Big money supercharges quantum computing *p. 1090*

Optogenetics sheds light on dopamine learning circuit *p. 1108*

How atmospheric aerosol particles form *p. 1119*

Science

$15
2 DECEMBER 2016
sciencemag.org
AAAS

SHORT CIRCUIT

Chronic activation rewires T cells for dysfunction

pp. 1104, 1160, & 1165

CONTENTS

1099
How bats spread viruses

2 DECEMBER 2016 • VOLUME 354 • ISSUE 6316

1090

NEWS

IN BRIEF

IN DEPTH

FEATURES

INSIGHTS

POLICY FORUM

PERSPECTIVES

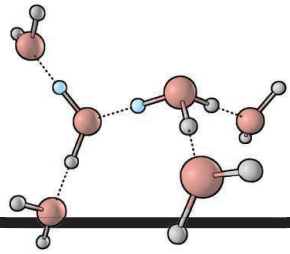
1110

SCIENCE PRIZE ESSAY

BOOKS *ET AL.*

Published by AAAS

CONTENTS



1101 & 1131
Pass the protons, please

2 DECEMBER 2016 • VOLUME 354 • ISSUE 6316

LETTERS

RESEARCH

IN BRIEF

REVIEW

RESEARCH ARTICLES

REPORTS

DEPARTMENTS

ON THE COVER

Illustration of a functional T cell (background) and an exhausted T cell (foreground). When responding to chronic infection or cancer, T cells become exhausted and lose many of their core functions. This process involves a fundamental "rewiring" of the cellular regulation of gene expression. Thus, exhausted T cells likely represent a distinct cell lineage. See pages 1104, 1160, and 1165.
Illustration: Valerie Altounian/Science

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Domestic individual membership and subscription (51 issues): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1522; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $89. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15.00 current issue, $20.00 back issue prepaid includes surface postage; bulk rates on request. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

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

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

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

Science serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority of conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

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

Published by AAAS

EDITORIAL

A comprehensive nuclear test ban

Ernest J. Moniz is the U.S. Secretary of Energy

A permanent end to nuclear explosive testing, combined with sustained reliable deterrence, is in the national security interest of the United States and its allies and friends. The next U.S. Administration and the Congress should revisit the Comprehensive Nuclear Test Ban Treaty (CTBT) in view of current realities and work together toward enhanced security through ratification and an international push for entry into force.

Nuclear weapons have been central to strategic stability among nations and alliances for more than half a century. Even as American and international leaders aspire to and work toward a stable world verifiably free of nuclear weapons, this outcome cannot be expected for a considerable time. Voluntary suspension of nuclear testing by the permanent members of the United Nations Security Council for over two decades has made an enormous contribution to stability and nuclear nonproliferation. India and Pakistan ceased testing nearly as long ago, adding a degree of stability in a challenging security environment. The international agreement that verifiably eliminates the risk of an Iranian nuclear weapon greatly mitigates the risk of a proliferation sprint in the Middle East. The glaring exception is North Korea with its nuclear testing as recently as September, demonstrating vividly the regional and global destabilizing effect of nuclear tests. All of this argues for a binding no-test regime.

"The time has come to revisit CTBT ratification."

In 1999, the Senate evaluated the CTBT, but did not consent to its ratification. Two principal objections needed to be addressed. The newly created science-based stockpile stewardship program, designed to replace nuclear test explosions with an integrated scientific program of computer simulations, data analysis, and laboratory tests of weapons subsystems, was at least a decade away from having its suite of new tools in place. There was also uncertainty about the ability to detect low-yield clandestine underground tests anywhere in the world. At that time, I argued within the Clinton Administration for a time-limited CTBT approval, providing a period during which the needed science and technology could be developed and deployed. Although the approval didn't happen, the science and technology did mature. The time has come to revisit CTBT ratification.

The stockpile stewardship innovation that took place in the Department of Energy's (DOE) national laboratories since the end of testing in 1992 is remarkable. Nothing was off the shelf, including a new supercomputing architecture and experimental facilities that explore previously unattainable pressures and temperatures. These capabilities have led to a continuously increasing understanding of nuclear weapons properties and dynamics that could not be approached within the testing paradigm. Each year, the Secretaries of Defense and of Energy provide the president with an assessment of the deterrent, building on a peer review carried out by the DOE national security laboratory directors and the commander of the Department of Defense's Strategic Command. All indications are that the stockpile will remain safe, secure, and reliable indefinitely without testing, even as the weapons age and are reduced in number. The new toolkits of supercomputer systems and lab facilities are unlocking additional benefits, such as simulating fluid dynamics for better engines and turbines, handling enormous data sets for cancer research and treatment, and probing energetic astrophysical phenomena.

The global monitoring network necessary for worldwide detection of nuclear explosions has been built. Incredibly sensitive measurements can now detect and identify radionuclide signatures associated with nuclear explosive tests, including low-yield or evasive tests. With the treaty in force, international experts will conduct an on-site inspection after a suspect event.

In short, the United States and its partners have the technology needed to make the CTBT work for deterrence and strategic stability. Clearly, ratification by the Senate will not itself bring the treaty into force, but it will provide a stronger foundation for achieving a global testing ban and for amplifying international pressure and sanctions on any country that does test.

–Ernest J. Moniz

Published online 30 November 2016; 10.1126/science.aal4864

PHOTOS: (INSET) CTBTO/FLICKR; (TOP RIGHT) U.S. DEPARTMENT OF ENERGY

NEWS

> “We feel relieved after these 2 months, which have been very stressful.”
>
> **Javier Castro Hernández, of the University Hospital of the Canary Islands on Tenerife in Spain.** Last week, Spanish military delivered his long-awaited order of 29 mice, stranded after two Spanish airlines stopped shipping lab animals.

IN BRIEF

Bleaching killed these table corals in the Great Barrier Reef.

PHOTO: GREG TORDA/ARC CENTRE OF EXCELLENCE FOR CORAL REEF STUDIES

Great Barrier Reef sees worst coral die-off ever

This year has seen the largest recorded die-off of corals in the Great Barrier Reef (GBR), Australian scientists reported this week. In the worst-hit section of the reef, a northern swath spanning about 700 kilometers, extensive coral bleaching—the loss of symbiotic algae—caused an average of 67% of the corals to perish. Bleaching occurs when unusually warm water results in corals losing their zooxanthellae, the colorful algae that live within coral tissue and provide nutrients. If waters cool quickly enough, zooxanthellae return; but prolonged bleaching causes death. After water temperatures soared in March, surveys found extensive bleaching along the northern GBR (*Science*, 1 April, p. 15). In October and November, investigators returned to the bleached reefs to confirm the death toll. There were some bright spots: The hot water did not extend to the southern reef, leaving it barely affected, and only 26% of the corals in the far north had died, possibly because local upwelling of cooler water provided relief.

AROUND THE WORLD

East Asia tops in math, science

BOSTON | Students from East Asia continue to lead the world in science and math, whereas U.S. students once again rank in the middle of the pack. Results of the 2015 Trends in International Mathematics and Science Study (TIMSS) released this week show a 48-point gap in eighth grade math scores between Japan—which ranks fifth behind Singapore; Hong Kong, China; South Korea; and Taiwan—and the rest of the world. (Among fourth graders, the math gap is 23 points.) In science, a few other countries join the East Asian nations in the top tier at both grade levels, and there is no large drop-off to the lower-ranked countries. The TIMSS also offers sobering news about the impact of schooling on student learning. In its first-ever longitudinal study, the group tracked the progress of students in nine participating countries over 8 years. Although the students were taking the most challenging math and science courses in their senior year, they still performed progressively worse as they moved from elementary to middle to high school. http://scim.ag/TIMSS2015

Alzheimer's drug fails

INDIANAPOLIS | A widely watched candidate drug for Alzheimer's disease failed to deliver on its promise in its final clinical trial, Indianapolis-based pharmaceutical giant Eli Lilly and Company announced on 23 November. Solanezumab is an antibody that targets amyloid plaque buildup around neurons—a hallmark of the disease, though not necessarily its cause. But in a trial of more than 2100 people with mild dementia, the intravenously administered drug did not significantly improve memory when compared with a placebo. Eli Lilly's billion-dollar investment was criticized for raising public hope that treatment might finally be within reach. Last spring, that criticism grew stronger after the drug company scrapped one of its criteria for efficacy (http://scim.ag/Lillychange), which was interpreted by some members of the scientific community and the drug development industry as a last-ditch effort to force the drug's success.

HHS, education nominees named

WASHINGTON, D.C. | U.S. President-elect Donald Trump on 29 November nominated orthopedic surgeon and six-term Representative Tom Price (R–GA) to be secretary of the Department of Health and Human Services, the parent agency of the National Institutes of Health (NIH), the Centers for Disease Control and Prevention, and the Food and Drug Administration (FDA). Price, 63, has been a leading critic of the Affordable Care Act and an advocate of streamlining FDA regulations to hasten approvals of new antibiotics and medical devices. He is an abortion opponent and has voted repeatedly against expanding the number of human embryonic stem cell lines available to NIH-funded researchers. Last week, Trump also announced his intention to nominate billionaire philanthropist Betsy DeVos to head the Department of Education. DeVos is a staunch advocate for school choice and charter schools but has no track record of opinion on higher education, leading academic researchers to express both fear and cautious optimism. **http://scim.ag/HHSEdnoms**

NIH keeps postdoc stipend hike

BETHESDA, MARYLAND | The National Institutes of Health (NIH) is raising stipends for postdoctoral fellows, despite a federal judge's decision last week to delay the nation-wide implementation of new overtime rules affecting low-paying jobs. The Fair Labor Standards Act (FLSA), which would have taken effect 1 December, requires employers to pay overtime to workers earning less than $47,476 annually. The court's injunction means universities are not required to pay overtime to postdocs earning below that threshold. But NIH said last week that it would stick to a plan it announced in August to hike the stipends of postdocs on its Kirschstein National Research Service Awards; starting salaries will increase from $43,692 to $47,484. Republican legislators may try to repeal the rule, which was finalized in May, when the new Congress convenes in January 2017.

More deaths in cancer trial

SEATTLE, WASHINGTON | Juno Therapeutics, the biotech company whose cancer immunotherapy trial was put briefly on hold in July after the deaths of three people being treated for leukemia (*Science*, 15 July, p. 203), revealed last week that two more people recently died after the same experimental treatment. The announcement prompted strong criticism of the company and of the U.S. Food and Drug Administration, which had halted the trial in July for just three business days before allowing it to proceed. All five patients, who had advanced leukemia, died from cerebral edema, an accumulation of fluid around the brain, after receiving chimeric antigen receptor T cell therapy, in which T cells are genetically engineered to fight cancer. The company had argued in July that the fatal side effect was likely due to the combination of the T cells and the chemotherapy drug fludarabine, which the patients received along with the immune treatment—but that theory is now in question because the two who died last month didn't receive that drug. As *Science* went to press, the trial is again on hold.

With a kiss of food, nurse ants can make either big or small workers.

PHOTO: ADRIA C. LEBOEUF

An ant's fate may be sealed with a kiss

Ants often appear to "kiss," quickly pressing their mouths together to feed each other and their larvae. But the insects are sharing more than food, scientists report this week in *eLife*: Via that exchange, called trophallaxis, they also shape one another and the colonies they live in. Social insects, such as ants and bees, store food in "social stomachs," from which they can regurgitate it later to transfer nutrients from foraging ants to nurse ants to larvae. But the trophallactic fluid of Florida carpenter ants contains a diverse array of chemicals, researchers say, such as proteins related to growth and digestion—and, surprisingly, a juvenile hormone important for regulating reproduction, development, and behavior. Scientists previously thought this hormone was only present in the ants' blood. To test the impact of transfering that hormone through trophallaxis, the researchers gave groups of worker ants a few larvae to rear, and provided some of the workers with food supplemented with the juvenile hormone. Larvae that received the hormone-laced food were twice as likely to mature into large worker adults, the team found. This, they say, suggests that nurse ants selectively feed larvae the hormone to meet the colony's changing needs—and help shape its long-term fate. **http://scim.ag/antkiss**

PHOTO: JASON SOUTH/FAIRFAX MEDIA/STRINGER/GETTY IMAGES

CONSERVATION

Dam-building threatens Mekong fisheries

With scores of dams planned, scientists debate measures to soften the impact

By **Richard Stone**

Every April, the pa nyawn catfish would make their way up the Mekong River to spawn, crowding through a narrow channel that skirts Khone Falls in southern Laos. Villagers netted the thumb-sized fish by the hundreds of thousands. Then, in 2014, work started on Don Sahong Dam, which straddles the channel. Although the dam won't be completed for another 2 years, construction has already cut off the migration and destroyed the fishing sites, says Zeb Hogan, a biologist with the Global Water Center at the University of Nevada in Reno. Sahong channel's pa nyawn fishery, he says, "no longer exists."

Its loss is a microcosm of a larger food security crisis threatening the Mekong delta, as Laos and its neighbors, hungry for electric power, embark on a dam-building spree. Don Sahong is the second dam under construction on the lower Mekong, the great artery of Southeast Asia that wends through Myanmar, Laos, Thailand, Cambodia, and Vietnam. Laos plans to start construction on a third main-stem dam—Pak Beng, near its border with Thailand—in early 2017. Several more, and dozens on major tributaries, are on the drawing board. By blocking migration routes and cutting sediment flow to the Mekong delta, the projects could wipe out more than a third of the lower Mekong Basin's annual haul of river fish, by one estimate—a serious blow to the region's 60 million people.

Hoping to forestall catastrophe, environmentalists and scientists are pressing the hydropower companies to incorporate "fish-friendly" turbines, ladders, and locks for migratory fish into dam designs. At a meeting last month in Vientiane, Laos, on fish passage approaches, some scientists saw reason for optimism. But others scoffed at the mitigation plans. They believe that the only way to spare the fisheries is to ax some projects. "I'm ready to sacrifice parts of the river if we can persuade the governments to leave the rest of it alone," says Ian Cowx, a fisheries biologist at the University of Hull in the United Kingdom. "If we can keep the lower Mekong flowing, at least you have a chance."

Laos, which government officials say aspires to become the "battery of Southeast Asia," has the most ambitious plans, envisioning nine dams on the lower Mekong. Work on the first, Xayaburi, began in 2010, despite concerns from Cambodia and Vietnam, Laos's downstream neighbors, that the dam would impede sediment and nutrient flow and pose a formidable—if not impassable—obstacle to migratory fish (*Science*, 12 August 2011, p. 814). Xayaburi Dam is expected to be completed in 2019.

Farther downstream, work is progressing fast on the Don Sahong Dam. Last month, the Save the Mekong coalition, an alliance of nongovernmental organizations, appealed to the developers to halt construction until "comprehensive information regarding the project's environmental and social impacts" is made public.

A project in Cambodia may pose the biggest threat. Lower Sesan 2 on the Sesan River, a major Mekong tributary, "is a major mistake," says Kim Geheb, Greater Mekong coordinator for the Consultative Group on International Agricultural Research in Vientiane. Now about 60% complete, that dam alone could slash fish catch in the Mekong Basin by about 9%, fisheries experts predict. "You just have to write off the Sesan and the Srepok," a river that flows into the Sesan just above the dam, Cowx says.

