposed research. We construct new data to track publication, citation, and patenting outcomes associated with more than 130,000 research project (R01) grants funded by the U.S. National Institutes of Health from 1980 to 2008. We find that better peer-review scores are consistently associated with better research outcomes and that this relationship persists even when we include detailed controls for an investigator's publication history, grant history, institutional affiliations, career stage, and degree types. A one-standard deviation worse peer-review score among awarded grants is associated with 15% fewer citations, 7% fewer publications, 19% fewer high-impact publications, and 14% fewer follow-on patents.

In 2014, the combined budgets of the U.S. National Institutes of Health (NIH), the U.S. National Science Foundation, and the European Research Council totaled almost \$40 billion. The majority of these funds were allocated to external researchers whose applications were vetted by committees of expert reviewers. But as funding has become more competitive and application award probabilities have fallen, some observers have posited that “the system now favors those who can guarantee results rather than those with potentially path-breaking ideas that, by definition, cannot promise success” (*1*). Despite its importance for guiding research investments, there have been few attempts to assess the efficacy of peer review.

Peer-review committees are unique in their ability to assess research proposals based on deep expertise but may be undermined by biases, insufficient effort, dysfunctional committee dynamics, or limited subject knowledge (*2, 3*). Disagreement about what constitutes important research may introduce randomness into the process (*4*). Existing research in this area has focused on understanding whether there is a correlation between good peer-review scores and successful research outcomes and yields mixed results (*5–7*). Yet raw correlations do not reveal whether reviewers are generating insight about the scientific merit of proposals. For example, if applicants from elite institutions generally produce more highly cited research, then a system that rewarded institutional rankings without even reading applications may appear effective at identifying promising research.

In this paper, we investigate whether peer review generates new insights about the scientific quality of grant applications. We call this ability peer review’s “value-added.” The value-added of NIH peer review is conceptually distinct from the value of NIH funding itself. For example, even if reviewers did a poor job of identifying the best applications, receiving a grant may still improve a researcher’s productivity by allowing her to main-

tain a laboratory and support students. Whereas previous work has studied the impact of receiving NIH funds on the productivity of awardees (*8, 9*), our paper asks whether NIH selects the most promising projects to support. Because NIH cannot possibly fund every application it receives, the ability to distinguish potential among applications is important for its success.

We say that peer review has high value-added if differences in grants’ scores are predictive of differences in their subsequent research output, after controlling for previous accomplishments of the applicants. This may be the case if reviewers generate additional insights about an application’s potential, but peer review may also have zero or even negative value-added if reviewers are biased, mistaken, or focused on different goals (*10*).

Because research outcomes are often skewed, with many low-quality or incremental contributions and relatively few ground-breaking discoveries (*2, 11*), we assess the value-added of peer review for identifying research that is highly influential or shows commercial promise. We also test the effectiveness of peer review in screening out applications that result in unsuccessful research (see the supplementary materials for full details on data and methods).

NIH is the world’s largest funder of biomedical research (*12*). With an annual budget of approximately \$30 billion, it supports more than 300,000 research personnel at more than 2500 institutions (*12, 13*). A funding application is assigned by topic to one of approximately 200 peer-review committees (known as study sections).

Our main explanatory variable is the “percentile score,” ranging from 0 to 100, which reflects an application’s ranking among all other applications reviewed by a study section in a given fiscal year; lower scores correspond to higher-quality applications. In general, applications are funded in order of their percentile score until the budget of their assigned NIH institute is exhausted. The average score in our sample is 14.2, with a standard deviation (SD) of 10.2; only about 1% of funded grants in our sample had a score worse than 50. Funding has become more competitive in recent years; only 14% of applications were funded in 2013.

*Corresponding author. E-mail: dli@hbs.edu (D.L.); lagha@bu.edu (L.A.) †Both authors contributed equally to this work.



Biological signatures in clumped isotopes of O₂

Laurence Y. Yeung *et al.*
Science **348**, 431 (2015);
DOI: 10.1126/science.aaa6284

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 23, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/431.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2015/04/22/348.6233.431.DC1.html>

A list of selected additional articles on the Science Web sites related to this article can be found at:

<http://www.sciencemag.org/content/348/6233/431.full.html#related>

This article cites 49 articles, 11 of which can be accessed free:
<http://www.sciencemag.org/content/348/6233/431.full.html#ref-list-1>

This article has been cited by 1 articles hosted by HighWire Press; see:
<http://www.sciencemag.org/content/348/6233/431.full.html#related-urls>

This article appears in the following subject collections:
Geochemistry, Geophysics

http://www.sciencemag.org/cgi/collection/geochem_phys

4. L. Y. Yeung, J. L. Ash, E. D. Young, *J. Geophys. Res.* **119**, 10 (2014).
5. D. A. Stolper et al., *Science* **344**, 1500–1503 (2014).
6. H. P. Affek, *Am. J. Sci.* **313**, 309–325 (2013).
7. B. H. Passy, G. A. Henkes, *Earth Planet. Sci. Lett.* **351**–352, 223–236 (2012).
8. D. A. Stolper et al., *Geochim. Cosmochim. Acta* **126**, 169–191 (2014).
9. W. Guo, J. L. Mosenfelder, W. A. Goddard III, J. M. Eiler, *Geochim. Cosmochim. Acta* **73**, 7203–7225 (2009).
10. J. Tang, M. Dietzel, A. Fernandez, A. K. Tripathi, B. E. Rosenheim, *Geochim. Cosmochim. Acta* **134**, 120–136 (2014).
11. H. P. Affek, S. Zaaarur, *Geochim. Cosmochim. Acta* **143**, 319–330 (2014).
12. S. Ono et al., *Anal. Chem.* **86**, 6487–6494 (2014).
13. R. D. Guy, M. L. Fogel, J. A. Berry, *Plant Physiol.* **101**, 37–47 (1993).
14. C. L. R. Stevens, D. Schultz, C. Van Baalen, P. L. Parker, *Plant Physiol.* **56**, 126–129 (1975).
15. Y. Helman, E. Barkan, D. Eisenstadt, B. Luz, A. Kaplan, *Plant Physiol.* **138**, 2292–2298 (2005).
16. H. C. Urey, L. J. Grief, *J. Am. Chem. Soc.* **57**, 321–327 (1935).
17. Materials and methods are available as supplementary materials on *Science Online*.
18. W. Hillier, T. Wydrynski, *Coord. Chem. Rev.* **252**, 306–317 (2008).
19. L. Rapatskiy et al., *J. Am. Chem. Soc.* **134**, 16619–16634 (2012).
20. A. M. Angeles-Boza et al., *Chem. Sci.* **5**, 1141 (2014).
21. A. M. Angeles-Boza, J. P. Roth, *Inorg. Chem.* **51**, 4722–4729 (2012).
22. B. Luz, E. Barkan, M. L. Bender, M. H. Thiemens, K. A. Boering, *Nature* **400**, 547–550 (1999).
23. A. Angert, S. Rachmilevitch, E. Barkan, B. Luz, *Global Biogeochem. Cycles* **17**, 1030 (2003).
24. M. Knox, P. D. Quay, D. Wilbur, *J. Geophys. Res.* **97** (C12), 20335–20343 (1992).
25. B. Luz, E. Barkan, *Geochim. Cosmochim. Acta* **69**, 1099–1110 (2005).
26. M. H. Cheah et al., *Anal. Chem.* **86**, 5171–5178 (2014).
27. K. E. Tempest, S. Emerson, *Mar. Chem.* **153**, 39–47 (2013).
28. R. S. Thurston, K. W. Mandernack, W. C. Shanks III, *Chem. Geol.* **269**, 252–261 (2010).
29. L. W. Juraneck, P. D. Quay, *Annu. Rev. Mar. Sci.* **5**, 503–524 (2013).
30. L. Y. Yeung, E. D. Young, E. A. Schauble, *J. Geophys. Res.* **117**, D18306 (2012).
31. B. M. Hoffman, D. Lukyanov, D. R. Dean, L. C. Seefeldt, *Acc. Chem. Res.* **46**, 587–595 (2013).
32. B. Kok, B. Forbush, M. McGloin, *Photochem. Photobiol.* **11**, 457–475 (1970).
33. T. Noguchi, *Phil. Trans. R. Soc. B* **363**, 1189–1195 (2008).
34. N. Cox et al., *Science* **345**, 804–808 (2014).

ACKNOWLEDGMENTS

We thank H. Hu and N. Levin for performing oxygen triple-isotope analyses of the terrarium water at Johns Hopkins University, and E. Schauble for helpful discussions during the course of this work. This research was supported in part by the National Science Foundation (EAR-1049655 and DGE-1144087), the National Aeronautics and Space Administration Cosmochemistry program, and the Deep Carbon Observatory. The data and model parameters used in this study are available in the supplementary materials (tables S1 to S3).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/431/suppl/DC1
 Materials and Methods
 Supplementary Text
 Figs. S1 to S5
 Tables S1 to S3
 References (35–50)

7 January 2015; accepted 13 March 2015
 10.1126/science.aaa6284

RESEARCH FUNDING

Big names or big ideas: Do peer-review panels select the best science proposals?

Danielle Li^{1*}† and Leila Agha^{2,3*}†

This paper examines the success of peer-review panels in predicting the future quality of proposed research. We construct new data to track publication, citation, and patenting outcomes associated with more than 130,000 research project (R01) grants funded by the U.S. National Institutes of Health from 1980 to 2008. We find that better peer-review scores are consistently associated with better research outcomes and that this relationship persists even when we include detailed controls for an investigator's publication history, grant history, institutional affiliations, career stage, and degree types. A one-standard deviation worse peer-review score among awarded grants is associated with 15% fewer citations, 7% fewer publications, 19% fewer high-impact publications, and 14% fewer follow-on patents.

In 2014, the combined budgets of the U.S. National Institutes of Health (NIH), the U.S. National Science Foundation, and the European Research Council totaled almost \$40 billion. The majority of these funds were allocated to external researchers whose applications were vetted by committees of expert reviewers. But as funding has become more competitive and application award probabilities have fallen, some observers have posited that “the system now favors those who can guarantee results rather than those with potentially path-breaking ideas that, by definition, cannot promise success” (*1*). Despite its importance for guiding research investments, there have been few attempts to assess the efficacy of peer review.

Peer-review committees are unique in their ability to assess research proposals based on deep expertise but may be undermined by biases, insufficient effort, dysfunctional committee dynamics, or limited subject knowledge (*2, 3*). Disagreement about what constitutes important research may introduce randomness into the process (*4*). Existing research in this area has focused on understanding whether there is a correlation between good peer-review scores and successful research outcomes and yields mixed results (*5–7*). Yet raw correlations do not reveal whether reviewers are generating insight about the scientific merit of proposals. For example, if applicants from elite institutions generally produce more highly cited research, then a system that rewarded institutional rankings without even reading applications may appear effective at identifying promising research.

In this paper, we investigate whether peer review generates new insights about the scientific quality of grant applications. We call this ability peer review’s “value-added.” The value-added of NIH peer review is conceptually distinct from the value of NIH funding itself. For example, even if reviewers did a poor job of identifying the best applications, receiving a grant may still improve a researcher’s productivity by allowing her to main-

tain a laboratory and support students. Whereas previous work has studied the impact of receiving NIH funds on the productivity of awardees (*8, 9*), our paper asks whether NIH selects the most promising projects to support. Because NIH cannot possibly fund every application it receives, the ability to distinguish potential among applications is important for its success.

We say that peer review has high value-added if differences in grants’ scores are predictive of differences in their subsequent research output, after controlling for previous accomplishments of the applicants. This may be the case if reviewers generate additional insights about an application’s potential, but peer review may also have zero or even negative value-added if reviewers are biased, mistaken, or focused on different goals (*10*).

Because research outcomes are often skewed, with many low-quality or incremental contributions and relatively few ground-breaking discoveries (*2, 11*), we assess the value-added of peer review for identifying research that is highly influential or shows commercial promise. We also test the effectiveness of peer review in screening out applications that result in unsuccessful research (see the supplementary materials for full details on data and methods).

NIH is the world’s largest funder of biomedical research (*12*). With an annual budget of approximately \$30 billion, it supports more than 300,000 research personnel at more than 2500 institutions (*12, 13*). A funding application is assigned by topic to one of approximately 200 peer-review committees (known as study sections).

Our main explanatory variable is the “percentile score,” ranging from 0 to 100, which reflects an application’s ranking among all other applications reviewed by a study section in a given fiscal year; lower scores correspond to higher-quality applications. In general, applications are funded in order of their percentile score until the budget of their assigned NIH institute is exhausted. The average score in our sample is 14.2, with a standard deviation (SD) of 10.2; only about 1% of funded grants in our sample had a score worse than 50. Funding has become more competitive in recent years; only 14% of applications were funded in 2013.

*Corresponding author. E-mail: dli@hbs.edu (D.L.); lagha@bu.edu (L.A.) †Both authors contributed equally to this work.

¹Harvard University, Cambridge, MA 02138, USA. ²Boston University, Boston, MA 02215, USA. ³National Bureau of Economic Research, Cambridge, MA 02138, USA.

Our sample consists of 137,215 research project (R01) grants funded from 1980 through 2008. R01s are project-based renewable grants that are NIH's primary grant mechanism, accounting for about half of its extramural grant spending. Of the grants in our sample, 56% are for new projects; the remaining successfully competed for renewal. We focus on funded grants because funding is likely to have direct effect on research productivity, making it difficult to infer the success of peer review by comparing funded and unfunded grants. Because our sample grants have the same funding status, we can attribute any remaining relationship between scores and outcomes to peer review, rather than funding. Because grants are almost always funded in order of their score, there is relatively little scope for selection on unobservables to introduce bias.

Our primary outcome variables are (i) the total number of publications that acknowledge grant support within 5 years of grant approval (via PubMed); (ii) the total number of citations that those publications receive through 2013 (via Web of Science); and (iii) patents that either directly cite NIH grant support or cite publications acknowledging grant support [via the *U.S. Patent and Trademark Office* (USPTO)]. These publication, citation, and patent outcomes are designed to reflect NIH's stated goals of rewarding research with high scientific and technical merit.

We also measure applicant-level characteristics: an investigator's publication and grant history, educational background, and institutional affiliation. We match investigators with publications using their full last name and their first and middle initials

(14). We track the number of articles an applicant published in the 5 years before submitting her application, as well as the impact of those publications as measured by the citations they have received by the time the application is evaluated. We identify "high-impact" publications as being among the top 0.1%, 1%, and 5% most cited, compared with articles published in the same year. To more precisely assess the quality of an applicant's ideas, we repeat this exercise for articles in which the applicant is a first or last author only. Our regression results include separate controls for each type of publication: any authorship position, and first or last author publications. By counting only citations received up to the date of grant review, we ensure that our measures contain only information available to reviewers at the time they evaluate the application.

Table 1. Do peer-review scores predict future citations and publications?

Each reported figure is the coefficient on scores from a single Poisson regression of grant outcomes on NIH peer-review scores; standard errors are reported in parentheses. The actual sample size used per regression depends on the number of nonzero observations for the dependent variable. The independent variable is the percentile score. "Future citations" refers to the total number of citations, to 2013, that accrue to all publications that acknowledge funding from a given grant. "Future publications" refers to the total number of such publications. Subject-year controls refer to study section

by fiscal year fixed effects, as well as NIH institute fixed effects. PI publication history includes controls for number of past publications, number of past citations, and number of past hit publications. PI career characteristics include controls for degrees and experience (time since highest degree). PI grant history controls for number of previous R01s and non-R01 NIH funding. PI institution and demographics control for the rank of the PI's institution, as well as gender and some ethnicity controls. Standard errors are clustered at the study section year level. *, statistical significance at the 10% level; **, 5% level; ***, 1% level.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
<i>Independent variable:</i>						
NIH percentile score	-0.0203*** (0.0006)	-0.0215*** (0.0008)	-0.0162*** (0.0007)	-0.0164*** (0.0007)	-0.0162*** (0.0007)	-0.0158*** (0.0007)
N	137,215	136,076	136,076	128,547	128,547	128,547
<i>Independent variable:</i>						
NIH percentile score	-0.0155*** (0.0003)	-0.0091*** (0.0003)	-0.0076*** (0.0003)	-0.0077*** (0.0003)	-0.0076*** (0.0003)	-0.0075*** (0.0003)
N	137,215	136,111	136,111	128,580	128,580	128,580
<i>Controls</i>						
Subject-year		X	X	X	X	X
PI publication history			X	X	X	X
PI career characteristics				X	X	X
PI grant history					X	X
PI institution/demographics						X

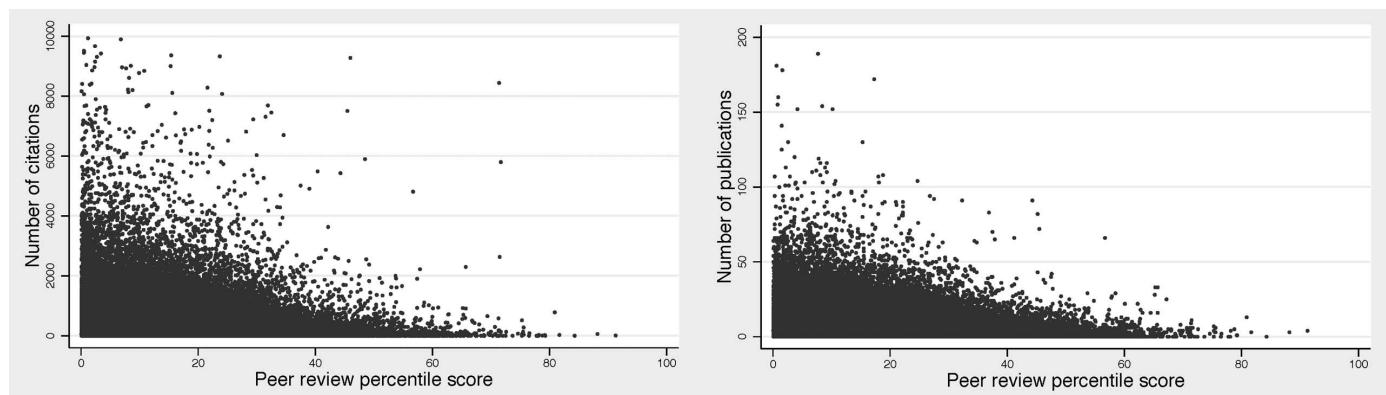


Fig. 1. Scatterplot of percentile scores and grant outcomes. The left panel plots the relationship between percentile scores and citations associated with a grant. Each dot represents a single grant. The right panel does the same for total publications. Extreme outliers with more than 10,000 citations or 200 publications are not displayed here.

We observe whether an applicant has an M.D., Ph.D., or both, as well as the year in which she received her final doctoral degree. We are missing degree and experience information for 0.45% and 7.16% of our sample, respectively; we include two separate indicators for missing these data. We measure whether this applicant previously received an R01 grant and whether the applicant has received any previous NIH funding. Using the name of the principal investigator (PI), we employ a probabilistic algorithm developed by Kerr to determine applicant gender and ethnicity (Hispanic or Asian) (15, 16, 17). We rank applicants' institutions by the number of NIH grants received over our study period and measure whether each applicant is from a top 5-, 10-, 20-, or 50-ranked institution. We are unable to determine the institutional affiliation of 14% of investigators; we include an indicator variable for missing institution information.

Consistent with previous work, there is substantial dispersion in research output even among

the relatively well-developed projects that receive NIH R01 funding (5). The median grant in our sample received 116 citations to publications acknowledging the grant; the mean is more than twice as high, 291, with an SD of 574. This variation in citations underscores the potential gains from being able to accurately screen grant applications on the basis of their research potential.

Our first set of results describes peer review's value-added for identifying research likely to result in many publications or citations. Table 1 reports results from Poisson regressions of future outcomes on peer-review scores, with different controls for an applicant's previous performance. The supplementary materials describe many additional robustness checks.

Model 1 of Table 1 reports, without any control variables, the percentage change in the number of citations and publications associated with a grant, given a one point increase in its percentile score. We find that NIH evaluations are statisti-

cally related to grant quality; our estimated coefficients indicate that a one percentile point worse peer-review score is associated with 1.6% fewer publications and 2% fewer citations. To consider the magnitude of these findings more clearly, we will describe our results by reporting how predicted outcomes change with a 1-SD (10.17 point) worse percentile score; in Model 1, a 1-SD worse score is associated with a 14.6% decrease in grant-supported research publications and a 18.6% decrease in citations to those publications ($P < 0.001$). This calculation is based on the overall SD in percentile score among funded grants, unconditional on PI characteristics (18). Figure 1 illustrates the raw relationship between scores and citations and publications in a scatterplot; the plot suggests a negative sloping relationship (recall that higher percentile scores indicate less favorably reviewed research).

There are potential concerns with interpreting the unadjusted relationship between scores and outcomes as a measure of peer review's value. Some grants may be expected to produce more citations or publications and thus appear higher quality, independent of their true quality. Older grants have more time to produce publications that in turn have more time to accrue citations. A publication with 100 citations may be average in one field but exceptional in another.

Model 2 of Table 1 addresses these concerns by including detailed fixed effects for study sections by year cells and NIH institutes. The inclusion of these fixed effects means that our estimates are based only on comparisons of scores and outcomes for grants evaluated in both the same fiscal year (to account for cohort effects) and in the same study section (to account for field effects). We also include NIH institute-level fixed effects to control for differences in citation and publication rates by fields, as defined by a grant's area of medical application. Controlling for cohort and field effects does not attenuate our main finding. For a 1-SD (10.17 point) worse score, we expect an 8.8% decrease in publications and a 19.6% decrease in citations (both $P < 0.001$). This suggests that scores for grants evaluated by the same study

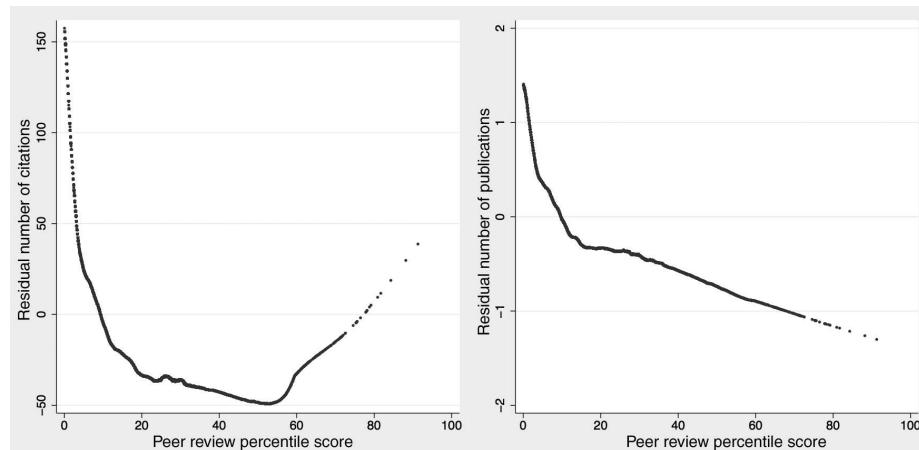


Fig. 2. Smoothed scatterplots of percentile scores and residual grant outcomes. These figures display smoothed scatterplots of the nonparametric relationship between unexplained variation in grant outcomes and percentile score, after accounting for differences in field of research, year, and applicant qualifications. The left panel plots the relationship between percentile scores and residual citations associated with a grant. The right panel does the same for residual publications.

Table 2. Do peer-review scores predict hit publications and follow-on patents? Each reported figure is the coefficient on scores from a single Poisson regression of grant outcomes on NIH peer-review scores; standard errors are in parentheses. High-impact publication is given by the count of publications acknowledging the grant that receive more citations than all but 0.1%, 1%, or 5% of publications from the same year. Direct patents are those that acknowledge funding from a grant; indirect patents are those that cite publications that acknowledge funding from a grant. We control for the same variables as described in Model 6 of Table 1.

	Dependent variable: High-impact publications			Dependent variable: Patents	
	Top 0.1% (1)	Top 1% (2)	Top 5% (3)	Direct (4)	Indirect (5)
Independent variable:					
NIH percentile score	-0.0246*** (0.0025)	-0.0209*** (0.0014)	-0.0172*** (0.0009)	-0.0153*** (0.0015)	-0.0149*** (0.0022)
N	88,795	118,245	125,021	122,850	92,893
Controls					
Subject-year	X	X	X	X	X
PI publication history	X	X	X	X	X
PI career characteristics	X	X	X	X	X
PI grant history	X	X	X	X	X
PI institution/demographics	X	X	X	X	X

section in the same year and assigned to the same NIH institute are better than randomly allocated.

We may observe this pattern, however, if reviewers simply give good scores to applicants with strong research credentials, and applicants with strong credentials generally tend to produce better research. Model 3 of Table 1 adds controls describing a PI's publication history in order to ask whether study section scores contain information about the quality of an application that could not be predicted by simply examining a PI's curriculum vita.

Specifically, we include the following additional control variables: (i) the number of articles published in the past 5 years; (ii) the total number of citations those articles have received up to the year of grant review; (iii) three variables describing the number of top 0.1%, 1%, and 5% articles that the PI has published in the previous 5 years; and (iv) alternate versions of these variables constructed only with the subset of publications for which the applicant was a first or last author. Controlling for publication history attenuates but does not eliminate the relationship: a 1-SD (10.17 point) worse score is associated with a 7.4% decrease in future publications and a 15.2% decrease in future citations (both $P < 0.001$).

The association between better scores and better outcomes could also be explained by the Matthew effect, a sociological phenomenon wherein credit and citations accrue to established investigators simply because they are established, regardless of the true quality of their work (19, 20). Were this the case, more connected applicants may receive better scores and more citations regardless of the true quality of their work. Our approach may thus credit peer review for responding to prestige, rather than the underlying quality of an applicant's ideas.

Model 4 controls for the PI's experience by adding indicators for whether the applicant has an M.D., Ph.D., or both, as well as a series of indicator variables capturing how many years have elapsed since receiving her terminal degree. If reviewers were simply giving better scores to candidates with more experience or skill writing grant proposals and publishing papers, then we would expect scores to become less predictive of future research output once we control for M.D./Ph.D. status and time since degree. Instead, our estimated relationship between peer-review scores and outcomes remains unchanged.

Model 5 considers the possibility that peer reviewers may be rewarding an applicant's grant

proposal writing skills rather than the underlying quality of her work. Specifically, we include variables controlling for whether the PI received NIH funding in the past, including four indicators for having previously received one R01 grant, two or more R01 grants, one NIH grant other than an R01, and two or more other NIH grants. To the extent that reviewers may be responding to an applicant's experience and skill with proposal writing, we would expect the inclusion of these variables reflecting previous NIH funding to attenuate our estimates of value-added. We find, however, that including these variables does not substantively affect our findings.

Finally, in Model 6, we also control for institutional quality, gender, and ethnicity, to capture other potentially unobserved aspects of prestige, connectedness, or access to resources that may influence review scores and subsequent research productivity. Our estimates again remain stable: comparing applicants with statistically identical backgrounds, the grant with a 1-SD worse score is predicted to have 7.3% fewer future publications and 14.8% fewer future citations (both $P < 0.001$).

Across Models 3 to 6, the estimated relationship between peer-review scores and outcomes remains remarkably stable, even as we add more covariates that describe an applicant's past accomplishments, prestige, proposal-writing skill, and professional connections. Although these variables certainly cannot capture every potential source of omitted variables bias, the stability of our results suggests that political connections and prestige are not a primary driver of peer review's value-added.

Next, we explore whether reviewers' expertise enables them to identify the strongest applications or to more efficiently screen out weaker applications. We use a local linear regression model to nonparametrically identify the relationship between peer-review score and research quality. This flexibility will allow the predictive power of peer-review scores to differ at each point along the score spectrum. We implement this approach in two steps, which are described in detail in the supplementary materials. First, we construct the residuals from a linear regression of research outcomes on all of the explanatory variables in Model 6, excluding the study section percentile score itself. These residuals represent the portions of grants' citations or publications that cannot be explained by applicants' previous qualifications or by application year or subject area (as detailed above). We then produce a locally weighted, linearly smoothed scatterplot relating peer-review scores to these residual citations and publications.

Figure 2 shows that peer reviewers add value by identifying the strongest research proposals. For all percentile scores less than 50 (the vast majority of awarded grants), worse scores are associated with lower expected residual citations and publications. The relationship is particularly steep at very low percentile scores, suggesting that study sections are particularly effective at discriminating quality among very well-reviewed applications.

One notable exception occurs for very poorly scored applications—those with percentile scores

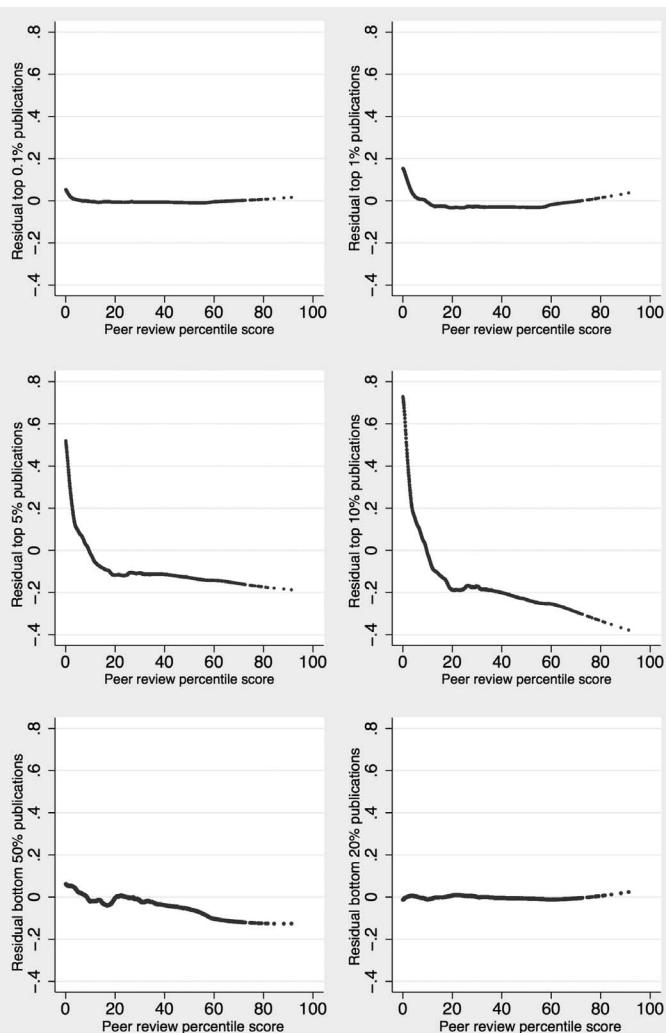


Fig. 3. Smoothed scatterplots of percentile scores and residual high- and low-citation publications. These figures display smoothed scatterplots of the nonparametric relationship between unexplained variation in grant outcomes and percentile score, after accounting for variation in field of research, year, and applicant qualifications. Each panel reports results on the number of residual publications in the indicated performance bin.

over 50—that were nonetheless funded. In this range, worse review scores are associated with higher citation counts. These applications constitute about 1% of funded applications and are highly unlikely to have met the standard award threshold but were instead funded “out of order.” We find higher average quality for this set of selected grants, suggesting that when program officers make rare exceptions to peer-review decisions, they are identifying a small fraction of applications that end up performing better than their initial scores would suggest.

Our final analysis explores whether peer reviewers’ value-added comes from being able to identify transformative science, science with considerable applied potential, or from being able to screen out very low-quality research. We define a “hit” publication as among the top 0.1%, 1%, or 5% most cited publications in its cohort, using all citations a publication receives through 2013. To explore whether reviewers have value-added in terms of identifying research with practical applications, we track the number of patents that explicitly acknowledge NIH funding. The majority of NIH grants, however, do not directly result in patents. Thus, we also count the number of patents that cite research funded by a grant (indirect patenting). We construct this variable by linking grants to publications using grant acknowledgement data and then applying a fuzzy matching algorithm that identifies publications cited by USPTO patents (21). This allows us to identify patents that cite publications that in turn acknowledge a grant. Importantly, this process (described further in the supplementary materials), allows us to identify patents regardless of whether those patents are assigned to the same investigator funded by the NIH grant. Indeed, most often these patents are held by private firms (22).