Several of the developers are trying to soften the impact on fisheries. The Xayaburi Power Company has reportedly committed $400 million to outfitting the dam with features that will reduce its ecological toll, including locks meant to enable fish to traverse the dam. That's "a huge number and a sign to other developers that they cannot ignore environmental considerations," says Hogan, who notes that it will be impera-

Ota Khami floats where his home stood before it was bulldozed to make way for the Lower Sesan 2 Dam in Cambodia.

tive to monitor the effectiveness of the mitigation measures. But Cowx says that after listening to a presentation on the plans in Vientiane, "I was completely gobsmacked. I've never seen such a disaster in my life." He doubts the "untested" system will work on the Mekong, where migratory fish span a huge range in lengths, from a few centimeters to a few meters. Xayaburi Power did not respond to a request for comment.

Even if the fish passages work, Cowx and others point out that the reservoirs behind dams like Xayaburi will destroy spawning grounds, and their stagnant waters could create a hostile habitat for larval fish. Guy Lanza, a biologist at the State University of New York College of Environmental Science and Forestry in Syracuse who has studied Xayaburi, predicts that the reservoir will trigger "a cascade of negative environmental effects," including fish kills and the release of methylated mercury and other toxic substances into the water column.

Near the mouth of the river, saltwater intrusion could also harm fisheries. Seas are rising because of climate change, but the dams are also contributing, by cutting off the flow of sediment that helps keep the delta above sea level. Because of China's dams on the upper Mekong, the sediment load from China "has basically stopped now," says Matti Kummu, a water expert at Aalto University in Espoo, Finland. Already, Cowx says, sediment-light water is scouring riverbanks in Laos and Thailand. "Erosion is horrendous," he says.

Dams on the lower Mekong will exacerbate the problem. They are abetted by a surprising factor: weaker tropical cyclones, which have slackened the river's flow. Kummu and colleagues reported in the 10 November issue of *Nature* that sediments deposited each year in the Mekong delta declined by about 52 million tons over a 25-year period ending in 2005. They attribute some 60% of that drop to less runoff from sparser monsoon rains.

A 2012 paper in *Global Environmental Change* predicted that if all 11 lower Mekong main-stem dams and 77 tributary dams slated for completion by 2030 are built, annual fish hauls from the lower Mekong could decline by up to 880,000 tons, or 42%. Biodiversity would also suffer in a region that is home to eight of the biggest freshwater fish species in the world, including the Mekong giant catfish—known to top 3 meters in length—the dog-eating catfish, and the giant Siamese carp. All are endangered, and all migrate hundreds of kilometers to spawn. "Dams could be the nail in the coffin," says Hogan, who studies the giant fish. "If we hope to be able to restore populations of these fish at some point in the future, we need to keep a few of them around and understand how these dams impact their life cycles."

Few observers expect a happy outcome in the lower Mekong Basin. The Mekong River Commission, an advisory body set up by lower Mekong nations, has provided a forum for airing grievances and prodding developers to tweak designs, but it has no authority to stop dams from moving forward. Geheb suggests it may be in the interest of Vietnam and Cambodia to pay Laos not to build more dams. In Vietnam, he says, "They're freaking out over what's happening on the main stem and the impact on the delta." If food security in the region becomes dire, Cowx predicts that Laos could get hauled before the International Court of Justice in the Hague. But by then, he says, "it will be too late" to save the river. ■

Power play on the Mekong

Even with fish-passage features, four hydroelectric dams now rising on the lower Mekong River and on a major tributary, and a fifth dam in the planning stages, are expected to harm the basin's fisheries—and erode food security.

	Operating date	Generating capacity (MW)
1	2019	1285
2	2019	260
3	2020	400
4	2024	912
5	In planning	2600

Planned
Under construction

CREDITS: (MAP) J. YOU/*SCIENCE*; (DATA) WLE GREATER MEKONG

U.S. POLICY

Congress votes on sweeping biomedical bill

NIH would get $4.8 billion for key research initiatives

By **Jocelyn Kaiser**, **Jeffrey Mervis**, *and* **Kelly Servick**

Congress this week was poised to begin voting on a sweeping biomedical innovation bill that includes nearly $5 billion in dedicated funding for a trio of major research initiatives at the National Institutes of Health (NIH). The bill also includes measures to speed the approval of new drugs and medical devices by the Food and Drug Administration (FDA), and it would create a mechanism for catalyzing efforts to streamline federal regulations that universities and academic researchers regard as burdensome.

The bipartisan bill, known as the 21st Century Cures Act, is the culmination of more than 2 years of lobbying by research, patient, and industry groups, and extensive negotiations between members of the Senate and the House of Representatives. The House was expected to approve the bill on 30 November, but its fate in the Senate is unclear. Senate Majority Leader Mitch McConnell (R–KY) has said completing Cures and sending it to President Barack Obama for signing is one of his highest priorities before Congress adjourns for the year, but disagreements over funding mechanisms or other issues could derail the legislation.

Still, biomedical research groups were cautiously optimistic that Cures will finally cross the finish line. The bill includes a long list of largely uncontroversial provisions, including calls for NIH to produce a comprehensive strategic plan, set up a special initiative for young scientists, establish a prize to incentivize certain kinds of research, and take new steps to encourage data sharing and ensure the reproducibility of NIH-funded research. And research lobbyists are delighted with provisions that set aside $4.8 billion over the next 10 years for three NIH initiatives: $1.4 billion for Obama's Precision Medicine Initiative, $1.8 billion for Vice President Joe Biden's cancer moonshot, and $1.6 billion for the White House's Brain Research through Advancing Innovative Neurotechnologies initiative. The bill also provides $30 million

over 3 years for regenerative medicine research using adult stem cells.

The NIH total is less than the $8.8 billion envisioned in earlier versions, but research advocates aren't complaining. They say Cures would put Congress on record as supporting sustained funding for key research areas, and specifies how the spending boost would be paid for. "We're absolutely thrilled," says Jon Retzlaff, director of science policy in the Washington, D.C., office of the American Association for Cancer Research, which has lobbied for Biden's cancer moonshot.

A 25 November Cures draft calls for raising the money by selling oil from the U.S. Strategic Petroleum Reserve and redirecting funds from a public health fund established by the Affordable Care Act (Obamacare). The money would flow into a new NIH Innovation Account that congressional appropriators would control. But that plan has drawn opposition from some senators, especially those concerned about the public health programs, and is a source of angst in the biomedical research community. The arrangement disappoints those who sought a dedicated stream of money for NIH not subject to the vagaries of the annual appropriations process, but satisfies some lawmakers' desire to keep Congress in control of spending.

"Our hope is that this fund will add a layer of security" for NIH's budget, says Tannaz Rasouli, a policy specialist at the Association of American Medical Colleges in Washington, D.C. But "there are no guarantees," warns Howard Garrison, a policy expert at the Federation of American Societies for Experimental Biology (FASEB) in Bethesda, Maryland. He and others want Congress to use the fund as a supplement to NIH's regular budget, but fear it could become an excuse to curb spending in other parts of the agency. Garrison also worries about other provisions that could burden NIH. In a bid to boost accountability, for instance, Congress wants the director of each institute to personally review and sign off on grants. "It's awkward, it's unnecessary, and it's burdensome," Garrison says. And FASEB worries a plan to impose renewable, 5-year term limits on institute directors could make it harder for NIH to recruit talent.

> ***"Our hope is that [the National Institutes of Health innovation act] will provide a layer of security" for the agency's budget.***
>
> **Tannaz Rasouli**, Association of American Medical Colleges

Getting a warmer welcome are provisions aimed at reducing the regulatory burden on institutions that get federal research dollars. Reporting requirements on grants have grown over the past few decades, for instance, as have rules meant to safeguard research subjects. The bill creates a Research Policy Board within the White House Office of Management and Budget. The board, which would be made up of representatives from up to 10 federal agencies and a similar number of research institutions, would study emerging regulatory problems as well as recommend how to harmonize existing policies. The idea came from a 2015 report by a committee of the National Academies of Sciences, Engineering, and Medicine. Another provision calls for a review of 5-year-old rules that tightened conflict of interest reporting by NIH-funded researchers. Institutions complain the rules have imposed big costs with little benefit.

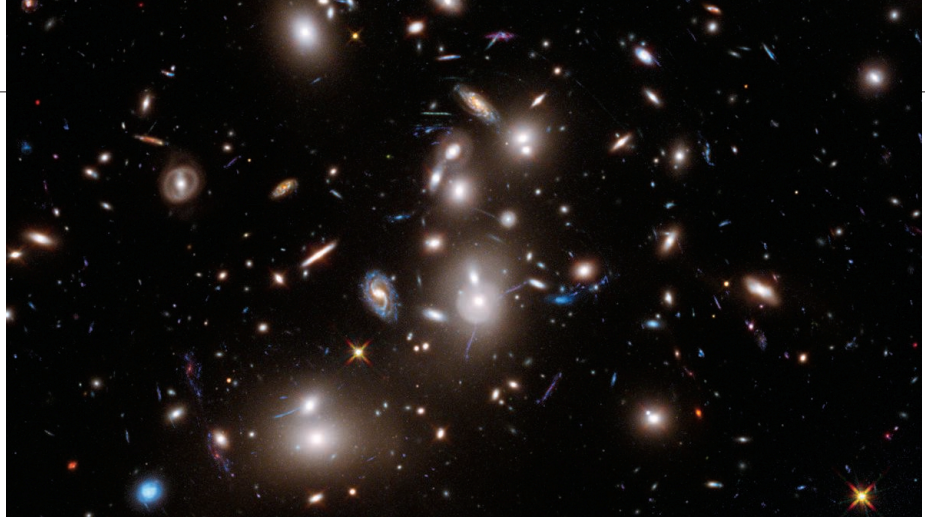
The bill's FDA provisions aim to accelerate the agency's review of some new drugs and medical devices. In certain cases, FDA would allow companies to run smaller clinical trials or rely on evidence collected outside of trials to support approval. In an apparent bid to boost stem cell and other experimental therapies, the current draft also directs FDA to give special attention to treatments designated as "regenerative advanced therapy." If a treatment meets the criteria—for example, if it is based on stem cells or other tissues and addresses an unmet medical need—regulators can offer a company faster review, or more flexibility in setting trial endpoints.

A set of potentially controversial provisions would expand FDA's priority review system, which attempts to create an incentive for companies to develop drugs for neglected tropical and pediatric diseases by doling out tradeable vouchers that entitle the companies to speedier agency reviews. Cures would add medical countermeasures—designed to respond to a chemical or biological attack, for example—to the treatments that would win vouchers for their developers. But critics say there is little evidence the vouchers have worked, and worry that creating even more of them will only reduce their potential value.

If Cures stalls in the Senate, backers fear it could be a lengthy wait for another chance to lock in NIH funding and tweak FDA rules. The bill, says Representative Fred Upton (R–MI), a leading sponsor, is a "once-in-a-generation, transformational opportunity to change the way we treat disease." ■

Rock star Roger Daltry (left) of The Who was among the many health activists who discussed the 21st Century Cures Act with Representatives Diane DeGette (D–CO, center) and Fred Upton (R–MI).

PHOTO: TOM WILLIAMS/CQ ROLL CALL VIA AP IMAGES



The gravity of the galaxy cluster Abell 2744 (left) magnifies light from some of the universe's oldest galaxies.

PHOTO: NASA, ESA, AND J. LOTZ, M. MOUNTAIN, A. KOEKEMOER, AND THE HFF TEAM (STSCI)

ASTRONOMY

Hubble uses galactic lens to study universe's first stars

Gravitational lensing aids search for source of radiation that ionized ancient cosmic gas

By **Daniel Clery**

In a just-completed observing program, astronomers fitted one of their most powerful instruments, the Hubble Space Telescope, with the cosmic equivalent of a telephoto lens. Their goal was to scour the oldest and most distant reaches of the cosmos for the small, dim galaxies that probably held most of the first stars. Hubble's vision isn't keen enough, so its handlers aimed it at six massive galaxy clusters—groupings of hundreds or thousands of galaxies—hoping for a boost from gravitational lensing, a phenomenon originally predicted by Albert Einstein. With their masses bending light from background objects, the thinking went, these clusters would bring fainter galaxies into view. "We're taking a trick from Einstein's book," says Jennifer Lotz of the Space Telescope Science Institute (STScI) in Baltimore, Maryland, leader of Hubble's Frontier Fields program.

The program aims to understand the so-called epoch of reionization (EoR)—the time when the universe was less than a billion years old and something blasted the neutral gas scattered through space after the big bang, stripping away its electrons. Most astronomers believe the culprit was ultraviolet light from the very first generation of stars. But were there even enough of them around at the time, and were they generating sufficient light of the sort needed to cause this cosmic metamorphosis? Astronomers are poring over data from Frontier Fields, but early results suggest there were.

Hubble has produced several earlier images of the most distant galaxies, made by staring at one small patch of sky for long periods. Four years ago, Matt Mountain, then-director of STScI, was considering whether to embark on another of these "deep field" efforts, which had revealed the largest and brightest of the primordial galaxies. An independent committee he asked for advice recommended trying the lensing strategy to see much fainter sources in the EoR. "Gravitational lensing was becoming a mature field. We could turn it into a real tool," says Mountain, now president of the Association of Universities for Research in Astronomy in Washington, D.C.

Over 3 years, some 900 of Hubble's 95-minute orbits, a large chunk of the STScI director's discretionary time, were devoted to observing the six galaxy clusters. "It was almost constant observing. Every week we had stuff coming down," Lotz says.

To interpret the images, which show a jumble of galaxies, far and near, Hubble had help from two other NASA space observatories, the Chandra x-ray telescope and the infrared Spitzer Space Telescope. Spitzer's cameras focused on the clusters for more than 1000 hours. Infrared light can pass through interstellar dust, so Spitzer data helped astronomers distinguish the truly distant, ancient galaxies from more recent, dusty ones, which can look similar.

Spitzer also helped get a fix on the sizes of galaxies and their ages. Early galaxies teemed with massive stars that burned fast and bright in the ultraviolet. Stretched out by expansion of the universe, the light of those stars is seen from Earth at visible wavelengths, easily picked up by Hubble. Light from cooler sunlike stars, meanwhile, gets stretched out to infrared, visible to Spitzer. Finding the mix of fast-burning young stars and old, slow burners tells you a lot about a galaxy. "From the ratio of optical to infrared light we can get the age of the galaxy and the number of stars," says Peter Capak of NASA's Jet Propulsion Laboratory in Pasadena, California, Spitzer lead in the project.

Chandra, meanwhile, helped the team decipher the distorted images produced by galaxy cluster gravitational lenses, which are lumpy and asymmetrical, not smooth like a telescope lens. To correct for the lens shape, astronomers have to map the distribution of matter—both normal and dark—within a galaxy cluster. Chandra pinpointed some of the densest parts of clusters, so-called active galactic nuclei, which shine brightly in x-rays. Still, working out how the lenses distort background objects remains something of a black art. "Each group has a different approach and uses different assumptions," Lotz says. Her team gave simulated images to five of the groups and compared their results to hone an overall approach. "It was very successful. We made a lot of progress," she says.

Frontier Fields finished its observations in September, and the team is now engaged in the mammoth task of processing the data to find the magnified ancient galaxies in each image. Data from two of the clusters—Abell 2744 and MACS J0416, both about 4 billion light-years away—have been released, and dozens of papers based on the data are already circulating as preprints.