As reported in Table 2, peer-review scores have value-added identifying hit publications and research with commercial potential. A 1-SD (10.17 points) worse score is associated with a 22.1%, 19.1%, and 16.0% reduction in the number of top 0.1%, 1%, and 5% publications, respectively. These estimates are larger in magnitude than our estimates of value-added for overall citations, especially as we consider the very best publications. The large value-added for predicting tail outcomes suggests that peer reviewers are more likely to reward projects with the potential for a very high-impact publication and have considerable ability to discriminate among strong applications.

A 1-SD worse percentile score predicts a 14% decrease in both direct and indirect patenting. Because of the heterogeneous and potentially long lags between grants and patents, many grants in our sample may one day prove to be commercially relevant even if they currently have no linked patents. This time-series truncation makes it more difficult to identify value-added with respect to commercialization of research and means that our estimates are likely downward biased.

Finally, we investigate the nonparametric relationship between percentile scores and publication outcomes, testing which score ranges are associated with the highest numbers of “hit”

publications, ranking at the top of the citation distribution, and which score ranges are associated with the highest numbers of “miss” publications, ranking near the bottom of the distribution. We follow the same local linear regression smoothing procedure outlined above and described in more detail in the supplementary materials.

Figure 3 shows that low percentile scores are consistently associated with higher residual numbers of hit publications, variation unexplained by the applicant’s background or field of study. The relationship between scores and residual research outcomes is steepest among the most well-reviewed applications. For example, funded grants with percentile scores near 0 are predicted to produce 0.05 more publications in the top 0.1% of the citation distribution, compared with applications scored near the 10th percentile (holding constant applicant qualifications and field).

Although this may seem like a modest increase, there is a small number of such hit publications, so a 0.05 increase in their number corresponds to a doubling of the mean number of top 0.1% publications arising from a grant. This relationship between scores and hit publications becomes weaker among applications with less competitive scores; a 10-percentile point difference in scores in the range of 20 to 30 would predict only a 0.0004 difference in the number of top 0.1% publications. This finding runs counter to the hypothesis that, in light of shrinking budgets and lower application success rates, peer reviewers fail to reward those risky projects that are most likely to be highly influential in their field (1, 2).

We don’t find evidence that the peer-review system adds value beyond previous publications and qualifications in terms of screening out low-citation papers. Better percentile scores are associated with slightly more publications in the bottom 50% of the citation distribution. There is no discernible relationship between residual publications in the bottom 20% and peer-review score among the funded grants in our sample, suggesting that while these less influential anticipated publications are not rewarded by the peer-review system, they are also not specifically penalized.

Our findings demonstrate that peer review generates information about the quality of applications that may not be available otherwise. This does not mean that the current NIH review system would necessarily outperform other allocation mechanisms that do not rely on expert peer evaluations. Our analysis focuses on the relationship between scores and outcomes among funded grants; for that reason, we cannot directly assess whether the NIH systematically rejects high-potential applications. Our results, however, suggest that this is unlikely to be the case, because we observe a positive relationship between better scores and higher-impact research among the set of funded applications.

Although our findings show that NIH grants are not awarded purely for previous work or elite affiliations and that reviewers contribute valuable insights about the quality of applications, mistakes and biases may still detract from the quality of funding decisions. We have not included an

accounting of the costs of peer review, most notably the time investment of the reviewers. These bibliometric outcomes may not perfectly capture NIH objectives or be the only measures relevant for evaluating social welfare; ideally, we would like to link grants with health and survival outcomes, but constructing those measures is difficult and beyond the scope of this paper. Future research may focus on whether the composition of peer-review committees is important to determining their success, including evaluator seniority and the breadth and depth of committee expertise.

REFERENCES AND NOTES

- B. Alberts, M. W. Kirschner, S. Tilghman, H. Varmus, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5773–5777 (2014).
- D. F. Horrobin, *JAMA* **263**, 1438–1441 (1990).
- J. M. Campanario, *Sci. Commun.* **19**, 181–211 (1998).
- S. Cole, J. R. Cole, G. A. Simon, *Science* **214**, 881–886 (1981).
- J. Berg, Productivity metrics and peer review scores: NIGMS feedback loop blog (2011); <https://loop.nigms.nih.gov/2011/06/productivity-metrics-and-peer-review-scores/>.
- J. Berg, Productivity metrics and peer review scores, continued: NIGMS feedback loop blog (2011); <https://loop.nigms.nih.gov/2011/06/productivity-metrics-and-peer-review-scores-continued>.
- N. Dantithi, C. O. Wu, P. Shi, M. Lauer, *Circ. Res.* **114**, 600–606 (2014).
- B. A. Jacob, L. Lefgren, *Res. Policy* **40**, 864–874 (2011).
- B. A. Jacob, L. Lefgren, *J. Public Econ.* **95**, 1168–1177 (2011).
- J. H. Tanne, *BMJ* **319**, 336 (1999).
- K. Arrow, The rate and direction of inventive activity: Economic and social factors (National Bureau of Economic Research, Cambridge, MA, 1962), pp. 609–626.
- About NIH Web site (2014); <http://www.nih.gov/about/>.
- E. R. Dorsey et al., *JAMA* **303**, 137–143 (2010).
- There is no further disambiguation, but we show that our results do not change when we restrict to investigators with rare names. See table S5 of the supplementary materials.
- W. R. Kerr, The ethnic composition of US inventors, Working Paper 08-006, Harvard Business School (2008); <http://www.people.hbs.edu/wkerr/Kerr%20WP08.EthMatch.pdf>.
- W. R. Kerr, *Rev. Econ. Stat.* **90**, 518 (2008).
- Due to the limitations of the name-based matching algorithm, we cannot reliably distinguish African-American investigators.
- For example, to calculate the 14.6% figure, we take the exponential of our estimated coefficient times the SD in scores, minus 1: $\exp(-0.0155 \times 10.17) - 1$.
- R. K. Merton, *Science* **159**, 56–63 (1968).
- P. Azoulay, T. Stuart, Y. Wang, *Manage. Sci.* **60**, 92–109 (2013).
- P. Azoulay, J. S. G. Zivin, B. N. Sampat, The diffusion of scientific knowledge across time and space: Evidence from professional transitions for the superstars of medicine, Tech. Rep., National Bureau of Economic Research (NBER, Cambridge, MA, 2011).
- P. Azoulay, J. Graff-Zivin, D. Li, B. Sampat, Public R&D investments and private sector patenting: Evidence from NIH funding rules, NBER working paper 20889 (2013); <http://irps.ucsd.edu/assets/001/506033.pdf>.

ACKNOWLEDGMENTS

We are grateful to P. Azoulay, M. Lauer, Z. Obermeyer, and B. Sampat for helpful comments, suggestions, and assistance with data. We also acknowledge assistance from M.-C. Chen, P. Kennedy, A. Manning, and especially R. Nakamura from the NIH Center for Scientific Review. This paper makes use of restricted-access data available from the National Institutes of Health. Those wishing to replicate its results may apply for access following the procedures outlined in the NIH Data Access Policy document available at <http://report.nih.gov/pdf/DataAccessPolicy.pdf>.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/434/suppl/DC1

Materials and Methods

Fig. S1

Tables S1 to S8

References (23–29)

8 October 2014; accepted 18 March 2015
10.1126/science.aaa0185



Big names or big ideas: Do peer-review panels select the best science proposals?

Danielle Li and Leila Agha
Science **348**, 434 (2015);
DOI: 10.1126/science.aaa0185

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 23, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/434.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2015/04/22/348.6233.434.DC1.html>

A list of selected additional articles on the Science Web sites related to this article can be found at:

<http://www.sciencemag.org/content/348/6233/434.full.html#related>

This article cites 16 articles, 6 of which can be accessed free:

<http://www.sciencemag.org/content/348/6233/434.full.html#ref-list-1>

This article appears in the following subject collections:

Science and Policy

http://www.sciencemag.org/cgi/collection/sci_policy

VACCINES

An Ebola whole-virus vaccine is protective in nonhuman primates

Andrea Marzi,^{1*} Peter Halfmann,^{2*} Lindsay Hill-Batorski,² Friederike Feldmann,³ W. Lesley Shupert,¹ Gabriele Neumann,² Heinz Feldmann,¹ Yoshihiro Kawaoka^{2,4,5†}

Zaire ebolavirus is the causative agent of the current outbreak of hemorrhagic fever disease in West Africa. Previously, we showed that a whole Ebola virus (EBOV) vaccine based on a replication-defective EBOV (EBOVΔVP30) protects immunized mice and guinea pigs against lethal challenge with rodent-adapted EBOV. Here, we demonstrate that EBOVΔVP30 protects nonhuman primates against lethal infection with EBOV. Although EBOVΔVP30 is replication-incompetent, we additionally inactivated the vaccine with hydrogen peroxide; the chemically inactivated vaccine remained antigenic and protective in nonhuman primates. EBOVΔVP30 thus represents a safe, efficacious, whole-EBOV vaccine candidate that differs from other EBOV vaccine platforms in that it presents all viral proteins and the viral RNA to the host immune system, which might contribute to protective immune responses.

The Ebola virus (EBOV) outbreak in West Africa has already claimed more than 5000 lives (1) and remains uncontrolled. One countermeasure to mitigate Ebola virus infections is vaccination. Several Ebola virus vaccine platforms have been developed over the last decades (2), three of which recently advanced to clinical trials: a DNA-based vaccine expressing different Ebola virus glycoproteins (GPs, the major Ebola virus immunogen) (3, 4), a replication-incompetent chimpanzee adenovirus expressing GP (5), and a live-attenuated vesicular stomatitis virus (VSV) expressing GP (5). The DNA platform completely protects nonhuman primates (the “gold standard” for Ebola virus research) only after multiple dosages of the DNA vaccine in combination with recombinant adenovirus (6), but has not been tested as a standalone vaccination strategy. The recombinant adenovirus platform (including the recently developed recombinant chimpanzee adenovirus) requires high vaccine doses and boosting to achieve complete and durable protection of nonhuman primates against lethal challenge with EBOV (7, 8). Complete protection of nonhuman primates against lethal EBOV challenge has also been accomplished with the VSV platform; however, the use of a replicating recombinant VSV (9–12) may be of concern because of issues related to vaccine safety. Hence, although

several platforms are being tested in clinical trials, additional options should be explored.

Whole-virus vaccines (either live attenuated or inactivated) have a long history as successful human vaccines, offering protection against potentially deadly viral diseases such as smallpox, influenza, mumps, and measles (13). Whole-virus vaccines present multiple viral proteins and the viral genetic material to the host immune system, which may trigger a broader and more robust immune response than vectored vaccines that present only single viral proteins. However, initial attempts to develop a gamma-irradiated, inactivated whole-EBOV vaccine failed to provide robust protection of nonhuman primates against challenge with a lethal dose of EBOV (14).

Previously, we developed a replication-defective EBOV (termed EBOVΔVP30) which is based on the Mayinga strain of EBOV and lacks the coding region for the essential viral transcription activator, VP30 (15). EBOVΔVP30 replicates to high titers in cell lines that stably express the VP30 protein, is genetically stable, and is nonpathogenic in rodents (15, 16). Mice and guinea pigs immunized twice with EBOVΔVP30 were fully

protected against a lethal challenge with mouse- or guinea pig-adapted EBOV, respectively (16). EBOVΔVP30 is a biosafety level-3 agent and exempt from “Select Agent” status; an EBOVΔVP30 vaccine could therefore be manufactured in existing biosafety level-3 facilities that operate under good manufacturing practices.

To assess the effectiveness of EBOVΔVP30 whole-virus vaccine in nonhuman primates, we inoculated groups of cynomolgus macaques (Table 1) intramuscularly (i.m.) with Dulbecco’s modified essential medium (DMEM) (control, group 1), a single dose of 10^7 focus-forming units (FFU) of EBOVΔVP30 (group 2), or two doses of 10^7 FFU of EBOVΔVP30 4 weeks apart (group 3). Previously, we demonstrated the genomic stability of EBOVΔVP30 by carrying out three independent experiments that each comprised seven consecutive passages of the virus in VeroVP30 cells. After the last passages, we sequenced the region surrounding the VP30 deletion site and did not detect any recombination events or mutations. Moreover, the passaged viruses did not grow in wild-type cells, further indicating the lack of recombination. Despite these findings, concerns have been raised that such an event could potentially affect vaccine safety. Recently, virus inactivation with hydrogen peroxide was shown to preserve the antigenicity of lymphocytic choriomeningitis (17, 18), vaccinia (17), West Nile (17, 19), and influenza (20) viruses. To increase the biosafety profile of EBOVΔVP30, we therefore treated it with hydrogen peroxide (H_2O_2 , 3% final concentration) for 4 hours on ice, followed by viral plaque assays in VP30-expressing cells, which confirmed complete virus inactivation. Nonhuman primates were then vaccinated twice with 10^7 FFU of the H_2O_2 -treated EBOVΔVP30 (group 4; two animals). Gamma-irradiation is an established procedure for Ebola virus inactivation, but irradiation conditions optimized for virus inactivation (rather than for antigenic epitope preservation) may alter antigenicity and therefore protective efficacy of Ebola virus vaccines (14). To test these concepts, we also vaccinated macaques twice with 10^7 FFU of wild-type EBOV gamma-irradiated in BSL-4 containment (group 5); again, the irradiation conditions used here ensured virus inactivation, but

Table 1. Overview of vaccination and challenge strategy.

Group	Vaccine	Inactivation	Vaccination Prime	Vaccination Boost	Protection	Euthanasia
Group 1	Mock*	—	—	—	0% [†] (n = 4)	Days 7 to 8 [‡]
Group 2	EBOVΔVP30	—	1×10^7 FFU	—	100% (n = 4)	N/A [§]
Group 3	EBOVΔVP30	—	1×10^7 FFU	1×10^7 FFU	100% (n = 4)	N/A
Group 4	EBOVΔVP30	Hydrogen peroxide	1×10^7 FFU	1×10^7 FFU	100% (n = 2)	N/A
Group 5	EBOV	Gamma-irradiation	1×10^7 FFU	1×10^7 FFU	0% (n = 4)	Days 6 to 9

*These authors contributed equally to this work. †Corresponding author. E-mail: kawaokay@svm.vetmed.wisc.edu
*DMEM. †Percentage of animals that survived challenge with a lethal dose of EBOV. [‡]Days after challenge. [§]N/A, not applicable.

were not optimized to preserve antigenicity. None of the vaccinated animals showed signs of illness, confirming our earlier data from mice and guinea pigs that EBOVΔVP30 is nonpathogenic in animals (16).

Four weeks after the last immunization, we challenged animals in BSL-4 containment i.m. with a lethal dose (1000 FFU) of the heterologous Kikwit strain of EBOV. While control macaques in group 1 had to be euthanized on day 7 or 8 after challenge according to established and approved humane endpoint criteria (21) (Table 1), all animals immunized once (group 2) or twice (group 3) with the EBOVΔVP30 vaccine survived the lethal challenge (Table 1). In addition, both animals immunized twice with H₂O₂-treated EBOVΔVP30 vaccine (group 4) survived infection with wild-type EBOV, indicating that H₂O₂-treated EBOVΔVP30 is immunogenic and elicits protective immune responses. In contrast, all macaques immunized with gamma-irradiated wild-type EBOV (group 5) developed signs of severe EBOV disease and had to be euthanized between days 6 and 9 after challenge (Table 1), supporting the concept that gamma-irradiation optimized for virus inactivation alters the immunogenicity of EBOV vaccines. The macaques that had to be euthanized after challenge with EBOV (groups 1 and 5) had high virus titers in their blood after challenge (Fig. 1). In contrast, no viremia was detected in animals immunized twice with untreated (group 3) or H₂O₂-treated EBOVΔVP30 (group 4) (Fig. 1), showing that H₂O₂-treated EBOVΔVP30 elicited a protective immune response. One of four animals that received a single immunization with EBOVΔVP30 [nonhuman primate (NHP) 8 in group 2] was viremic on days 3 and 6 after challenge, but cleared the virus on day 9 (Fig. 1). In addition, a different animal in group 2 (NHP 7) had a fever on day 6 after challenge (table S1). These data indicate that a single vaccination with EBOVΔVP30 does not always prevent EBOV replication or signs of illness (fever), but does protect the host from death upon EBOV challenge. Together, our findings demonstrate the vaccine potential of a whole-EBOV vaccine based on EBOVΔVP30.

To better understand the correlates of protection, we measured the immune responses 2 and 4 weeks after the last immunization (i.e., 2 weeks and immediately before EBOV challenge). Two weeks after the last vaccination (day -14), macaques immunized twice with EBOVΔVP30 (group 3) had a high immunoglobulin G (IgG) antibody response to the viral GP based on a GP-specific enzyme-linked immunosorbent assay (ELISA) assay (Fig. 2). Two immunizations with H₂O₂-treated EBOVΔVP30 (group 4) resulted in a slightly lower, but still robust, immune response (Fig. 2). In macaques immunized once with EBOVΔVP30 (group 2), we detected a low, but measurable, IgG antibody response (Fig. 2). Serum samples from animals that succumbed to EBOV challenge—namely, those mock-immunized (group 1) or immunized twice with gamma-irradiated wild-type EBOV (group 5)—did not possess measurable IgG titers to GP (Fig. 2). The IgG titers to

EBOV GP on the day of challenge (day 0, Fig. 2) followed the same trend, but were low. The IgG titers to EBOV GP closely mirrored survival rates and virus titers (Table 1 and Fig. 1); these data indicate that immunization with EBOVΔVP30 elicits an antibody response to GP that is important for protection against EBOV infection. A similar correlation between a GP-specific antibody response and protection has been demonstrated with other experimental EBOV vaccine platforms (22, 23).

The antibody repertoire was further characterized by assessing the levels of neutralizing antibodies to GP as measured by plaque reduction neutralization (PRNT) assays. The serum dilution that reduced the titer of VSV-expressing EBOV GP by ≥50% (plaque reduction neutralization titer 50, PRNT₅₀) was 1:20 to 1:40 for samples obtained from animals immunized twice with EBOVΔVP30 (group 3; table S2); no statistically significant decline in neutralizing antibody levels was detected between day -14 (2 weeks before

challenge) and day 0 (table S2). In contrast, we detected slightly lower PRNT₅₀ titers of ~1:10 for sera obtained from animals immunized once with untreated or H₂O₂-treated EBOVΔVP30 (groups 2 or 4, respectively; table S2). No neutralizing antibodies were detected in control animals or animals immunized twice with gamma-irradiated EBOV (groups 1 or 5, respectively; table S2). Overall, the neutralizing antibody titers were low, but similar to those detected upon vaccination of animals with VSV-expressing EBOV GP (II).

Most experimental Ebola virus vaccine platforms provide only the viral GP as antigen, expressed from recombinant viruses or protein expression plasmids; in contrast, the EBOVΔVP30 vaccine presents all viral proteins plus the viral genetic material to the host. Early studies with EBOV-like particles (VLPs) suggested that the viral matrix protein (VP40) and nucleoprotein (NP) are also immunogenic (24), prompting us to carry out ELISAs specific for these two viral proteins. Two weeks after the last vaccination

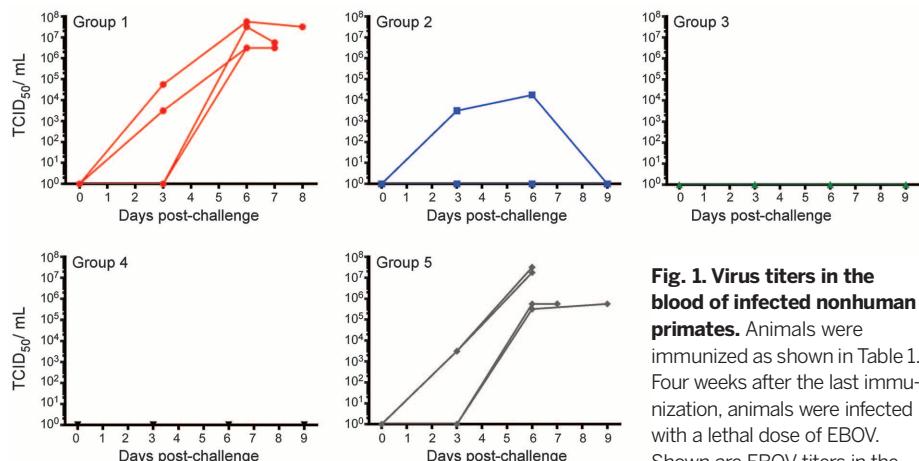


Fig. 1. Virus titers in the blood of infected nonhuman primates. Animals were immunized as shown in Table 1. Four weeks after the last immunization, animals were infected with a lethal dose of EBOV. Shown are EBOV titers in the blood of individual nonhuman primates from each group. Virus titers are shown as 50% tissue culture infective dose (TCID₅₀).

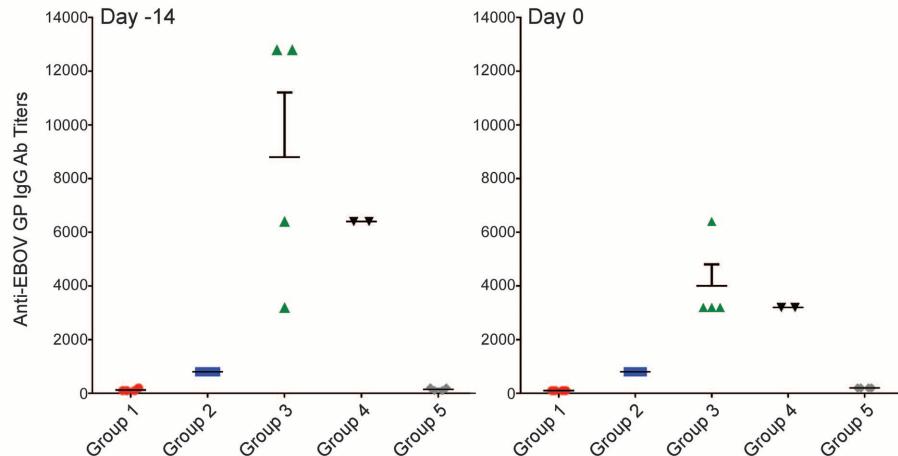


Fig. 2. Immune responses in vaccinated nonhuman primates. IgG antibody responses to EBOV GP 2 weeks after the last vaccination (day -14) and on the day of challenge (day 0). Antibody titers were measured with an ELISA specific for EBOV GP. Titers shown are the highest reciprocal dilution that resulted in an optical density (OD) of ≥0.2.

(day -14), macaques immunized twice with untreated (group 3) or H_2O_2 -treated (group 4) EBOV Δ VP30 had high NP and VP40 antibody titers (fig. S1). Lower, but still robust, NP and VP40 antibody titers were observed in macaques immunized once with EBOV Δ VP30 (group 1). Contrary to the GP antibody titers, we also detected NP and VP40 antibodies in animals immunized twice with gamma-irradiated EBOV (group 5), suggesting that gamma-irradiation under conditions optimized for virus inactivation has a greater effect on the antigenicity of GP epitopes than on that of NP and VP40 epitopes. Collectively, these data demonstrate that antibodies to NP and VP40 are elicited after vaccination with EBOV Δ VP30 and that the levels of these antibodies are higher in protected animals than in those that succumbed to infection. However, the importance of NP and VP40 antibodies to protection from EBOV infection is not yet known.

In addition to the antibody response, we also measured the cellular immune response by examining the number of mononuclear cells producing interferon- γ (IFN- γ). On day -14 (2 weeks before challenge), animals in groups 2 and 3, immunized one or twice with EBOV Δ VP30, respectively, had the highest number of IFN- γ -producing cells (fig. S2). Although treatment of EBOV Δ VP30 with H_2O_2 (group 4) reduced the number of IFN- γ -producing cells, more IFN- γ -producing cells were detected in these animals compared with those immunized twice with gamma-irradiated EBOV (group 5; fig. S2) or left untreated (group 1; fig. S2).

Data from Geisbert *et al.* (14) and our present findings suggest that gamma-irradiation optimized to inactivate EBOV destroys the antigenicity of wild-type EBOV, particularly in EBOV GP. H_2O_2 -treated EBOV Δ VP30, however, elicited a robust IgG response and protected nonhuman primates against lethal EBOV challenge, although H_2O_2 treatment resulted in a slight reduction of anti-

genicity compared with untreated virus (Fig. 2). Hence, H_2O_2 treatment of EBOV Δ VP30 appears to preserve key antigenic epitopes, as has been demonstrated for other viruses (17–20). To examine potential differences in antigenicity between gamma-irradiated and H_2O_2 -treated virus, we performed an ELISA-based assay, using a panel of 19 monoclonal antibodies (mAbs) directed against GP. Most mAbs showed levels of binding comparable to that of GP; however, four (mAbs 12, 21, 226, and 662) reacted more efficiently with H_2O_2 -treated than with gamma-irradiated virus (Fig. 3). Most likely, gamma-irradiation affected the conformation of the epitopes recognized by these antibodies, resulting in the lack of protection upon immunization with gamma-irradiated virus. Hence, the epitopes recognized by mAbs 12, 21, 226, and 662 may play an important role in antibody-mediated protection in immunized macaques and potentially in humans; indeed, mAb 226 is known to have virus-neutralizing properties (25). One mAb (1031) interacted more efficiently with gamma-irradiated than with H_2O_2 -treated virus, while a polyclonal antiserum reacted similarly with both virus preparations tested (Fig. 3).

When EBOV was first discovered over 35 years ago, whole-virus vaccines inactivated by formalin or gamma-irradiation were tested, but failed to elicit complete protection in nonhuman primates (14). The development of whole-virus vaccines was therefore abandoned, and VLPs composed of GP and VP40 (and NP) were explored as a safe and immunogenic platform to present several viral proteins to the host immune system (2, 26–28). These VLPs are immunogenic, but three vaccinations with adjuvanted VLPs were required to achieve protective efficacy in nonhuman primates (24). Here, we present a vaccine strategy that offers several advantages: (i) It provides protection from a lethal challenge of EBOV in nonhuman primates after a single immunization, although one animal became viremic and another animal

developed a fever; (ii) it is highly immunogenic, as shown by robust antibody responses elicited upon vaccination; (iii) it is amenable to large-scale production, because EBOV Δ VP30 grows to titers of $>10^7$ FFU/ml in VP30-expressing cells (15); (iv) it is safe, owing to its inability to replicate outside VP30-expressing cells (15); and (v) it presents all viral proteins and its genomic RNA to the host, similar to whole-virus vaccines and VLPs. It should be noted that NHPs immunized once with EBOV Δ VP30 (group 2) were protected from a lethal EBOV challenge, although two of the four animals showed signs of illness (fever was detected in NHP 7, and viremia was detected in NHP 8; table S1). However, all four animals in group 2 (NHPs 5 to 8) showed similar immune responses (table S2 and summarized in table S3).

To address any potential concerns over recombination events that would restore the replicative ability of EBOV Δ VP30, we also chemically inactivated it with H_2O_2 . Hydrogen peroxide treatment causes breaks in single- and double-stranded DNA or RNA (17) and thus inactivates viruses without affecting their antigenicity. By contrast, gamma-irradiation (used to generate the first experimental whole EBOV vaccine) causes the (de) hydroxylation of amino acids, the cleavage of polypeptide backbones (29), and the generation of free radicals that could cause the destruction of the antigenic properties of some epitopes. These differences in mechanism may explain why viruses treated with H_2O_2 are more immunogenic than those irradiated with gamma rays; however, optimization of irradiation conditions may improve the immunogenicity of vaccine candidates.

In summary, our data indicate that EBOV Δ VP30 is an effective whole-EBOV vaccine that warrants further assessment.

REFERENCES AND NOTES

- www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/index.html
- A. Marzi, H. Feldmann, *Expert Rev. Vaccines* **13**, 521–531 (2014).
- U. N. Sarwar *et al.*, *J. Infect. Dis.* **211**, 549–557 (2015).
- J. E. Martin *et al.*, *Clin. Vaccine Immunol.* **13**, 1267–1277 (2006).
- www.niaid.nih.gov/news/newsreleases/2014/Pages/EbolaVaxCandidate.aspx.
- N. J. Sullivan, A. Sanchez, P. E. Rollin, Z. Y. Yang, G. J. Nabel, *Nature* **408**, 605–609 (2000).
- N. J. Sullivan *et al.*, *Nat. Med.* **17**, 1128–1131 (2011).
- D. A. Stanley *et al.*, *Nat. Med.* **20**, 1126–1129 (2014).
- T. W. Geisbert *et al.*, *J. Virol.* **83**, 7296–7304 (2009).
- T. W. Geisbert *et al.*, *Vaccine* **26**, 6894–6900 (2008).
- S. M. Jones *et al.*, *Nat. Med.* **11**, 786–790 (2005).
- X. Qiu *et al.*, *PLOS ONE* **4**, e5547 (2009).
- I. J. Amanna, M. K. Slifka, *Antiviral Res.* **84**, 119–130 (2009).
- T. W. Geisbert *et al.*, *Emerg. Infect. Dis.* **8**, 503–507 (2002).
- P. Halfmann *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1129–1133 (2008).
- P. Halfmann *et al.*, *J. Virol.* **83**, 3810–3815 (2009).
- I. J. Amanna, H. P. Raué, M. K. Slifka, *Nat. Med.* **18**, 974–979 (2012).
- J. M. Walker, H. P. Raué, M. K. Slifka, *J. Virol.* **86**, 13735–13744 (2012).
- A. K. Pinto *et al.*, *J. Virol.* **87**, 1926–1936 (2013).
- J. L. Dembinski *et al.*, *J. Virol. Methods* **207**, 232–237 (2014).
- D. L. Brining *et al.*, *Comp. Med.* **60**, 389–395 (2010).
- A. Marzi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1893–1898 (2013).
- J. E. Blaney *et al.*, *PLOS Pathog.* **9**, e1003389 (2013).
- K. L. Warfield *et al.*, *J. Infect. Dis.* **196** (suppl. 2), S430–S437 (2007).
- A. Takada *et al.*, *J. Virol.* **77**, 1069–1074 (2003).
- K. L. Warfield *et al.*, *J. Infect. Dis.* **196** (suppl. 2), S421–S429 (2007).

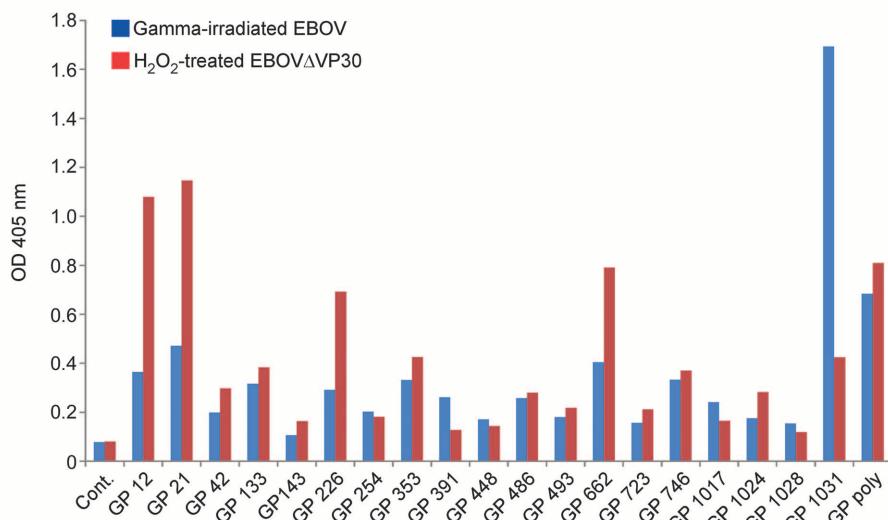


Fig. 3. Effects of H_2O_2 -treatment and gamma-irradiation on the antigenicity of EBOV GP. Using a panel of 19 mAbs (1 μ g/ml) directed against EBOV GP, we performed an ELISA to examine the antigenicity of gamma-irradiated EBOV (blue) and H_2O_2 -treated EBOV Δ VP30 (red).