Some previous studies had suggested that the early universe held so much gas that the first stars could not have done the job of ionizing it. But the galaxies that Frontier Fields is beginning to reveal may have been enough to do the job. "We are seeing enough stars," Capak says. Still, questions remain, he says: "Do the stars have the right properties? How much of their light is escaping the galaxies?"

It will take Hubble's successor, the James Webb Space Telescope, scheduled to launch in 2018, to fully answer these puzzles, but Frontier Fields is getting a head start. "It's good to be as prepared as possible," Capal says. With Frontier Fields, "we can take the primary steps now, not in the first year of James Webb." ■

INFECTIOUS DISEASE

One year later, Zika scientists prepare for a long war

Researchers still don't understand the risks, but worry about a possible second wave of birth defects

By **Gretchen Vogel**

It may have sounded like good news. Two weeks ago, the World Health Organization (WHO) announced that the Zika epidemic was no longer a Public Health Emergency of International Concern, a designation reserved for acute threats that Zika received in February. WHO's downgrade came after several months of dropping case numbers.

But as the agency was keen to point out, the change just means that Zika has evolved from a cross-border emergency to another long-term public health challenge that needs the world's attention. The virus is now firmly entrenched in Latin America, which is already battling several other mosquito-borne diseases.

Meanwhile, fundamental gaps remain in scientists' knowledge. They don't know how many people Zika infects silently, whether infection leads to long-term immunity, or how big a threat it poses to babies in the womb. And although several candidate vaccines have entered initial human trials, a widely available product is still years away. As Albert Ko, an infectious disease expert at Yale University, puts it, "We have no prevention, no treatment, and no effective ways for counseling women on risks."

Heartbreaking pictures from Brazil heralded the crisis in late 2015: babies born with tiny heads, their brains apparently deformed by a virus known for more than a half-century in Africa and Southeast Asia but never considered a serious threat. Within a few months, Zika had raced through Latin America and the Caribbean and had confronted researchers with a puzzle that was both terrifying and fascinating.

The decline in reported Zika cases is due in part to seasonal drops in populations of mosquitoes, Zika's main vector. (Researchers have also learned that it can hide in semen for months and be transmitted during sex.) Cases could pick up again soon, as the South American summer gets underway. "There are still many places that are going to witness an epidemic," says Ko, who helped document the steep rise in severe birth defects in northeast Brazil a year ago.

In the lab, researchers have worked out how the virus could harm fetuses. In cell cultures and in animal models, it zeroes in on nervous system cells, especially immature ones needed for brain growth. An unusually small head size, called microcephaly, is just one of the birth defects linked to Zika. Doctors have also documented problems with babies' eyes, ears, and limbs, as well as more subtle brain damage.

But a key question has proven difficult to answer: Why have the effects of Zika been so pronounced in northeast Brazil? Doctors in Pernambuco, Bahia, and Paraíba states reported more than 4000 cases of microcephaly between July 2015 and February (see graphic, below); Zika was eventually confirmed as the most likely cause for about one-third of them. But in other areas that suffered outbreaks, the impact has been much less severe. The southern city of São Paulo has seen plenty of birth defects, including cysts in the brain and eye and ear problems, but very few microcephaly cases, says virologist Maurício Lacerda Nogueira of the Faculty of Medicine of São José do Rio Preto. The rate of problems is similar, he says. "What is different is the severity."

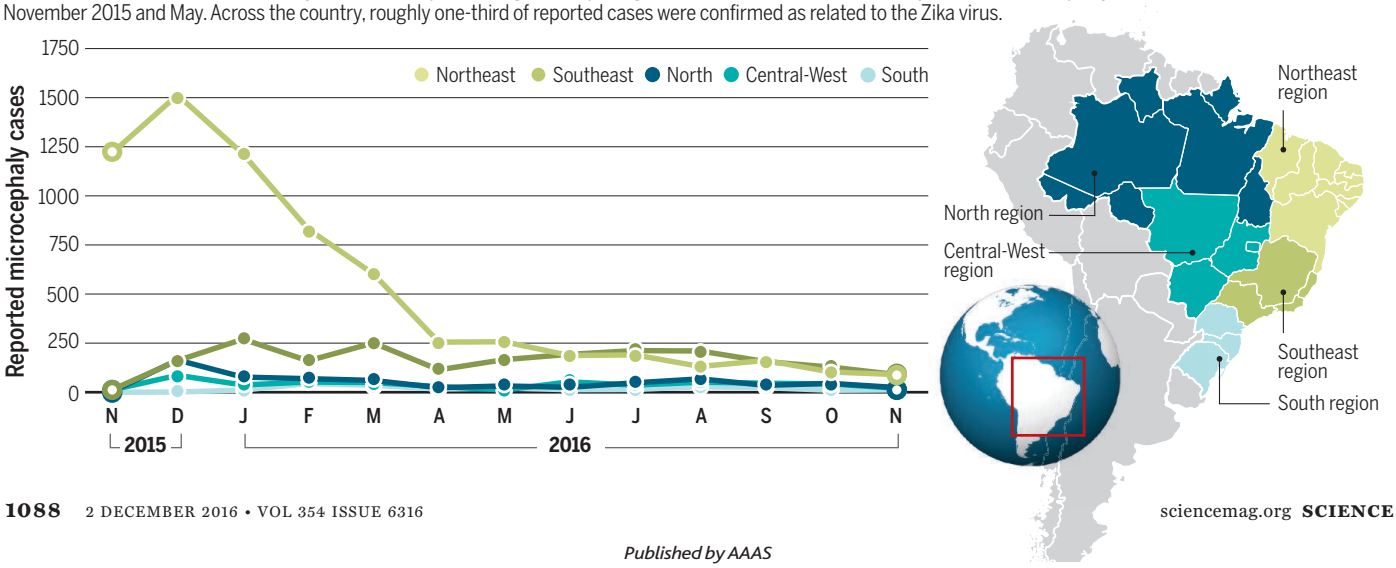
Other Zika-affected countries have fared better as well. Colombia, which had an estimated 100,000 Zika cases, including more than 17,000 in pregnant women, has reported only 58 cases of microcephaly. Mexico has monitored more than 3000 pregnant women believed to be infected but found not a single microcephaly case.

Part of the explanation—besides differences in reporting—may be that after northeast Brazil was hit, women began postponing pregnancy or having abortions if they feared they had been infected. "In northeast Brazil, the virus caught everyone by surprise. Now, everybody knows," says Laura Rodrigues of the London School of Hygiene & Tropical Medicine (LSHTM).

Researchers also wonder whether a cofactor may have exacerbated the virus's effects in the northeast. Initially, some suggested previous exposure to the dengue virus, a close relative of Zika, might trigger antibodies that interfere with the body's response to Zika. But São Paulo has dengue rates similar to those in the northeast, Nogueira says. An-

What happened in Northeast Brazil?

Researchers still don't understand why Northeast—a political region comprising nine Brazilian states—had so many cases of microcephaly and related birth defects between November 2015 and May. Across the country, roughly one-third of reported cases were confirmed as related to the Zika virus.



CREDITS: (GRAPHIC) G. GRULLÓN/*SCIENCE*; (DATA) BRAZIL MINISTRY OF HEALTH

other possibility is that yellow fever vaccination, which is more common in the southeast than in the northeast, might offer some protection against Zika because the two viruses are also related. Malnutrition in the poverty-stricken northeast may play a role as well; more severe birth defects seem to cluster in poorer areas.

Another theory is that pyriproxyfen, an insecticide used to target mosquitoes, might be involved. In 2014, municipalities in Pernambuco started treating drinking water tanks in some regions with the larvicide in an effort to fight dengue. WHO has approved the treatment as safe, but some researchers have pointed out that pyriproxyfen's chemical structure resembles that of retinoic acid, which plays a key role in both fetal and placental development. Yet in a study published in October, Rodrigues and her colleagues found no statistically significant difference in rates of microcephaly between regions where pyriproxyfen was used and those that used a different chemical, Bt toxin.

Perhaps the explanation is simply that northeast Brazil experienced a much larger Zika outbreak than other regions did, Rodrigues says. The virus was probably spreading for more than a year before doctors detected it in early 2015; perhaps the second season led to more infections and more birth defects. That would bode ill for regions that have had only 1 year of Zika, Ko adds: "We are concerned that there may be a second wave that is going to come this next season."

Assessing the real risk to pregnant women is difficult, however, because Zika remains hard to diagnose. "We don't have a denominator," says Rosanna Peeling, an expert on diagnostics at LSHTM. The virus can be detected in blood or other body fluids during the 10 days or so that an infection lasts, but most people don't visit a doctor in that period because symptoms are mild. Detecting past infections is difficult because the Zika virus is so similar to several other viruses—including dengue, which is rampant in Latin America—that current antibody tests can't distinguish between them. Several labs are working on Zika-specific antibody tests, but they will take years to develop, Peeling says.

The lack of a test also makes it hard to tell whether past Zika infection provides lasting protection, and how likely the virus is to cause fresh outbreaks in parts of the world where it has long been endemic, such as Asia and Africa. In the meantime, cohort studies that seek out patients with Zika infections as they happen and follow them can provide some answers. Several such studies are underway, but the first results are not expected for a year or more. Until then, Zika will keep researchers guessing. ■

NEUROSCIENCE

Energy pulses reveal possible new state of memory

Transcranial magnetic stimulation brings latent working memories back to direct consciousness

By **Jessica Boddy**

Memory researchers have shone light into a cognitive limbo. A new memory—the name of someone you've just met, for example—is held for seconds in so-called working memory, as your brain's neurons continue to fire. If the person is important to you, the name will over a few days enter your long-term memory, preserved by permanently altered neural connections. But where does it go during the in-between hours, when it has left your standard working memory and is not yet embedded in long-term memory?

On p. 1136, a research team shows that memories can be resurrected from this limbo. Their observations point to a new form of working memory, which they dub prioritized long-term memory, that exists without elevated neural activity. Consistent with other recent work, the study suggests that information can somehow be held among the synapses that connect neurons, even after conventional working memory has faded (*Science*, 14 March 2008, p. 1543).

> ***"This is a really fundamental find—it's like the dark matter of memory."***
> **Geoffrey Woodman**, Vanderbilt University

"This is a really fundamental find—it's like the dark matter of memory," says Geoffrey Woodman, a cognitive neuroscientist at Vanderbilt University in Nashville who was not involved with the work. "It's hard to really see it or measure it in any clear way, but it has to be out there. Otherwise, things would fly apart."

Cognitive neuroscientist Nathan Rose and colleagues at the University of Wisconsin (UW) in Madison initially had subjects watch a series of slides showing faces, words, or dots moving in one direction. They tracked the resulting neural activity using functional magnetic resonance imaging (fMRI) and, with the help of a machine learning algorithm, showed they could classify the brain activity associated with each item. Then the subjects viewed the items in combination—a word and face, for example—but were cued to focus on just one item. At first, the brain signatures of both items showed up, as measured in this round with electroencephalography (EEG). But neural activity for the uncued item quickly dropped to baseline, as if it had been forgotten, whereas the EEG signature of the cued item remained, a sign that it was still in working memory. Yet subjects could still quickly recall the uncued item when prompted to remember it a few seconds later.

Rose, who recently left UW for the University of Notre Dame in South Bend, Indiana, and his colleagues then turned to transcranial magnetic stimulation (TMS), a noninvasive method that uses rapidly changing magnetic fields to deliver a pulse of electrical current to the brain. They had subjects perform the same cued memory task, then applied a broad TMS pulse just after the signature of the uncued memory item had faded. The appropriate neural activity for that "forgotten" item spiked, showing the memory was reactivated into immediate consciousness from its latent state. What's more, when the TMS directly targeted the brain areas that were initially active for the uncued item, the reactivation response was even stronger.

The study doesn't address how synapses or other neuronal features can hold this second level of working memory, or how much information it can store. "It's a primitive early step in understanding how we bring things into mind," says UW cognitive neuroscientist Bradley Postle, a study co-author.

Woodman agrees. "Good studies tend to bring to light more questions than they answer," he says. "This work absolutely does that." Ultimately, he says, this new memory state could have a range of practical implications, from helping college students learn more efficiently to assisting people with memory-related neurological conditions such as amnesia, epilepsy, and schizophrenia. ■

Chris Monroe examines equipment that keeps ions trapped in long-lasting quantum states.

QUEST FOR QUBITS

How small startups are vying with corporate behemoths for quantum supremacy

By **Gabriel Popkin**

It's a Sunday afternoon in September, and the two co-founders of ionQ, a quantum computing startup, are meeting for a strategy session with their first hire: their new CEO. Sitting in comfy leather chairs in the Physical Sciences Complex at the University of Maryland (UMD) in College Park, the two founders are experiencing a touch of culture clash. Lifelong research scientists, UMD physicist Chris Monroe and Jungsang Kim, an electrical engineer at Duke University in Durham, North Carolina, are relaxed and talkative about their company's plans, even in the presence of a reporter. They tick off reasons why trapped ions, their specialty, will make for a great quantum computer—perfect reproducibility, long lifetimes, and good controllability with lasers.

Their CEO, David Moehring, whom Monroe and Kim have just hired away from the U.S. Intelligence Advanced Research Projects Activity, is more guarded. He warns Monroe and Kim against divulging information that he thinks a startup should keep secret—including exactly how much money they received from the venture capital firm New Enterprise Associates. (He is willing to confirm that it's several million dollars.) Kim nods at Moehring and chuckles. "At some point this guy will implement a policy that we need to get his approval to talk."

These unlikely partners share a common conviction: that quantum computing—which aims to harness quantum mechanics to vastly accelerate computation—is ready for prime time. They are not alone. Tech giants Intel, Microsoft, IBM, and Google are all plowing tens of millions of dollars into quantum computing. Yet the contenders are betting on different technological horses: No one yet knows what type of quantum logic bit, or qubit, will power a practical quantum computer.

Google, often considered the field's leader, has signaled its choice: tiny, superconducting circuits. Its group has built a nine-qubit machine and hopes to scale up to 49 within a year—an important threshold. At about 50 qubits, many say a quantum computer could achieve "quantum supremacy," a term coined by John Preskill, a physicist at the California Institute of Technology in Pasadena, to denote a quantum computer that can do something beyond the ken of a classical computer, such as simulate molecular structures in chemistry and materials science, or tackle certain problems in cryptography or machine learning.

IonQ's team isn't ruffled by Google's success. "I'm not worried that Google will declare next month that the game's over," Kim says. "Or maybe they can declare it, but it won't be over." Still, ionQ, which lacks a building or even a website, is a decided underdog. The startup is sticking with trapped ions—the technology behind the world's very first quantum logic gates, which Monroe himself helped create in 1995. With precisely tuned laser pulses, Monroe can knock ions into quantum states that last for seconds—far longer than Google's qubits. Kim has developed a modular scheme for connecting groups of ions together, which might allow ionQ to scale up faster than many expect. But so far, its leaders have joined just five qubits into a programmable machine. Trapped ions are "a bit of a black sheep right now," Monroe admits, "but I think in the coming years people will be flocking to it."

One thing is certain: Building a quantum computer has gone from a far-off dream of a few university scientists to an immediate goal for some of the world's biggest companies. And Monroe and his colleagues are among many who hope to cash in. Although superconducting qubits may have taken a momentary lead among industry players, experts agree that it's far too early to declare a winner. "It's a good thing that these different technologies are being developed in parallel," says Preskill, an unofficial dean of quantum information science. "Because there could be surprises that really change the game."