27. D. L. Swenson et al., *Vaccine* **23**, 3033–3042 (2005).
 28. K. L. Warfield et al., *J. Immunol.* **175**, 1184–1191 (2005).
 29. E. S. Kempner, *J. Pharm. Sci.* **90**, 1637–1646 (2001).

ACKNOWLEDGMENTS

We thank E. Ollmann-Saphire (Scripps Research Institute, La Jolla, CA) for purified EBOV NP. We also thank S. Watson for editing the manuscript, T. Armbrust for excellent technical assistance, and staff of the Rocky Mountain Veterinary Branch for assistance with animal work. Y.K. and G.N. are inventors on a patent (held by the

University of Wisconsin Alumni Research Foundation) for EBOV reverse genetics; therefore, a Material Transfer Agreement (MTA) is required to obtain this system. Funding for this research was provided by the Region V “Great Lakes” Regional Center of Excellence (GLRCE; U54 AI 5153) and by Health and Labour Sciences Research Grants, Japan. The study was partially funded by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, NIH. Raw data can be found at https://docs.google.com/spreadsheets/d/1dBgzt5_z4rp-qOuxXcl_FbUz8wNqMvHy6kV_PtWOMY/edit?usp=sharing.

SUPPLEMENTARY MATERIALS

www.science.org/content/348/6233/439/suppl/DC1
 Materials and Methods
 Tables S1 to S3
 Figs. S1 and S2

14 December 2014; accepted 13 March 2015
 Published online 26 March 2015;
 10.1126/science.aaa4919

GENOME EDITING

The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations

Valentino M. Gantz* and Ethan Bier*

An organism with a single recessive loss-of-function allele will typically have a wild-type phenotype, whereas individuals homozygous for two copies of the allele will display a mutant phenotype. We have developed a method called the mutagenic chain reaction (MCR), which is based on the CRISPR/Cas9 genome-editing system for generating autocatalytic mutations, to produce homozygous loss-of-function mutations. In *Drosophila*, we found that MCR mutations efficiently spread from their chromosome of origin to the homologous chromosome, thereby converting heterozygous mutations to homozygosity in the vast majority of somatic and germline cells. MCR technology should have broad applications in diverse organisms.

It is often desirable to generate recessive loss-of-function mutations in emergent model organisms; however, identifying such mutations in the heterozygous condition is challenging. Taking advantage of the CRISPR/

Cas9 genome-editing method (1, 2), we have developed a strategy to convert a *Drosophila* heterozygous recessive mutation into a homozygous condition manifesting a mutant phenotype. We reasoned that autocatalytic insertional mutants

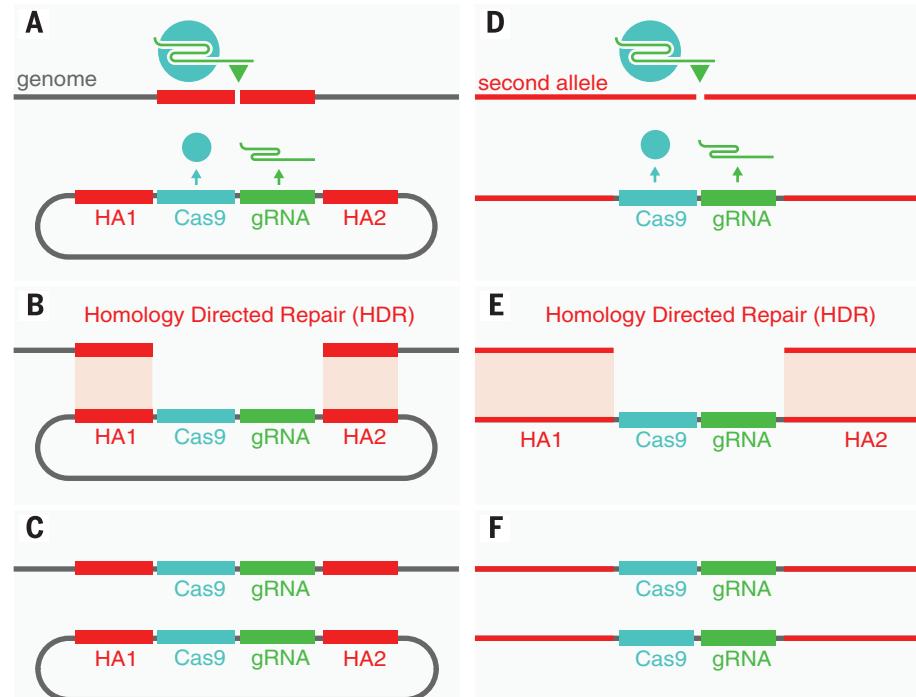
could be generated with a construct having three components: (i) A Cas9 gene (expressed in both somatic and germline cells), (ii) a guide RNA (gRNA) targeted to a genomic sequence of interest, and (iii) homology arms flanking the Cas9-gRNA cassettes that match the two genomic sequences immediately adjacent to either side of the target cut site (Fig. 1A). In such a tripartite construct, Cas9 should cleave the genomic target at the site determined by the gRNA (Fig. 1A) and then insert the Cas9-gRNA cassette into that locus via homology-directed repair (HDR) (Fig. 1, B and C). Cas9 and the gRNA produced from the insertion allele should then cleave the opposing allele (Fig. 1D), followed by HDR-driven propagation of the Cas9-gRNA cassette to the companion chromosome (Fig. 1, E and F). We refer to this trans-acting mutagenesis scheme as a mutagenic chain reaction (MCR).

We expected that autocatalytic allelic conversion by MCR should be very efficient in both somatic and germline precursor cells, given the high frequency and specificity of mutagenesis (3)

Section of Cell and Developmental Biology, University of California, San Diego, La Jolla, CA 92095, USA.

*Corresponding author. E-mail: vgantz@ucsd.edu (V.M.G.); ebier@ucsd.edu (E.B.)

Fig. 1. Scheme outlining the mutagenic chain reaction (MCR). (A to C) A plasmid consisting of a core cassette carrying a Cas9 transgene, a gRNA targeting a genomic sequence of interest, and flanking homology arms corresponding to genomic sequences abutting the target cleavage site (A) inserts the core Cas9-gRNA cassette into the targeted locus via HDR [(B) and (C)]. (D to F) In turn, the inserted cassette expresses both Cas9 and the gRNA, leading to cleavage (D) and HDR-mediated insertion of the cassette into the second allele, thereby rendering the mutation homozygous [(E) and (F)]. HA1 and HA2 denote the two homology arms that directly flank the gRNA-directed cut site.





An Ebola whole-virus vaccine is protective in nonhuman primates
 Andrea Marzi *et al.*
Science **348**, 439 (2015);
 DOI: 10.1126/science.aaa4919

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 23, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/439.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2015/03/25/science.aaa4919.DC1.html>

This article cites 27 articles, 12 of which can be accessed free:

<http://www.sciencemag.org/content/348/6233/439.full.html#ref-list-1>

This article appears in the following subject collections:

Epidemiology

<http://www.sciencemag.org/cgi/collection/epidemiology>

27. D. L. Swenson et al., *Vaccine* **23**, 3033–3042 (2005).
 28. K. L. Warfield et al., *J. Immunol.* **175**, 1184–1191 (2005).
 29. E. S. Kempner, *J. Pharm. Sci.* **90**, 1637–1646 (2001).

ACKNOWLEDGMENTS

We thank E. Ollmann-Saphire (Scripps Research Institute, La Jolla, CA) for purified EBOV NP. We also thank S. Watson for editing the manuscript, T. Armbrust for excellent technical assistance, and staff of the Rocky Mountain Veterinary Branch for assistance with animal work. Y.K. and G.N. are inventors on a patent (held by the

University of Wisconsin Alumni Research Foundation) for EBOV reverse genetics; therefore, a Material Transfer Agreement (MTA) is required to obtain this system. Funding for this research was provided by the Region V “Great Lakes” Regional Center of Excellence (GLRCE; U54 AI 5153) and by Health and Labour Sciences Research Grants, Japan. The study was partially funded by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, NIH. Raw data can be found at https://docs.google.com/spreadsheets/d/1dBgzt5_z4rp-qOuxXcl_FbUz8wNqMvHy6kVP_tpWOMY/edit?usp=sharing.

SUPPLEMENTARY MATERIALS

www.science.org/content/348/6233/439/suppl/DC1
 Materials and Methods
 Tables S1 to S3
 Figs. S1 and S2

14 December 2014; accepted 13 March 2015
 Published online 26 March 2015;
 10.1126/science.aaa4919

GENOME EDITING

The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations

Valentino M. Gantz* and Ethan Bier*

An organism with a single recessive loss-of-function allele will typically have a wild-type phenotype, whereas individuals homozygous for two copies of the allele will display a mutant phenotype. We have developed a method called the mutagenic chain reaction (MCR), which is based on the CRISPR/Cas9 genome-editing system for generating autocatalytic mutations, to produce homozygous loss-of-function mutations. In *Drosophila*, we found that MCR mutations efficiently spread from their chromosome of origin to the homologous chromosome, thereby converting heterozygous mutations to homozygosity in the vast majority of somatic and germline cells. MCR technology should have broad applications in diverse organisms.

It is often desirable to generate recessive loss-of-function mutations in emergent model organisms; however, identifying such mutations in the heterozygous condition is challenging. Taking advantage of the CRISPR/

Cas9 genome-editing method (1, 2), we have developed a strategy to convert a *Drosophila* heterozygous recessive mutation into a homozygous condition manifesting a mutant phenotype. We reasoned that autocatalytic insertional mutants

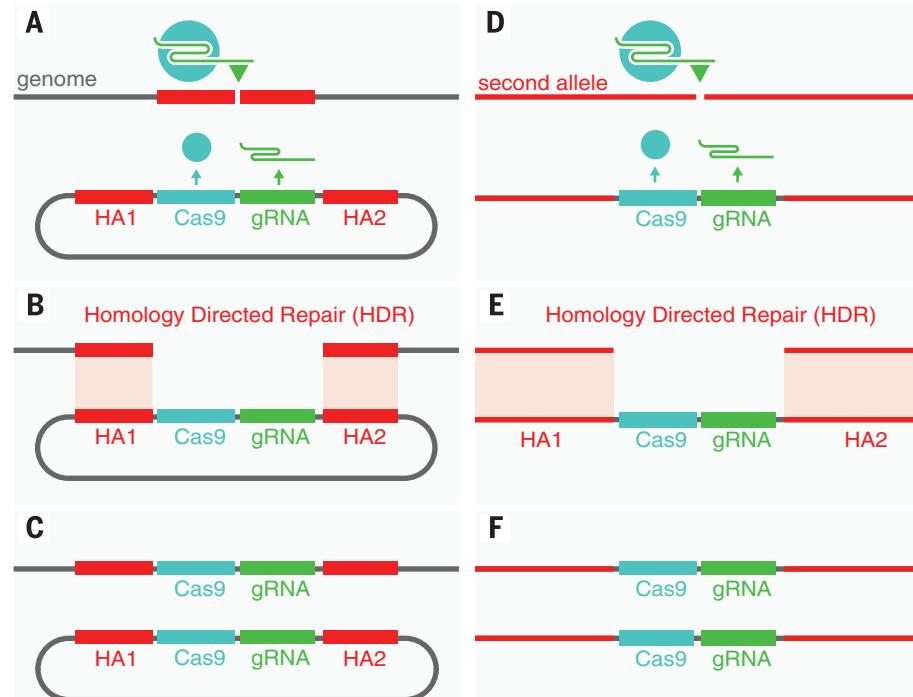
could be generated with a construct having three components: (i) A Cas9 gene (expressed in both somatic and germline cells), (ii) a guide RNA (gRNA) targeted to a genomic sequence of interest, and (iii) homology arms flanking the Cas9-gRNA cassettes that match the two genomic sequences immediately adjacent to either side of the target cut site (Fig. 1A). In such a tripartite construct, Cas9 should cleave the genomic target at the site determined by the gRNA (Fig. 1A) and then insert the Cas9-gRNA cassette into that locus via homology-directed repair (HDR) (Fig. 1, B and C). Cas9 and the gRNA produced from the insertion allele should then cleave the opposing allele (Fig. 1D), followed by HDR-driven propagation of the Cas9-gRNA cassette to the companion chromosome (Fig. 1, E and F). We refer to this trans-acting mutagenesis scheme as a mutagenic chain reaction (MCR).

We expected that autocatalytic allelic conversion by MCR should be very efficient in both somatic and germline precursor cells, given the high frequency and specificity of mutagenesis (3)

Section of Cell and Developmental Biology, University of California, San Diego, La Jolla, CA 92095, USA.

*Corresponding author. E-mail: vgantz@ucsd.edu (V.M.G.); ebier@ucsd.edu (E.B.)

Fig. 1. Scheme outlining the mutagenic chain reaction (MCR). (A to C) A plasmid consisting of a core cassette carrying a Cas9 transgene, a gRNA targeting a genomic sequence of interest, and flanking homology arms corresponding to genomic sequences abutting the target cleavage site (A) inserts the core Cas9-gRNA cassette into the targeted locus via HDR [(B) and (C)]. (D to F) In turn, the inserted cassette expresses both Cas9 and the gRNA, leading to cleavage (D) and HDR-mediated insertion of the cassette into the second allele, thereby rendering the mutation homozygous [(E) and (F)]. HA1 and HA2 denote the two homology arms that directly flank the gRNA-directed cut site.



and efficacy of homology-based integration (4) mediated by separate genome-encoded Cas9 and gRNA genes observed in previous studies. We tested this prediction in *D. melanogaster* with the use of a characterized efficient target sequence (*y1*) (5) in the X-linked *yellow* (*y*) locus as the gRNA target and a *vasa-Cas9* transgene as a source of Cas9 (Fig. 2C) because it is expressed in both germline and somatic cells (4). As the defining element of our MCR scheme, we also included two homology arms, ~1 kb each, flanking the central elements (Fig. 2C) that precisely abut the gRNA-directed cut site. Wild-type (*y+*) embryos were injected with the *y*-MCR element (see supplementary materials), and emerging F₀ flies were crossed to a *y+* stock. According to Mendelian inheritance, all F₁ female progeny of such a cross should have a *y+* phenotype (i.e., F₁ females inherit a *y+* allele from their wild-type parent).

From two independent F₀ male (δ) \times *y+* female (φ) crosses and 7 F₀ φ \times *y+* δ crosses, we recovered *y-* F₁ φ progeny, which should not happen according to Mendelian inheritance of a recessive allele. Six such *y*^{MCR} F₁ φ were crossed

individually to *y+* δ , resulting in 95 to 100% (average = 97%) of their F₂ progeny exhibiting a full-bodied *y-* phenotype (Fig. 2, E and G, and table S1), in contrast to the expected rate of 50% (i.e., only in males). We similarly tested MCR transmission via the germline in two *y-* F₁ δ recovered from an F₀ φ cross that also yielded *y-* female siblings. These *y-* F₁ δ were considered candidates for carrying the *y*-MCR construct and were crossed to *y+* females. All but one of their F₂ female progeny had a full-bodied *y-* phenotype (Fig. 2, E and F). Occasionally among *y*^{MCR} F₂ φ we also recovered mosaics (~4%) with a few small *y+* patches as well as a lone example of a 50% chimeric female (Fig. 2H), and in two instances, we recovered *y+* male progeny from a *y*^{MCR} F₁ φ mother (Fig. 2E and table S1). These infrequent examples of imperfect *y*-MCR transmission indicate that although HDR is highly efficient at this locus in both somatic and germline lineages, the target occasionally evades conversion.

Polymerase chain reaction (PCR) analysis of the *y* locus in individual *y-* F₁ progeny confirmed the precise gRNA- and HDR-directed genomic insertion of the *y*-MCR construct in all flies giving

rise to *y-* female F₂ progeny (Fig. 2D). Males carried only this single allele, as expected, whereas females in addition possessed a band corresponding to the size of the wild-type *y* locus (Fig. 2D, lane 4), which varied in intensity between individuals, indicating that females were mosaic for MCR conversion. The left and right *y*-MCR PCR junction fragments were sequenced from *y-* F₁ progeny from five independent F₀ parents. All had the precise expected HDR-driven insertion of the *y*-MCR element into the chromosomal *y* locus. In addition, sequence analysis of a rare nonconverted *y+* allele recovered in a male offspring from a *y*^{MCR} F₁ φ (Fig. 2E) revealed a single-nucleotide change at the gRNA cut site (resulting in a T→I substitution), which most likely resulted from nonhomologous end-joining repair, as well as an in-frame insertion-deletion (indel) in a *y+* φ sibling of this male (fig. S1 and table S1). The high recovery rate of full-bodied *y-* F₁ and F₂ female progeny from single parents containing a *y*^{MCR} allele detectable by PCR indicates that the conversion process is remarkably efficient in both somatic and germline lineages. Phenotypic evidence of mosaicism in a small percentage of MCR-carrying females and the presence of *y* locus-derived PCR products of wild-type size in all tested *y-* F₁ females suggest that females may all be mosaic to varying degrees. In summary, both genetic and molecular data reveal that the *y*-MCR element efficiently drives allelic conversion in somatic and germline lineages.

MCR technology should be applicable to different model systems and a broad array of situations, such as enabling mutant F₁ screens in pioneer organisms, accelerating genetic manipulations and genome engineering, providing a potent gene drive system for delivery of transgenes in disease vector or pest populations, and potentially serving as a disease-specific delivery system for gene therapy strategies. We provide an example in this study of an MCR element causing a viable insertional mutation within the coding region of a gene. It should also be possible, however, to efficiently generate viable deletions of coding or noncoding DNA by including two gRNAs in the MCR construct targeting separated sequences and appropriate flanking homology arms. Using the simple core elements tested in this study, MCR is applicable to generating homozygous viable mutations, creating regulatory mutations of essential genes, or targeting other nonessential sequences. The method may also be adaptable to targeting essential genes if an in-frame recoded gRNA-resistant copy of the gene providing sufficient activity to support survival is included.

In addition to these positive applications of MCR technology, we are also keenly aware of the substantial risks associated with this highly invasive method. Failure to take stringent precautions could lead to the unintentional release of MCR organisms into the environment. The supplementary material includes a stringent, institutionally approved barrier containment protocol that we developed and are currently adhering to for MCR experiments. Since this study was

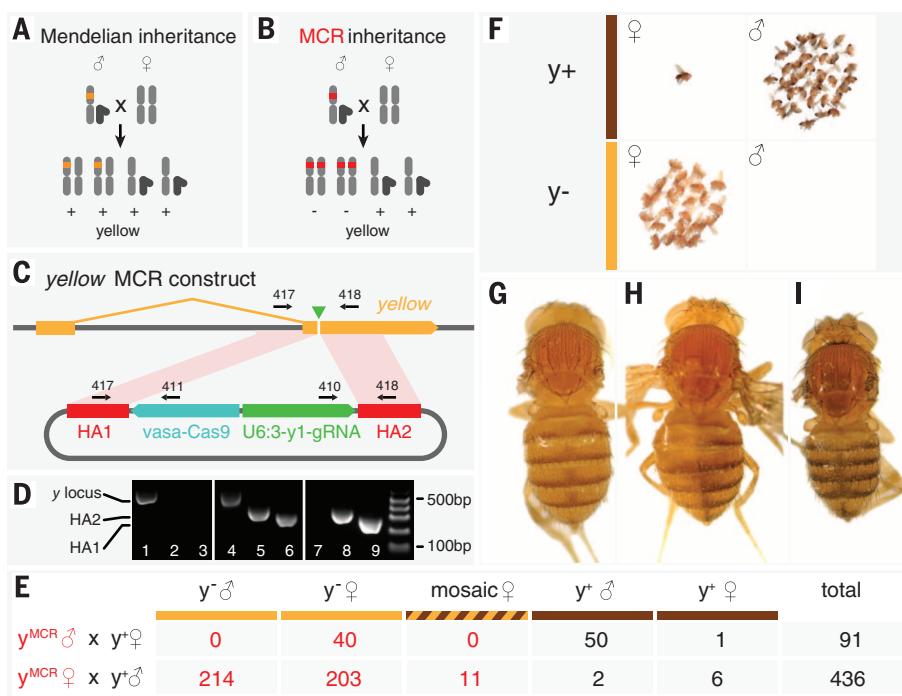


Fig. 2. Experimental demonstration of MCR in *Drosophila*. (A) Mendelian male inheritance of an X-linked trait. (B) Theoretical MCR-based inheritance results in the initially heterozygous allele converting the second allele, thereby generating homozygous female progeny. (C) Diagram of *y*-MCR construct. Two *y* locus homology arms flanking the *vasa-Cas9* and *y*-gRNA transgenes are indicated, as are the locations of the PCR primers used for analysis of the genomic insertion site (see supplementary materials). (D) PCR analysis of *y+* MCR-derived F₂ δ (lanes 1 to 3; see fig. S1 for sequence), *y*^{MCR} F₁ φ (lanes 4 to 6), and *y*^{MCR} F₁ δ (lanes 7 to 9) showing junctional bands corresponding to *y*-MCR insertion into the chromosomal *y* locus (lanes 2, 3, 5, 6, 8, and 9) and the presence (lanes 1 and 4) or absence (lane 7) of a PCR band derived from the *y* locus. Although the *y*^{MCR} F₁ δ (carrying a single X chromosome) displays only MCR-derived PCR products (lanes 8 and 9), *y*^{MCR} F₁ φ s generate both MCR and noninsertional amplification products. (E) Summary of F₂ progeny obtained from crosses described in table S1. (F) Low-magnification view of F₂ progeny flies from an *y*^{MCR} δ \times *y+* φ cross. Nearly all female progeny display a *y-* phenotype. (G) High-magnification view of a full-bodied *y*^{MCR} F₁ φ . (H) A rare 50% left-right mosaic female. (I) A *y+* control fly.

submitted for publication, a preprint has been posted on the bioRxiv web server showing that a split Cas9-gRNA gene drive system efficiently biases inheritance in yeast (6). The split system was used to avoid accidental escape of the gene drives. The use of a similar strategy in future MCR organisms would reduce, but not eliminate, risks associated with accidental release. We therefore concur with others (7, 8) that a dialogue on this topic should become an immediate high-priority issue. Perhaps, by analogy to the famous Asilomar meeting of 1975 that assessed the risks of recombinant DNA technology, a similar conference could be convened to consider biosafety measures and institutional policies appropriate for limiting the risk of engaging in MCR research while affording workable opportunities for positive applications of this concept.

REFERENCES AND NOTES

1. F. Zhang, Y. Wen, X. Guo, *Hum. Mol. Genet.* **23**, R40–R46 (2014).
2. P. D. Hsu, E. S. Lander, F. Zhang, *Cell* **157**, 1262–1278 (2014).
3. F. Port, H. M. Chen, T. Lee, S. L. Bullock, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2967–E2976 (2014).
4. S. J. Gratz et al., *Genetics* **196**, 961–971 (2014).
5. A. R. Bassett, C. Tibbit, C. P. Ponting, J. L. Liu, *Cell Rep.* **4**, 220–228 (2013).
6. J. E. DiCarlo, A. Chavez, S. L. Dietz, K. M. Esvelt, G. M. Church, <http://biorex.org/content/early/2015/01/16/013896> (2015).
7. K. A. Oye et al., *Science* **345**, 626–628 (2014).
8. K. M. Esvelt, A. L. Smidler, F. Catteruccia, G. M. Church, *eLife* **10.7554/eLife.03401** (2014).

ACKNOWLEDGMENTS

We thank M. Yanofsky, W. McGinnis, S. Wasserman, R. Kolodner, H. Bellen, and members of the Bier lab for helpful discussions and comments on the manuscript; M. Harrison, K. O'Connor-Giles, J. Wildonger, and S. Bullock for providing CRISPR/Cas9 reagents and information; and J. Vinetz and A. Lubar for granting us

access to their BSL2 Insectary. Supported by NIH grants R01 GM067247 and R56 NS029870 and by a generous gift from S. Sandell and M. Marshall. E.B. and V.M.G. are authors on a patent applied for by the University of California, San Diego (provisional patent application number 62075534) that relates to the mutagenic chain reaction. MCR fly stocks and DNA constructs are available from E.B. under a material transfer agreement from UCSD. This protocol for use and containment of our MCR stocks in a BSL2 barrier insectary also used for containment of malaria-infected mosquitoes was reviewed and approved by the UCSD Institutional Biosafety Committee (BUA R461).

SUPPLEMENTARY MATERIALS

www.science.org/content/348/6233/442/suppl/DC1
Materials and Methods
Supplementary Text
Fig. S1
Table S1

31 December 2014; accepted 10 March 2015
Published online 19 March 2015;
10.1126/science.aaa5945

PROTEIN FOLDING

Translational tuning optimizes nascent protein folding in cells

Soo Jung Kim,¹ Jae Seok Yoon,¹ Hideki Shishido,¹ Zhongying Yang,¹ LeeAnn A. Rooney,¹ Jose M. Barral,^{2,3} William R. Skach^{1,4*}

In cells, biosynthetic machinery coordinates protein synthesis and folding to optimize efficiency and minimize off-pathway outcomes. However, it has been difficult to delineate experimentally the mechanisms responsible. Using fluorescence resonance energy transfer, we studied cotranslational folding of the first nucleotide-binding domain from the cystic fibrosis transmembrane conductance regulator. During synthesis, folding occurred discretely via sequential compaction of N-terminal, α -helical, and α/β -core subdomains. Moreover, the timing of these events was critical; premature α -subdomain folding prevented subsequent core formation. This process was facilitated by modulating intrinsic folding propensity in three distinct ways: delaying α -subdomain compaction, facilitating β -strand intercalation, and optimizing translation kinetics via codon usage. Thus, de novo folding is translationally tuned by an integrated cellular response that shapes the cotranslational folding landscape at critical stages of synthesis.

Most proteins must acquire a defined three-dimensional structure in order to function. Folding pathways that generate these structures have primarily been characterized by using model substrates that fold rapidly, spontaneously, and reversibly in vitro (1, 2). In cells, however, protein folding is kinetically coupled to synthesis as the nascent polypeptide emerges from the ribosome. Whereas certain small proteins may remain unstructured during synthesis (3), many complex proteins exhibit length-dependent folding intermediates whose structural properties (4) and/or folding efficien-

cies (5) deviate from those observed in vitro. In such cases, the folding energy landscape, as well as folding outcome, can be influenced by ribosome effects (4, 6, 7), polypeptide elongation rate (8–10), molecular crowding (11, 12), and cotranslational interactions with cellular chaperones (13, 14). Indeed, cotranslational constraints can bias kinetically competing folding events to generate alternate stable structures with different functional properties (8, 15, 16). Despite improved computational methods, few principles have been established experimentally to explain how biosynthetic parameters influence specific folding events and outcome (3, 4, 17–19).

To address this issue, we used fluorescence resonance energy transfer (FRET) to examine structural transitions of ribosome-bound folding intermediates generated through in vitro translation of truncated RNA transcripts. This approach derives from the principle that during folding, certain residues distant in primary structure are brought into close proximity, increasing the

FRET efficiency between donor and acceptor fluorophores that are cotranslationally incorporated into the nascent polypeptide (Fig. 1A) (18, 19). Here, the donor fluorophore, cyan fluorescent protein (CFP), was fused to the N terminus of the first nucleotide-binding domain (NBD1) from the cystic fibrosis transmembrane conductance regulator (CFTR), and a small acceptor dye was incorporated at surface-exposed residues (UAG codons) by using a synthetic suppressor tRNA (figs. S1 and S2). FRET measurements obtained at sequential nascent chain lengths thus provide conformational snapshots into the equilibrium ensemble of stably arrested ribosome-bound nascent chains in the context of their native biosynthetic machinery (Fig. 1A) (3, 17–20).

Using this system, we defined the cotranslational folding pathway of CFTR NBD1, whose defective folding causes cystic fibrosis (21–24). NBD1 contains three subdomains (N-terminal, α -helical, and parallel-F1-type- β -sheet core) and exhibits a complex vectorial topography that limits CFTR maturation (22, 25) and prevents reversible folding in vitro. To examine its cotranslational folding pathway, FRET acceptor sites were chosen within 4 to 9 Å of the CFP fusion (Thr³⁸⁹) (26) to report on the positioning of strands S3, S6, S7, and S8 in the β -sheet core (Fig. 1, B and C). Analysis of sequentially stalled polypeptides yielded a characteristic length-dependent rise and plateau in FRET for each acceptor site (Fig. 1D). This rise in FRET reports on acquisition of a native-like fold (19) and reflects the earliest biosynthetic stage at which the acceptor dye and its corresponding β -strand are optimally positioned within NBD1. Results show that S3, S6, S7, and S8 could therefore reach a native-like structure when the ribosome has synthesized residues 550, 624, 654, and 674, respectively (Fig. 1D), although actual folding intermediates will depend on relative folding kinetics and translation elongation rate.

Despite their proximity, S6 exhibited a more gradual rise in FRET and was optimally positioned at a later stage of synthesis than was S3 (Fig. 1D). Because the ribosome exit tunnel sequesters approximately 40 residues, optimal S6 positioning

¹Department of Biochemistry and Molecular Biology, Oregon Health and Science University (OHSU), Portland, OR 97239, USA. ²Department of Neuroscience and Cell Biology, University of Texas Medical Branch, Galveston, TX 77550-0620, USA. ³Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX 77550-0620, USA. ⁴Cystic Fibrosis Foundation Therapeutics, Bethesda, MD 20814, USA.

*Corresponding author. E-mail: skachw@ohsu.edu



The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations

Valentino M. Gantz and Ethan Bier
Science **348**, 442 (2015);
DOI: 10.1126/science.aaa5945

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 23, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/442.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2015/03/18/science.aaa5945.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/348/6233/442.full.html#related>

This article **cites 7 articles**, 5 of which can be accessed free:

<http://www.sciencemag.org/content/348/6233/442.full.html#ref-list-1>

This article appears in the following **subject collections**:

Development

<http://www.sciencemag.org/cgi/collection/development>

submitted for publication, a preprint has been posted on the bioRxiv web server showing that a split Cas9-gRNA gene drive system efficiently biases inheritance in yeast (6). The split system was used to avoid accidental escape of the gene drives. The use of a similar strategy in future MCR organisms would reduce, but not eliminate, risks associated with accidental release. We therefore concur with others (7, 8) that a dialogue on this topic should become an immediate high-priority issue. Perhaps, by analogy to the famous Asilomar meeting of 1975 that assessed the risks of recombinant DNA technology, a similar conference could be convened to consider biosafety measures and institutional policies appropriate for limiting the risk of engaging in MCR research while affording workable opportunities for positive applications of this concept.

REFERENCES AND NOTES

1. F. Zhang, Y. Wen, X. Guo, *Hum. Mol. Genet.* **23**, R40–R46 (2014).
2. P. D. Hsu, E. S. Lander, F. Zhang, *Cell* **157**, 1262–1278 (2014).
3. F. Port, H. M. Chen, T. Lee, S. L. Bullock, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2967–E2976 (2014).
4. S. J. Gratz et al., *Genetics* **196**, 961–971 (2014).
5. A. R. Bassett, C. Tibbit, C. P. Ponting, J. L. Liu, *Cell Rep.* **4**, 220–228 (2013).
6. J. E. DiCarlo, A. Chavez, S. L. Dietz, K. M. Esvelt, G. M. Church, <http://biorex.org/content/early/2015/01/16/013896> (2015).
7. K. A. Oye et al., *Science* **345**, 626–628 (2014).
8. K. M. Esvelt, A. L. Smidler, F. Catteruccia, G. M. Church, *eLife* **10.7554/eLife.03401** (2014).