QUBITS OUTMUSCLE classical computer bits thanks to two uniquely quantum effects: superposition and entanglement. Superposition allows a qubit to have a value of not just 0 or 1, but both states at the same time, enabling simultaneous computation. Entanglement enables one qubit to share its state with others separated in space, creating a sort of super-superposition, whereby processing capability doubles with every qubit. An algorithm using, say, five entangled qubits can effectively do 2^5, or 32, computations at once, whereas a classical computer would have to do those 32 computations in succession. As few as 300 fully entangled qubits could, theoretically, sustain more parallel computations than there are atoms in the universe.

This massive parallelism would not help with many tasks—nobody thinks quantum computers will revolutionize word processing or email. But it could dramatically speed up algorithms designed to explore vast numbers of different paths simultaneously, and solve problems that include searching through large data sets, discovering new chemical catalysts, and factoring large numbers used to encrypt data. Quantum computers may even find a role simulating black holes and other phenomena in physics.

There is a major catch, however. Quantum superpositions and entangled states are exquisitely fragile. They can be destroyed by slight perturbations from the environment—or by attempts to measure them. A quantum computer needs protection from what Robert Schoelkopf, a physicist at Yale University, calls "a sea of classical chaos."

Though theoretical ideas started appearing in the early 1980s, experimental quantum computing got going only in 1995, after Peter Shor, a mathematician at Bell Labs in Murray Hill, New Jersey, showed that a quantum computer could quickly factor large numbers—a capability that would render much of modern cryptography obsolete. Shor and others also showed that it was theoretically possible to keep fragile qubits stable indefinitely by using neighboring qubits to correct their errors.

Suddenly, physicists and their funders had both a concrete reason to build a quantum computer and a sign that the machine

PHOTO: © CAMERON DAVIDSON

wouldn't dissolve into a pile of cascading errors. David Wineland, a Nobel Prize–winning physicist at a National Institute of Standards and Technology (NIST) laboratory in Boulder, Colorado, had already pioneered methods to use lasers to cool ions and control their internal quantum states. Within a year of Shor's discoveries, Wineland and Monroe, a NIST staff scientist at the time, built the first quantum mechanical logic gate, using lasers to manipulate electron states in a beryllium ion. Because of Wineland's experience with ions, the chance to seize the lead in early quantum computing experiments "fell in our laps," Monroe says.

As millions of government research dollars began flowing to quantum physics groups around the world, other kinds of qubits began to appear. By the early 2010s, trapped ions faced a strong challenge from a new darling: circuit loops made out of superconductors—metallic materials that can carry an oscillating electric current without resistance when chilled nearly to absolute zero. The 0 and 1 of the qubit correspond to different current strengths. Adding to their appeal, the loops can be seen with the naked eye, controlled with simple microwave electronics rather than finicky lasers, and fabricated using techniques from conventional computer chip manufacturing. They also operate very quickly.

At least at first, however, superconductors had a fatal weakness: Environmental noise, even from the electronics used to control them, can disrupt their quantum superpositions in a small fraction of a microsecond. But engineering refinements have improved the circuits' stability by more than a million times, so that they now can remain in a superposition state for tens of microseconds—though they still collapse far faster than ions.

IN 2007, D-Wave Systems, a startup company in Burnaby, Canada, surprised just about everybody by announcing that it had built a quantum computer, with 16 superconducting qubits. D-Wave's machine didn't entangle all the qubits, and it couldn't be programmed qubit by qubit. Instead, it relied on a technique called quantum annealing, in which qubits are entangled only with near neighbors and interact to produce not a set of parallel computations, but a single overall quantum state. D-Wave developers hoped to map complicated mathematical problems onto such states and use quantum effects to find minimum points, a promising technique for solving optimization problems such as efficiently routing air traffic.

Almost instantly critics cried foul: D-Wave did not even attempt to do certain things that many thought essential to quantum computing, such as error correction. But several companies, including Google and Lockheed Martin, bought and tested D-Wave devices. A tentative consensus emerged: They did something quantum, and, for certain specialized tasks, they might perform faster than a conventional computer. Quantum or not, D-Wave jolted the private sector awake. "It was really eye-opening," Monroe says. "[D-Wave] showed that there's a market, there's a hunger for these devices." Within a few years, companies started lining up behind technologies that aligned with their in-house expertise.

Intel made one of the biggest bets, announcing in 2015 that it would invest $50 million into research at QuTech, an offshoot of Delft University of Technology in the Netherlands. The company is focusing on silicon quantum dots, often called "artificial atoms." A quantum dot qubit is a small chunk of material in which, as in an atom, the quantum states of an electron can represent 0 and 1. Unlike ions or atoms, however, a quantum dot doesn't need lasers to trap it.

> ***"... there's a market, there's a hunger for these devices."***
> **Chris Monroe**, University of Maryland

Early quantum dots were made from near-perfect crystals of gallium arsenide, but researchers have turned to silicon, hoping to leverage the massive manufacturing infrastructure of the semiconductor industry. "I think [Intel's] heart is with silicon," says Leo Kouwenhoven, scientific director of QuTech. "That's what they're good at." But silicon-based qubits are well behind those based on ions or superconductors, with the first two-qubit logic gate reported only last year by a group at the University of New South Wales in Sydney, Australia.

Microsoft went for what many consider an even longer shot: topological qubits based on nonabelian anyons. These aren't objects at all—they're quasiparticles, traveling along the boundary between two different materials—and their quantum states are encoded in the different braiding paths they follow in time. Because the shapes of the braided paths lead to the qubit superpositions, they would be "topologically protected" from collapse, similar to how a shoelace stays tied even if nudged or bumped.

This meant that theoretically, a topological quantum computer wouldn't need to devote so many qubits to error correction. As early as 2005, a Microsoft-led team proposed a way to build a topologically protected qubit in hybrid semiconductor-superconductor structures, and Microsoft has funded several groups to try to make one. Recent papers from these groups and from a separate effort at Bell Labs have shown hints of the crucial anyon in the patterns of electrical currents that flow in their specialized circuitry, and the scientists are close to demonstrating an actual qubit, Preskill says. "I think in a year or two, we can consider it to be nailed: Topological qubits exist."

Google, for its part, recruited John Martinis, a superconducting qubit expert at the University of California, Santa Barbara (UCSB), who had studied D-Wave's operation and shortcomings. In 2014, the company swallowed his UCSB research team whole, hiring about a dozen people. Soon afterward, Martinis's team announced they had built a nine-qubit machine at UCSB, one of the largest programmable quantum computers so far, and they are now trying to scale up. To avoid creating an unwieldy jumble of wires, they are rebuilding the system into a 2D array that will sit on top of a wafer with control wires etched into it.

In July, Martinis's team—now up to about 30 scientists and engineers—used three superconducting qubits to simulate the ground state energy of a hydrogen molecule, demonstrating that quantum computers can simulate simple quantum systems as well as classical computers. The result points to the coming power of a machine with quantum supremacy, he says. Martinis calls the 1-year timetable for reaching a 49-qubit computer a "stretch goal," but he believes it may be possible.

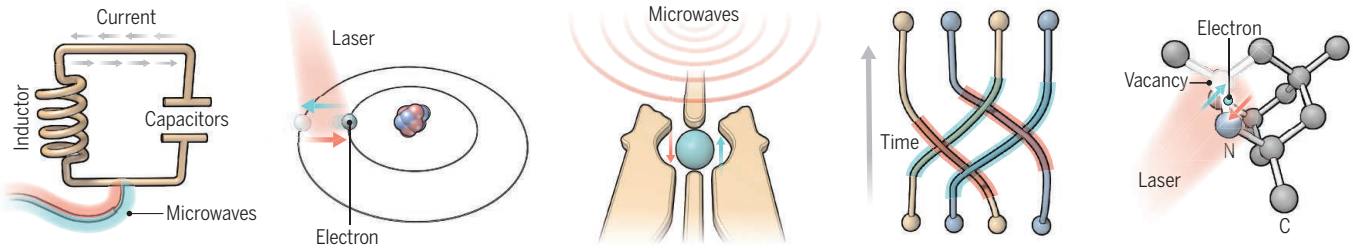
MEANWHILE, Monroe is grappling with the challenges that come with trapped ions. As qubits, they can remain stable for seconds, thanks to vacuum chambers and electrodes that stabilize them even in the presence of external noise. Yet that isolation also means it is a challenge to get the qubits to interact. Monroe recently entangled 22 ytterbium ions in a linear chain, but so far he is not able to control or query all ion pairs in the chain, as a quantum computer will require.

The complexity of controlling the ensemble rises with the number of ions squared, so adding many more is impractical. The way forward, Monroe believes, is to go modular and use fiber optics to link traps holding perhaps 20 ions each. In such a scheme, certain qubits within each module would act as hubs, reading out information from the rest of the qubits and sharing it with other modules; this way, most qubits could remain shielded from external interference.

On a recent afternoon, Monroe toured his six lab spaces at UMD. In his three older labs, electrical wires and vacuum lines descend

A bit of the action

In the race to build a quantum computer, companies are pursuing many types of quantum bits, or qubits, each with its own strengths and weaknesses.



	Superconducting loops	Trapped ions	Silicon quantum dots	Topological qubits	Diamond vacancies
	A resistance-free current oscillates back and forth around a circuit loop. An injected microwave signal excites the current into superposition states.	Electrically charged atoms, or ions, have quantum energies that depend on the location of electrons. Tuned lasers cool and trap the ions, and put them in superposition states.	These "artificial atoms" are made by adding an electron to a small piece of pure silicon. Microwaves control the electron's quantum state.	Quasiparticles can be seen in the behavior of electrons channeled through semiconductor structures. Their braided paths can encode quantum information.	A nitrogen atom and a vacancy add an electron to a diamond lattice. Its quantum spin state, along with those of nearby carbon nuclei, can be controlled with light.
Longevity (seconds)	0.00005	>1000	0.03	N/A	10
Logic success rate	99.4%	99.9%	~99%	N/A	99.2%
Number entangled	9	14	2	N/A	6
Company support	Google, IBM, Quantum Circuits	ionQ	Intel	Microsoft, Bell Labs	Quantum Diamond Technologies
Pros	Fast working. Build on existing semiconductor industry.	Very stable. Highest achieved gate fidelities.	Stable. Build on existing semiconductor industry.	Greatly reduce errors.	Can operate at room temperature.
Cons	Collapse easily and must be kept cold.	Slow operation. Many lasers are needed.	Only a few entangled. Must be kept cold.	Existence not yet confirmed.	Difficult to entangle.

Note: Longevity is the record coherence time for a single qubit superposition state, logic success rate is the highest reported gate fidelity for logic operations on two qubits, and number entangled is the maximum number of qubits entangled and capable of performing two-qubit operations.

in tangles from above. On oversize tables, a bewildering array of lenses and mirrors shape and direct laser light toward portals in small steel vacuum chambers containing the all-important ions. Overhead heating, ventilation, and air conditioning (HVAC) equipment—necessary to keep dust down and stabilize the lab temperature—gives off a steady drone. "I'm passionate about HVAC," Monroe says.

The three newer labs are, by contrast, tidy and eerily empty. Instead of Rube Goldberg optics tables, most of the lasers are integrated, plug-and-play units from companies like Honeywell—prototypes for the kinds of turnkey systems that ionQ needs to perfect if it is going to succeed. "The lasers we use now have only one knob, and it's 'on,'" Monroe says. He is antsy to get ionQ's labs up and running, so he can transition his highly paid research scientists onto ionQ's payroll and set them to perfecting technologies they've developed at UMD—which, thanks to an unusual agreement, ionQ can license exclusively and royalty-free. Next year he will take his first-ever sabbatical to focus on building ionQ. The private sector money flowing into quantum research, he says, "is the biggest deal in my career."

EVEN AS MONEY HAS POURED IN, quantum computing is a long way from becoming a secretive commercial field. The major research groups—even those affiliated with big companies—are still publishing results and presenting at conferences. They say they have a mutual interest in publicizing their advances, not least so that potential customers can think about how they could use a quantum computer. "We all need a market," Monroe says.

What's more, nobody knows enough about quantum computing yet to go it alone with a single qubit type. Every approach needs refining before quantum computers can be scaled up. Superconductor- and silicon-based qubits need to be manufactured with more consistency, and the refrigerators that chill them need streamlining. Trapped ions need faster logic gates and more compact lasers and optics. Topological qubits still need to be invented.

A future quantum computer could well be a hybrid, with ultrafast superconducting qubits running algorithms, then dumping output to more stable ion memory, while photons shuttle information among different parts of the machine or between nodes of a quantum internet. "One can imagine we'll have an environment in which several types of qubits exist and play different roles," says Krysta Svore, a Microsoft researcher in Redmond, Washington.

A quantum computer is so new, and so strange, that even the world's top quantum physicists and computer engineers do not know what a commercial one will ultimately look like. Physicists will need to simply build the most complex computer possible with current technology, then confront the new challenges that are sure to crop up, Svore says. Build, study, and repeat. "We like to joke that once we have a quantum computer, we're going to use it to design the next quantum computer." ■

Gabriel Popkin is a freelance journalist in Mount Rainier, Maryland.

Li Hui extracts DNA to help nab modern killers—and probe ancient myths.

BRINGING LEGENDS TO LIFE

Geneticist Li Hui believes a DNA database can authenticate mythical figures from before the dawn of China

By **Kathleen McLaughlin**, *in Chengtoushan, China*

Inside a museum built over Stone Age ruins here, two workers last month attached an industrial-sized suction cup to a vacuum-sealed glass case. Air hissed as the seal was cracked and the lid was gently lifted away, exposing an eroded human skeleton. Li Hui, a geneticist from Fudan University in Shanghai, China, had his prize. He leaned over the remains, once a major figure in Neolithic Chengtoushan, an early site of rice cultivation in southern China's Hunan province. The bones are alone in their glass casket, but when excavated in the mid-1990s they were draped with a jade necklace and surrounded by sculptures of a phoenix.

As a ghoulish green light bathed the skeleton, Li wrapped his hand in a plastic bag to minimize contamination and delicately plucked a bone fragment that had chipped off the femur. Then he nodded, and the workers sealed the display case back up. Back in his laboratory, Li will extract and analyze DNA from the bones, using techniques from the burgeoning field of ancient DNA. But unlike others in this hot field who want to understand ancient populations and migrations, Li is seeking scientific support for some of China's most cherished legends.

Scientists here believe the skeleton could be as many as 6500 years old, 2 millennia older than China's first historical dynasty, the Xia—a time, according to legend, when deities known as the Three Sovereigns ruled the land. Chinese credit the sage-kings with laying the foundations of their culture: inventing silk and medicine, for instance, and fashioning China's written characters. Based on the luxurious relics buried with the skeleton—which archaeologists dubbed "the mayor"—Li believes he may have been a chieftain of a clan associated with Fuxi, one of the mythical sovereigns. Fuxi is credited with rice cultivation, and the phoenix is a symbol associated with that sovereign.

By comparing DNA from the bone chip to sequences in a vast database of DNA samples gathered around China, Li hopes to probe that tantalizing possibility. It is the latest and boldest of his efforts to turn myths into history with the help of DNA. "We are retracing and rebuilding history to understand the development and adaptation of Chinese people," Li says, "so we can imagine the future of Chinese people—what will we evolve into?"