ACKNOWLEDGMENTS

We thank M. Yanofsky, W. McGinnis, S. Wasserman, R. Kolodner, H. Bellen, and members of the Bier lab for helpful discussions and comments on the manuscript; M. Harrison, K. O'Connor-Giles, J. Wildonger, and S. Bullock for providing CRISPR/Cas9 reagents and information; and J. Vinetz and A. Lubar for granting us

access to their BSL2 Insectary. Supported by NIH grants R01 GM067247 and R56 NS029870 and by a generous gift from S. Sandell and M. Marshall. E.B. and V.M.G. are authors on a patent applied for by the University of California, San Diego (provisional patent application number 62075534) that relates to the mutagenic chain reaction. MCR fly stocks and DNA constructs are available from E.B. under a material transfer agreement from UCSD. This protocol for use and containment of our MCR stocks in a BSL2 barrier insectary also used for containment of malaria-infected mosquitoes was reviewed and approved by the UCSD Institutional Biosafety Committee (BUA R461).

SUPPLEMENTARY MATERIALS

www.science.org/content/348/6233/442/suppl/DC1
Materials and Methods
Supplementary Text

Fig. S1
Table S1

31 December 2014; accepted 10 March 2015
Published online 19 March 2015;
10.1126/science.aaa5945

PROTEIN FOLDING

Translational tuning optimizes nascent protein folding in cells

Soo Jung Kim,¹ Jae Seok Yoon,¹ Hideki Shishido,¹ Zhongying Yang,¹ LeeAnn A. Rooney,¹ Jose M. Barral,^{2,3} William R. Skach^{1,4*}

In cells, biosynthetic machinery coordinates protein synthesis and folding to optimize efficiency and minimize off-pathway outcomes. However, it has been difficult to delineate experimentally the mechanisms responsible. Using fluorescence resonance energy transfer, we studied cotranslational folding of the first nucleotide-binding domain from the cystic fibrosis transmembrane conductance regulator. During synthesis, folding occurred discretely via sequential compaction of N-terminal, α -helical, and α/β -core subdomains. Moreover, the timing of these events was critical; premature α -subdomain folding prevented subsequent core formation. This process was facilitated by modulating intrinsic folding propensity in three distinct ways: delaying α -subdomain compaction, facilitating β -strand intercalation, and optimizing translation kinetics via codon usage. Thus, de novo folding is translationally tuned by an integrated cellular response that shapes the cotranslational folding landscape at critical stages of synthesis.

Most proteins must acquire a defined three-dimensional structure in order to function. Folding pathways that generate these structures have primarily been characterized by using model substrates that fold rapidly, spontaneously, and reversibly in vitro (1, 2). In cells, however, protein folding is kinetically coupled to synthesis as the nascent polypeptide emerges from the ribosome. Whereas certain small proteins may remain unstructured during synthesis (3), many complex proteins exhibit length-dependent folding intermediates whose structural properties (4) and/or folding efficien-

cies (5) deviate from those observed in vitro. In such cases, the folding energy landscape, as well as folding outcome, can be influenced by ribosome effects (4, 6, 7), polypeptide elongation rate (8–10), molecular crowding (11, 12), and cotranslational interactions with cellular chaperones (13, 14). Indeed, cotranslational constraints can bias kinetically competing folding events to generate alternate stable structures with different functional properties (8, 15, 16). Despite improved computational methods, few principles have been established experimentally to explain how biosynthetic parameters influence specific folding events and outcome (3, 4, 17–19).

To address this issue, we used fluorescence resonance energy transfer (FRET) to examine structural transitions of ribosome-bound folding intermediates generated through in vitro translation of truncated RNA transcripts. This approach derives from the principle that during folding, certain residues distant in primary structure are brought into close proximity, increasing the

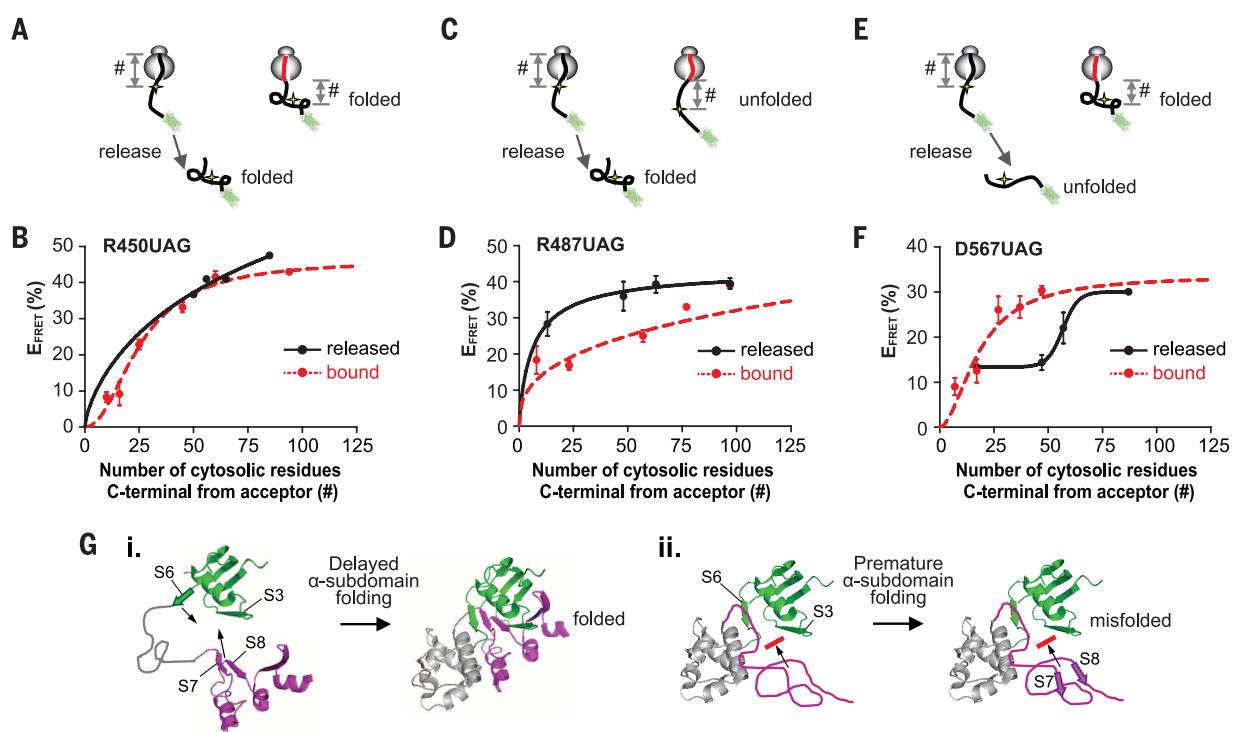
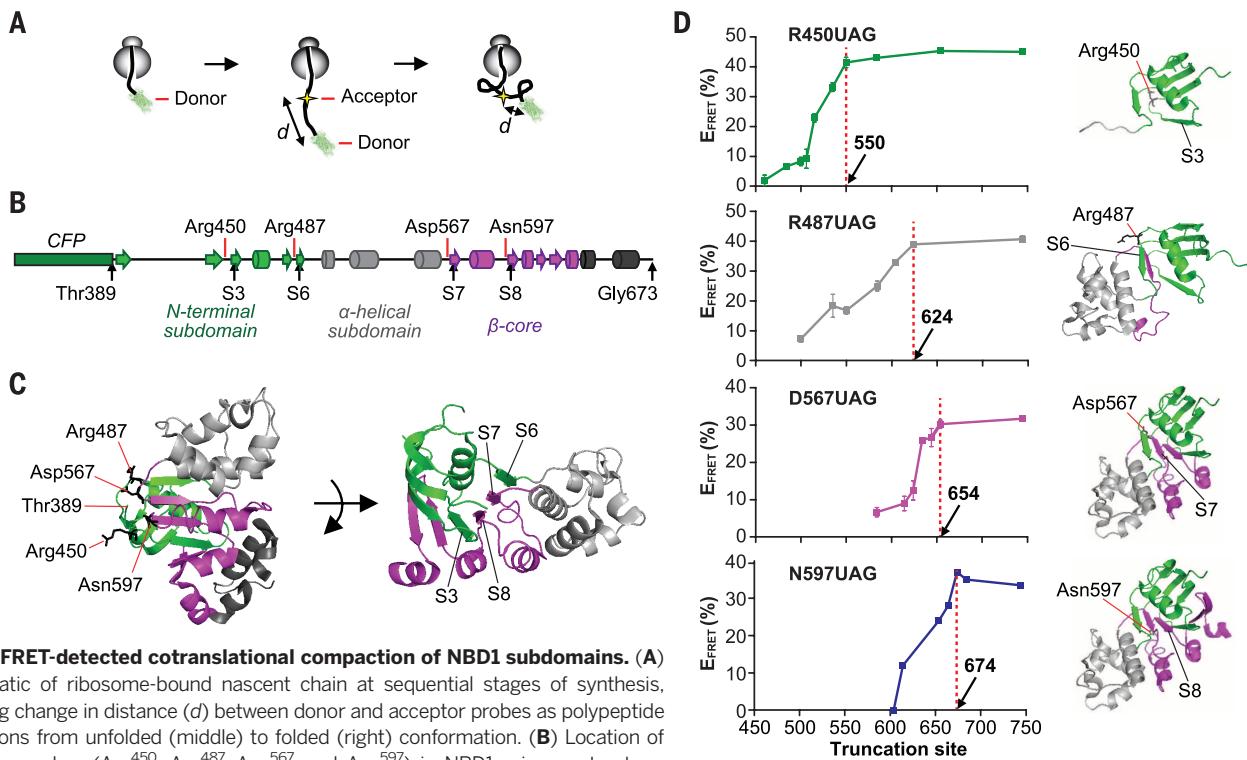
FRET efficiency between donor and acceptor fluorophores that are cotranslationally incorporated into the nascent polypeptide (Fig. 1A) (18, 19). Here, the donor fluorophore, cyan fluorescent protein (CFP), was fused to the N terminus of the first nucleotide-binding domain (NBD1) from the cystic fibrosis transmembrane conductance regulator (CFTR), and a small acceptor dye was incorporated at surface-exposed residues (UAG codons) by using a synthetic suppressor tRNA (figs. S1 and S2). FRET measurements obtained at sequential nascent chain lengths thus provide conformational snapshots into the equilibrium ensemble of stably arrested ribosome-bound nascent chains in the context of their native biosynthetic machinery (Fig. 1A) (3, 17–20).

Using this system, we defined the cotranslational folding pathway of CFTR NBD1, whose defective folding causes cystic fibrosis (21–24). NBD1 contains three subdomains (N-terminal, α -helical, and parallel-F1-type- β -sheet core) and exhibits a complex vectorial topography that limits CFTR maturation (22, 25) and prevents reversible folding in vitro. To examine its cotranslational folding pathway, FRET acceptor sites were chosen within 4 to 9 Å of the CFP fusion (Thr³⁸⁹) (26) to report on the positioning of strands S3, S6, S7, and S8 in the β -sheet core (Fig. 1, B and C). Analysis of sequentially stalled polypeptides yielded a characteristic length-dependent rise and plateau in FRET for each acceptor site (Fig. 1D). This rise in FRET reports on acquisition of a native-like fold (19) and reflects the earliest biosynthetic stage at which the acceptor dye and its corresponding β -strand are optimally positioned within NBD1. Results show that S3, S6, S7, and S8 could therefore reach a native-like structure when the ribosome has synthesized residues 550, 624, 654, and 674, respectively (Fig. 1D), although actual folding intermediates will depend on relative folding kinetics and translation elongation rate.

Despite their proximity, S6 exhibited a more gradual rise in FRET and was optimally positioned at a later stage of synthesis than was S3 (Fig. 1D). Because the ribosome exit tunnel sequesters approximately 40 residues, optimal S6 positioning

¹Department of Biochemistry and Molecular Biology, Oregon Health and Science University (OHSU), Portland, OR 97239, USA. ²Department of Neuroscience and Cell Biology, University of Texas Medical Branch, Galveston, TX 77550-0620, USA. ³Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX 77550-0620, USA. ⁴Cystic Fibrosis Foundation Therapeutics, Bethesda, MD 20814, USA.

*Corresponding author. E-mail: skachw@ohsu.edu



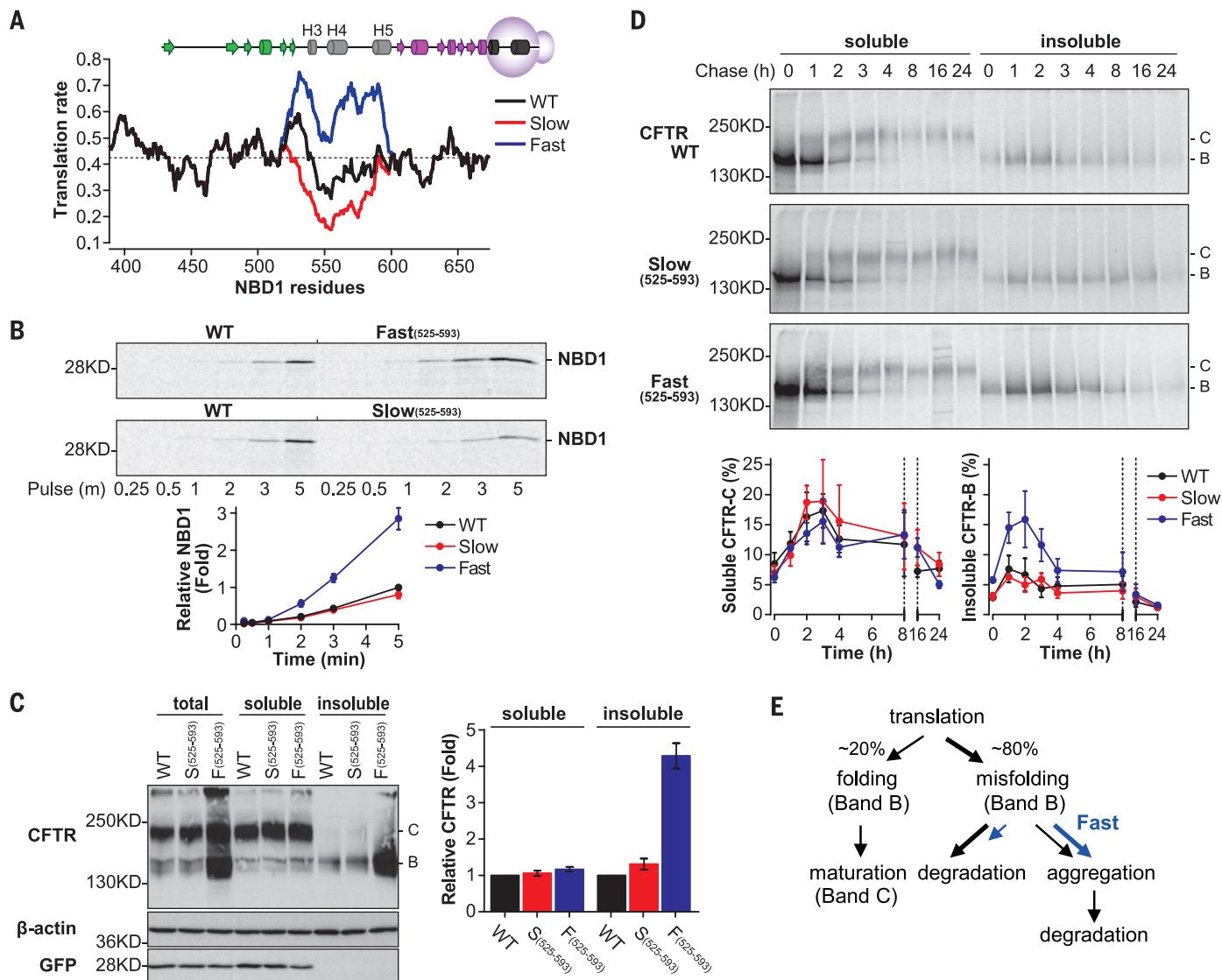


Fig. 3. CFTR folding dependence on NBD1 translation elongation rate.

(A) Predicted translation elongation rate (λ_0) calculated as a 15-amino acid moving window average, for WT NBD1 (black), Slow₍₅₂₅₋₅₉₃₎ (red), and Fast₍₅₂₅₋₅₉₃₎ (blue) aligned with NBD1 secondary structural elements as they emerge from the ribosome. (B) Autoradiogram of [35 S]Met-labeled WT, Slow₍₅₂₅₋₅₉₃₎, and Fast₍₅₂₅₋₅₉₃₎ NBD1. Graph shows fold change in protein compared with that of WT at 5 min ($n = 3$ replicates \pm SEM). (C) CFTR immunoblot from human embryonic kidney

(HEK) 293T cell lysate showing core glycosylated (band B) and mature (band C) CFTR from total, radioimmunoprecipitation assay (RIPA)-soluble, and insoluble lysate fractions. Graph shows fold increase over that of WT ($n \geq 5$ replicates \pm SEM). (D) [35 S]Met-labeled CFTR immunoprecipitated from RIPA soluble and insoluble fractions. Graphs show band C (left) and band B (right) CFTR as percentage of total protein recovered at time (T) = 0 ($n = 3$ replicates \pm SEM). (E) Schematic showing proposed effect of Fast₍₅₂₅₋₅₉₃₎ substitutions on CFTR processing (blue arrows).

was therefore delayed until the entire α -subdomain and S7 emerged into the cytosol (Fig. 1D). To understand this delay, we compared the conformation of S3 and S6 in ribosome-bound versus ribosome-released polypeptides, where each construct contained equivalent cytosolically exposed residues (Fig. 2) (19, 27). Under these conditions, ribosome attachment had no detectable impact on N-terminal subdomain folding (Fig. 2, A and B, and fig. S3A). In contrast, S6 exhibited a higher FRET efficiency and achieved a native-like conformation at shorter chain lengths (by nearly 50 residues) in the ribosome-free state (Fig. 2, C and D, and fig. S3B). Ribosome attachment therefore delayed S6 positioning during α -subdomain synthesis.

In adenosine 5'-triphosphate-binding cassette transporters, S7 and S8 are synthesized after the α -subdomain and must intercalate between S3 and S6 to form the four-strand hydrophobic β -sheet core (Fig. 1, B and C). Yet, S6, S7, and S8 are all positioned into the native NBD fold at a similar stage of synthesis (Fig. 1D). Because optimal S7 positioning requires S8 and is a prerequisite for CFTR folding (fig. S4), we tested whether ribosome-mediated delay of S6 positioning might facilitate β -sheet core formation. Indeed, ribosome release before core folding prevented S7 insertion, as evidenced by a reduction in FRET efficiency (Fig. 2, E and F, and fig. S3, C and D). The ribosome therefore exhibits two opposite but interrelated effects on nascent chain conformation

that directly influence folding outcome: delaying spontaneous folding at one stage of synthesis (S6 positioning), while enabling folding at another (S7/S8 intercalation). Because premature polypeptide release uncouples these events, the ribosome appears to facilitate NBD1 folding, potentially in concert with associated chaperones, by maintaining a relatively open α -subdomain conformation to ensure orderly and sequential insertion of S6, S7, and S8 during a critical translational window (Fig. 2G).

These results predict that β -sheet core formation depends on the timing and/or kinetics of α -subdomain folding. Because the predicted translation elongation rate (λ_0) slows dramatically during α -subdomain synthesis (Fig. 3A and table

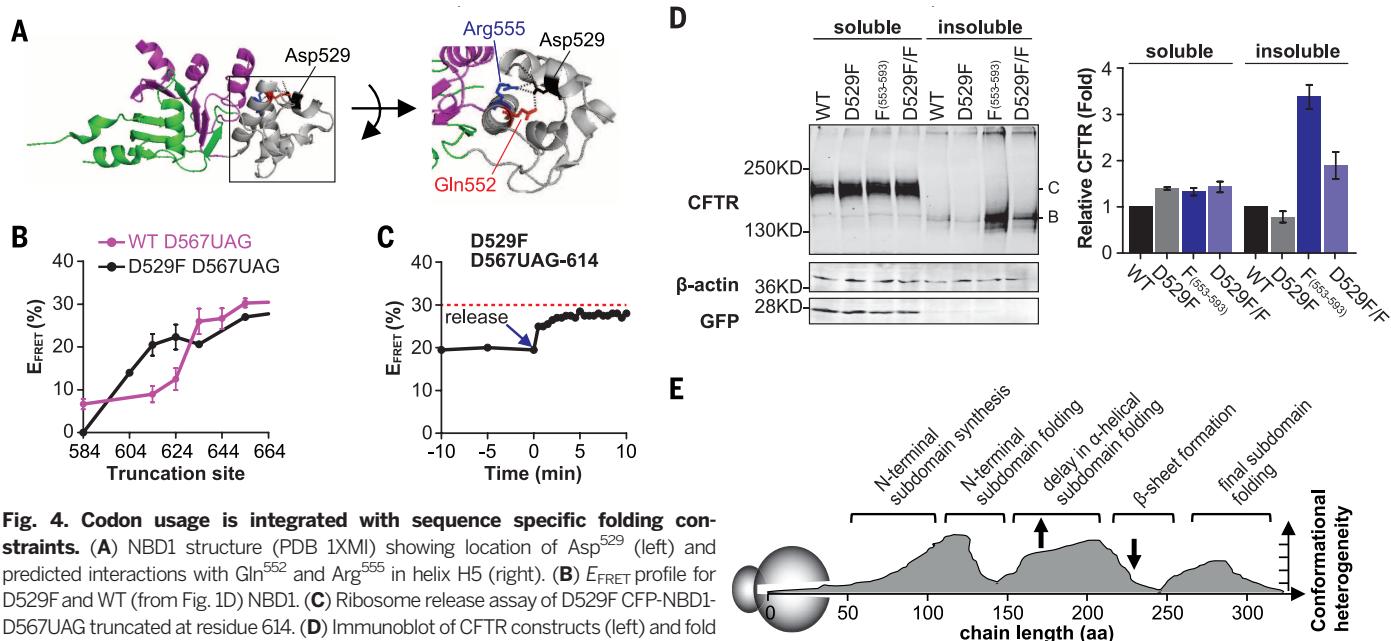


Fig. 4. Codon usage is integrated with sequence specific folding constraints. (A) NBD1 structure (PDB 1XMI) showing location of Asp⁵²⁹ (left) and predicted interactions with Gln⁵⁵² and Arg⁵⁵⁵ in helix H5 (right). (B) E_{FRET} profile for D529F and WT (from Fig. 1D) NBD1. (C) Ribosome release assay of D529F CFP-NBD1-D567UAG truncated at residue 614. (D) Immunoblot of CFTR constructs (left) and fold increase over WT (right) ($n \geq 4$ replicates \pm SEM) as in Fig. 3C. (E) Illustration showing that NBD1 cotranslational folding is both delayed (upward arrows) and enhanced (downward arrows) by cellular biosynthetic machinery to optimize folding outcome.

S1), we introduced synonymous codon substitutions predicted to either maximize or minimize the translation rate within this region (Fig. 3A and fig. S5). Metabolic pulse labeling verified that Fast₍₅₂₅₋₅₉₃₎ synonymous codons did indeed increase the rate of NBD1 synthesis in cells (Fig. 3B and fig. S6). Remarkably, despite generating identical amino acid sequences (fig. S7 and table S2), Fast₍₅₂₅₋₅₉₃₎ codon substitutions resulted in aggregation of full-length CFTR (Band B) (Fig. 3C) as well as isolated NBD1 (fig. S8) 1 to 4 hours after synthesis, suggesting either delayed conversion into the insoluble fraction or less efficient degradation (Fig. 3, D and E). In contrast, Slow₍₅₂₅₋₅₉₃₎ codon substitutions had no detectable effect on either NBD1 or full-length wild-type (WT) or ΔF508 CFTR (Fig. 3, B to D, and figs. S8 and S9). Similarly, Fast substitutions within codons 501 to 540 had no effect on CFTR aggregation (fig. S10).

Thus, the translation elongation rate appears to be tailored to folding needs of specific peptide regions and, in this case, kinetically couples α -subdomain and β -sheet core folding as H3 and H4 emerge from the ribosome (Fig. 3A). This region is home to numerous suppressor mutations that improve NBD1 stability, CFTR folding efficiency, and ΔF508 CFTR processing (24, 28). One such variant, D529F, improves NBD1 and full-length CFTR folding without affecting NBD1 thermal stability (25), suggesting that it may act along the folding pathway. (Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. In the mutants, other amino acids were substituted at certain locations; for example,

D529F indicates that aspartic acid at position 529 was replaced by phenylalanine.) Asp⁵²⁹ makes several polar interactions with helix H5 (Gln⁵⁵² and Arg⁵⁵⁵) (Fig. 4A) that are adjacent to the hydrophobic core of the α -subdomain (Ile⁵⁰², Cys⁵²⁴, Leu⁵²⁶, Leu⁵⁴¹, and Ile⁵⁵⁶). The substitution at this site might therefore affect timing of local α -subdomain collapse. Consistent with this hypothesis, D529F resulted in S7 positioning at an earlier stage of synthesis without affecting S6 (Fig. 4B and fig. S11A). D529F also eliminated ribosome dependence for β -sheet core folding (Fig. 4C) and suppressed aggregation of full-length CFTR caused by Fast₍₅₂₅₋₅₉₃₎ codon substitutions (Fig. 4D), thus restoring kinetic coupling between α -helical subdomain and β -sheet core.

This study delineates cotranslational folding of a topologically complex protein as a series of dynamically modulated folding events that can be viewed as a function of chain length (Fig. 4E). As the nascent polypeptide emerges from the ribosome, formation of low FRET-associated open conformers is interrupted by discrete intervals of peptide compaction. Although each of these folding events could be theoretically analogous to *in vitro* folding of an equivalent peptide domain (3), this does not appear to be the case. Rather, folding occurs in sequential, coupled steps, the timing of which is both positively and negatively influenced by biosynthetic machinery. Rapid cotranslational folding of the N-terminal subdomain likely provides a template for subsequent interfacial interactions that assist domain assembly (29). In contrast, optimal folding of the non-contiguously synthesized β -strands is achieved by actively delaying placement of a presynthesized N-terminal strand (S6) until C-terminal strands (S7 and S8) are available. This process is coordinated by maintaining the nascent polypeptide in a folding-competent conformation (4, 27) both by direct ribosome effects (4, 7) and the translation rate as dictated by codon usage (11, 12). Cotranslationally recruited chaperones (30), not examined here, may also contribute to the delay in S6 placement and α -subdomain collapse (13). We refer to this overall process as “translational tuning,” in which multiple mechanisms are simultaneously integrated during synthesis to modulate intrinsic folding properties of the nascent chain. Translational tuning also integrates conserved codon usage with biophysical properties imposed by amino acid sequence, both of which are tailored to optimize outcome based on specific folding requirements (31, 32).

Thus, the translation elongation rate appears to be tailored to folding needs of specific peptide regions and, in this case, kinetically couples α -subdomain and β -sheet core folding as H3 and H4 emerge from the ribosome (Fig. 3A). This region is home to numerous suppressor mutations that improve NBD1 stability, CFTR folding efficiency, and ΔF508 CFTR processing (24, 28). One such variant, D529F, improves NBD1 and full-length CFTR folding without affecting NBD1 thermal stability (25), suggesting that it may act along the folding pathway. (Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. In the mutants, other amino acids were substituted at certain locations; for example,

REFERENCES AND NOTES

1. E. Braselmann, J. L. Chaney, P. L. Clark, *Trends Biochem. Sci.* **38**, 337–344 (2013).
2. K. S. Hingorani, L. M. Giersch, *Curr. Opin. Struct. Biol.* **24**, 81–90 (2014).
3. C. Eichmann, S. Preissler, R. Riek, E. Deuerling, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9111–9116 (2010).
4. C. M. Kaiser, D. H. Goldman, J. D. Chodera, I. Tinoco Jr., C. Bustamante, *Science* **334**, 1723–1727 (2011).
5. K. G. Ugrinov, P. L. Clark, *Biophys. J.* **98**, 1312–1320 (2010).
6. E. P. O'Brien, J. Christodoulou, M. Vendruscolo, C. M. Dobson, *J. Am. Chem. Soc.* **133**, 513–526 (2011).
7. A. Kosolapov, C. Deutsch, *Nat. Struct. Mol. Biol.* **16**, 405–411 (2009).
8. C. Kimchi-Sarfaty et al., *Science* **315**, 525–528 (2007).
9. G. Zhang, M. Hubalewska, Z. Ignatova, *Nat. Struct. Mol. Biol.* **16**, 274–280 (2009).
10. P. S. Spencer, E. Siller, J. F. Anderson, J. M. Barral, *J. Mol. Biol.* **422**, 328–335 (2012).
11. S. B. Zimmerman, A. P. Minton, *Annu. Rev. Biophys. Biomol. Struct.* **22**, 27–65 (1993).
12. S. R. McGuffee, A. H. Elcock, *PLOS Comput. Biol.* **6**, e1000694 (2010).
13. F. Willmund et al., *Cell* **152**, 196–209 (2013).
14. A. Mashaghi et al., *Nature* **500**, 98–101 (2013).

15. I. M. Sander, J. L. Chaney, P. L. Clark, *J. Am. Chem. Soc.* **136**, 858–861 (2014).
16. M. Zhou *et al.*, *Nature* **495**, 111–115 (2013).
17. L. D. Cabrita, S. T. Hsu, H. Launay, C. M. Dobson, J. Christodoulou, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22239–22244 (2009).
18. C. A. Woolhead, P. J. McCormick, A. E. Johnson, *Cell* **116**, 725–736 (2004).
19. A. Khushoo, Z. Yang, A. E. Johnson, W. R. Skach, *Mol. Cell* **41**, 682–692 (2011).
20. P. K. Devaraneni *et al.*, *Cell* **146**, 134–147 (2011).
21. S. H. Cheng *et al.*, *Cell* **63**, 827–834 (1990).
22. B.-H. Qu, E. H. Strickland, P. J. Thomas, *J. Biol. Chem.* **272**, 15739–15744 (1997).
23. C. Wang *et al.*, *Protein Sci.* **19**, 1932–1947 (2010).
24. A. A. Aleksandrov *et al.*, *J. Mol. Biol.* **419**, 41–60 (2012).
25. J. L. Mendoza *et al.*, *Cell* **148**, 164–174 (2012).
26. H. A. Lewis *et al.*, *J. Mol. Biol.* **396**, 406–430 (2010).
27. D. A. Kelkar, A. Khushoo, Z. Yang, W. R. Skach, *J. Biol. Chem.* **287**, 2568–2578 (2012).
28. J. L. Teem *et al.*, *Cell* **73**, 335–346 (1993).
29. J. Frydman, H. Erdjument-Bromage, P. Tempst, F. U. Hartl, *Nat. Struct. Biol.* **6**, 697–705 (1999).
30. M. Chartier, F. Gaudreault, R. Najmanovich, *Bioinformatics* **28**, 1438–1445 (2012).
31. S. Pechmann, J. Frydman, *Nat. Struct. Mol. Biol.* **20**, 237–243 (2013).
32. F. Gloge, A. H. Becker, G. Kramer, B. Bukau, *Curr. Opin. Struct. Biol.* **24**, 24–33 (2014).

ACKNOWLEDGMENTS

We thank L. David, B. Conti, and members of the Skach laboratory for helpful discussions; CFTR Folding Consortium for providing 3G11 CFTR antibody. This research was funded by NIH grants R01GM53457 and R01DK51818 (to W.R.S.) and Cystic Fibrosis Foundation Therapeutics grants KIM10FO (to S.J.K.) and

SKACH05XO (to W.R.S.). Mass spectrometric analysis was performed by the OHSU Proteomics Shared Resource with support from P30EYE010572, P30CA069533, S10OD012246, and S10RR025571. U.S. patent application 13/664,252 covers the methodology to assess and redesign the NBD1 coding sequences, claiming priority to provisional application 61/553,861, filed on 31 October 2011. No other counterpart applications have been filed.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/444/suppl/DC1
Material and Methods
Figs. S1 to S11
Table S1 and S2
References (33–44)
30 November 2014; accepted 12 February 2015
10.1126/science.aaa3974

INFECTIOUS DISEASE

Life-threatening influenza and impaired interferon amplification in human IRF7 deficiency

Michael J. Ciancanelli,¹ Sarah X. L. Huang,^{2,3*} Priya Luthra,^{4*} Hannah Garner,^{5*} Yuval Itan,¹ Stefano Volpi,^{6,7} Fabien G. Lafaille,¹ Céline Trouillet,⁵ Miroc Schmolke,⁴ Randy A. Albrecht,^{4,8} Elisabeth Israelsson,⁹ Hye Kyung Lim,¹ Melina Casadio,¹ Tamar Hermesh,¹ Lazaro Lorenzo,^{10,11} Lawrence W. Leung,⁴ Vincent Pedergnana,^{10,11} Bertrand Boisson,¹ Satoshi Okada,^{1,12} Capucine Picard,^{1,10,11,13} Benedicte Ringuier,¹⁴ Françoise Troussier,¹⁵ Damien Chaussabel,^{9,16†} Laurent Abel,^{1,10,11†} Isabelle Pelleri,^{17,18†} Luigi D. Notarangelo,^{6†} Adolfo García-Sastre,^{4,8,19†} Christopher F. Basler,^{4†} Frédéric Geissmann,^{5†} Shen-Ying Zhang,^{1,10,11†} Hans-Willem Snoeck,^{2,3†} Jean-Laurent Casanova^{1,10,11,20,21†}

Severe influenza disease strikes otherwise healthy children and remains unexplained. We report compound heterozygous null mutations in *IRF7*, which encodes the transcription factor interferon regulatory factor 7, in an otherwise healthy child who suffered life-threatening influenza during primary infection. In response to influenza virus, the patient's leukocytes and plasmacytoid dendritic cells produced very little type I and III interferons (IFNs). Moreover, the patient's dermal fibroblasts and induced pluripotent stem cell (iPSC)-derived pulmonary epithelial cells produced reduced amounts of type I IFN and displayed increased influenza virus replication. These findings suggest that *IRF7*-dependent amplification of type I and III IFNs is required for protection against primary infection by influenza virus in humans. They also show that severe influenza may result from single-gene inborn errors of immunity.

Both seasonal and pandemic influenza viruses typically cause self-limiting respiratory disease but occasionally cause life-threatening acute respiratory distress syndrome (ARDS).

The frequency of severe disease depends on the viral strain (1). Known host risk factors to severe influenza consist of a few acquired comorbidities, such as chronic pulmonary disease (2, 3). The pathogenesis of most cases of life-threatening influenza remains unexplained, especially among children (4). The lack of severe influenza in patients with known primary immunodeficiencies, including inborn errors of T and/or B cell immunity that predispose to a variety of related infections such as severe parainfluenza (5), is also unexplained (6). Only haploinsufficiency for *GATA2*, resulting in abnormal

hematopoietic cell development, was reported in a few patients with severe influenza and other infections (7). Monogenic inborn errors of immunity may underlie life-threatening, isolated diseases in otherwise healthy children during primary infection by a few other viruses (8). We therefore hypothesized that severe influenza striking otherwise healthy children may also result from single-gene inborn errors of immunity.

We performed whole-exome sequencing (WES) in a 7-year-old girl ("P"), one of 22 individuals sequenced (including only three children <5 years old) proven to have developed influenza in the course of primary infection (table S1). P suffered life-threatening ARDS during infection with laboratory-confirmed pandemic H1N1 (pH1N1) 2009 influenza A virus (IAV) in January 2011 at

the age of 2.5 years, prior to any influenza vaccination. Serum drawn shortly after hospitalization showed protective antibody titers against A/Netherlands/602/2009 (H1N1) IAV (fig. S1A) but not against A/Perth/16/2009 (H3N2) or B/Brisbane/60/2008 (fig. S1B), indicating that this was her first encounter with IAV. She did not suffer from severe infections caused by other viruses (supplementary case report and fig. S1C). The patient had no known comorbidity and no detectable immunological abnormalities suggestive of any T or B cell deficit (table S2). She was born to nonconsanguineous parents of French descent (Fig. 1A and fig. S2A). WES analysis of the trio revealed, and Sanger sequencing confirmed, two compound heterozygous *IRF7* mutations—p.Phe410Val (F410V) and p.Gln421X (Q421X)—with

¹St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY, USA. ²Columbia Center for Translational Immunology, Columbia University Medical Center, New York, NY, USA. ³Department of Medicine, Columbia University Medical Center, New York, NY, USA. ⁴Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁵Centre for Molecular and Cellular Biology of Inflammation (CMCBI), King's College London, London SE1 1UL, UK. ⁶Division of Immunology and Manton Center for Orphan Disease Research, Children's Hospital, Harvard Medical School, Boston, MA, USA. ⁷Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, 16132 Genoa, Italy. ⁸Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁹Department of Systems Immunology, Benaroya Research Institute at Virginia Mason, Seattle, WA, USA.

¹⁰Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM UMR1163, Paris, France. ¹¹University Paris Descartes, Imagine Institute, Paris, France. ¹²Department of Pediatrics, Hiroshima University Graduate School of Biomedical & Health Sciences, Hiroshima, Japan. ¹³Study Centre for Primary Immunodeficiencies, AP-HP, Necker Hospital, Paris, France. ¹⁴Pediatric Intensive Care Unit, University Hospital, Angers, France. ¹⁵General Pediatrics Unit, University Hospital, Angers, France. ¹⁶Department of Systems Biology, Sidra Medical and Research Center, Doha, Qatar. ¹⁷Pediatric Immunology, Hematology and Oncology Unit, University Hospital Centre of Angers, Angers, France. ¹⁸INSERM U892, CNRS U6299, Angers, France. ¹⁹Department of Medicine, Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ²⁰Pediatric Immuno-Hematology Unit, Necker Hospital for Sick Children, AP-HP, Paris, France.

²¹Howard Hughes Medical Institute, New York, NY, USA. *These authors contributed equally to this work. †These authors contributed equally to this work. [‡]Corresponding author. E-mail: jean-laurent.casanova@rockefeller.edu



Translational tuning optimizes nascent protein folding in cells

Soo Jung Kim *et al.*
Science **348**, 444 (2015);
 DOI: 10.1126/science.aaa3974

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 23, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/444.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2015/04/22/348.6233.444.DC1.html>

A list of selected additional articles on the Science Web sites related to this article can be found at:

<http://www.sciencemag.org/content/348/6233/444.full.html#related>

This article cites 44 articles, 11 of which can be accessed free:
<http://www.sciencemag.org/content/348/6233/444.full.html#ref-list-1>

This article has been cited by 1 articles hosted by HighWire Press; see:
<http://www.sciencemag.org/content/348/6233/444.full.html#related-urls>

This article appears in the following subject collections:
 Biochemistry

<http://www.sciencemag.org/cgi/collection/biochem>

15. I. M. Sander, J. L. Chaney, P. L. Clark, *J. Am. Chem. Soc.* **136**, 858–861 (2014).
16. M. Zhou *et al.*, *Nature* **495**, 111–115 (2013).
17. L. D. Cabrita, S. T. Hsu, H. Launay, C. M. Dobson, J. Christodoulou, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22239–22244 (2009).
18. C. A. Woolhead, P. J. McCormick, A. E. Johnson, *Cell* **116**, 725–736 (2004).
19. A. Khushoo, Z. Yang, A. E. Johnson, W. R. Skach, *Mol. Cell* **41**, 682–692 (2011).
20. P. K. Devaraneni *et al.*, *Cell* **146**, 134–147 (2011).
21. S. H. Cheng *et al.*, *Cell* **63**, 827–834 (1990).
22. B.-H. Qu, E. H. Strickland, P. J. Thomas, *J. Biol. Chem.* **272**, 15739–15744 (1997).
23. C. Wang *et al.*, *Protein Sci.* **19**, 1932–1947 (2010).
24. A. A. Aleksandrov *et al.*, *J. Mol. Biol.* **419**, 41–60 (2012).
25. J. L. Mendoza *et al.*, *Cell* **148**, 164–174 (2012).
26. H. A. Lewis *et al.*, *J. Mol. Biol.* **396**, 406–430 (2010).
27. D. A. Kelkar, A. Khushoo, Z. Yang, W. R. Skach, *J. Biol. Chem.* **287**, 2568–2578 (2012).
28. J. L. Teem *et al.*, *Cell* **73**, 335–346 (1993).
29. J. Frydman, H. Erdjument-Bromage, P. Tempst, F. U. Hartl, *Nat. Struct. Biol.* **6**, 697–705 (1999).
30. M. Chartier, F. Gaudreault, R. Najmanovich, *Bioinformatics* **28**, 1438–1445 (2012).
31. S. Pechmann, J. Frydman, *Nat. Struct. Mol. Biol.* **20**, 237–243 (2013).
32. F. Gloge, A. H. Becker, G. Kramer, B. Bukau, *Curr. Opin. Struct. Biol.* **24**, 24–33 (2014).

ACKNOWLEDGMENTS

We thank L. David, B. Conti, and members of the Skach laboratory for helpful discussions; CFTR Folding Consortium for providing 3G11 CFTR antibody. This research was funded by NIH grants R01GM53457 and R01DK51818 (to W.R.S.) and Cystic Fibrosis Foundation Therapeutics grants KIM10FO (to S.J.K.) and

SKACH05XO (to W.R.S.). Mass spectrometric analysis was performed by the OHSU Proteomics Shared Resource with support from P30EYE010572, P30CA069533, S10OD012246, and S10RR025571. U.S. patent application 13/664,252 covers the methodology to assess and redesign the NBD1 coding sequences, claiming priority to provisional application 61/553,861, filed on 31 October 2011. No other counterpart applications have been filed.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/444/suppl/DC1
Material and Methods
Figs. S1 to S11
Table S1 and S2
References (33–44)
30 November 2014; accepted 12 February 2015
10.1126/science.aaa3974

INFECTIOUS DISEASE

Life-threatening influenza and impaired interferon amplification in human IRF7 deficiency

Michael J. Ciancanelli,¹ Sarah X. L. Huang,^{2,3*} Priya Luthra,^{4*} Hannah Garner,^{5*} Yuval Itan,¹ Stefano Volpi,^{6,7} Fabien G. Lafaille,¹ Céline Trouillet,⁵ Miroc Schmolke,⁴ Randy A. Albrecht,^{4,8} Elisabeth Israelsson,⁹ Hye Kyung Lim,¹ Melina Casadio,¹ Tamar Hermesh,¹ Lazaro Lorenzo,^{10,11} Lawrence W. Leung,⁴ Vincent Pedergnana,^{10,11} Bertrand Boisson,¹ Satoshi Okada,^{1,12} Capucine Picard,^{1,10,11,13} Benedicte Ringuier,¹⁴ Françoise Troussier,¹⁵ Damien Chaussabel,^{9,16†} Laurent Abel,^{1,10,11†} Isabelle Pelleri,^{17,18†} Luigi D. Notarangelo,^{6†} Adolfo García-Sastre,^{4,8,19†} Christopher F. Basler,^{4†} Frédéric Geissmann,^{5†} Shen-Ying Zhang,^{1,10,11†} Hans-Willem Snoeck,^{2,3†} Jean-Laurent Casanova^{1,10,11,20,21†}

Severe influenza disease strikes otherwise healthy children and remains unexplained. We report compound heterozygous null mutations in *IRF7*, which encodes the transcription factor interferon regulatory factor 7, in an otherwise healthy child who suffered life-threatening influenza during primary infection. In response to influenza virus, the patient's leukocytes and plasmacytoid dendritic cells produced very little type I and III interferons (IFNs). Moreover, the patient's dermal fibroblasts and induced pluripotent stem cell (iPSC)-derived pulmonary epithelial cells produced reduced amounts of type I IFN and displayed increased influenza virus replication. These findings suggest that *IRF7*-dependent amplification of type I and III IFNs is required for protection against primary infection by influenza virus in humans. They also show that severe influenza may result from single-gene inborn errors of immunity.

Both seasonal and pandemic influenza viruses typically cause self-limiting respiratory disease but occasionally cause life-threatening acute respiratory distress syndrome (ARDS).

The frequency of severe disease depends on the viral strain (1). Known host risk factors to severe influenza consist of a few acquired comorbidities, such as chronic pulmonary disease (2, 3). The pathogenesis of most cases of life-threatening influenza remains unexplained, especially among children (4). The lack of severe influenza in patients with known primary immunodeficiencies, including inborn errors of T and/or B cell immunity that predispose to a variety of related infections such as severe parainfluenza (5), is also unexplained (6). Only haploinsufficiency for *GATA2*, resulting in abnormal

hematopoietic cell development, was reported in a few patients with severe influenza and other infections (7). Monogenic inborn errors of immunity may underlie life-threatening, isolated diseases in otherwise healthy children during primary infection by a few other viruses (8). We therefore hypothesized that severe influenza striking otherwise healthy children may also result from single-gene inborn errors of immunity.

We performed whole-exome sequencing (WES) in a 7-year-old girl ("P"), one of 22 individuals sequenced (including only three children <5 years old) proven to have developed influenza in the course of primary infection (table S1). P suffered life-threatening ARDS during infection with laboratory-confirmed pandemic H1N1 (pH1N1) 2009 influenza A virus (IAV) in January 2011 at

the age of 2.5 years, prior to any influenza vaccination. Serum drawn shortly after hospitalization showed protective antibody titers against A/Netherlands/602/2009 (H1N1) IAV (fig. S1A) but not against A/Perth/16/2009 (H3N2) or B/Brisbane/60/2008 (fig. S1B), indicating that this was her first encounter with IAV. She did not suffer from severe infections caused by other viruses (supplementary case report and fig. S1C). The patient had no known comorbidity and no detectable immunological abnormalities suggestive of any T or B cell deficit (table S2). She was born to nonconsanguineous parents of French descent (Fig. 1A and fig. S2A). WES analysis of the trio revealed, and Sanger sequencing confirmed, two compound heterozygous *IRF7* mutations—p.Phe410Val (F410V) and p.Gln421X (Q421X)—with

¹St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY, USA. ²Columbia Center for Translational Immunology, Columbia University Medical Center, New York, NY, USA. ³Department of Medicine, Columbia University Medical Center, New York, NY, USA. ⁴Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁵Centre for Molecular and Cellular Biology of Inflammation (CMCBI), King's College London, London SE1 1UL, UK. ⁶Division of Immunology and Manton Center for Orphan Disease Research, Children's Hospital, Harvard Medical School, Boston, MA, USA. ⁷Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, 16132 Genoa, Italy. ⁸Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁹Department of Systems Immunology, Benaroya Research Institute at Virginia Mason, Seattle, WA, USA.

¹⁰Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM UMR1163, Paris, France. ¹¹University Paris Descartes, Imagine Institute, Paris, France. ¹²Department of Pediatrics, Hiroshima University Graduate School of Biomedical & Health Sciences, Hiroshima, Japan. ¹³Study Centre for Primary Immunodeficiencies, AP-HP, Necker Hospital, Paris, France. ¹⁴Pediatric Intensive Care Unit, University Hospital, Angers, France. ¹⁵General Pediatrics Unit, University Hospital, Angers, France. ¹⁶Department of Systems Biology, Sidra Medical and Research Center, Doha, Qatar. ¹⁷Pediatric Immunology, Hematology and Oncology Unit, University Hospital Centre of Angers, Angers, France. ¹⁸INSERM U892, CNRS U6299, Angers, France. ¹⁹Department of Medicine, Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ²⁰Pediatric Immuno-Hematology Unit, Necker Hospital for Sick Children, AP-HP, Paris, France.

²¹Howard Hughes Medical Institute, New York, NY, USA. *These authors contributed equally to this work. †These authors contributed equally to this work. [‡]Corresponding author. E-mail: jean-laurent.casanova@rockefeller.edu

Fig. 1. Autosomal recessive IRF7 deficiency from compound heterozygous mutations.

(A) Familial segregation of *IRF7* mutations in a nonconsanguineous French family. A pedigree chart shows inheritance of the F410V mutation from a heterozygous parent (P) to a homozygous affected child (F410V/Q421X). **(B)** Schematic illustration of *IRF7A* featuring DNA binding domain (DBD), constitutive activation domain (CAD), virus-activated domain (VAD), inhibitory domain (ID), and signal response domain (SRD). A potential nuclear localization signal (NLS) lies between amino acids 417 and 440, and the nuclear export signal (NES) between amino acids 448 and 462. Phosphorylation sites (P) Ser⁴⁷⁷ and Ser⁴⁷⁹ lie at the C terminus. Mutations are shown in red. **(C)** Wild-type, F410V, or Q421X *IRF7* activation of *IFNB*, *IFNA4*, or *IFNA6* promoter-driven reporter assay. Cells are uninfected (UI) or infected with SeV. Means \pm SD of three independent experiments are shown. *P < 0.01, **P < 0.005, ***P < 0.001 as determined by t test. **(D)** Phosphorylation of HA-tagged wild-type (WT), F410V, or Q421X *IRF7* coexpressed with FLAG-tagged TBK1 as assessed by Western blot with phospho-specific *IRF7* antibody (P-*IRF7*); GAPDH was used as a loading control. This result is representative of two experiments.

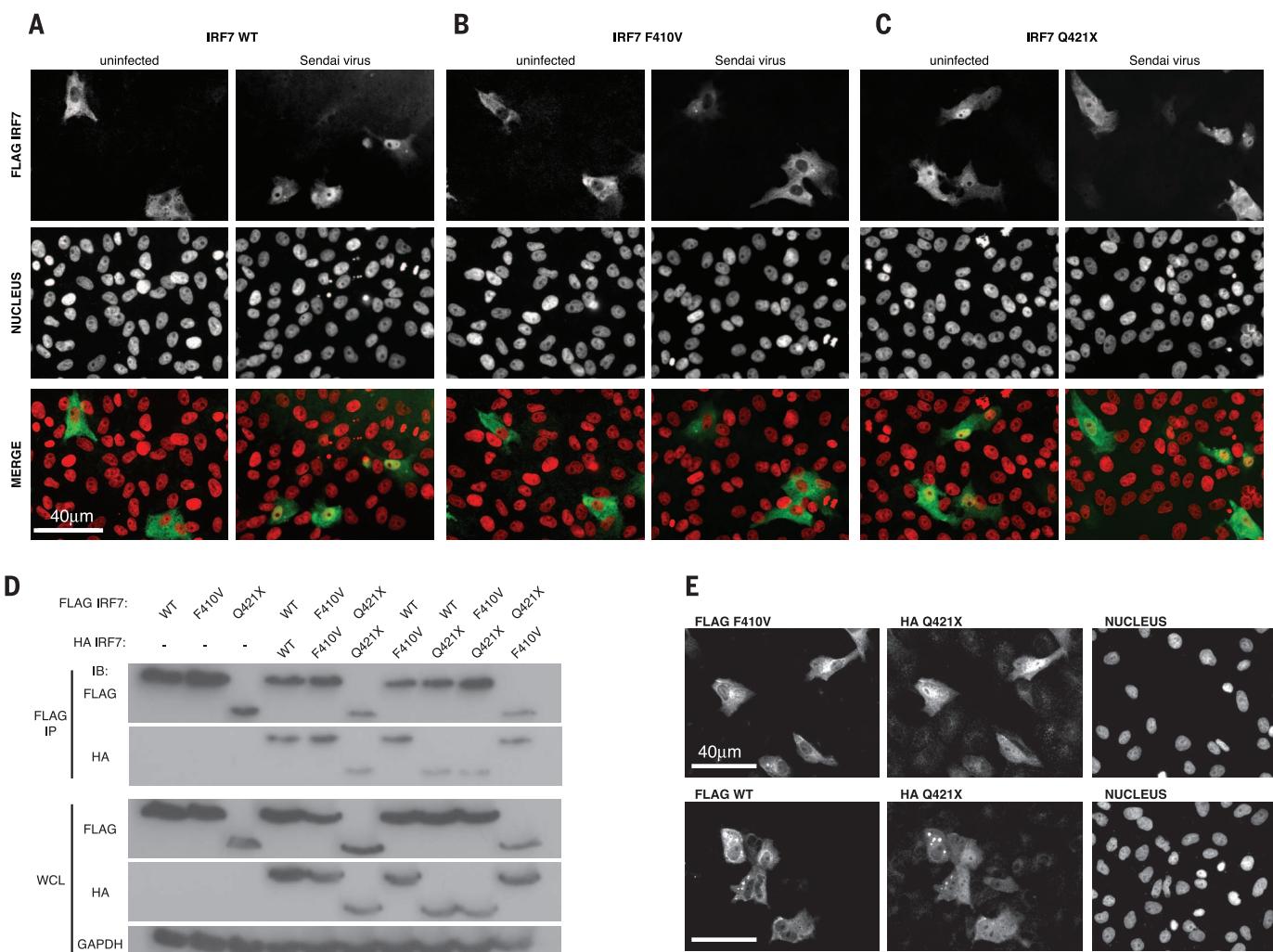
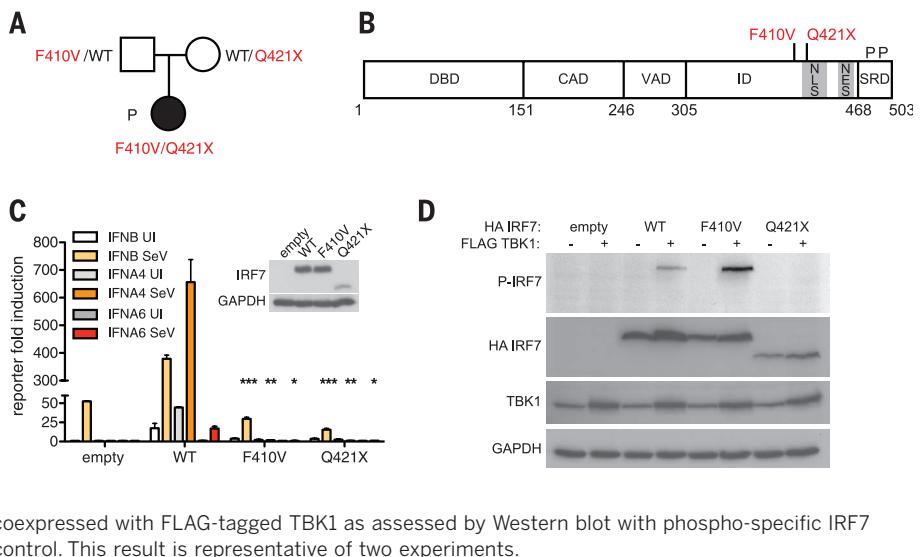


Fig. 2. P's *IRF7* alleles are loss-of-function by different mechanisms. **(A to C)** Localization of FLAG-tagged wild-type (A), F410V (B), or Q421X (C) *IRF7* in uninfected or Sendai virus–infected Vero cells by immunofluorescence imaging. This result is representative of two experiments. **(D)** Wild-type and mutant *IRF7* dimerization by immunoprecipitation (IP) with antibody to FLAG followed by Western blot with antibodies to FLAG and HA. WCL, whole-cell lysate. **(E)** Localization of FLAG F410V (top) and FLAG wild-type *IRF7* (bottom) cotransfected with HA Q421X *IRF7* in Vero cells as assessed by immunofluorescence imaging. This result is representative of two experiments.

each parent being heterozygous for a single mutated allele (Fig. 1A and fig. S2B), which defined the best candidate genotype in this patient (9).

Interferon regulatory factor 7 (IRF7) is a transcription factor that amplifies interferon (IFN) production in response to viruses (10–12). Specifically, IRF7 is involved in the amplification of mouse and human type I (13 *IFNA*, *IFNB*, *IFNE*, *IFNK*, *IFNW*) and type III (*IL29*, *IL28A*, *IL28B*) IFN genes (11–13). The missense F410V substitution is predicted to be damaging and absent in public databases. The nonsense Q421X (Fig. 1B and fig. S2B) is found as a heterozygous variant in 1 out of 118,062 chromosomes of the Exome Aggregation Consortium (ExAC) cohort, yielding a minor allele frequency of 0.000008. There are

currently no homozygous or compound heterozygous nonsynonymous mutations found in our in-house (table S3) and public databases (ExAC). Thus, autosomal recessive IRF7 deficiency by compound heterozygosity may underlie severe influenza in this child.

Each mutation was loss-of-function in reporter assays driven by *IFNB*, *IFNA4*, or *IFNA6* promoters (Fig. 1C) (11). None of the other five heterozygous *IRF7* variants tested was loss-of-function (table S4 and fig. S3). IRF7 can be activated by RIG-I recognition of IAV genomic RNA, resulting in C-terminal serine phosphorylation by the IKK-related kinases TBK1 and IKK- ϵ (11, 14–20). Wild-type IRF7 was phosphorylated upon overexpression of TBK1, as was F410V (Fig. 1D). The truncated

Q421X product lacks the C-terminal serine residues and was not phosphorylated (Fig. 1D). Phosphorylation induces IRF7 nuclear accumulation and transcription of type I and III IFNs (11, 14–16). Wild-type IRF7 accumulated in the nuclei of all transfected Vero cells by 8 hours post-infection (hpi) with Sendai virus (SeV) (Fig. 2A), whereas F410V was cytoplasmic (Fig. 2B) and Q421X was nuclear with or without SeV infection (Fig. 2C). This suggested that F410V disrupts a potential nuclear localization signal (amino acids 417 to 440) and was consistent with Q421X missing the nuclear export signal (amino acids 448 to 462) (15).

The two mutant *IRF7* alleles led to loss of function by different mechanisms: F410V did not accumulate in the nucleus despite phosphorylation,

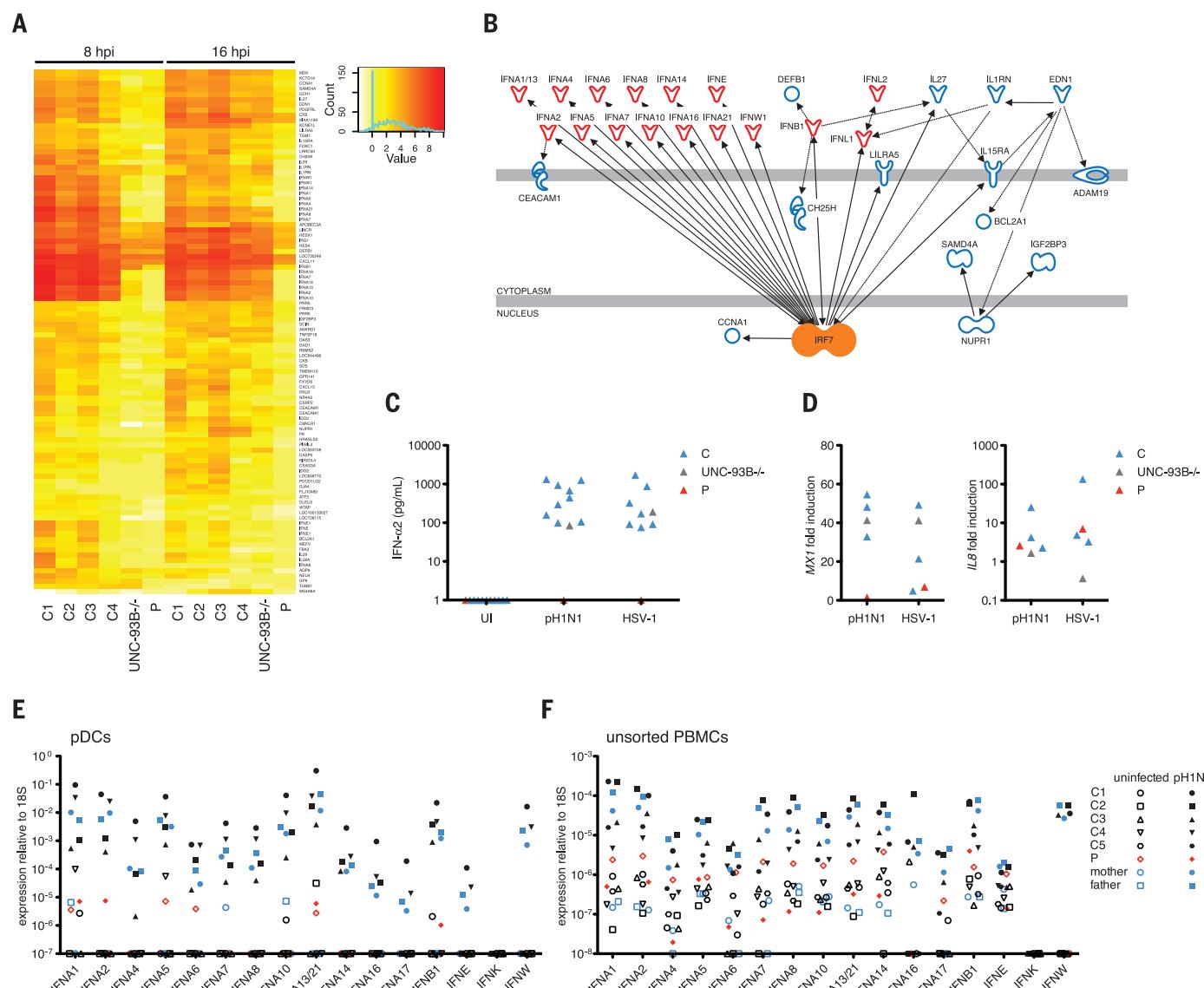


Fig. 3. Impaired IRF7-dependent innate immunity in leukocytes and pDCs. (A) The top 5% of genes, in PBMCs infected with pH1N1 at MOI (multiplicity of infection) = 2, whose relative changes were >2 (up or down) in controls and ≤ 1 in the patient, P, as analyzed by gene array. (B) Causal network analysis of the differentially regulated genes. (C) IFN- α production at 24 hpi with pH1N1 IAV or HSV-1 in pDCs from four healthy controls, P, and an

UNC-93B-deficient individual (UNC-93B^{-/-}), a control for TLR responses. (D) MX1 and *IL8* were measured in cells from (C) by qPCR. (E and F) The expression of all indicated type I IFN genes was measured by qPCR in purified pDCs (E) and unsorted PBMCs (F) separately infected with pH1N1 IAV at MOI = 1. The probe for IFNA13 also detects IFNA21 mRNA. All data shown in (C) to (F) are representative of two independent experiments.

whereas Q421X resided in the nucleus without phosphorylation. The mutant IRF7 proteins could homo- or heterodimerize (Fig. 2D), which suggests that coexpression might enable nuclear unphosphorylated Q421X to shuttle phosphorylated cytoplasmic F410V to the nucleus, where

IRF7 mutant heterodimers could up-regulate IFNs. However, there was no nuclear relocalization of F410V or the wild type in the presence of Q421X

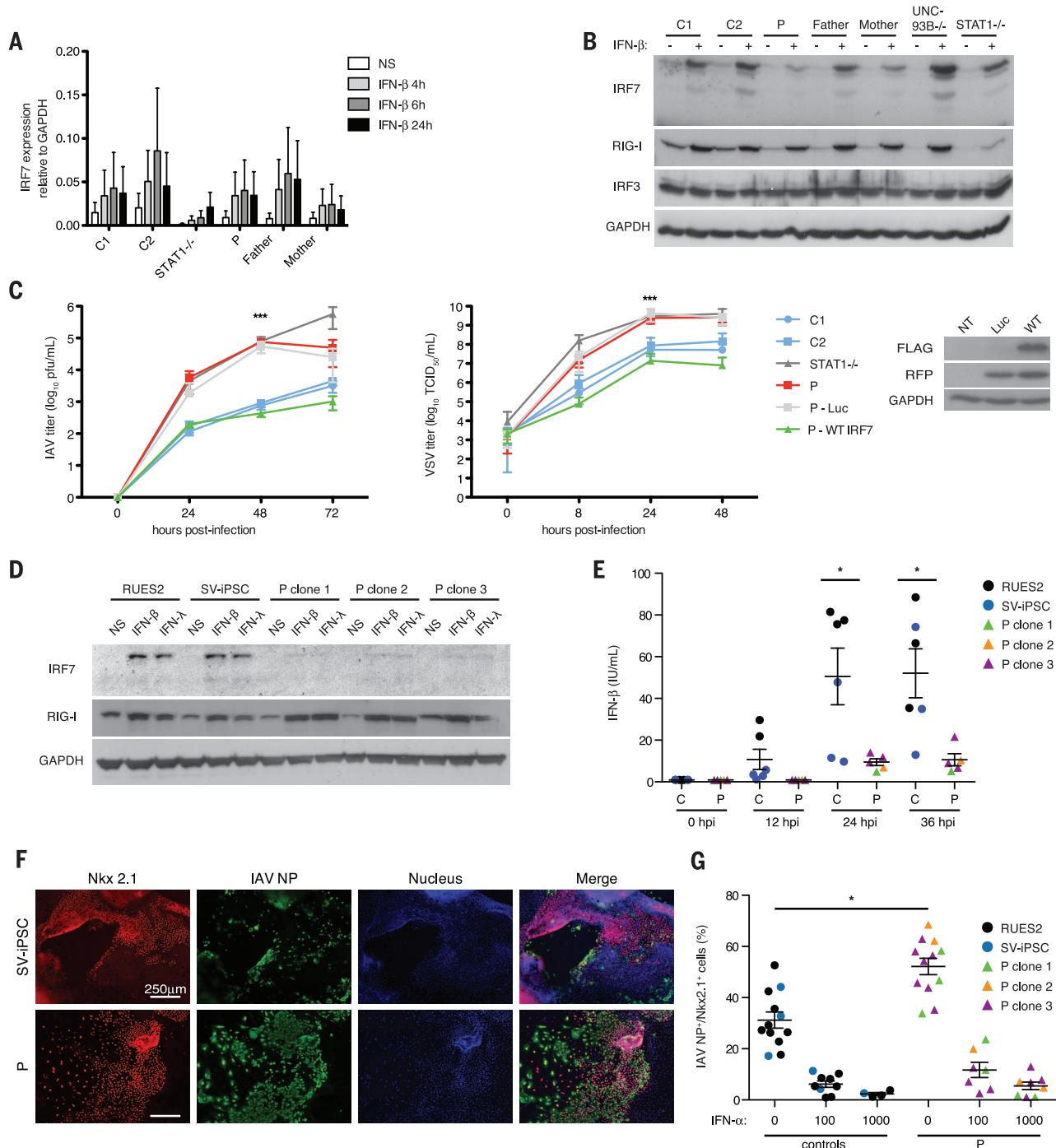


Fig. 4. IRF7-dependent intrinsic immunity is required for control of IAV infection. (A) IRF7 mRNA induction at indicated time points after IFN- β treatment in F-SV40. Means \pm SD of three replicates are shown. (B) IRF7 expression in F-SV40 measured by Western blot 18 hours after treatment with IFN- β . (C) Virus titers in F-SV40 from P stably transfected luciferase (Luc) or wild-type IRF7 with an internal ribosome entry site-expressed red fluorescent protein, after infection with pHIN1 (MOI = 10) or VSV (MOI = 3). Means \pm SD for pHIN1 IAV ($n = 3$) and VSV ($n = 7$) are shown. *** $P < 0.001$ between controls and P as determined by t test. (D) IRF7 expression in IFN- β - or IFN- λ -treated PECs derived from healthy control ESCs (RUES2), SV-iPSCs, or three individual clones

(clones 1, 2, and 3) of P's iPSCs as detected by Western blot. (E) IFN- β production in PECs infected with A/PR/8/34-GFP as measured by ELISA at indicated time points. Means \pm SD of two independent experiments are shown. * $P < 0.05$ as determined by t test. (F) Staining of IAV nucleoprotein (NP) (green) and Nkx2.1 (red) in PECs derived from SV-iPSC control and P infected at MOI = 1 with pHIN1 IAV. Cells derived from a single representative clone of P's iPSCs are shown. (G) Percentage of Nkx2.1 $^{+}$ cells scored as positive for IAV NP, in PECs untreated or treated with IFN- α (100 or 1000 U/ml) for 18 hours and infected with pHIN1 for 24 hours. Means \pm SD of two independent experiments are shown. * $P < 1 \times 10^{-6}$ as determined by χ^2 analysis.