Li's quest has won cautious praise from some scholars. Paleoanthropologist Chris

PHOTO: LI HUI

Stringer of the Natural History Museum in London calls Li's work "ambitious" and notes that DNA has proven to be a more powerful tool for elucidating prehistory than he and others could have imagined even a few years ago. Others have qualms. "I can't think of any historian who would be interested in this, except as a way of analyzing people in China today and how they think about the past," says Sigrid Schmalzer, a historian at the University of Massachusetts in Amherst and author of *The People's Peking Man: Popular Science and Human Identity in Twentieth-Century China*. She and some colleagues see a skein of nationalism running through the attempt to establish a cohesive Chinese culture predating the Xia dynasty. "There is a very strong interest in China in laying claim to a Chinese identity just as far back as you can take it," Schmalzer says.

Li shrugs and says, simply, that there is a gulf between historians and geneticists, and that he is following where the facts lead him. He has debunked some historical claims, he notes, but he says he already has support for the Three Sovereigns legend: studies of the Y chromosome—the male sex chromosome—that trace about half of modern Chinese men back to three ancient groups. "What makes me believe that [the Three Sovereigns legend] is true is the perfect match of the mythology, archaeology, and genetics in timescales and geographic distribution," he says. Chengtoushan's mayor, Li says, may even have been the historical figure revered as Fuxi.

OVER THE PAST 2 DECADES, scientists in Li's home base, Fudan University's Key Laboratory of Contemporary Anthropology, have collected blood samples from more than 400,000 people across China to map ethnic groups. "Chinese people want to know their links to each other," Li says, which made it easy for him and his mentors to gather samples. "Before DNA, they liked to study their family records. They want to know who they are related to."

The Fudan group's original goal was not to probe myths, but to build a genetic tree for the people speaking China's 400-plus dialects and languages. Having sampled DNA of individuals who collectively speak 324 tongues, the team is about three-quarters of the way toward reaching its goal. The lab's founder, Fudan University Vice President Li Jin, believes the map will give a more precise picture of China's diversity than does the government's classification of Chinese people according to 56 official ethnicities.

Li Hui, 38, joined Li Jin's lab as an undergraduate in 1998. He made his first big splash a decade later, when he used the growing DNA collection to discredit a theory that indigenous Taiwanese—the original occupants of the island, where the Nationalists fled in the waning days of China's civil war in the late 1940s—descended from Micronesians and Polynesians. Rather, their genes could be traced to the Dai people of southern China, he reported in *Evolutionary Biology* in 2008. That finding was in line with the Chinese government's view that the island has always been Chinese.

Fudan's DNA trove serves the state's interests in practical ways as well. Police call on the group once a week, on average, to help solve crimes. Because most of China's 1.4 billion people share just 100 surnames, and until recently these clans stayed put in their home villages, the database allows the geneticists to go beyond police DNA forensics and pinpoint a suspect's surname and hometown. The geneticists have identified victims or provided leads in about 50 cases so far, Li says.

The team's latest forensic success came last summer, after police in Gansu province sent them DNA swabbed from a crime scene. Over 14 years, 11 women and girls had been raped and killed in western Gansu, and the police deduced that the cases were related. Analyzing the DNA, Li's group ascertained the serial killer's surname likely was Gao, and zeroed in on his likely home village in Gansu. That was enough for police to go on; in August they captured a man named Gao Chengyong, who confessed to the crimes.

LI HUI'S FIRST FORAY into Chinese mythology came in 2013, when he analyzed DNA from a relative of the Emperor Cao Cao, a renowned warlord who ruled during the second century C.E. For centuries, rumor had it that Cao Cao's grandfather was a eunuch, meaning the emperor's father had been adopted. On national TV, Li debunked that claim. The DNA showed that Cao Cao's grand-uncle and his descendants were related—evidence of an unbroken family line.

Now, Li is taking on the Three Sovereigns, possibly China's most cherished legend. By examining Y chromosome sequences in their database, he and his Fudan colleagues have built a circumstantial case for the existence of the mythical clans. Y chromosome haplogroups—sets of DNA variations that tend to be inherited together—indicate that about half of Chinese men today are descended from three lineages: the Three Sovereigns, Li says.

This intricate rubbing depicts Fuxi, the sage-king of Chinese mythology credited with developing rice cultivation.

PHOTO: COURTESY OF JIANG HONGWEI

Li's assumption that the match between the genetic pattern and the mythical threesome is more than coincidence raises some eyebrows. "I am not sure where one starts out with assumptions about genetic connections that necessarily relates to the mythical founders of Chinese civilization," says Rowan Flad, an archaeologist at Harvard University, who studies the emergence of complex society in Stone Age China. "Population genetics and genealogy rooted in origin stories don't initially strike me as the most compatible data sets."

Li will try to prove otherwise. In his Shanghai lab, cluttered incongruously with DNA sequencing machines and stuffed monkeys, he will extract the mayor's DNA and compare it to sequences in the database. Li suspects the DNA will reveal that the mayor belonged to an ethnic group linked to early rice cultivation—and that ethnic group would be the Fuxi clan.

Li sums up his quest in a poem, written in Tang dynasty style, that he composed to honor Chengtoushan. It's an ode to the secrets that the site may reveal. "Ancient histories so far from the books. … Who built this first city. … Please, let me hear your distant tune." ■

Kathleen McLaughlin is a writer in Beijing.

POLICY FORUM

ENERGY AND ENVIRONMENTAL ECONOMICS

Reforming the U.S. coal leasing program

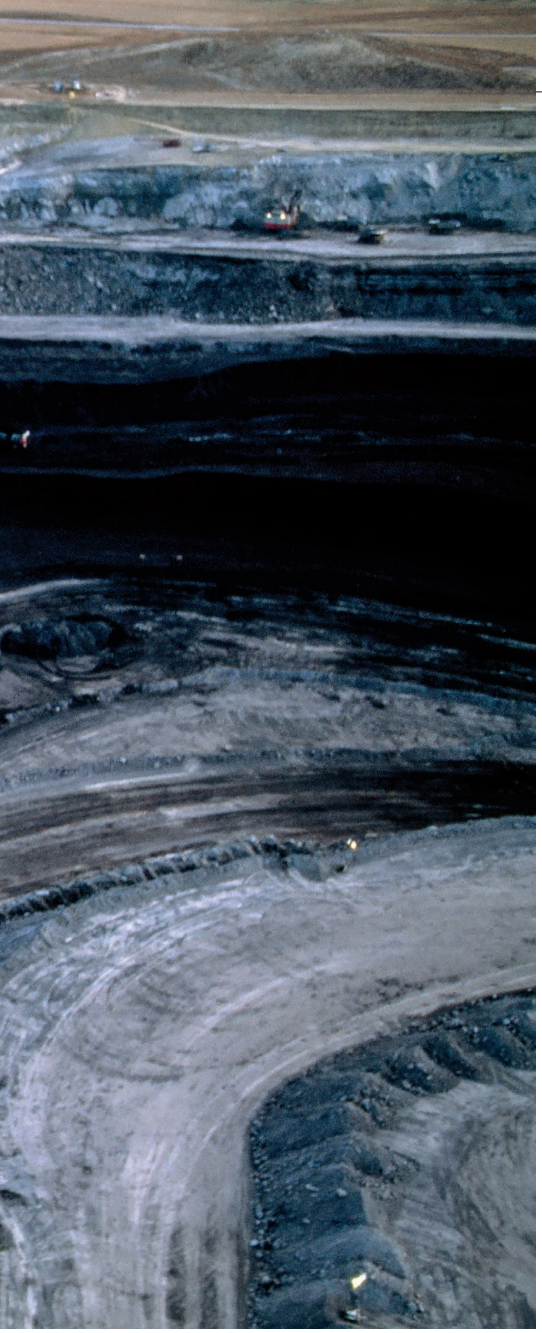
Royalty rates and auction practices do not reflect the social costs of coal

By **Kenneth Gillingham,[1] James Bushnell,[2] Meredith Fowlie,[3] Michael Greenstone,[4] Charles Kolstad,[5] Alan Krupnick,[6] Adele Morris,[7] Richard Schmalensee,[8] James Stock[9]**

About 40% of all coal mined in the United States is extracted from lands owned by the federal government, under leases managed by the U.S. Department of the Interior (DOI). Burning that coal accounts for 13% of U.S. energy-related greenhouse gas (GHG) emissions (*1*). With the largest and lowest-cost reserves in the United States, federal coal alone—estimated at nearly 10% of the world's known reserves—has potential to contribute substantially to atmospheric CO_2 concentrations (*2*). In response to calls for reform, DOI has issued a moratorium on new leases while it develops a Programmatic Environmental Impact Statement to guide the first major reform of the program since 1982. We review existing knowledge of key issues relevant to reform, highlighting the social costs of coal extraction, the extent of substitution away from federal coal induced by raising additional leasing revenue, the lack of competition in the leasing auctions, and the incentives inherent in the current leasing program structure. We then turn to critical areas of research that can be done in the near term and would contribute to more informed debate and policy development.

In light of the U.S. pledge under the Paris Agreement and the tangible actions the

PHOTO: JOEL SARTORE/NATIONAL GEOGRAPHIC CREATIVE



This aerial view of a nonlignite coal strip mine in the Powder River Basin of Wyoming is representative of many operations on federal lands. A reformed federal coal leasing system could provide a more fair return to the taxpayer and reduce environmental damages.

United States is taking to reduce CO_2 emissions, there have been increasing calls from scientists and environmental organizations to incorporate climate considerations into federal coal leasing policy [e.g., (*3*)]. In parallel, there is a conceptually distinct call for reforming the program to bring in greater revenue in order to "provide a fair return to the taxpayer" (*4*). The current program raises revenue through (i) auctions for the right to a 10-year lease (with a confidential minimum bid set by DOI), (ii) royalty payments per ton produced based on the first transaction price (i.e., the price at the first sale of the coal), and (iii) negligible annual land rental payments. The revenue is roughly evenly shared between federal and state governments (*5*).

SOCIAL COSTS, LEAKAGE, INCENTIVES

Nearly all federal coal production occurs in a few Western states. Most is in Wyoming and Montana, which have vast reserves being exploited by low-cost surface mines in the Powder River Basin (PRB) and produce >80% of all federal coal (see the photo). PRB coal sells for much less than other U.S. coal. For example, the average mine-mouth spot price of PRB coal on 28 October 2016 was $0.51 per one million British thermal units (MMBtu), whereas the average price in other major coal-producing basins varied from $1.41 to $1.80 per MMBtu (*6*). PRB coal also has low sulfur content, often lessening the need for equipment to control SO_2 emissions and implying that PRB coal is higher value to some power plants. Although there are differences in transportation costs, low-cost federal coal from the PRB puts downward pressure on coal prices nationwide (*7*).

Low prices for energy are beneficial for consumers but are problematic when the prices do not take into account the full social costs of extraction and combustion. These costs include private costs from extracting and transporting the energy, as well as the external costs that all of society bears, such as those stemming from emissions of coal-bed methane, water pollution, and land degradation from mining activities; CO_2 and other GHG emissions from transporting and combusting coal; and water pollution from coal combustion waste materials (*8*–*11*). When these environmental costs are large, the market price for energy will be considerably below the social cost; the consequence is that more coal is used than is optimal socially.

A striking illustration of how large one of these external costs—from combustion emissions of CO_2—is for coal relative to the market price of coal is seen in the figure below. In contrast to oil and gas, which are also extracted from federal lands, the monetized climate damages from PRB coal combustion are about six times the spot price of coal. This underscores that PRB coal is substantially underpriced relative to its social costs. It also points to the importance of addressing climate externalities of coal use, whether through new regulation to price carbon and other GHG emissions, downstream regulation like the U.S. Environmental Protection Agency's (EPA's) Clean Power Plan (CPP) that seeks to reduce CO_2 emissions from power plants, reforms of the federal coal program, or some combination of these.

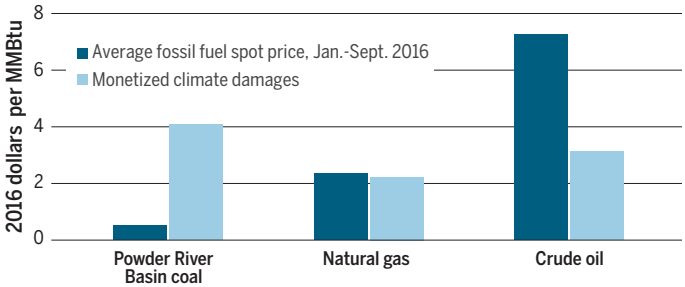
Over the past decade, there has been an abundance of research examining the potential for emissions "leakage," a shift of emissions due to a substitution of activity from regulated to unregulated sources, sectors, or regions. This is especially relevant in the context of reforming the coal leasing program for two reasons. First, the CPP is not likely to fully address the external costs of climate change, in part because of design features that allow for leakage of emissions to unregulated and less-regulated electricity production (*12*). Although reforming the coal leasing program would not likely affect leakage under the CPP, leasing reform combined with regulation of stationary sources of GHGs may help address coal's large climate costs more efficiently. Second, an increase in royalty or lease payments for federal coal can lead to leakage via substitution from federal to nonfederal coal production, such as from Appalachia and the Midwest. If the increase of nonfederal coal production is large enough, it could offset the decline in federal coal production, an offset that implies lower government revenues, with little change in nationwide coal production, prices, or emissions. If the increase in nonfederal coal production is small, government revenues would increase and nationwide production and emissions would decline.

There is some analysis of the degree of leakage from federal to nonfederal coal. One well-known energy model (the Integrated Planning Model) implies that increases in royalties on federal coal would increase government revenue and reduce nationwide coal production and emissions (*1*). In the absence of the CPP, an additional charge on coal at the mine mouth, set so that the CO_2 in the coal is monetized at the U.S. government's central estimate of the social cost of carbon, could achieve roughly three-quarters of the emissions reductions expected from the CPP, a finding that is particularly relevant as the CPP is being litigated (*1*). Another analysis explored the effects of basing royalties on the final delivered price of coal—the price paid by the party combusting the coal—rather than the first transaction price, which often is at the mine mouth. This analysis found that substitution of nonfederal coal for federal would be sufficiently small that revenues from the royalty program would increase substantially (*13*).

These studies focus on the effects of raising royalty rates rather than bringing in more revenue from leasing auctions. This is due to the structure of the auctions, which often lack any genuine competition: More than 90% of leases from 1990 to 2012 had a single bidder, largely because all but a few leases expand existing mines that already have infrastructure (*5*). Consequently, bidding is often uneconomic for other firms, leaving the operator of the existing mine without meaningful competition. Not surprisingly, royalty

payments ($796 million in 2012) have been more than double the auction revenue in most years in the past decade (*5*).

Recent work has analyzed additional programmatic shortcomings. For example, because royalties are based on the first transaction price of the coal, mine operators have an incentive to sell coal for low rates to subsidiaries of the coal company. If the first transaction price is with the final consumer (usually a power plant), mine operators are permitted to reduce the royalty payments by taking self-reported deductions for transportation and washing, rather than deductions based on market prices, which may be substantially below the self-reported estimates (*5*). Finally, mine operators can engage in "take-or-pay" contracts with electric utilities that have lower transaction prices (allowing for lower royalty payments), which are compensated for by penalty payments associated with purchases less than a stated (usually large) benchmark (*14*). These shortcomings highlight the possibility that reforms could increase government revenues and allow the federal government to reduce other taxes, pay down the federal debt, or invest in projects with high social value.



Average fossil fuel spot prices are at the mine mouth or production rig. Costs are heat-rate adjusted. Average annual coal, oil, and gas spot prices are the final delivered prices in 2016 dollars [from www.eia.gov]. Monetized CO_2 climate damages of $44 per ton of CO_2 are based on the median U.S. government social cost of carbon at a 3% discount rate adjusted for inflation to 2016 dollars (*9*). Carbon emission factors come from (*10*) and do not include GHG emissions from the extraction process. Heating values are from (*6*) and (*10*). These calculations are based on the methodology from (*11*). See supplementary materials for data and calculations.

RESEARCH NEEDS

Four major understudied issues could inform policy debate. First is the potential extent of leakage from federal to nonfederal coal. Two different energy sector models project a relatively small degree of replacement of federal with nonfederal coal (*1*, *13*). But these two models capture only some of the many relationships in the energy sector and have analyzed only a few of the many possible reforms. Others have posited that substitution of nonfederal coal for federal may be much greater (*15*). Thus, empirical work and further energy modeling can refine our understanding of how government revenues, nationwide coal production, and emissions might change with alternative reforms of the coal leasing program.

Second, any reform of the leasing program would occur in the context of many other downstream policies for mitigating climate change, which could affect the outcomes of the reform. There is modeling of coal leasing reforms with and without the most important downstream policy, the CPP (*1*), but there are myriad ways states can choose to achieve their emissions targets under the CPP. Further complicating matters are the many local and state-level climate policies. Interactions among all of these policies will determine the efficiency of federal coal leasing reform for addressing the costs of climate change and will also determine how coal production and government revenues will change with a reform.

Third, how will reform change the effects of the leasing program on the global, national, and regional coal markets? Switching a generating unit away from PRB coal to nonfederal coal often entails nontrivial costs, potentially leading to fuel switching away from coal altogether. This could enhance existing trends: With low natural gas prices, an aging fleet of coal power plants, and new regulations for air pollutants like mercury, more than 80% of electric generation retired in 2015 was conventional thermal coal. Although the U.S. Energy Information Administration (EIA) projects that coal will remain a major energy source for the next two decades (*16*), a switch away from PRB coal may affect particular mining regions. Understanding the effects on employment, mineral-related state revenues, and the price of electricity to consumers will help make policy trade-offs transparent.

Fourth, despite evidence of programmatic shortcomings in the coal leasing program, there is remarkably little research on overcoming these shortcomings. Research to redesign the auctions to be more competitive—such as through a regional leasing model—or to reform how royalty payments are structured could help address environmental considerations by bringing the market price closer to the social cost, while bringing in additional government revenue.

Although there is still uncertainty and work to be done, there is more than enough evidence to point us in the general direction of reform. Going forward, a research agenda based on energy modeling and data-intensive empirical work by economists, engineers, and other social scientists should be a high priority to help answer questions that are directly relevant to reforming the federal coal leasing system. More broadly, there is a strong case for exploring and researching similar reforms of the federal oil and natural gas leasing programs. ■

[1]Yale University, New Haven, CT 06520, USA. [2]University of California, Davis, CA 95616, USA. [3]University of California, Berkeley, CA 94720, USA. [4]University of Chicago, Chicago, IL 60637, USA. [5]Stanford University, Stanford, CA 94305, USA. [6]Resources for the Future, Washington, DC 20036, USA. [7]Brookings Institution, Washington, DC 20036, USA. [8]Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [9]Harvard University, Cambridge, MA 02138, USA. Email: kenneth.gillingham@yale.edu

REFERENCES AND NOTES

1. T. Gerarden, W.S. Reeder, J. Stock, "Federal coal program reform, The Clean Power Plan, and the Interaction of upstream and downstream climate policies" (Working paper 22214, National Bureau of Economic Research, Cambridge, MA, 2016).
2. J. A. Luppens *et al.*, "Coal geology and assessment of coal resources and reserves in the Powder River Basin, Wyoming and Montana" (Professional paper 1809, U.S. Geological Survey, Reston, VA, 2015).
3. K. Caldeira *et al.*, Letter to U.S. Department of Interior and Bureau of Land Management to end coal leasing on public lands, 26 July 2016; https://www.biologicaldiversity.org.
4. M. Squillace, *Nat. Resour. Environ.* **27** (3), art7 (2013).
5. U.S. Government Accountability Office, "Coal leasing: BLM could enhance appraisal process, more explicitly consider coal exports, and provide more public information" (Report 14-140, GAO, Washington, DC, 2013).
6. EIA, Coal Markets Report (EIA, Washington, DC, 2016); https://www.eia.gov/coal/markets/.
7. Council of Economic Advisers, "The economics of coal leasing on federal lands: Ensuring a fair return to the taxpayer" (White House, Washington, DC, 2016).
8. National Research Council, *Hidden Costs of Energy: Unpriced Consequences of Energy Production and Use* (National Academies Press, Washington, DC, 2010).
9. Interagency Working Group on Social Cost of Greenhouse Gases, "Technical support document: Technical update of the social cost of carbon for regulatory impact analysis: Under Executive Order 12866" (White House, Washington, DC, 2016); http://bit.ly/2fkbBJn.
10. EPA, "Emission factors for greenhouse gas inventories" (EPA, Washington, DC, 2014); http://bit.ly/EPA2gdp70j.
11. M. Greenstone, D. Stuart, "Fossil fuel trends and the social cost of carbon" (Energy Policy Institute, Univ. of Chicago, Chicago, 2015); https://epic.uchicago.edu/technical-document.
12. M. Fowlie *et al.*, *Science* **346**, 815 (2014).
13. M. Haggerty, M. Lawson, J. Pearcy, "Steam coal at an arm's length: An evaluation of proposed reform options for U.S. coal used in power generation" (Headwaters Economics Working paper, Bozeman, MT, 2016).
14. I. Peterson, *Geo. J. L. Public Policy* **13** (1), 165 (2015).
15. A. Krupnick, J. Darmstadter, N. Richardson, K. McLaughlin, "Putting a carbon charge on federal coal: Legal and economic issues" (Discussion paper 15-13, Resources for the Future, Washington, DC, 2015).
16. EIA, "Annual energy outlook 2016" (Report 0383, EIA, Washington, DC, 2016).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1096/suppl/DC1

10.1126/science.aak9851

GRAPHIC: G. GRULLÓN/*SCIENCE*

PERSPECTIVES

Vampire bats (such as this common vampire bat, *Desmodus rotundus*) live in colonies in South and Central America.

EPIDEMIOLOGY

As the bat flies

Can virus transmission from bats to humans be predicted?

By **David T. S. Hayman**

The importance of predicting the timing and location of infectious disease emergence events from animal into human populations is highlighted by the effect of Ebola virus in West Africa. Such predictions are, however, usually hampered by a dearth of data. In a recent analysis of rabies viruses derived from vampire bats (see the photo) in Peru, Streicker *et al*. show that with sufficient data on both pathogen and host and with accurate models, predictions can be made to inform surveillance and public health efforts (*1*).

Bats are the hosts of numerous lethal zoonotic viruses, including Ebola and Marburg viruses, severe acute respiratory syndrome (SARS) coronavirus relatives, and rabies virus (*2*). Despite their importance, most bat-virus systems are studied unsystematically and too rarely through ecological frameworks. In the absence of rich data sets, sound theoretical frameworks allow the synthesis of empirical data from different fields through a range of modeling approaches. They thus help to elucidate data gaps and increase understanding of viral dynamics and emergence.

Bats are unique among mammals in that they fly. Flight, the often small size of bats, and their nocturnal behavior contribute to the difficulties researchers have in studying them. Yet flight has allowed the ~1300 recorded bat species to become widespread around the world, occupying all continents (except Antarctica) and remote landmasses such as New Zealand and Hawai'i. Bats' diets range from insects to fruit to fish to blood, and they can exist as solitary migrants or highly colonial animals. This extraordinary diversity appears to be matched by the diversity of viruses that bats host (*2*).

In their analysis, Streicker *et al*. used models that combine evolutionary (phylogenetic) relationships and geographic distributions of genetic lineages to estimate how fast different rabies virus lineages spread. Surveillance data of vampire bat–derived rabies among livestock helped to support the findings; gene flow among vampire colonies across the Andes from rabies-infected to rabies-free regions allowed prediction of viral spread. The authors then used landscape resistance models to project where rabies virus would spread and when it would arrive.

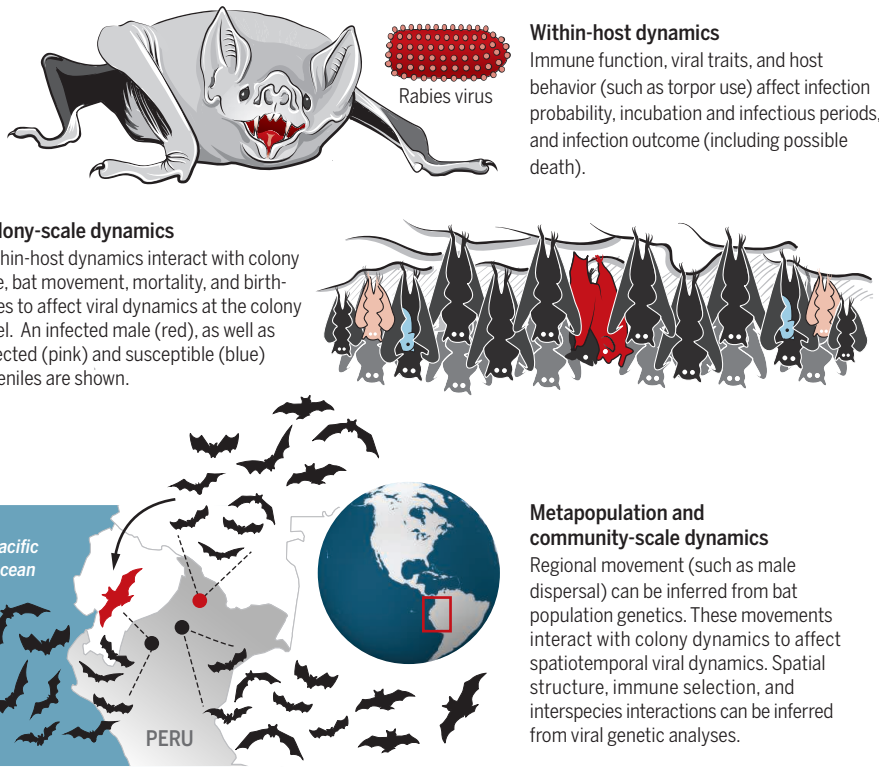
Streicker and colleagues have previously used similarly rich data sets to estimate rates of rabies virus spillover among bat species by focusing on species rather than geographic location (*3*). However, more limited data can suffice for reconstructing epidemiological processes through phylogenetic analyses. For example, Biek *et al*. were able to identify a distinct spatiotemperal pattern for Ebola virus outbreaks from analysis of just 13 gene fragments from Ebola virus outbreaks from 1976 to 2003; later analyses that included bat-derived Ebola virus gene fragments confirmed the genetic relatedness and temporal structure (*4*). Such analyses allow predictions of the spread of viruses in time and space based on just a few viral sequence fragments (see the figure).

Analysis of host relatedness can inform understanding of viral dynamics and spillover. Streicker *et al*. conclude that male bats are key to the dispersal in Peruvian vampire bat population dynamics and rabies virus transmission (*1*), and rabies virus is more frequently transmitted between related species than between unrelated ones (*3*). Yet genetic analyses of bat population structure can be used to make inductive arguments about broad-scale pathogen transmission in the absence of detailed virological data. For example, the lack of barriers to gene flow (panmixia) in sub-Saharan African straw-colored fruit bats might facilitate viral transmission (*5*) (see the figure). Comparative analyses suggest that

Molecular Epidemiology and Public Health Laboratory, Hopkirk Research Institute, Massey University, Private Bag 11-222, Palmerston North, New Zealand. Email: d.t.s.hayman@massey.ac.nz

Viral dynamics vary in space and time

A multiscale approach is needed to predict where and when humans or other animals are at risk of infection from viruses carried by bats. Data compiled from three levels—the individual, colony, and region—help to understand viral bat host dynamics and to guide predictions.



Within-host dynamics
Immune function, viral traits, and host behavior (such as torpor use) affect infection probability, incubation and infectious periods, and infection outcome (including possible death).

Colony-scale dynamics
Within-host dynamics interact with colony size, bat movement, mortality, and birth-rates to affect viral dynamics at the colony level. An infected male (red), as well as infected (pink) and susceptible (blue) juveniles are shown.

Metapopulation and community-scale dynamics
Regional movement (such as male dispersal) can be inferred from bat population genetics. These movements interact with colony dynamics to affect spatiotemporal viral dynamics. Spatial structure, immune selection, and interspecies interactions can be inferred from viral genetic analyses.

other bat species are also panmictic across the Congo Basin to Upper Guinea region, possibly facilitating Ebola virus transmission across large spatial areas (*6*).

It remains unclear whether bats respond to viral infections in different ways from other mammals (*7*, *8*). It is important to understand such differences, not least because scientists studying viruses often rely on antibody production through immune responses to help understand viral dynamics in the absence of virus detection (*9*, *10*) (see the figure). Two bat-rabies virus studies estimated the proportion of bats that become infectious and succumb to rabies virus exposure in nature. Both estimated similarly low proportions of exposures that led to disease, most infections were immunizing. Yet the supported mechanisms for rabies virus persistence within colonies differed. Blackwood *et al.* suggest that rabies virus cannot persist long-term in typically sized vampire bat colonies without viral reintroduction from other colonies (*9*). However, George *et al.* have reported different findings for North American big brown bats, which require overwinter hibernation to survive. Their model results suggest that prolonged seasonal bat torpor is necessary to stall the progression of rabies virus and to extend already long incubation periods so that fatal rabies virus infection can persist in colonies each year. The extended incubation time allows rabies virus to infect susceptible juvenile bats born after hibernation (*11*). These contrasting findings highlight differences in viral persistence mechanisms for the same virus in bat species that differ in their life histories.

For many other bat-virus systems, the mechanisms of inter- and intraspecies transmission and within-host dynamics, such as infectious period, remain unknown (*7*). Examples include coronaviruses, which include the precursors of pandemic SARS-coronavirus that are ubiquitous in bats (*2*). Drexler *et al.* reported detection of a coronavirus, six astroviruses, and an adenovirus in feces under a European Greater mouse–eared bat maternity colony (*12*). They found increased viral amplification of coronaviruses and astroviruses, but not adenoviruses, during colony formation and post-parturition. These observations suggest differing transmission dynamics at individual and population levels, driven by viral traits and breeding cycles (see the figure). Dynamic models that incorporate such cycles can be developed to formulate hypotheses. One example is the hypothesis that seasonal birthing might facilitate filovirus persistence in bats. Increasing availability of relevant data for Marburg virus in colonial, cave-dwelling Egyptian fruit bats allows this hypothesis to be tested (*7*, *10*, *13*).