(Fig. 2E). Consistently, there was no rescue in reporter assays when the two mutants were coexpressed (fig. S4A). Moreover, neither allele was dominant negative in terms of IRF7 or IRF3 function (fig. S4, B to D). This is consistent with the lack of infectious phenotype, including severe influenza, in P's heterozygous parents. Overall, the patient's severe influenza may result from lack of functional IRF7 homodimers, or IRF3-IRF7 heterodimers, or both.

We investigated the genome-wide impact of IAV infection in peripheral blood mononuclear cells (PBMCs) by microarray, quantitative polymerase chain reaction (qPCR), and enzyme-linked immunosorbent assay (ELISA). At baseline, PBMCs from P displayed significant down-regulation of innate immune genes (fig. S5). We observed a robust induction of type I IFN genes *IFNA14*, -16, -7, -2, -10, -13, and -21 (in order of relative change); *IFNE*; and *IFNW1*, as well as type III IFN genes *IL29* and *IL28A*, in healthy donors but not in P, at 8 and 16 hpi (Fig. 3A). Among IFN genes, the only exception was a factor of 2 induction of *IFNB* in P, albeit this was less than in controls by a factor of 4. Ingenuity pathway analysis software independently predicted IRF7 as an upstream regulator of these genes ($P = 2.02 \times 10^{-13}$) (Fig. 3B). IFN-stimulated genes (ISGs) known to inhibit IAV replication—*MX1* (21), *RSAD2* (22), *BST2* (23), and *SERPINE1* (24)—were up-regulated normally in P, perhaps via stimulation by IFN- β . Her PBMCs displayed a profound or complete defect of IFN- α 2 production after infection with 11 other viruses or stimulation with Toll-like receptor (TLR) agonists (fig. S6). IFN- β and IFN- λ 1 levels were impaired but less so than IFN- α , whereas IL-6 production was normal (fig. S6, B to D). Overall, we observed an overwhelming and selective defect of type I and III IFN induction in P's PBMCs.

In mice and humans, IRF7 is constitutively expressed in plasmacytoid dendritic cells (pDCs), resulting in pDCs being the major type I IFN-producing cells (25, 26). We measured IFN- α 2 production in response to pH1N1 infection in P's pDCs (Fig. 3C), which were found at normal frequency (table S5). We observed no IFN- α 2 production in pDCs from P at 24 hpi with pH1N1 and herpes simplex virus-1 (HSV-1) (Fig. 3C). *MX1* induction was abolished, while that of *IL8* was normal (Fig. 3D). The patient's heterozygous mother produced IFN- α 2 like controls (fig. S7A). We quantified the induction of all 20 human type I and type III IFN genes by qPCR in purified pDCs and unsorted PBMCs after infection with pH1N1 (Fig. 3, E and F). At 8 hpi, P was deficient for type I IFN (including the 13 *IFNA*, *IFNE*, *IFNK*, and *IFNW*) and type III IFN (*IL29*, *IL28A*, and *IL28B*) genes in both cell preparations (Fig. 3, E and F, and fig. S7B). *IFNB* was, however, mildly induced in P's pDCs (Fig. 3E), and the induction of *MX1* in P at 8 hpi was normal (fig. S7C). Collectively, a small amount of IRF7-independent IFN- β triggered early ISG up-regulation at 8 hpi in pDCs, but the IRF7-dependent amplification of type I and III IFNs was lacking for sustained ISG induction at 24 hpi.

We investigated the impact of IRF7 deficiency on cell-intrinsic, nonhematopoietic immunity, using P's SV40-immortalized fibroblasts (F-SV40). Basal and IFN- β -induced *IRF7* mRNA expression in F-SV40 from P was normal (Fig. 4A). In contrast, IRF7 protein expression was diminished, even after IFN- β treatment (Fig. 4B). We observed approximately 2-log higher titers of IAV at 48 hpi relative to healthy controls (Fig. 4C). Stable transfection of wild-type IRF7 complemented this phenotype (Fig. 4C). Wild-type IRF7 similarly rescued enhanced replication of vesicular stomatitis virus (VSV) (Fig. 4C). Highly pathogenic avian H5N1 IAV also replicated to high titers in P's F-SV40, indicating that the phenotype was not IAV strain-specific (fig. S8A). Further, treatment with exogenous IFN- α 2b protected the fibroblasts from IAV and VSV replication (fig. S8, B and C). IFN- β production by fibroblasts was impaired after IAV infection (fig. S8D); however, it was normal after stimulation with extracellular or intracellular polyinosine-polycytidine (synthetic double-stranded RNA) (fig. S8E). This is consistent with intact IRF3-dependent signaling (fig. S4D) and detectable IFN- β production in P's PBMCs (fig. S6A).

IAVs first target the entire respiratory tract, with pH1N1 2009 viral antigen present in type I and type II pneumocytes in humans (2, 27). We generated patient-specific pulmonary epithelial cells (PECs) from induced pluripotent stem cells (iPSCs) derived from P's primary fibroblasts (28). IAV replication and IFN- β induction were compared in PECs derived from an embryonic stem cell line (RUES2), a SeV-reprogrammed healthy control iPSC line (SV-iPSC), and three independent IRF7-deficient iPSC clones. IRF7 expression in response to IFN- α , IFN- β , and IFN- λ 1 treatment and IFN- β production in response to IAV were impaired in P lines (Fig. 4, D and E). Overall titers of IAV appeared equal regardless of *IRF7* genotype, largely because of efficient replication in cells negative for Nkx2.1, a pulmonary epithelium marker in these cultures. However, when we scored the infected Nkx2.1⁺ cells for IAV nucleoprotein (NP) antigen (Fig. 4F), 52.4% of P's PECs were double positive, versus 27.2% in controls (Fig. 4G). This phenotype was rescued by treatment with IFN- α 2b (Fig. 4G), IFN- β , or IFN- λ 1 (fig. S9). These data suggest that impaired intrinsic immunity in the pulmonary epithelium may have contributed to P's ARDS. Impaired IFN production by P's pDCs may also have caused disease, as evidenced by the pulmonary cell rescue with exogenous IFN- α 2b.

A single-gene inborn error of IRF7 immunity can underlie life-threatening, isolated influenza in humans during primary infection, broadening the range of human infections that result from genetic lesions (8, 9, 29, 30). IRF7 deficiency disrupts the main function of pDCs, the production of antiviral IFNs. This is distinct from deficiencies of IRF8 and GATA2 that impair development of all circulating monocytes and DCs and of multiple myeloid and lymphoid subsets, respectively (7, 31, 32). IRF7 deficiency also affects cell-autonomous, intrinsic immunity in PECs. The lack of IRF7-dependent type I and

III IFN amplification by pDCs, PECs, or possibly other cell types likely underlies the patient's severe influenza. Interestingly, IRF7 is redundant for vaccine-mediated immunity to influenza viruses, as the child has been healthy for 5 years with annual influenza vaccination as the sole secondary prevention. The IRF7-deficient child, now 7 years old, also displays a narrow infectious phenotype, restricted to severe influenza, at odds with the broad role of mouse IRF7 in antiviral immunity (10, 30). Human IRF7 seems to be largely redundant in host defense against viruses. It will be important to search for deficiencies in *IRF7* and related genes in children with influenza and other severe viral illnesses. Our study provides proof of principle that single-gene inborn errors of immunity can cause severe childhood influenza. IFN-based, patient-tailored therapeutic strategies could be helpful in life-threatening influenza of childhood (33, 34).

REFERENCES AND NOTES

1. P. Palese, M. L. Shaw, in *Fields Virology*, D. Knipe, P. Howley, Eds. (Lippincott Williams & Wilkins, Philadelphia, ed. 5, 2007), pp. 1647–1689.
2. W.-J. Shieh et al., *Am. J. Pathol.* **177**, 166–175 (2010).
3. F. S. Dawood et al., *Pediatrics* **128**, e27–e32 (2011).
4. Centers for Disease Control and Prevention, FluView: Influenza-Associated Hospitalization Surveillance Network (2013); <http://gis.cdc.gov/grasp/fluview/FluHospChars.html>.
5. A. R. Gennery, A. J. Cant, *J. Clin. Pathol.* **54**, 191–195 (2001).
6. W. Al-Herz et al., *Front. Immunol.* **5**, 162 (2014).
7. V. Bigley et al., *J. Exp. Med.* **208**, 227–234 (2011).
8. J. L. Casanova, L. Abel, *Annu. Rev. Genomics Hum. Genet.* **14**, 215–243 (2013).
9. J. L. Casanova, M. E. Conley, S. J. Seligman, L. Abel, L. D. Notarangelo, *J. Exp. Med.* **211**, 2137–2149 (2014).
10. K. Honda et al., *Nature* **434**, 772–777 (2005).
11. I. Marié, J. E. Durbin, D. E. Levy, *EMBO J.* **17**, 6660–6669 (1998).
12. P. I. Osterlund, T. E. Pietilä, V. Veckman, S. V. Kotenko, I. Julkunen, *J. Immunol.* **179**, 3434–3442 (2007).
13. M. Sato et al., *FEBS Lett.* **441**, 106–110 (1998).
14. A. Caillaud, A. G. Hovanessian, D. E. Levy, I. J. Marié, *J. Biol. Chem.* **280**, 17671–17677 (2005).
15. R. Lin, Y. Mamane, J. Hiscott, *J. Biol. Chem.* **275**, 34320–34327 (2000).
16. I. Marié, E. Smith, A. Prakash, D. E. Levy, *Mol. Cell. Biol.* **20**, 8803–8814 (2000).
17. H. Kato et al., *J. Exp. Med.* **205**, 1601–1610 (2008).
18. J. Rehwinkel et al., *Cell* **140**, 397–408 (2010).
19. H. Kato et al., *Nature* **441**, 101–105 (2006).
20. M. Yoneyama et al., *Nat. Immunol.* **5**, 730–737 (2004).
21. P. Staeheli, O. Haller, W. Boll, J. Lindenmann, C. Weissmann, *Cell* **44**, 147–158 (1986).
22. X. Wang, E. R. Hinson, P. Cresswell, *Cell Host Microbe* **2**, 96–105 (2007).
23. B. Mangeat et al., *J. Biol. Chem.* **287**, 22015–22029 (2012).
24. M. Dittmann et al., *Cell* **160**, 631–643 (2015).
25. M. Kerkmann et al., *J. Immunol.* **170**, 4465–4474 (2003).
26. A. Izaguirre et al., *J. Leukoc. Biol.* **74**, 1125–1138 (2003).
27. Y. Itoh et al., *Nature* **460**, 1021–1025 (2009).
28. S. X. Huang et al., *Nat. Biotechnol.* **32**, 84–91 (2014).
29. M. E. Conley, J. L. Casanova, *Curr. Opin. Immunol.* **30**, 17–23 (2014).
30. J. L. Casanova, L. Abel, L. Quintana-Murci, *Cold Spring Harb. Symp. Quant. Biol.* **78**, 157–172 (2013).
31. S. Hambleton et al., *N. Engl. J. Med.* **365**, 127–138 (2011).
32. M. Pasquet et al., *Blood* **121**, 822–829 (2013).
33. S. Baron, A. Isaacs, *BMJ* **1**, 18–20 (1962).
34. C. Agrati et al., *J. Infect. Dis.* **202**, 681–689 (2010).

ACKNOWLEDGMENTS

We warmly thank our patient and her family. We thank S. Boucherit, T. Kochetkov, M. Fenner, M. Duchamp, A. Abhyankar, A. Belkadi, L. Shang, Y. Liang, L. Amar, and Y. Nemirovskaya for their

contributions and all members of the laboratory for fruitful discussions. Plasmids used to generate recombinant influenza viruses are subject to a material transfer agreement. We thank members of the French Memo-Flu-ARDS study group (PIs: B. Autran, Inserm U945, and C.-E. Luyt, Service de réanimation médicale, Hôpital Pitie-Salpêtrière, Paris) for recruiting adult severe flu patients. Supported by National Center for Research Resources and National Center for Advancing Translational Sciences grant 8UL1TR000043; NIH grants 5R01NS072381 (J.-L.C. and L.D.N.), 5R01AI00887 (L.D.N.), and 1U19AI109945 (C.F.B.); the Rockefeller University, the St. Giles Foundation, the French National Research Agency under the "Investments for the Future" program grant ANR-10-IAHU-01, the

Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (ANR-10-LABX-62-IBED), Inserm, Paris Descartes University, and ERC grant 2010-StG-261299 (F.G.); and Center for Research on Influenza Pathogenesis and the NIAID-funded Center of Excellence for Influenza Research and Surveillance contract HHSN272201400008C (A.G.-S.). The data presented in this manuscript are tabulated in the main paper and in the supplementary materials. Raw data for the microarray analyses performed in this study are available from the public repository of GEO DataSets (accession no. GSE66486). The raw sequence data are available on the Sequence Read Archive (SRA) database: Bioproject SRP055919.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/448/suppl/DC1
Case Report
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 to S7
References (35–50)

24 October 2014; accepted 12 March 2015
Published online 26 March 2015;
10.1126/science.aaa1578

NEUROBIOLOGY

SARM1 activation triggers axon degeneration locally via NAD⁺ destruction

Josiah Gerdts,¹ E.J. Brace,² Yo Sasaki,¹ Aaron DiAntonio,^{2,3} Jeffrey Milbrandt*,^{1,3}

Axon degeneration is an intrinsic self-destruction program that underlies axon loss during injury and disease. Sterile alpha and TIR motif-containing 1 (SARM1) protein is an essential mediator of axon degeneration. We report that SARM1 initiates a local destruction program involving rapid breakdown of nicotinamide adenine dinucleotide (NAD⁺) after injury. We used an engineered protease-sensitized SARM1 to demonstrate that SARM1 activity is required after axon injury to induce axon degeneration. Dimerization of the Toll-interleukin receptor (TIR) domain of SARM1 alone was sufficient to induce locally mediated axon degeneration. Formation of the SARM1 TIR dimer triggered rapid breakdown of NAD⁺, whereas SARM1-induced axon destruction could be counteracted by increased NAD⁺ synthesis. SARM1-induced depletion of NAD⁺ may explain the potent axon protection in Wallerian degeneration slow (Wld^s) mutant mice.

Cells undergo regulated self-destruction during development and in response to stresses (1). Axons, the longest cellular structures in the body, have a locally mediated self-destruction program that removes damaged axons but also promotes axon loss in the setting of neurological disorders (2). Axon degeneration is antagonized by the Wallerian degeneration slow (Wld^s) chimeric protein (3). The active moiety of Wld^s is the enzyme nicotinamide mononucleotide adenyltransferase 1 (Nmnat1), which synthesizes the essential cofactor nicotinamide adenine dinucleotide (NAD⁺) (4), but the function of Nmnat1 and NAD⁺ in axon protection remains unclear (2). The protein SARM1 (sterile alpha and TIR motif-containing 1; TIR, Toll-interleukin receptor) is an essential mediator of axon degeneration (5, 6). SARM1 is a negative regulator of Toll-like receptor-activated transcriptional programs (7), but its mechanism for axon degeneration is unknown.

To investigate whether SARM1 functions before or after injury, we engineered a system to

inactivate SARM1 with pharmacologic control. Protease-sensitized SARM1 (SARM^{ps}) contains a tobacco etch virus (TEV) protease consensus sequence between the sterile alpha motif (SAM) and TIR domains, which are both essential for SARM1 function (6). SARM^{ps} is thus cleaved and inactivated by TEV protease. SARM^{ps} was fused to the rapamycin-binding domain Frb and the N-terminal portion of split TEV protease (Ntev) (8) and coexpressed with C-terminal split TEV fused to FK866 binding protein (Fkbp-Ctev), allowing rapamycin-induced cleavage (Fig. 1A and fig. S1). In dorsal root ganglion (DRG) neurons, cleavage of SARM^{ps} was mostly complete within 60 min of rapamycin treatment (Fig. 1B and fig. S2A). SARM^{ps} functionality was verified by expression of SARM^{ps} in isolated Sarm1^{-/-} DRG neurons. When Sarm1^{-/-} axons were severed (diagrammed in Fig. 1C), they remained intact after 24 hours, whereas axons of neurons expressing SARM^{ps} showed degeneration measured by axon morphometry (Fig. 1D), similar to wild-type axons. SARM^{ps} function was lost upon cleavage triggered by rapamycin in the presence of Fkbp-Ctev (Fig. 1, D and E) or by expression of full-length TEV (fig. S2B). Cleavage of SARM^{ps} initiated 12 hours before or up to 2 hours after axon transection fully suppressed axon degeneration measured 24 hours after axotomy. Because cleavage of SARM^{ps} after axons were disconnected from

cell bodies resulted in protection, SARM1 must function after injury to promote degeneration.

SARM1 has no predicted enzymatic function but contains a TIR domain, which is the effector domain of Toll-like receptors (TLRs). Activation of TLRs results in dimerization of TIR domains that transmit a signal to cytosolic effector proteins (9). We tested whether multimerization of the TIR domain of SARM1 (sTIR) might induce axon degeneration. A minimal region of human SARM1 comprising sTIR and the adjacent multimerization (SAM) domains, but lacking the auto-inhibitory N terminus (SAM-TIR), is constitutively active and promotes cell and axon destruction in cultured DRG neurons (6). Expression of this activated form of SARM1 in vivo in *Drosophila* motor (Fig. 2A) or sensory neurons (fig. S3) also caused cell and axon destruction. This degeneration was not observed in *Drosophila* expressing SAM-TIR harboring a disruptive sTIR mutation.

To evaluate the sufficiency of sTIR dimerization in axon destruction, we engineered a pharmacologically controlled dimerizable sTIR by fusing it to the rapamycin-binding domains Frb and Fkbp (Fig. 2B) (10). We expressed Frb-sTIR and Fkbp-sTIR in DRG neurons and found that sTIR dimerization by rapamycin induced axon fragmentation within 12 hours (Fig. 2C) and neuronal cell death within 24 hours (Fig. 2D). sTIR-induced toxicity did not require the inhibition of mammalian target of rapamycin (mTOR), because the rapamycin analog AP20187, which does not target mTOR, also stimulated axon degeneration in cells expressing the homodimerizable Fkbp^{F36V}-sTIR (10). SARM1 activation is thus sufficient to elicit axonal and neuronal destruction. Cell and axon degeneration were not induced upon dimerization of the TIR domains of TLR4 or the adaptor MYD88 (Fig. 2E).

We tested whether SARM1 promotes axon degeneration through a local mechanism. sTIR-induced degeneration does not require a physical connection between the axon and soma: Sarm1^{-/-} axons persisted after severing; however, sTIR dimerization by AP20187 caused fragmentation of these severed segments within 12 hours (Fig. 2F). Dimerization of sTIR locally within axons also led to selective axon destruction. We grew DRG neurons in adjacent fluid compartments: one containing the cell bodies and proximal axons and the other containing only distal axons (Fig. 2G). Application of AP20187 to both compartments led to destruction of proximal and distal axons, whereas selective application to the distal

¹Department of Genetics, Washington University Medical School, Saint Louis, MO, USA. ²Department of Developmental Biology, Washington University Medical School, Saint Louis, MO, USA. ³Hoppe Center for Neurological Disorders, Saint Louis, MO, USA.

*Corresponding author. E-mail: jmilbrandt@wustl.edu

contributions and all members of the laboratory for fruitful discussions. Plasmids used to generate recombinant influenza viruses are subject to a material transfer agreement. We thank members of the French Memo-Flu-ARDS study group (PIs: B. Autran, Inserm U945, and C.-E. Luyt, Service de réanimation médicale, Hôpital Pitie-Salpêtrière, Paris) for recruiting adult severe flu patients. Supported by National Center for Research Resources and National Center for Advancing Translational Sciences grant 8UL1TR000043; NIH grants 5R01NS072381 (J.-L.C. and L.D.N.), 5R01AI00887 (L.D.N.), and 1U19AI09945 (C.F.B.); the Rockefeller University, the St. Giles Foundation, the French National Research Agency under the "Investments for the Future" program grant ANR-10-IAHU-01, the

Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (ANR-10-LABX-62-IBED), Inserm, Paris Descartes University, and ERC grant 2010-StG-261299 (F.G.); and Center for Research on Influenza Pathogenesis and the NIAID-funded Center of Excellence for Influenza Research and Surveillance contract HHSN272201400008C (A.G.-S.). The data presented in this manuscript are tabulated in the main paper and in the supplementary materials. Raw data for the microarray analyses performed in this study are available from the public repository of GEO DataSets (accession no. GSE66486). The raw sequence data are available on the Sequence Read Archive (SRA) database: Bioproject SRP055919.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/448/suppl/DC1
Case Report
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 to S7
References (35–50)

24 October 2014; accepted 12 March 2015
Published online 26 March 2015;
10.1126/science.aaa1578

NEUROBIOLOGY

SARM1 activation triggers axon degeneration locally via NAD⁺ destruction

Josiah Gerdts,¹ E.J. Brace,² Yo Sasaki,¹ Aaron DiAntonio,^{2,3} Jeffrey Milbrandt*,^{1,3}

Axon degeneration is an intrinsic self-destruction program that underlies axon loss during injury and disease. Sterile alpha and TIR motif-containing 1 (SARM1) protein is an essential mediator of axon degeneration. We report that SARM1 initiates a local destruction program involving rapid breakdown of nicotinamide adenine dinucleotide (NAD⁺) after injury. We used an engineered protease-sensitized SARM1 to demonstrate that SARM1 activity is required after axon injury to induce axon degeneration. Dimerization of the Toll-interleukin receptor (TIR) domain of SARM1 alone was sufficient to induce locally mediated axon degeneration. Formation of the SARM1 TIR dimer triggered rapid breakdown of NAD⁺, whereas SARM1-induced axon destruction could be counteracted by increased NAD⁺ synthesis. SARM1-induced depletion of NAD⁺ may explain the potent axon protection in Wallerian degeneration slow (Wld^s) mutant mice.

Cells undergo regulated self-destruction during development and in response to stresses (1). Axons, the longest cellular structures in the body, have a locally mediated self-destruction program that removes damaged axons but also promotes axon loss in the setting of neurological disorders (2). Axon degeneration is antagonized by the Wallerian degeneration slow (Wld^s) chimeric protein (3). The active moiety of Wld^s is the enzyme nicotinamide mononucleotide adenyltransferase 1 (Nmnat1), which synthesizes the essential cofactor nicotinamide adenine dinucleotide (NAD⁺) (4), but the function of Nmnat1 and NAD⁺ in axon protection remains unclear (2). The protein SARM1 (sterile alpha and TIR motif-containing 1; TIR, Toll-interleukin receptor) is an essential mediator of axon degeneration (5, 6). SARM1 is a negative regulator of Toll-like receptor-activated transcriptional programs (7), but its mechanism for axon degeneration is unknown.

To investigate whether SARM1 functions before or after injury, we engineered a system to

inactivate SARM1 with pharmacologic control. Protease-sensitized SARM1 (SARM^{ps}) contains a tobacco etch virus (TEV) protease consensus sequence between the sterile alpha motif (SAM) and TIR domains, which are both essential for SARM1 function (6). SARM^{ps} is thus cleaved and inactivated by TEV protease. SARM^{ps} was fused to the rapamycin-binding domain Frb and the N-terminal portion of split TEV protease (Ntev) (8) and coexpressed with C-terminal split TEV fused to FK866 binding protein (Fkbp-Ctev), allowing rapamycin-induced cleavage (Fig. 1A and fig. S1). In dorsal root ganglion (DRG) neurons, cleavage of SARM^{ps} was mostly complete within 60 min of rapamycin treatment (Fig. 1B and fig. S2A). SARM^{ps} functionality was verified by expression of SARM^{ps} in isolated Sarm1^{-/-} DRG neurons. When Sarm1^{-/-} axons were severed (diagrammed in Fig. 1C), they remained intact after 24 hours, whereas axons of neurons expressing SARM^{ps} showed degeneration measured by axon morphometry (Fig. 1D), similar to wild-type axons. SARM^{ps} function was lost upon cleavage triggered by rapamycin in the presence of Fkbp-Ctev (Fig. 1, D and E) or by expression of full-length TEV (fig. S2B). Cleavage of SARM^{ps} initiated 12 hours before or up to 2 hours after axon transection fully suppressed axon degeneration measured 24 hours after axotomy. Because cleavage of SARM^{ps} after axons were disconnected from

cell bodies resulted in protection, SARM1 must function after injury to promote degeneration.

SARM1 has no predicted enzymatic function but contains a TIR domain, which is the effector domain of Toll-like receptors (TLRs). Activation of TLRs results in dimerization of TIR domains that transmit a signal to cytosolic effector proteins (9). We tested whether multimerization of the TIR domain of SARM1 (sTIR) might induce axon degeneration. A minimal region of human SARM1 comprising sTIR and the adjacent multimerization (SAM) domains, but lacking the auto-inhibitory N terminus (SAM-TIR), is constitutively active and promotes cell and axon destruction in cultured DRG neurons (6). Expression of this activated form of SARM1 in vivo in *Drosophila* motor (Fig. 2A) or sensory neurons (fig. S3) also caused cell and axon destruction. This degeneration was not observed in *Drosophila* expressing SAM-TIR harboring a disruptive sTIR mutation.

To evaluate the sufficiency of sTIR dimerization in axon destruction, we engineered a pharmacologically controlled dimerizable sTIR by fusing it to the rapamycin-binding domains Frb and Fkbp (Fig. 2B) (10). We expressed Frb-sTIR and Fkbp-sTIR in DRG neurons and found that sTIR dimerization by rapamycin induced axon fragmentation within 12 hours (Fig. 2C) and neuronal cell death within 24 hours (Fig. 2D). sTIR-induced toxicity did not require the inhibition of mammalian target of rapamycin (mTOR), because the rapamycin analog AP20187, which does not target mTOR, also stimulated axon degeneration in cells expressing the homodimerizable Fkbp^{F36V}-sTIR (10). SARM1 activation is thus sufficient to elicit axonal and neuronal destruction. Cell and axon degeneration were not induced upon dimerization of the TIR domains of TLR4 or the adaptor MYD88 (Fig. 2E).

We tested whether SARM1 promotes axon degeneration through a local mechanism. sTIR-induced degeneration does not require a physical connection between the axon and soma: Sarm1^{-/-} axons persisted after severing; however, sTIR dimerization by AP20187 caused fragmentation of these severed segments within 12 hours (Fig. 2F). Dimerization of sTIR locally within axons also led to selective axon destruction. We grew DRG neurons in adjacent fluid compartments: one containing the cell bodies and proximal axons and the other containing only distal axons (Fig. 2G). Application of AP20187 to both compartments led to destruction of proximal and distal axons, whereas selective application to the distal

¹Department of Genetics, Washington University Medical School, Saint Louis, MO, USA. ²Department of Developmental Biology, Washington University Medical School, Saint Louis, MO, USA. ³Hoppe Center for Neurological Disorders, Saint Louis, MO, USA.