Trait-based approaches can be used to gain insights about transmission mechanisms in the absence of temporal data. Transmission of rabies viruses among North American bat species is mostly limited to closely related bat species. Yet, rates of viral transmission from solitary migrants, hoary and eastern red bats, are high, suggesting that migratory species can facilitate viral transmission (*3*). Analyses of bat-virus community networks showed that gregarious bat species are important for viral sharing, whereas regional migrants are predicted to be viral spreaders (*14*). These studies suggest that there may be general life-history traits, such as being migratory, that facilitate viral sharing among bats (*1*, *5*, *14*). Trait-based analyses can also inform field studies, such as through identifying potentially undiscovered filovirus hosts (*15*).

Despite the enormous number of viruses discovered in bat populations, many open questions remain about within-host, population, metapopulation, and community dynamics and about spillover transmission (see the figure). Theoretical frameworks and phylodynamic, epidemiological, and ecological approaches will allow us to understand bat-virus systems and to model how perturbations such as land-use change and climate change will affect their infection dynamics. Challenging longitudinal field and laboratory studies will be essential. But if sufficient, appropriate data are collected, we will be able to predict the spatiotemporal dynamics of bat-virus systems and determine who is most at risk of infection and where and when. ■

REFERENCES AND NOTES

1. D. G. Streicker *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10926 (2016).
2. D. Hayman, *Annu. Rev. Virol.* **3**, 77 (2016).
3. D. G. Streicker *et al.*, *Science* **329**, 676 (2010).
4. R. Biek, P. D. Walsh, E. M. Leroy, L. A. Real, *PLOS Pathog.* **2**, e90 (2006).
5. A. J. Peel *et al.*, *Nat. Commun.* **4**, 2770 (2013).
6. A. Hassanin *et al.*, *C. R. Biol.* **339**, 517 (2016).
7. B. R. Amman *et al.*, *J. Wildl. Dis.* **51**, 113 (2015).
8. P. Zhou *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2696 (2016).
9. J. C. Blackwood, D. G. Streicker, S. Altizer, P. Rohani, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20837 (2013).
10. D. T. S. Hayman, *Proc. R. Soc. B* **282**, 20142591 (2015).
11. D. B. George *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10208 (2011).
12. J. F. Drexler *et al.*, *Emerg. Infect. Dis.* **17**, 449 (2011).
13. B. R. Amman *et al.*, *PLOS Pathog.* **8**, e1002877 (2012).
14. A. D. Luis *et al.*, *Ecol. Lett.* **18**, 1153 (2015).
15. B. A. Han *et al.*, *PLOS Negl. Trop. Dis.* **10**, e0004815 (2016).

ACKNOWLEDGMENTS

I acknowledge funding from the Royal Society of New Zealand Marsden Fund (MAU1503). I am grateful to N. French, D. Wilkinson, Matt Knox, T. O'Shea, and reviewers for useful comments.

10.1126/science.aaj1818

GRAPHIC: K. SUTLIFF/*SCIENCE*

CHEMICAL PHYSICS

Spying on the neighbors' pool

Spectral signatures are obtained for the movement of protons in cold water clusters

By **Sotiris S. Xantheas**

The structure and properties of the proton in water are of fundamental importance in many areas of chemistry and biology. The high mobility of the proton in an aqueous solution is understood in terms of its "hopping" between neighboring water molecules, as suggested by the two-century-old Grotthuss mechanism. The barrier for this process intimately depends on the proton's surrounding environment, which is manifested by the connectivity of the immediate hydrogen-bonding network as well as its dynamics caused by thermal fluctuations. On page 1131 of this issue, Wolke *et al.* (*1*) shed new light on the role that the proton's water neighbors play toward facilitating positive charge translocation within a hydrogen-bonded network in a cold water cluster.

Understanding the speciation and reactivity of the proton in an aqueous environment begins with acids and bases, which can transfer (either donate or accept) a proton, according to Brønsted and Lowry. This process was further explained by Lewis in terms of changes in acids' and bases' electronic structure in an attempt to offer a generalization of the Arrhenius theory. Simple proton transfers or the ones coupled to an electron transfer determine speciation, valence, and reactivity in aqueous media (*2*) and explain electrochemical processes (*3*), whereas voltage-gated proton channels play an essential role in the function of many cells (*4*).

The water environment plays a role in the molecular-level description of the proton, the two limiting cases being the Eigen-type $H_3O^+(H_2O)_m$ (*5*) and the Zundel-type $H_5O_2^+(H_2O)_n$ (*6*) cations in water clusters of varying size (see the figure). Infrared (IR) vibrational spectroscopy is a powerful experimental tool for identifying the spectral signatures associated with the underlying water network structure. With the aid of theoretical calculations, spectral bands can be decoded and assigned to the causal molecular vibrations.

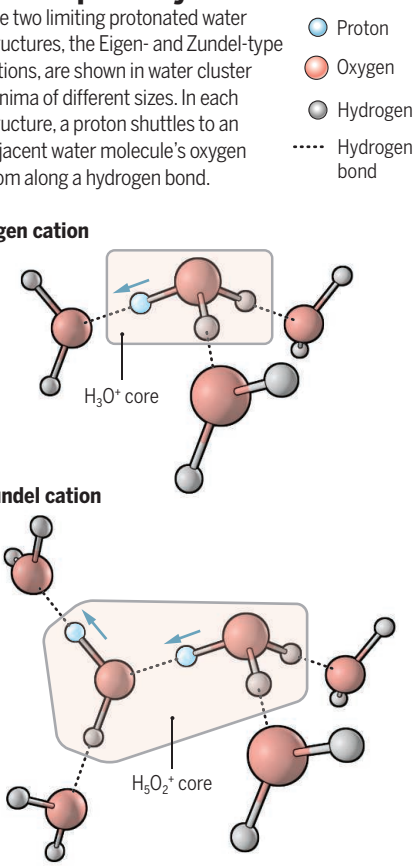
The challenge is that these bands are often quite broad in condensed-phase environments at room temperature and smear out the fundamental vibrations occurring at the molecular level. By selectively tagging the proton's water neighbors, placed at known distinct positions within a cluster, Wolke *et al.* showed that they could isolate each neighbor's different response to the positive charge (*7*). Isotopic substitution with deuterium in neighboring water molecules enables identification of spectral patterns arising from their interaction with the proton. Thus, how each neighbor was altered by its local environment could be accounted for quantitatively. The present study shows that it is now possible to identify the spectroscopic signatures along the proton transfer pathway and quantify the correlation between the hydrogen-bonded OH stretching frequency and its surrounding environment in a cold aqueous cluster.

This study represents an important step toward understanding the proton's structural motifs and associated hopping process in aqueous cluster networks and ultimately in aqueous solution at room temperature. The current "bottom-up" (cluster) approach is complementary to recent "top-down" (solution) experimental ultrafast two-dimensional IR (2D-IR) spectroscopic measurements used to probe the spectral correlations between the stretching and bending vibrations in the constituent water molecules of the Zundel cation in a concentrated (4M) aqueous hydrochloric acid solution (*8*). The analysis of the 2D-IR spectra obtained during that study suggested an unexpected large concentration of Zundel-type cations, further reinforcing their role in the proton transfer mechanism.

Theory can help bridge the gap between these "bottom-up" and "top-down" approaches that aim at understanding proton speciation and dynamical properties in aqueous environments of varying size, composition, and external conditions (e.g., temperature and pressure). The interpretation of the measured spectral features can be enhanced by theory, even if existing theoretical approaches are currently challenged (*9*) when called on to accurately describe the vibrations of even the fundamental units of those cations in cold aqueous clusters (*10*). However, more approximate methods that are currently available to treat the collective motions in extended systems [such as density functional theory models, multistate valence bond (*11*) models, or both] cannot yet offer a first principles–based approach to the problem of accurately describing both network structure, its fluctuations, and the corresponding spectral signatures. New theoretical methodologies are needed that accurately account for the network's collective interactions and fluctuations, as well as approaches for decoding the spectral patterns associated with the underlying molecular motions in liquids. ■

Proton pathways

The two limiting protonated water structures, the Eigen- and Zundel-type cations, are shown in water cluster minima of different sizes. In each structure, a proton shuttles to an adjacent water molecule's oxygen atom along a hydrogen bond.



Physical Sciences Division, Pacific Northwest National Laboratory, 902 Battelle Boulevard, MS K1-83, Richland, WA 99352, USA. Email: sotiris.xantheas@pnnl.gov

GRAPHIC: C. BICKEL/*SCIENCE*

REFERENCES AND NOTES

1. C. T. Wolke *et al.*, *Science* **354**, 1131 (2016).
2. J. M. Mayer, *Annu. Rev. Phys. Chem.* **55**, 363 (2004).
3. C. Costentin, M. Robert, J.-M. Savéant, *Acc. Chem. Res.* **43**, 1019 (2010).
4. T. E. DeCoursey, *Physiol. Rev.* **93**, 599 (2013).
5. M. Eigen, *Angew. Chem. Int. Ed.* **3**, 1 (1964).
6. M. L. Huggins, *J. Phys. Chem.* **40**, 723 (1936).
7. C. T. Wolke *et al.*, *J. Chem. Phys.* **144**, 074305 (2016).
8. M. Thämer, L. De Marco, K. Ramasesha, A. Mandal, A. Tokmakoff, *Science* **350**, 78 (2015).
9. S. S. Xantheas, *Nature* **457**, 673 (2009).
10. O. Vendrell, F. Gatti, H.-D. Meyer, *Angew. Chem. Int. Ed.* **48**, 352 (2009).
11. Y. Wu, H. Chen, F. Wang, F. Paesani, G. A. Voth, *J. Phys. Chem. B* **112**, 467 (2008).

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences and Biosciences. Pacific Northwest National Laboratory (PNNL) is a multiprogram national laboratory operated for DOE by Battelle.

10.1126/science.aal1413

ASTRONOMY

Galaxy formation through cosmic recycling

Direct observations reveal how massive galaxies formed in the early universe

By **Nina Hatch**

Extremely massive galaxies are seen in the young universe, but their presence is puzzling because we do not yet understand how they became so massive so quickly. How do they get enough fuel to form stars so rapidly? The raw fuel for forming stars is cold molecular gas, and although this gas is common within young galaxies (*1*), we do not know how it is replenished once the first reservoirs are converted into stars. On page 1128 of this issue, Emonts *et al.* (*2*) report observations that may provide our first clue to this fueling problem. They have detected a giant reservoir of recycled molecular gas that is replenishing the fuel supply of one of the most massive galaxies in the young universe.

School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK. Email: nina.hatch@nottingham.ac.uk

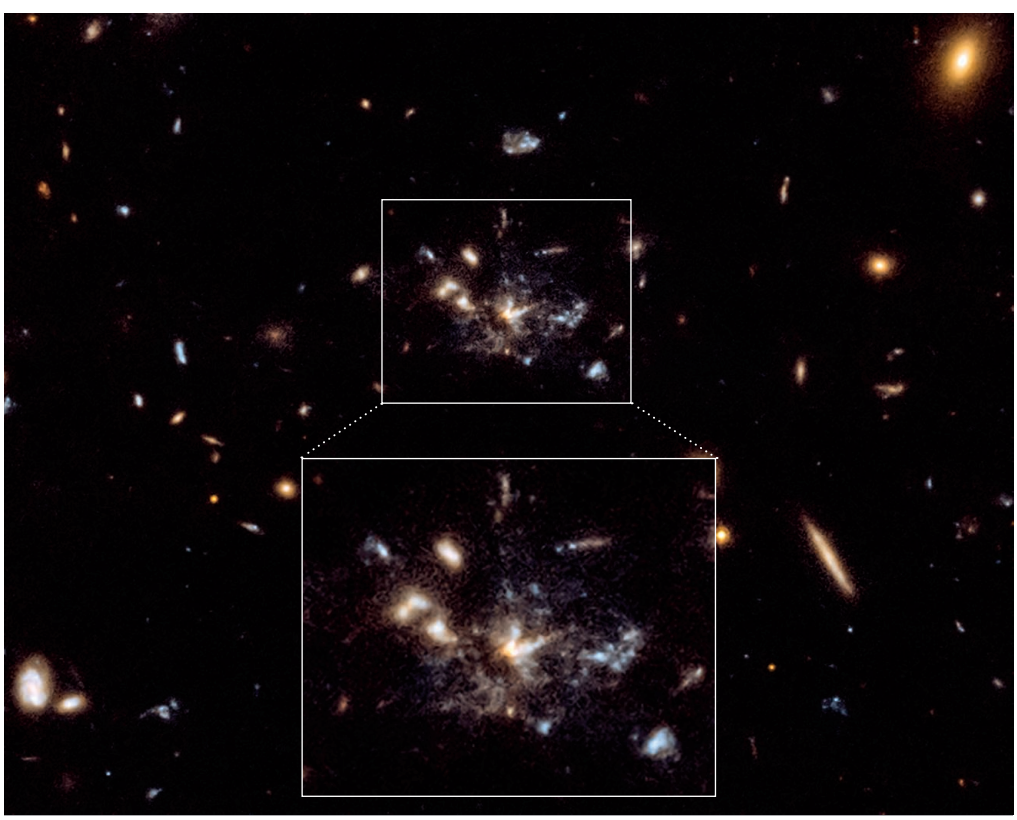
Massive galaxies form through a short period of rapid stellar production, followed by a prolonged period when they obtain stars by cannibalizing numerous smaller galaxies (*3*). How much of the galaxy is produced in the initial phase of star condensation versus the amount of growth through galaxy mergers is debated, with tension between recent theories and observations. Observations have shown that massive galaxies have only slightly increased in mass over the past 9 billion years (*4*), arguing that they form primarily through rapid star formation during the first few billion years. However, simulations of galaxy formation have trouble replicating this early period of rapid star formation (*5*).

The gap in our understanding is how to feed the galaxies with enough raw material to form. A constant supply of cold gas is required to fuel the rapid formation of stars, but this is difficult to achieve in simulations. Cold primordial gas is prevented from simply streaming through to fuel star formation in massive galaxies. When gas falls toward the galaxies, it is heated by shocks to a few million degrees, forming hot halos that may take billions of years to cool sufficiently to form stars.

Several theories have been postulated to get around this problem, including cooler streams of primordial gas being able to punch through the hot halo surrounding the galaxy (*6*); precipitation of gas from the surrounding hot halo (*7*); or recycling of old stellar gas that is driven out of the galaxy by winds from massive stars (*8*). Progress will require observations that can help explain whether these processes occur.

To observe how massive galaxies accrete cold gas, Emonts *et al.* turned their attention to the Spiderweb galaxy. This is one of the most massive galaxies in the universe, even though it is so far away that we observe it when the universe was only 3 billion years old. Its morphology resembles that of a spider's web, with a mature galaxy at the center (*9*), surrounded by a 100-kpc web of diffuse ultraviolet light, and tens of smaller galaxies that resemble flies caught in the web (*10*). It has all the hallmarks of a galaxy caught in the act of forming: multiple small merging galaxies and hundreds of new stars being created per year (see the photo).