*Corresponding author. E-mail: jmilbrandt@wustl.edu

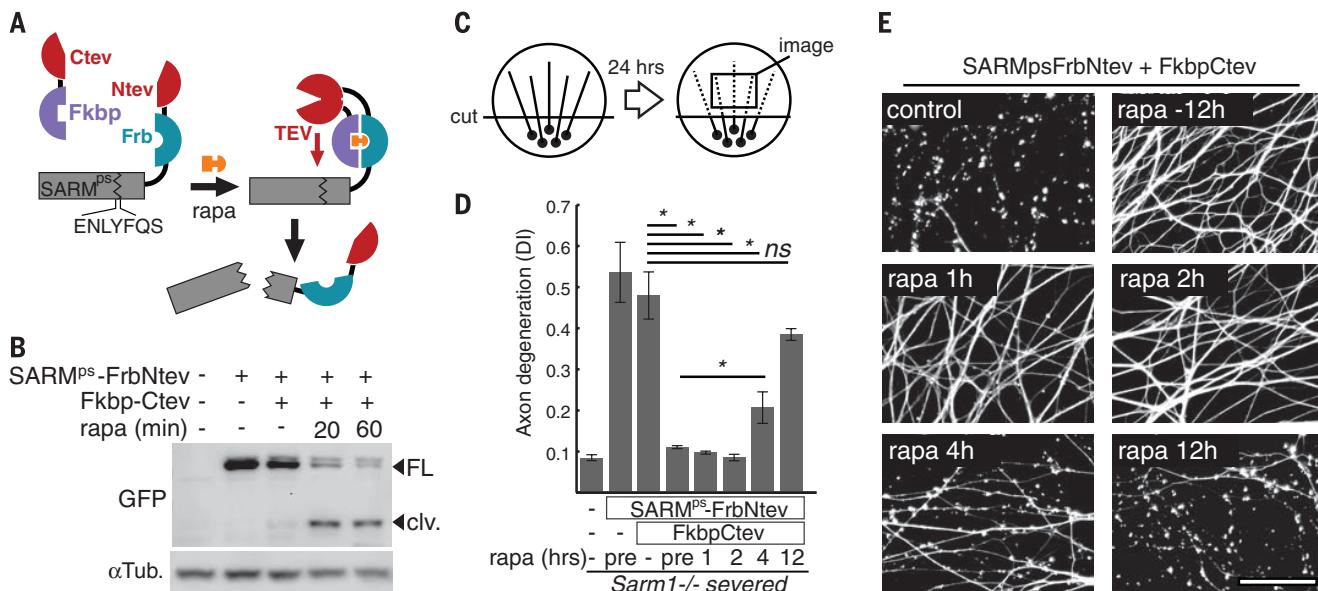


Fig. 1. SARM1 functions after axon injury to promote destruction. (A) Schematic showing how expression of SARM^{ps}-Frbs-Ntev with Fkbp-Ctev allows rapamycin-induced complementation of split TEV and concomitant SARM^{ps} cleavage. (B) Gel electrophoresis with anti-GFP immunoblot showing SARM^{ps} cleavage in DRG neurons induced by 100 nM rapamycin (rapa); FL, full length SARM^{ps}-Frbs-Ntev-Cerulean; clv., cleaved form. α -Tubulin (α Tub) was a loading control. (C) Diagram of in vitro injury model: Isolated DRG neurons were severed, and axon degeneration was quantified from axon images after 24 hours. (D) Requirement

for SARM1 activity after axotomy to induce axon degeneration. Axon degeneration is reported as the degeneration index (DI), a morphometric ratio of fragmented axon area to total axon area (13). Sarm1^{-/-} DRG neurons treated with expression lentiviruses (control, SARM^{ps}-FrbsNtev, and Fkbp-Ctev) were severed and treated with 100 nM rapamycin at various times (pre = 12 hours pre-injury). (E) Micrographs show representative α -tubulin-stained axons corresponding to select treatment groups in (D). Scale bar, 50 μ m. Error bars, SEM; *P < 0.01; one-way analysis of variance (ANOVA) with Tukey's post-hoc test.

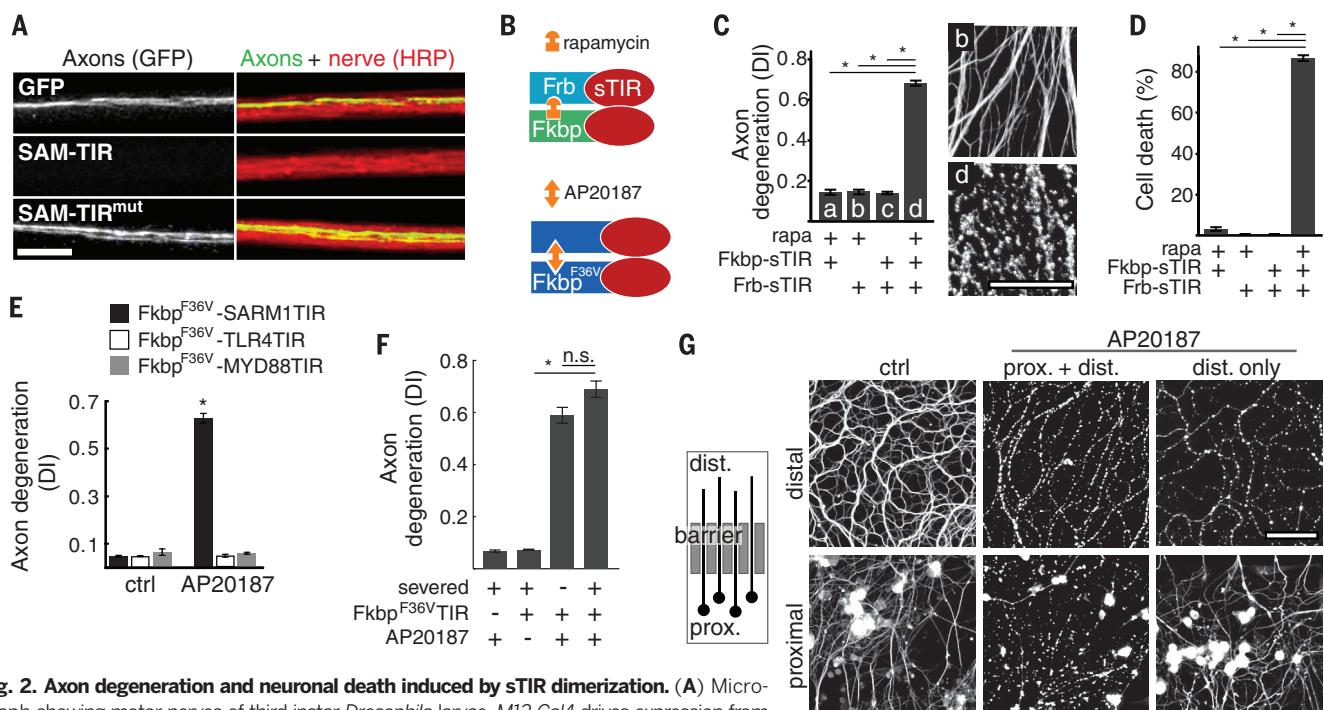


Fig. 2. Axon degeneration and neuronal death induced by sTIR dimerization. (A) Micrograph showing motor nerves of third-instar *Drosophila* larvae. M12-Gal4 drives expression from mCD8-GFP (green) alone or with either UAS-SAM-TIR or UAS-SAM-TIR^{mut} in single motor axons in each nerve (red, HRP). UAS-SAM-TIR expression caused axon loss in 49 out of 49 (49/49) nerves as shown, whereas SAM-TIR with a disruptive TIR mutation led to degeneration in 0/70 nerves ($\chi^2 = 119$; P < 0.001); scale bar = 10 μ m. (B) Schematic showing sTIR dimerization by rapamycin or AP20187. (C) Effect of sTIR, dimerized sTIR, and rapamycin on axon degeneration. α -Tubulin stained axons correspond to bars b and d. (D) Effect of sTIR dimerization on neuronal viability quantified by ethidium homodimer exclusion after 24 hours. (E) Effects of dimerization of sTIR or TIR domains of MYD88 or TLR4 on axon degeneration. (F) Effects of sTIR dimerization on degeneration of Sarm1^{-/-} axons physically disconnected from cell bodies. (G) (Left) Diagram of axons growing through a diffusion barrier into an isolated fluid compartment. (Right) Micrographs of isolated distal axon segments after application of AP20187 globally or selectively to distal axons. Scale bar, 50 μ m. Error bars, SEM; *P < 0.01; one-way ANOVA with Tukey's post-hoc test.

chamber elicited selective distal axon degeneration after 24 hours (Fig. 2G).

SARM1 TIR dimerization elicited rapid pathophysiological changes: Axon degeneration and neuronal death were evident within 1.5 and 6 hours, respectively (fig. S4, A and B), and neuronal mitochondrial membrane potential dissipated and calcium accumulated with similar kinetics (fig. S4, C to E). These measurements indicate early energetic failure. We thus focused on biochemical events leading from SARM1 activation to axonal demise. Axon degeneration is antagonized by the NAD⁺ synthetic enzyme Nmnat1, which, like SARM1, functions locally within axons (17). Injured axons exhibit declining levels of NAD⁺ before morphologic changes (12), but it is unknown whether this is a cause or consequence of axon destruction. Although *Wld^s*/Nmnat1 does not increase the steady-state abundance of NAD⁺ (13), in the setting of acute NAD⁺ depletion it might maintain sufficient levels of NAD⁺ for viability (diagrammed in Fig. 3A). We thus tested whether SARM1 activation leads to depletion of NAD⁺.

To test whether endogenous SARM1 is necessary for axonal loss of NAD⁺ after axotomy, we isolated axons from cultured wild-type and *Sarm1*^{-/-} DRG neurons 3 and 4 hours after injury, a time when they remain morphologically intact, and measured the abundance of NAD⁺ using high-

performance liquid chromatography (HPLC). The abundance of NAD⁺ decreased after injury in wild-type axons but remained stable in *Sarm1*^{-/-} axons (Fig. 3B). Loss of adenosine triphosphate (ATP), an expected consequence of NAD⁺ depletion, was also SARM1-dependent (fig. S5A). To determine whether SARM1 is also necessary for axotomy-induced loss of NAD⁺ in vivo, we compared concentrations of NAD⁺ in distal sciatic nerve segments from adult wild-type and *Sarm1*^{-/-} mice. At 30 hours after injury, amounts of NAD⁺ were decreased in wild-type nerves but remained stable in *Sarm1*^{-/-} nerves (Fig. 3C). At this time, injured nerves remained morphologically intact (fig. S5C), and amounts of ATP were stable (fig. S5B).

We tested whether SARM1 activation was sufficient to elicit loss of NAD⁺ by measuring neuronal NAD⁺ after sTIR dimerization. sTIR dimerization by the addition of AP20187 caused rapid loss of NAD⁺; within 15 min, the abundance of NAD⁺ was reduced by 66%, and by 90 min, 90% of the NAD⁺ was lost (Fig. 3D). The abundance of ATP also declined after sTIR dimerization, but its depletion was slower than that of NAD⁺.

Together these data implicate NAD⁺ loss as a critical step in SARM1-mediated axon destruction. We therefore examined whether increased NAD⁺ synthesis could counteract the destruction program activated by sTIR dimerization. In DRG

neurons, both axon degeneration and cell death initiated by sTIR dimerization were completely blocked by the expression of Nmnat1 and nicotinamide phosphoribosyltransferase (Nampt), which together synthesize NAD⁺ (Fig. 3A). Protection appeared to require NAD⁺ synthesis, because concurrent treatment with the Nampt inhibitor FK866 blocked the protection afforded by these enzymes (Fig. 3, E and F). Similarly, sTIR-induced axon degeneration and cell death were blocked by supplementation with the cell-permeant NAD⁺ precursor nicotinamide riboside (NR) (Fig. 3, G and H) (14). *Drosophila* larvae expressing the dimerizable Fkbp^{F36V}-sTIR in motor neurons that were fed AP20187 showed extensive axon degeneration that was blocked by coexpression of cytosolic Nmnat1 (Fig. 3I).

To extend our analysis of biochemical events after SARM1 activation, we created a heterologous human embryonic kidney (HEK293T) cell line (HTir) that stably expresses Frb-sTIR and Fkbp-sTIR. After 12 hours of sTIR dimerization in HTir cells, toxicity was evident, as indicated by the loss of ATP (fig. S6A) and altered morphology (fig. S6B). Both effects were blocked by NR supplementation. Inhibition of NAD⁺ synthesis with FK866 increased the loss of ATP, whereas FK866 was not toxic in the absence of sTIR dimerization (fig. S6, A and B).

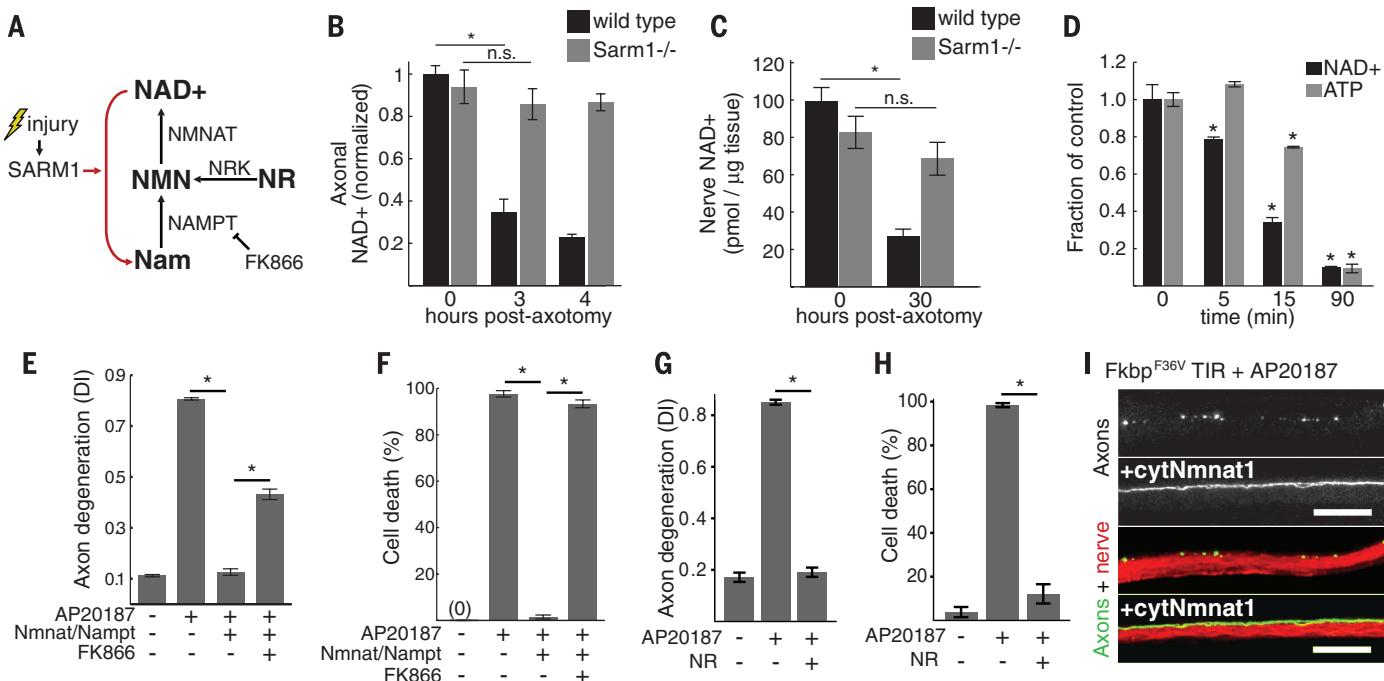


Fig. 3. Loss of NAD⁺ underlies SARM1-induced destruction. (A) Diagram of NAD⁺ synthesis and inhibition by FK866; Nrk, nicotinamide riboside kinase; NMN, nicotinamide mononucleotide. (B) Axonal NAD⁺ concentration in cultured wild-type and *Sarm1*^{-/-} DRG neurons after axotomy; normalized to wild-type control. (C) NAD⁺ concentration in distal sciatic nerve segments from wild-type or *Sarm1*^{-/-} animals after transection; wild-type $n = 5$; *Sarm1*^{-/-} $n = 9$. (D) Neuronal NAD⁺ and ATP concentrations after sTIR dimerization by AP20187; comparisons are made to 0 min control. (E and F) Axon degeneration (E) and neuronal cell death (F) induced by sTIR homodimerization (AP20187) and inhibition by NAD⁺ synthetic enzymes with or without the Nampt inhibitor

FK866 (10 nM); measured 24 hours after sTIR dimerization and FK866 application. (G and H) Effect of NR (1 mM) on axon degeneration (G) and neuronal cell death (H) induced by sTIR homodimerization (AP20187) for 24 hours with or without NR. (I) Micrographs showing sTIR-induced motor axon fragmentation in third-instar *Drosophila* larvae blocked by cytosolic Nmnat1 (cytNmnat1) expression. *M12-Gal4* drives expression from UAS-mCD8-GFP (green) and UAS-Fkbp^{F36V}-TIR with or without UAS-cytNmnat1 in single motor axons in each nerve (red, HRP). Degeneration score = 76 ± 4% (control) versus 11 ± 2% (cytNmnat1); $P < 0.001$ (*t* test); scale bar, 20 μm. Error bars, SEM; * $P < 0.01$; one-way ANOVA with Tukey's post-hoc test.

To evaluate whether NAD⁺ depletion alone is sufficient to induce axon destruction, we stimulated direct intracellular breakdown of NAD⁺ by dimerization of the poly-ADP-ribose polymerase (PARP) domain of Tankyrase 1 (Tnkp; diagrammed in Fig. 4A). We generated dimerizable Fkbp^{F36V}-Tnkp and showed that AP20187 treatment of cells expressing this construct led to loss of NAD⁺ and formation of PAR (fig. S7, A and B). In the presence of FK866, which inhibits de novo NAD⁺ synthesis, Tnkp dimerization in dividing cells led to rapid energetic failure (ATP loss) that was blocked by the Tankyrase inhibitor XAV939 (fig. S7C). NR supplementation blocked toxicity but not PAR formation, indicating that NAD⁺ loss and not PAR formation caused cell death (fig. S7, A and C). In neurons, Tnkp-induced depletion of NAD⁺ caused degeneration of uninjured wild-type and Sarm1^{-/-} axons (Fig. 4, B and C). Moreover, NAD⁺ depletion from isolated (presevered) Sarm1^{-/-} axons led to degeneration (Fig. 4C). Thus, rapid NAD⁺ depletion is sufficient to cause rapid axon loss.

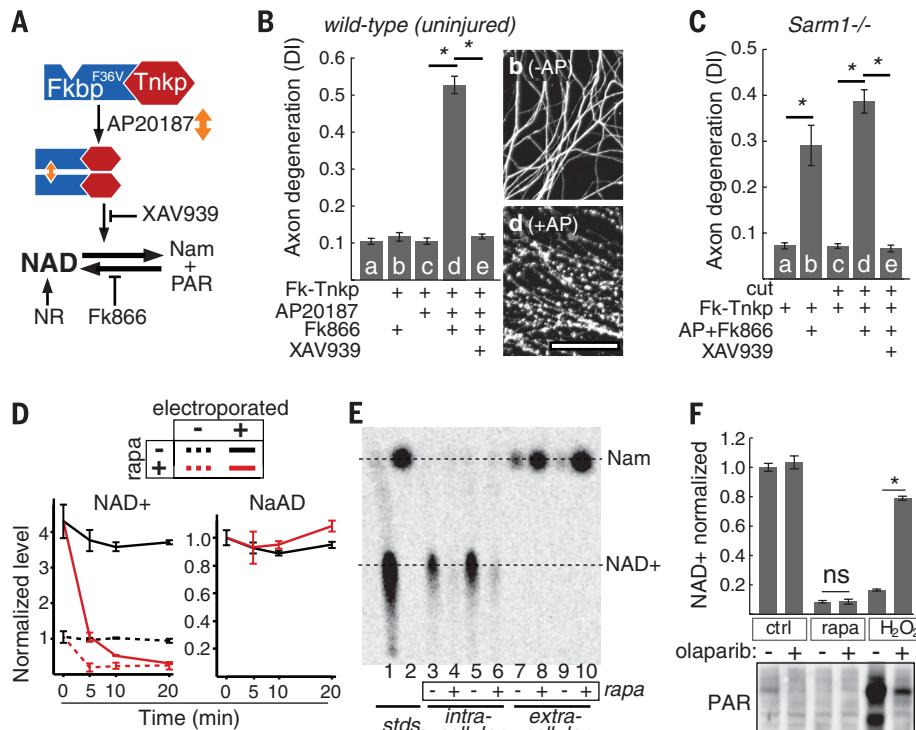


Fig. 4. Effects of NAD⁺ breakdown on axon degeneration. (A) Diagram of NAD⁺ manipulation using Tnkp dimerization. NAD⁺ loss induced by Fkbp^{F36V}-Tnkp dimerization is blocked by Tankyrase inhibitor XAV939 or NR and is exacerbated by FK866. (B) Axon degeneration in response to NAD⁺ depletion by dimerized Tnkp and FK866 after 24 hours (bar d) and inhibition by Tankyrase inhibitor XAV939 (100 nM; bar e). Representative α-tubulin-stained axons corresponding to bars b and d are shown; scale bar, 50 μm. (C) Effect of NAD⁺ depletion by dimerized Tnkp + FK866 on axon degeneration in Sarm1^{-/-} uninjured axons (bar b) or isolated (cut) Sarm1^{-/-} axons (bar d). (D) Effect of sTIR dimerization on endogenous (dotted lines) and exogenously introduced (solid lines) NAD⁺ or NaAD (control) in HTir cells. NaAD is undetectable in nonelectroporated cells. (E) Conversion of ¹⁴C-NAD⁺ in HTir cells to Nam 15 min after SARM1 TIR dimerization. NAD⁺ and Nam from cell extracts and extracellular media were resolved by thin-layer chromatography. (F) (Top) Effect of the PARP inhibitor olaparib (100 nM) on NAD⁺ loss induced by 1 mM H₂O₂ (10 min) or sTIR dimerization (10 min) in HTir cells. (Bottom) PAR formation after H₂O₂ treatment or sTIR dimerization in HTir cells expressing PARG shRNA and inhibition by olaparib. Error bars, SEM; *P < 0.01; one way ANOVA with Tukey's post-hoc test.

To define whether SARM1-mediated depletion of NAD⁺ results from increased consumption or decreased synthesis of NAD⁺, we introduced exogenous NAD⁺ and, as a control, nicotinic acid adenine dinucleotide (NaAD) into HTir cells by electroporation (15), followed by sTIR dimerization. Control cells showed rapid loss of endogenous NAD⁺ within 5 min in response to sTIR dimerization. Electroporation in the presence of NAD⁺ increased the concentration of NAD⁺ by a factor of 4.3, but NAD⁺ was rapidly consumed upon sTIR dimerization. The specificity of this reaction is highlighted by the stability of the closely related analog NaAD (Fig. 4D). sTIR-induced loss of NAD⁺ thus involves the active consumption of NAD⁺. We next demonstrated that the consumed NAD⁺ is converted to nicotinamide (Nam). When radiolabeled ¹⁴C-NAD⁺ was introduced into cells, 15 min of sTIR dimerization elicited loss of ¹⁴C-NAD⁺ and concomitant increases in ¹⁴C Nam as detected by thin-layer chromatography (Fig. 4E). Similarly, sTIR dimerization in nonelectroporated cells

also elicited Nam release as detected by HPLC (fig. S8).

Rapid breakdown of NAD⁺ induced by SARM1 TIR is similar to that observed when PARP is activated in response to DNA damage (16). However, NAD⁺ breakdown induced by sTIR is PARP-independent. The PARP inhibitor olaparib reduced NAD⁺ loss induced by H₂O₂, but had no effect on SARM1-induced loss of NAD⁺ (Fig. 4F). Furthermore, H₂O₂ led to PARP-dependent accumulation of PAR, whereas no PAR was detected after sTIR dimerization (Fig. 4F). Finally, sTIR dimerization in Parp1^{-/-} cells induced loss of NAD⁺, axon degeneration, and cell death (fig. S9). These cell-destruction phenotypes were also unaffected by genetic ablation of the NAD⁺ glycohydrolase CD38, another major consumer of NAD⁺ (17) (fig. S9). SARM1 therefore initiates an NAD⁺ breakdown program that drives axon destruction and cell death independently of PARP1 and CD38.

SARM1 and its orthologs promote axonal degeneration (5, 6) as well as neuronal (18–20) and non-neuronal (21, 22) cell death. SARM1-induced breakdown of NAD⁺ links axon degeneration to the axon-protective Wld^s protein. The presence of Wld^s or other sources of axonal Nmnat may allow for rapid resynthesis of NAD⁺ and the maintenance of metabolic function, thereby counteracting the destructive effects of NAD⁺ degradation by SARM1. Identification of a class of neuroprotective drugs that increase NAD⁺ biosynthesis through effects on Nampt has highlighted the therapeutic potential of augmented NAD⁺ synthesis in neurological disorders (23). Our study provides further biological rationale for NAD⁺ augmentation as a therapeutic approach. Inhibition of SARM1-mediated NAD⁺ loss may be an alternative or synergistic therapeutic strategy for the treatment of neurologic disorders.

REFERENCES AND NOTES

- Y. Fuchs, H. Steller, *Cell* **147**, 742–758 (2011).
- L. Conforti, J. Gilley, M. P. Coleman, *Nat. Rev. Neurosci.* **15**, 394–409 (2014).
- M. P. Coleman et al., *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9985–9990 (1998).
- T. Araki, Y. Sasaki, J. Milbrandt, *Science* **305**, 1010–1013 (2004).
- J. M. Osterloh et al., *Science* **337**, 481–484 (2012).
- J. Gerdts, D. W. Summers, Y. Sasaki, A. DiAntonio, J. Milbrandt, *J. Neurosci.* **33**, 13569–13580 (2013).
- L. A. O'Neill, K. A. Fitzgerald, A. G. Bowie, *Trends Immunol.* **24**, 286–289 (2003).
- M. C. Wehr et al., *Nat. Methods* **3**, 985–993 (2006).
- J. Y. Kang, J.-O. Lee, *Annu. Rev. Biochem.* **80**, 917–941 (2011).
- A. Fegan, B. White, J. C. T. Carlson, C. R. Wagner, *Chem. Rev.* **110**, 3315–3336 (2010).
- Y. Sasaki, J. Milbrandt, *J. Biol. Chem.* **285**, 41211–41215 (2010).
- J. Wang et al., *J. Cell Biol.* **170**, 349–355 (2005).
- Y. Sasaki, B. P. S. Vohra, F. E. Lund, J. Milbrandt, *J. Neurosci.* **29**, 5525–5535 (2009).
- A. Nikiforov, C. Dölle, M. Niere, M. Ziegler, *J. Biol. Chem.* **286**, 21767–21778 (2011).
- M. C. Elia, L. E. Motyka, T. D. Stamato, *Anal. Biochem.* **192**, 329–333 (1991).
- M. Y. Kim, T. Zhang, W. L. Kraus, *Genes Dev.* **19**, 1951–1967 (2005).
- P. Aksoy, T. A. White, M. Thompson, E. N. Chini, *Biochem. Biophys. Res. Commun.* **345**, 1386–1392 (2006).
- Y. Kim et al., *J. Exp. Med.* **204**, 2063–2074 (2007).

19. P. Mukherjee, T. A. Woods, R. A. Moore, K. E. Peterson, *Immunity* **38**, 705–716 (2013).
20. D. W. Summers, A. DiAntonio, J. Milbrandt, *J. Neurosci.* **34**, 9338–9350 (2014).
21. E. S. Blum, M. C. Abraham, S. Yoshimura, Y. Lu, S. Shaham, *Science* **335**, 970–973 (2012).
22. P. Panneerselvam et al., *Cell Death Differ.* **20**, 478–489 (2013).
23. G. Wang et al., *Cell* **158**, 1324–1334 (2014).

ACKNOWLEDGMENTS

Supported by National Institutes of Health grants R01DA020812 (A.D.), R01AG013730 (J.M.), R01NS065053, R01NS087632, and R01NS078007 (J.M. and A.D.), and F31NS074517 (J.G.); and a grant from Vertex Pharmaceuticals. We thank ChromaDex (Irvine, California) for providing nicotinamide riboside; A. Strickland, T. Farhner, and N. Panchenko for technical assistance; and members of the Milbrandt and DiAntonio labs for fruitful discussions.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/453/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
References (24–27)

7 July 2014; accepted 12 March 2015
10.1126/science.1258366

RIBOSOME

Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo

Daniel H. Goldman,^{1*} Christian M. Kaiser,^{2,3†} Anthony Milin,^{1,3} Maurizio Righini,² Ignacio Tinoco Jr.,¹ Carlos Bustamante^{1,2,4,5,6,7†}

Protein synthesis rates can affect gene expression and the folding and activity of the translation product. Interactions between the nascent polypeptide and the ribosome exit tunnel represent one mode of regulating synthesis rates. The SecM protein arrests its own translation, and release of arrest at the translocon has been proposed to occur by mechanical force. Using optical tweezers, we demonstrate that arrest of SecM-stalled ribosomes can indeed be rescued by force alone and that the force needed to release stalling can be generated in vivo by a nascent chain folding near the ribosome tunnel exit. We formulate a kinetic model describing how a protein can regulate its own synthesis by the force generated during folding, tuning ribosome activity to structure acquisition by a nascent polypeptide.

The ribosome translates mRNA into amino acid sequences that contain the information needed for the polypeptide to attain its native structure. Differential usage of synonymous codons and structural elements in the mRNA modulate polypeptide elongation rates. Such rate variations may be required for proper folding and processing of nascent proteins (1). Moreover, interactions of specific nascent chain sequences (2, 3) with the ribosome exit tunnel (4) result in reduced rates of elongation. The bacterial SecM protein represents an example of a stalling sequence that interacts with the ribosome exit tunnel and allosterically represses the peptidyl transferase activity of the ribosome (4–7). Translation of SecM regulates expression of SecA, the motor component of the bacterial Sec translocon (2). Release of stalling in vivo requires interactions between nascent SecM and the translocon machinery (8, 9). It has been suggested that mechanical

force exerted by the translocon relieves elongation arrest and leads to translation restart (10).

To investigate the effect of force on the release of SecM-stalled ribosome-nascent chains (RNCs), we adapted a single-molecule optical tweezers assay (11) (Fig. 1A), enabling the application of defined forces to single ribosome-associated nascent polypeptides. We generated stalled RNCs that contained the C-terminal domain of human calmodulin (CaM) (figs. S1 and S2). CaM provides a mechanical fingerprint (12) in our experiments by exhibiting equilibrium folding and unfolding (“hopping”) at ~7 pN (Fig. 1B and supplementary materials). To detect release of stalled ribosomes, we used the antibiotic puromycin. Puromycin binds to the ribosomal A site and is incorporated into the nascent polypeptide, leading to its release from the ribosome (13). SecM-arrested ribosomes, containing a prolyl-tRNA^{Pro} stably bound in the A site, are refractory to treatment with puromycin, but become sensitive after arrest release, proline incorporation, and translocation (14) (figs. S3 and S4). In the presence of puromycin and EF-G, arrest release will become apparent as a rupture of the tether (Fig. 1B and fig. S4).

We applied a defined, constant force to the molecule in the range of 10 to 30 pN and measured the time required to restart translation, as measured by tether rupture. The mean restart times decreased with increasing force (Fig. 1C). We calculated the rate of stalling rescue as a function of the applied force (Fig. 1, C and

D, and figs. S5 and S6). By fitting the force-dependent rates to Bell’s model (15), we estimated a distance to the transition state (Δx^\ddagger) of 0.4 nm [95% confidence interval (CI): 0.1 nm, 0.8 nm] and a zero-force rupture rate (k_0) of $3 \times 10^{-4} \text{ s}^{-1}$ (95% CI: $0.5 \times 10^{-4} \text{ s}^{-1}$, $20 \times 10^{-4} \text{ s}^{-1}$). This rate is in agreement with biochemical ensemble experiments, in which no force was applied (Fig. 1D, blue dot and fig. S3). In the force range of our experiments, release of SecM-mediated arrest is accelerated by more than an order of magnitude (Fig. 1D), supporting the hypothesis that SecM arrest is relieved by the mechanical force generated by the SecA adenosine triphosphatase (ATPase).

Cotranslational insertion of transmembrane helices via the translocon can release SecM-mediated stalling, presumably by generating force (16). We wondered whether folding of a nascent globular protein domain could generate a force capable of modulating elongation by acting on peptide-tunnel interactions. Such interactions could serve to tune elongation rates to folding transitions (3, 4). The exit tunnel is too narrow to accommodate folded protein domains (17); therefore, as a nascent polypeptide emerges from the exit tunnel and folds in close proximity to the ribosome, it will be sterically excluded from the tunnel. This steric exclusion might generate a force that pulls on the nascent chain within the exit tunnel, which could modulate ribosome activity.

Having established that force accelerates SecM arrest release, we used SecM as a sensor to detect if nascent protein folding outside the ribosome can release the arrest. We constructed a library of plasmids encoding fusion proteins in which the stalling sequence (SecM17) is separated from the de novo-designed protein Top7 (18) by flexible linker sequences of various lengths, followed by a reporter green fluorescent protein (GFP) (Fig. 2A, fig. S7, and supplementary materials). Top 7 folds rapidly against an applied force in close proximity to the ribosome (fig. S8). The GFP coding sequence is translated only upon successful release of the SecM17-mediated translation arrest. Variations in the length of the linker separating Top7 and SecM17 would affect the translation outcome of these constructs (Fig. 2B). Short linker sequences will not allow folding of Top7 because the C terminus of the protein will be sequestered in the exit tunnel (Fig. 2B, top). Intermediate-length linkers will allow more of the Top7 sequence to emerge from the ribosome tunnel and for the protein to fold and produce the steric exclusion folding force (Fig. 2B, middle). And while longer

¹Department of Chemistry, University of California, Berkeley, CA 94720, USA. ²Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720, USA.

³Department of Biology, Johns Hopkins University, Baltimore, MD 21218, USA. ⁴Department of Physics, University of California, Berkeley, CA 94720, USA. ⁵Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. ⁶Kavli Energy Nanosciences Institute at Berkeley, Berkeley, CA 94720, USA. ⁷Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA.

*These authors contributed equally to this work. †Corresponding author. E-mail: carlos@alce.berkeley.edu (C.B.), kaiser@jhu.edu (C.M.K.)



SARM1 activation triggers axon degeneration locally via NAD⁺ destruction

Josiah Gerds *et al.*

Science **348**, 453 (2015);

DOI: 10.1126/science.1258366

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 23, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/453.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2015/04/22/348.6233.453.DC1.html>

This article cites 27 articles, 15 of which can be accessed free:

<http://www.sciencemag.org/content/348/6233/453.full.html#ref-list-1>

This article appears in the following subject collections:

Neuroscience

<http://www.sciencemag.org/cgi/collection/neuroscience>

19. P. Mukherjee, T. A. Woods, R. A. Moore, K. E. Peterson, *Immunity* **38**, 705–716 (2013).
20. D. W. Summers, A. DiAntonio, J. Milbrandt, *J. Neurosci.* **34**, 9338–9350 (2014).
21. E. S. Blum, M. C. Abraham, S. Yoshimura, Y. Lu, S. Shaham, *Science* **335**, 970–973 (2012).
22. P. Panneerselvam et al., *Cell Death Differ.* **20**, 478–489 (2013).
23. G. Wang et al., *Cell* **158**, 1324–1334 (2014).

ACKNOWLEDGMENTS

Supported by National Institutes of Health grants R01DA020812 (A.D.), R01AG013730 (J.M.), R01NS065053, R01NS087632, and R01NS078007 (J.M. and A.D.), and F31NS074517 (J.G.); and a grant from Vertex Pharmaceuticals. We thank ChromaDex (Irvine, California) for providing nicotinamide riboside; A. Strickland, T. Farhner, and N. Panchenko for technical assistance; and members of the Milbrandt and DiAntonio labs for fruitful discussions.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/453/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
References (24–27)

7 July 2014; accepted 12 March 2015
10.1126/science.1258366

RIBOSOME

Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo

Daniel H. Goldman,^{1*} Christian M. Kaiser,^{2,3†} Anthony Milin,^{1,3} Maurizio Righini,² Ignacio Tinoco Jr.,¹ Carlos Bustamante^{1,2,4,5,6,7†}

Protein synthesis rates can affect gene expression and the folding and activity of the translation product. Interactions between the nascent polypeptide and the ribosome exit tunnel represent one mode of regulating synthesis rates. The SecM protein arrests its own translation, and release of arrest at the translocon has been proposed to occur by mechanical force. Using optical tweezers, we demonstrate that arrest of SecM-stalled ribosomes can indeed be rescued by force alone and that the force needed to release stalling can be generated in vivo by a nascent chain folding near the ribosome tunnel exit. We formulate a kinetic model describing how a protein can regulate its own synthesis by the force generated during folding, tuning ribosome activity to structure acquisition by a nascent polypeptide.

The ribosome translates mRNA into amino acid sequences that contain the information needed for the polypeptide to attain its native structure. Differential usage of synonymous codons and structural elements in the mRNA modulate polypeptide elongation rates. Such rate variations may be required for proper folding and processing of nascent proteins (1). Moreover, interactions of specific nascent chain sequences (2, 3) with the ribosome exit tunnel (4) result in reduced rates of elongation. The bacterial SecM protein represents an example of a stalling sequence that interacts with the ribosome exit tunnel and allosterically represses the peptidyl transferase activity of the ribosome (4–7). Translation of SecM regulates expression of SecA, the motor component of the bacterial Sec translocon (2). Release of stalling in vivo requires interactions between nascent SecM and the translocon machinery (8, 9). It has been suggested that mechanical

force exerted by the translocon relieves elongation arrest and leads to translation restart (10).

To investigate the effect of force on the release of SecM-stalled ribosome-nascent chains (RNCs), we adapted a single-molecule optical tweezers assay (11) (Fig. 1A), enabling the application of defined forces to single ribosome-associated nascent polypeptides. We generated stalled RNCs that contained the C-terminal domain of human calmodulin (CaM) (figs. S1 and S2). CaM provides a mechanical fingerprint (12) in our experiments by exhibiting equilibrium folding and unfolding (“hopping”) at ~7 pN (Fig. 1B and supplementary materials). To detect release of stalled ribosomes, we used the antibiotic puromycin. Puromycin binds to the ribosomal A site and is incorporated into the nascent polypeptide, leading to its release from the ribosome (13). SecM-arrested ribosomes, containing a prolyl-tRNA^{Pro} stably bound in the A site, are refractory to treatment with puromycin, but become sensitive after arrest release, proline incorporation, and translocation (14) (figs. S3 and S4). In the presence of puromycin and EF-G, arrest release will become apparent as a rupture of the tether (Fig. 1B and fig. S4).

We applied a defined, constant force to the molecule in the range of 10 to 30 pN and measured the time required to restart translation, as measured by tether rupture. The mean restart times decreased with increasing force (Fig. 1C). We calculated the rate of stalling rescue as a function of the applied force (Fig. 1, C and

D, and figs. S5 and S6). By fitting the force-dependent rates to Bell’s model (15), we estimated a distance to the transition state (Δx^\ddagger) of 0.4 nm [95% confidence interval (CI): 0.1 nm, 0.8 nm] and a zero-force rupture rate (k_0) of $3 \times 10^{-4} \text{ s}^{-1}$ (95% CI: $0.5 \times 10^{-4} \text{ s}^{-1}$, $20 \times 10^{-4} \text{ s}^{-1}$). This rate is in agreement with biochemical ensemble experiments, in which no force was applied (Fig. 1D, blue dot and fig. S3). In the force range of our experiments, release of SecM-mediated arrest is accelerated by more than an order of magnitude (Fig. 1D), supporting the hypothesis that SecM arrest is relieved by the mechanical force generated by the SecA adenosine triphosphatase (ATPase).

Cotranslational insertion of transmembrane helices via the translocon can release SecM-mediated stalling, presumably by generating force (16). We wondered whether folding of a nascent globular protein domain could generate a force capable of modulating elongation by acting on peptide-tunnel interactions. Such interactions could serve to tune elongation rates to folding transitions (3, 4). The exit tunnel is too narrow to accommodate folded protein domains (17); therefore, as a nascent polypeptide emerges from the exit tunnel and folds in close proximity to the ribosome, it will be sterically excluded from the tunnel. This steric exclusion might generate a force that pulls on the nascent chain within the exit tunnel, which could modulate ribosome activity.

Having established that force accelerates SecM arrest release, we used SecM as a sensor to detect if nascent protein folding outside the ribosome can release the arrest. We constructed a library of plasmids encoding fusion proteins in which the stalling sequence (SecM17) is separated from the de novo-designed protein Top7 (18) by flexible linker sequences of various lengths, followed by a reporter green fluorescent protein (GFP) (Fig. 2A, fig. S7, and supplementary materials). Top 7 folds rapidly against an applied force in close proximity to the ribosome (fig. S8). The GFP coding sequence is translated only upon successful release of the SecM17-mediated translation arrest. Variations in the length of the linker separating Top7 and SecM17 would affect the translation outcome of these constructs (Fig. 2B). Short linker sequences will not allow folding of Top7 because the C terminus of the protein will be sequestered in the exit tunnel (Fig. 2B, top). Intermediate-length linkers will allow more of the Top7 sequence to emerge from the ribosome tunnel and for the protein to fold and produce the steric exclusion folding force (Fig. 2B, middle). And while longer

¹Department of Chemistry, University of California, Berkeley, CA 94720, USA. ²Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720, USA.

³Department of Biology, Johns Hopkins University, Baltimore, MD 21218, USA. ⁴Department of Physics, University of California, Berkeley, CA 94720, USA. ⁵Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. ⁶Kavli Energy Nanosciences Institute at Berkeley, Berkeley, CA 94720, USA. ⁷Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA.

*These authors contributed equally to this work. †Corresponding author. E-mail: carlos@alce.berkeley.edu (C.B.), kaiser@jhu.edu (C.M.K.)

linker sequences would also allow Top7 to fold, increased separation between the folding domain and the ribosome should abolish the proposed force-generating steric exclusion release mechanism (Fig. 2B, bottom).

We transformed *Escherichia coli* with the plasmid library containing linker lengths varying from 4 to 28 amino acids. When grown under inducing conditions, a fraction of the colonies exhibited green fluorescence, indicating accumulation of GFP (Fig. 2C) and suggesting that SecM17-mediated stalling had been rescued in some of the transformants. We isolated and sequenced plasmid DNA from 63 fluorescent colonies. Plasmids isolated from fluorescent colonies contained linker sequences between 15

and 22 amino acids in length (Fig. 2D and fig. S9). Given that the SecM17 sequence contributes 16 amino acids to the polypeptide and the ribosome tunnel can accommodate 30 to 35 residues (17), a linker length of 15 to 22 amino acids corresponds to having the protein sequence barely outside the tunnel exit. These results suggest that nascent chain folding near the ribosome tunnel exit can result in release of SecM arrest by stretching the polypeptide in the tunnel.

When Top 7 folds near the tunnel exit, it does so against the steric exclusion force that it generates in the process. The protein must be able to fold against this force and remain folded for a sufficiently long period of time to release stalling by SecM. To estimate the forces gen-

erated by the protein, we performed optical tweezers force spectroscopy measurements with single Top7 molecules tethered by their termini (fig. S10). We measured the distributions of lifetimes of both the unfolded and folded states (Fig. 3, A to C). From these distributions, we extracted the force-dependent rates of folding and unfolding events (Fig. 3D and supplementary materials) (19). Folding rates decrease with increasing force applied to the protein, and unfolding rates increase. The intersection of the folding and unfolding distributions occurs at ~12 pN and represents the force at which the protein has equal probability of being folded or unfolded, a mean lifetime of 28 s for both states. Thus, as it emerges on the surface of the ribosome

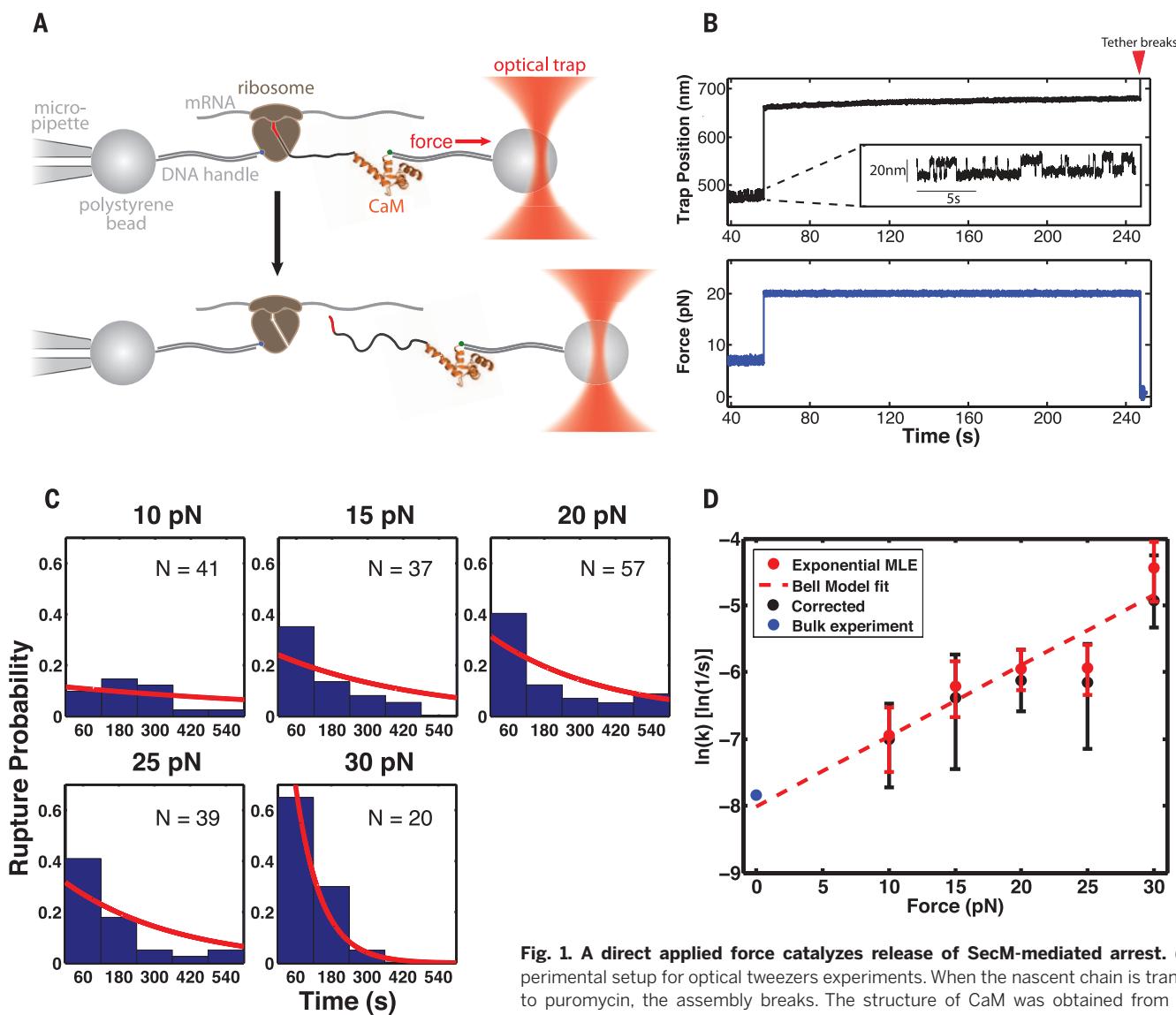


Fig. 1. A direct applied force catalyzes release of SecM-mediated arrest. (A) Experimental setup for optical tweezers experiments. When the nascent chain is transferred to puromycin, the assembly breaks. The structure of CaM was obtained from Protein Data Bank (PDB) ID 1CLL. (B) Example trace for restart experiment. After the “hopping”

signature of CaM is observed (inset) at 7 pN, the force is raised to 20 pN. Red arrow: The tether breaks after ~3 min at 20 pN. (C) Restart lifetimes at each force. Red lines: Distributions returned by the right-censoring MLE. (D) Force-dependent rates for restart of SecM-stalled RNCs in the optical tweezers. Rates are determined as shown in (C), with error bars representing 95% CIs returned by the MLE. Red dotted line: Fit of Bell's model to optical tweezers data. Δx^\ddagger : 0.4 nm (95% CI: 0.1 nm, 0.8 nm) and k_0 : $3.3 \times 10^{-4} \text{ s}^{-1}$ (95% CI: $0.5 \times 10^{-4} \text{ s}^{-1}$, $20 \times 10^{-4} \text{ s}^{-1}$). Black points: Rates determined with a method to account for nonspecific tether rupture (fig. S6). Error determined by bootstrapping. Blue dot: Lifetime obtained from bulk experiment (fig. S3).

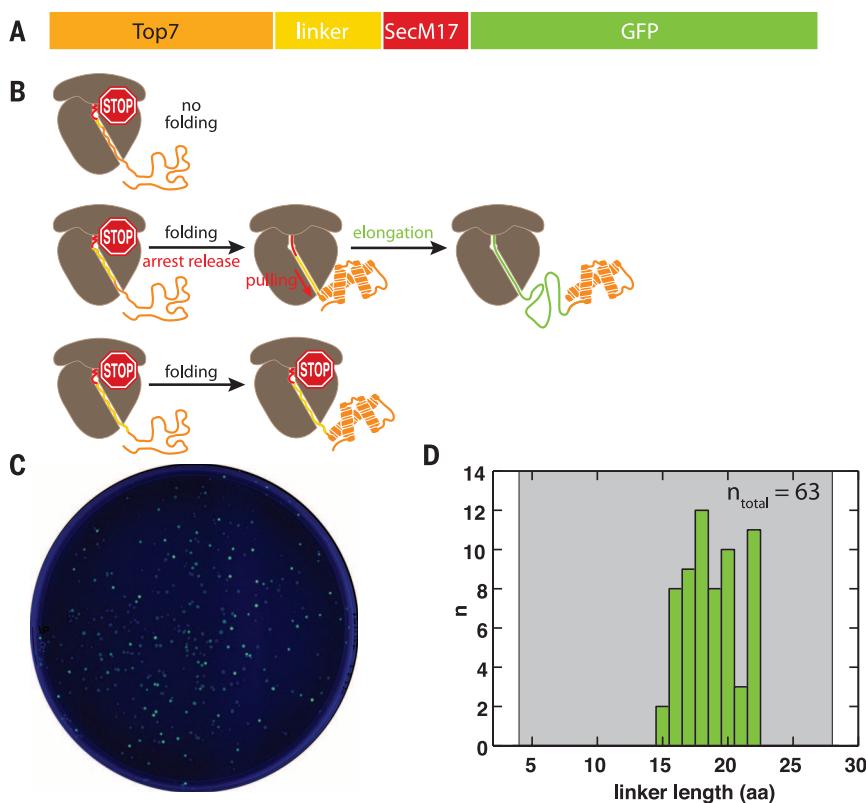


Fig. 2. Nascent protein folding near the ribosome tunnel exit can rescue SecM-mediated stall-ing. (A) Primary sequence of the construct used in the GFP reporter assay. (B) Schematic illustrating the translation outcome for a short (top), intermediate (middle), and long (bottom) linker. (C) Ultraviolet-illuminated image of colonies transformed with the linker library and grown under inducing conditions. (D) Histogram of linker lengths recovered by sequencing of fluorescent colonies. Gray shaded area: library range.

and folds, Top7 can exert a force of at least 12 pN for many seconds on the nascent chain still in the tunnel, before it unfolds. On the basis of our single-molecule results, we propose a kinetic scheme that describes how folding can modulate arrest release rates (Fig. 4A):

$$\begin{array}{c} k_N(F) \quad k_R(F) \\ R \leftarrow U \leftrightarrow N \rightarrow R \\ k_0 \quad k_U(F) \end{array}$$

U and N are the unfolded and natively folded states of Top7; F is the force, and R the stall-released ribosome state, which can be accessed from the unfolded (U) state (at the basal, force-independent rate, k_0). $k_N(F)$ and $k_U(F)$ represent the force-dependent folding and unfolding rates of Top7 obtained from single-molecule experiments (fig. SII).

The effectiveness with which the pulling force catalyzes stall release depends on the force generated upon folding, the probability that the protein folds at that force, and the lifetime of the folded state. To determine how translation stall release rates depend on these factors, we solved the kinetic scheme above for the effective stall release rate, $k_{R_{\text{eff}}}(F)$ (supplementary materials). An approximate solution that assumes equilibrium between N and U (good for $k_U \gg k_R$) yields (see supplementary materials for the exact solution):

$$k_{R_{\text{eff}}}(F) = f_0 \left(k_R + \frac{k_U k_0}{k_N} \right)$$

where f_0 is the fraction of natively folded protein assuming equilibrium with the unfolded

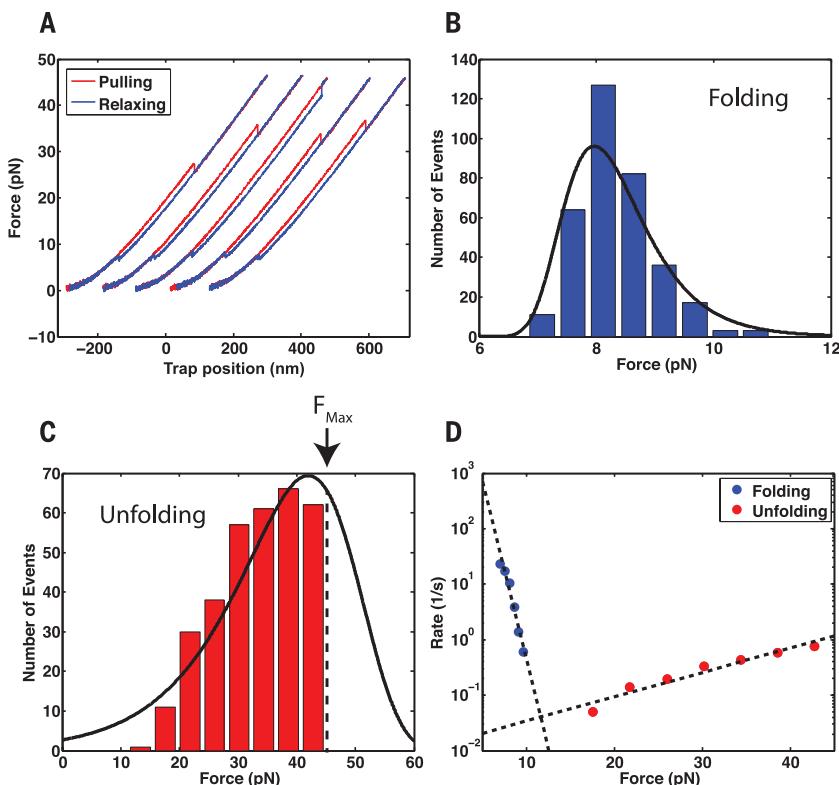


Fig. 3. Top7 refolds against an applied mechanical load. (A) Example force ramp cycles for a single Top7 molecule. Pulling is shown in red, relaxing in blue. Successive cycles are offset along the x axis for display purposes. (B and C) Folding and unfolding force distributions, respectively, for Top7 at a pulling speed of 100 nm/s. Black line: Distributions reconstructed from the force-dependent rates in (D). The unfolding-force distribution in (C) is right-censored because the maximum force in pulling experiments was set at 45 pN to avoid tether rupture. (D) Force-dependent rates of folding and unfolding extracted from the distributions in (B and C). Dashed lines: fit of Bell's model to the force-dependent rates. For folding, Δx^\pm : 6 nm (95% CI: 4 nm, 8 nm) and k_0 : $1 \times 10^6 \text{ s}^{-1}$ (95% CI: $0.04 \times 10^6 \text{ s}^{-1}$, $30 \times 10^6 \text{ s}^{-1}$). For unfolding, Δx^\pm : 0.4 nm (95% CI: 0.3 nm, 0.6 nm) and k_0 : 0.01 s^{-1} (95% CI: 0.003 s^{-1} , 0.03 s^{-1}) (supplementary materials).

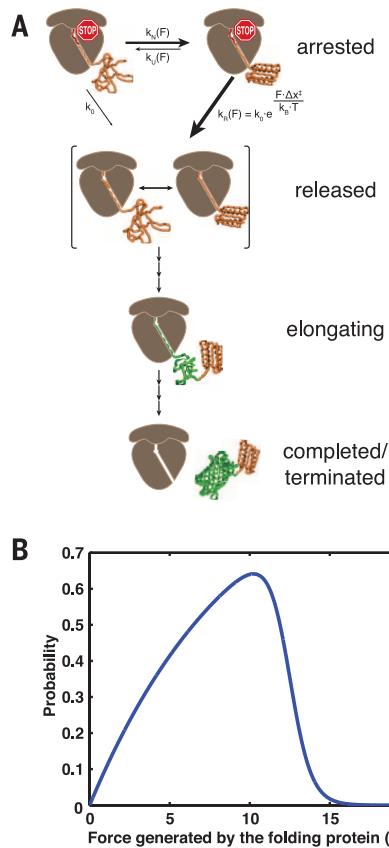


Fig. 4. Kinetic model for folding-induced release of stalled ribosomes. (A) Kinetic scheme illustrating the pathway to release of translation arrest. The nascent polypeptide can transit reversibly between the native and unfolded states, with rate constants $k_N(F)$ and $k_U(F)$. Once folded, the nascent protein both generates and experiences a force, “ F ,” which can drive it either irreversibly to the “released” state, with rate constant $k_R(F)$, or back to the unfolded state. In addition, the stall can be released via the irreversible spontaneous process from the unfolded state, with rate constant k_0 , which is independent of force. PDB: 1QYS and 1EMA. (B) The probability of force-catalyzed stall release is plotted as a function of the folding force.

state. f_0 decreases with force as the equilibrium is tilted toward the unfolded state; in contrast, the term in parentheses increases with force as the release rate increases and the protein is biased toward the unfolded state (fig. S12). Owing mainly to the difference in experimental geometry between the dual-tethered Top7 in the

optical tweezers and the singly tethered Top7 in the in vivo experiments, the force-dependent rates likely represent an underestimate to the true rates (supplementary materials). Although the experimentally observed quantity, $k_{\text{R}_{\text{eff}}}(F)$, represents the composite kinetics of both the spontaneous and force-dependent release rates, we calculated the probability that release proceeds by the mechanical, as opposed to the spontaneous, process. This function provides the most likely force at which folding of the nascent protein leads to the release of the ribosome stall and can be expressed as:

$$P(F) = \frac{k_{\text{R}_{\text{eff}}}(F) - k_0}{k_{\text{R}_{\text{eff}}}(F)}$$

A plot of $P(F)$ versus force shows that the probability of force-catalyzed stall release is a maximum near 10 pN (Fig. 4B). At forces below 12 pN, Top7 is mostly folded (N), so release proceeds largely through the mechanical path, whose rate, k_R , increases with increasing force; at forces higher than 12 pN, the protein spends shorter and shorter times in the folded state, so release occurs more and more via the slower, spontaneous path, k_0 , from the unfolded state.

Our results provide evidence that the translocon must generate at least 10 pN of force to relieve SecM-induced ribosomal arrest. It appears that another translocase, ClpX, operates in a similar force regime (20, 21). Given that a number of polypeptide sequences are known to stall the ribosome (3, 22–24), our results suggest that force can play a generally important role as a regulator of elongation. In instances where ribosome–nascent chain interactions are less robust than the SecM system, folding could play an important role in modulating translation elongation and vice versa. For example, if such interactions occur near the C terminus of a newly synthesized domain, elongation would slow down when the polypeptide segment just outside the ribosome begins to acquire stable structure, permitting the folding of this segment to be completed before more of the polypeptide is synthesized. Likewise, folding of the domain, and the force so generated, may provide the signal for speeding up elongation through and beyond the regulatory signal. A force could be generated not only by nascent chain folding, but also by the binding of partners to the nascent chain outside the ribosome. In eukaryotic cells, ribosome profiling experiments have shown that the chaperone Hsp70 relieves global stalling of ribosomes near the beginning of genes (25, 26), perhaps by generating a pulling force similar to the scenario of Hsp70 binding to protein aggregates (27). Thus, force generated by either nascent chain

folding or chaperone binding could constitute an important feedback mechanism to tune elongation to folding.

REFERENCES AND NOTES

- D. A. Nissley, E. P. O'Brien, *JACS* **136**, 17892–17898 (2014).
- K. Ito, S. Chiba, *Annu. Rev. Biochem.* **82**, 171–202 (2013).
- Deutsch, in *Regulatory Nascent Polypeptides* (Springer, Tokyo, 2014), pp. 61–86.
- D. N. Wilson, R. Beckmann, *Curr. Opin. Struct. Biol.* **21**, 274–282 (2011).
- A. Tsai, G. Kornberg, M. Johansson, J. Chen, J. D. Puglisi, *Cell Rep.* **7**, 1521–1533 (2014).
- H. Nakatogawa, K. Ito, *Cell* **108**, 629–636 (2002).
- J. Gumbart, E. Schreiner, D. N. Wilson, R. Beckmann, K. Schulten, *Biophys. J.* **103**, 331–341 (2012).
- M. N. Yap, H. D. Bernstein, *Mol. Microbiol.* **81**, 540–553 (2011).
- K. Nakamori, S. Chiba, K. Ito, *FEBS Lett.* **588**, 3098–3103 (2014).
- M. E. Butkus, L. B. Prundeanu, D. B. Oliver, *J. Bacteriol.* **185**, 6719–6722 (2003).
- C. M. Kaiser, D. H. Goldman, J. D. Chodera, I. Tinoco Jr., C. Bustamante, *Science* **334**, 1723–1727 (2011).
- J. P. Junker, F. Ziegler, M. Rief, *Science* **323**, 633–637 (2009).
- P. Nathans, *Natl. Acad. Sci. U.S.A.* **51**, 585–592 (1964).
- H. Muto, H. Nakatogawa, K. Ito, *Mol. Cell* **22**, 545–552 (2006).
- G. I. Bell, *Science* **200**, 618–627 (1978).
- N. Ismail, R. Hedman, N. Schiller, G. von Heijne, *Nat. Struct. Mol. Biol.* **19**, 1018–1022 (2012).
- N. R. Voss, M. Gerstein, T. A. Steitz, P. B. Moore, *J. Mol. Biol.* **360**, 893–906 (2006).
- B. Kuhlman et al., *Science* **302**, 1364–1368 (2003).
- O. K. Dudko, G. Hummer, A. Szabo, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15755–15760 (2008).
- R. A. Maillard et al., *Cell* **145**, 459–469 (2011).
- M. E. Aubin-Tam, A. O. Olivares, R. T. Sauer, T. A. Baker, M. J. Lang, *Cell* **145**, 257–267 (2011).
- C. J. Woolstenhulme et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, E878–E887 (2013).
- S. Ude et al., *Science* **339**, 82–85 (2013).
- L. K. Doerfel et al., *Science* **339**, 85–88 (2013).
- R. Shalgi et al., *Mol. Cell* **49**, 439–452 (2013).
- B. Liu, Y. Han, S. B. Qian, *Mol. Cell* **49**, 453–463 (2013).
- P. De Los Rios, A. Ben-Zvi, O. Slutsky, A. Azem, P. Goloubinoff, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6166–6171 (2006).

ACKNOWLEDGMENTS

We thank G. Chistol and L. Alexander for guidance regarding data analysis and the kinetic model, and members of the Bustamante and Tinoco laboratories for helpful discussions on the manuscript. D.H.G acknowledges support from the NIH K99 Award 5K99GM086516 and the QB3 Institute, Berkeley (Distinguished Postdoctoral Fellowship); M.R. acknowledges support from the Damon Runyon Cancer Research Foundation (DRG-2096-11). C.B. acknowledges support from NIH grant 5R01GM32543; I.T. acknowledges support from NIH grant GM10840.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6233/457/suppl/DC1
Materials and Methods
Figs. S1 to S12
References (28–30)

30 September 2014; accepted 9 March 2015
10.1126/science.1261909



Mechanical force releases nascent chain-mediated ribosome arrest *in vitro* and *in vivo*

Daniel H. Goldman *et al.*
Science **348**, 457 (2015);
 DOI: 10.1126/science.1261909

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 24, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6233/457.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2015/04/22/348.6233.457.DC1.html>

A list of selected additional articles on the Science Web sites related to this article can be found at:

<http://www.sciencemag.org/content/348/6233/457.full.html#related>

This article cites 28 articles, 12 of which can be accessed free:
<http://www.sciencemag.org/content/348/6233/457.full.html#ref-list-1>

This article has been cited by 1 articles hosted by HighWire Press; see:
<http://www.sciencemag.org/content/348/6233/457.full.html#related-urls>

This article appears in the following subject collections:
 Biochemistry

<http://www.sciencemag.org/cgi/collection/biochem>