To locate the reservoir of molecular gas that feeds this galaxy, Emonts *et al.* use two independent interferometric radio observations, taken at different spatial resolutions. Collecting data for 90 hours with a compact array creates a very sensitive image, but with low spatial resolution. So they couple this with a further 8 hours of observations in a wide telescope configuration, which creates a high–spatial resolution image. Taken together, these observations show that only a third of the total molecular gas emission comes from the central galaxy. Their surprising conclusion is that most of the cold gas is located



Recycling and growth. The massive Spiderweb galaxy is caught in the act of forming when the universe was only 3 billion years old. Most of the cold molecular gas that Emonts *et al.* detect with radio telescopes lies between the galaxies in the inset, but it is invisible in this Hubble Space Telescope image. This recycled gas can provide the central galaxy with enough fuel to form the next generation of stars.

PHOTO: NASA, ESA, GEORGE MILEY, AND RODERIK OVERZIER/LEIDEN OBSERVATORY/NETHERLANDS

between the galaxies in a vast reservoir that extends across 70 kpc. Stars are condensing out of this vast reservoir at a rate of at least hundreds per year (*11*), which will fall onto the galaxy to grow its outer layers.

Some of the gas within this reservoir must be recycled material that was once situated within stars. This is because Emonts *et al.* detect the carbon monoxide molecule. Whereas hydrogen was formed a few minutes after the Big Bang throughout the universe, the heavier nuclei of carbon and oxygen are only formed in vast quantities within the inner regions of stars. Because the gas reservoir contains carbon and oxygen molecules, at least some fraction of it must have once been inside earlier generations of stars.

The presence of heavy nuclei in the reservoir has important consequences. Gas expelled from stars is initially very hot and joins the hot halo in an ionized state. The cooling rate of gas is much faster when elements heavier than hydrogen are present, so the recycled gas ensures that cold material for stellar production can more readily precipitate out of the surrounding hot halo to form the cold molecular reservoir.

The detection of this vast cloud of recycled molecular gas may help solve the problem of how massive galaxies grow, but it also

"To observe how massive galaxies accrete cold gas, Emonts* et al. *turned their attention to the Spiderweb galaxy."

throws up new questions. What expels the recycled gas out of the galaxies? How does this gas cool down to form molecules? And how widespread is this mode of galaxy growth? The Spiderweb galaxy is just one galaxy in the young universe, and further recycled reservoirs will need to be observed to determine whether this is a common route by which massive galaxies grow. ■

REFERENCES

1. E. Daddi *et al.*, *Astrophys. J.* **673**, 21 (2008).
2. B. H. C. Emonts *et al.*, *Science* **354**, 1128 (2016).
3. L. Oser *et al.*, *Astrophys. J.* **725**, 2312 (2010).
4. C. A. Collins *et al.*, *Nature* **458**, 603 (2009).
5. B. De Lucia, J. Blaizot, *Mon. Not. R. Astron. Soc.* **375**, 2 (2007).
6. A. Dekel *et al.*, *Nature* **457**, 451 (2009).
7. A. R. Fabian, *Astron. Astrophys.* **32**, 277 (1994).
8. D. Narayanan *et al.*, *Nature* **525**, 496 (2015).
9. L. Pentericci *et al.*, *Astron. Astrophys.* **326**, 580 (1997).
10. G. K. Miley *et al.*, *Astrophys. J.* **650**, 29 (2006).
11. N. A. Hatch *et al.*, *Mon. Not. R. Astron. Soc.* **383**, 931 (2008).

10.1126/science.aah6255

STEM CELLS

Metabolic cues for hematopoietic stem cells

Manipulating mitophagy and dietary valine may lead to stem cell therapies

By **Pia Sommerkamp**[1] *and* **Andreas Trumpp**[1,2]

Hematopoietic stem cells (HSCs) are at the helm of the hierarchically organized hematopoietic system that ensures the lifelong production of all blood cells. HSCs depend on metabolic cues to secure their protective quiescent status and to enable rapid activation and replenishment of the blood system in response to stressful situations such as infections, excessive bleeding, or chemotherapy-induced myeloablation (*1*–*3*). On pages 1156 and 1152 of this issue, Ito *et al.* (*4*) and Taya *et al.* (*5*), respectively, uncover important roles for the degradation of defective mitochondria (mitophagy) and the amino acid valine in HSC maintenance and function.

Phenotypic HSCs isolated by flow cytometry from mouse bone marrow represent a rather heterogenous cell population (*6*, *7*). Ito *et al.* genetically engineered mice to express green fluorescent protein under the control of regulatory elements that govern the expression of the *Tie2* gene. TIE2 is a receptor tyrosine kinase that responds to the growth factor angiopoietin-1, and the angiopoietin-1–TIE2 pathway is important for HSC maintenance (*8*). Remarkably, 68% of TIE2-expressing (TIE2$^+$) HSCs showed reconstitution capacity after transplantation as individual cells. These HSCs are located near arteriolar structures in the bone marrow, rather than at sinusoids (*9*). They also exhibit enhanced cell cycle quiescence, sit atop the hematopoietic cell hierarchy, and preferentially undergo symmetric stem cell division.

To identify the gene networks that regulate TIE2$^+$ HSCs, Ito *et al.* performed single-cell gene expression analysis combined with genetic deletion studies. They found increased expression of genes encoding the transcription factors promyelocytic leukemia (PML) and peroxisome proliferator–activated receptor-delta (PPARδ), which regulate fatty acid oxidation (FAO). This links to earlier data showing that the PML-PPARδ-FAO axis controls HSC maintenance (*10*). Ito *et al.* further extend this pathway to PPARδ-controlled mitophagy, which is mediated by the E3 ubiquitin ligase subunit PARKIN and the PTEN-induced putative protein kinase 1 (PINK1) (*4*). Mitophagy enables symmetric cell division and replenishment of the hematopoietic system even under stress conditions. This metabolic cue—minimizing oxidative stress by increasing mitophagy—is thus characteristic of highly potent TIE2$^+$ HSCs and is also conserved in human HSCs. The authors further show that targeting the PPARδ-FAO-mitophagy axis with PPARδ agonists enabled the expansion of TIE2$^+$ HSCs in culture and in mice, suggesting potential clinical applications (see the figure).

During homeostasis, HSCs are maintained in a protective dormant state. However, during normal blood production and especially in response to physiological stress, HSCs must produce progeny while maintaining their functional and genomic integrity to prevent malignant transformation (*1*). The findings of Ito *et al.* suggest that cycling HSCs can eliminate damaged mitochondria and thus limit the production of mutagenic reactive oxygen species. Indeed, limited mitochondrial biogenesis has been cited as crucial for HSC maintenance (*11*). Overall, these data suggest a model in which HSCs can switch to various metabolic programs depending on the physiological needs. By targeting these metabolic cues, cell fate manipulation and HSC expansion might be feasible.

The metabolic characteristics of HSCs have recently attracted attention (*2*), but little is known about the role of single amino acids for HSC function. Taya *et al.* show that bone marrow contains 100-fold higher concentrations of all 20 amino acids compared to peripheral blood. To test the requirement of each amino acid for in vitro and in vivo HSC maintenance, they used single amino acid–depleted culture media or mouse chow. These analyses revealed that mouse HSC proliferation and maintenance depends on the branched-chain amino acid valine. It is secreted by vascular endothelial stromal cells, suggesting that valine may be a critical component of the HSC niche. Indeed, after 4

[1]*Division of Stem Cells and Cancer, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.* [2]*Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM gGmbH), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. Email: a.trumpp@dkfz.de*

weeks of a valine-restricted diet, long-term repopulating HSCs from the mouse bone marrow are lost. After only 2 weeks of valine starvation, a sufficient number of HSC niches become vacant and allow chemoirradiation-free engraftment of transplanted donor-HSCs in congenic mice.

Because myeloablative chemoirradiation has severe side effects (infertility, poor overall health, premature aging), there is a medical need for alternative stem cell transplantation approaches. Along these lines, the short-term dietary valine restriction is comparatively well tolerated and, besides the desired hypocellularity in hematopoietic tissues, results only in decreased numbers of hair follicles and increased brown fat tissue in mice. Importantly, and in contrast to chemoirradiation, valine-restricted mice remained fertile, showed good overall health, and exhibited no obvious signs of premature aging. However, fast refeeding with a valine-containing diet caused the death of about 50% of mice owing to the development of a typical refeeding syndrome (*12*). These complications could be avoided by applying a gradual change in the valine-restricted diet. Nonetheless, concerns about the safety of this strategy remain. Previously, it was shown that reduced expression of the KIT receptor tyrosine kinase (whose ligand is stem cell factor) was also sufficient to allow stem cell engraftment (*13*, *14*). It would thus be interesting to examine whether the effects of valine depletion on HSCs are mediated by decreasing KIT expression.

Taya *et al.* also observed that human HSCs cultured in vitro or analyzed in xenografts depended on valine and, in contrast to mice, also on leucine. Valine and leucine are branched-chain amino acids that, when metabolized, influence α-ketoglutarate (aKG) concentrations. aKG controls embryonic stem cell pluripotency by acting as a substrate for DNA- and histone-demethylases that control the cell's epigenetic state (*15*). More studies are needed to explore such a mechanism and the molecular processes by which valine and leucine control HSC maintenance. This should help to unravel how our daily diet controls stem cell function and thus our regenerative organ systems. ■

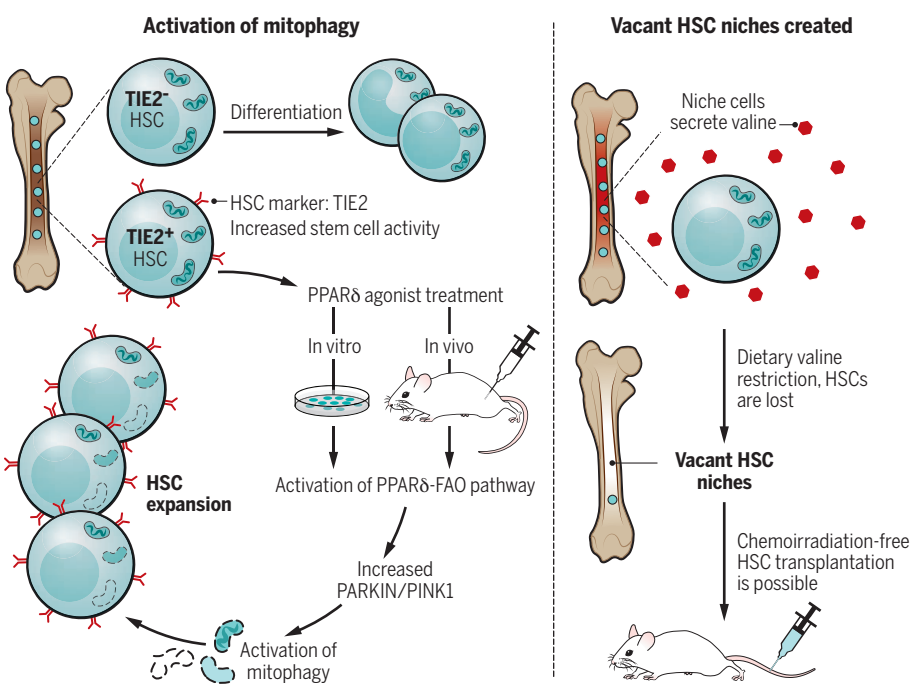
REFERENCES

1. D. Walter *et al.*, *Nature* **520**, 549 (2015).
2. K. Ito, T. Suda, *Nat. Rev. Mol. Cell Biol.* **15**, 243 (2014).
3. A. Trumpp *et al.*, *Nat. Rev. Immunol.* **10**, 201 (2010).
4. K. Ito *et al.*, *Science* **354**, 1156 (2016).
5. Y. Taya *et al.*, *Science* **354**, 1152 (2016).
6. A. Wilson *et al.*, *Cell* **135**, 1118 (2008).
7. R. Yamamoto *et al.*, *Cell* **154**, 1112 (2013).
8. F. Arai *et al.*, *Cell* **118**, 149 (2004).
9. J. Y. Chen *et al.*, *Nature* **530**, 223 (2016).
10. K. Ito *et al.*, *Nat. Med.* **18**, 1350 (2012).
11. C. Mantel *et al.*, *Cell Cycle* **9**, 2008 (2010).
12. H. M. Mehanna *et al.*, *BMJ* **336**, 1495 (2008).
13. A. Chhabra *et al.*, *Sci. Transl. Med.* **8**, 351ra105 (2016).
14. K. N. Cosgun *et al.*, *Cell Stem Cell* **15**, 227 (2014).
15. B. W. Carey *et al.*, *Nature* **518**, 413 (2015).

10.1126/science.aal3466

Metabolic characteristics of HSCs

Activation of mitophagy through a pathway that activates fatty acid oxidation (FAO) promotes TIE2⁺ HSC expansion by enhanced symmetric divisions. Valine is required for HSC maintenance, and lack of dietary valine empties bone marrow HSC niches, enabling chemoirradiation-free donor-HSC engraftment.



GRAPHIC: K. SUTLIFF/*SCIENCE*

IMMUNOLOGY

Can T cells be too exhausted to fight back?

The transcriptional network becomes less malleable in persistently activated T cells

By **Stephen J. Turner** *and* **Brendan E. Russ**

When T cells are persistently activated by antigen, such as during chronic infection or in cancer, they can become functionally incapable of performing their effector activities, a condition called T cell exhaustion. Exhaustion therefore thwarts optimal immune control of infection and tumors. There is a need to learn more about the molecular factors that drive T cell exhaustion and just how malleable T cell immunity is once exhaustion is established. On pages 1165 and 1160 of this issue, Sen *et al.* (*1*) and Pauken *et al.* (*2*), respectively, demonstrate that T cell exhaustion represents a stable differentiation state, underpinned by the apparently irreversible installation of an exhaustion-specific genetic landscape. This implies that perhaps in a majority of cases of persistent immune activation, T cells are too exhausted to fight back against cancer or pathogens.

Cancer immunotherapy aims to commandeer the T cell arm of our immune system to detect and remove tumors (*3*, *4*). This stems from observations that during an immune response, the immune system limits collateral damage to host tissues by engaging inhibitory pathways that suppress T cell function (*5*, *6*). Unfortunately, many tumors engage these same inhibitory checkpoints to evade antitumor immunity (*7*). Given that several of these immune checkpoints are mediated by receptor–ligand signaling, they have proven amenable to administration of antibodies that interrupt ligand binding (called "checkpoint blockade"), allowing T cell reactivation and tumor targeting (*3*, *4*). Indeed, checkpoint blockade has generated excitement owing to some spectacular clinical results, including complete tumor regression and remission in patients with

Department of Microbiology, Biomedical Discovery Institute, Monash University, Australia.
Email: stephen.j.turner@monash.edu

weeks of a valine-restricted diet, long-term repopulating HSCs from the mouse bone marrow are lost. After only 2 weeks of valine starvation, a sufficient number of HSC niches become vacant and allow chemoirradiation-free engraftment of transplanted donor-HSCs in congenic mice.

Because myeloablative chemoirradiation has severe side effects (infertility, poor overall health, premature aging), there is a medical need for alternative stem cell transplantation approaches. Along these lines, the short-term dietary valine restriction is comparatively well tolerated and, besides the desired hypocellularity in hematopoietic tissues, results only in decreased numbers of hair follicles and increased brown fat tissue in mice. Importantly, and in contrast to chemoirradiation, valine-restricted mice remained fertile, showed good overall health, and exhibited no obvious signs of premature aging. However, fast refeeding with a valine-containing diet caused the death of about 50% of mice owing to the development of a typical refeeding syndrome (*12*). These complications could be avoided by applying a gradual change in the valine-restricted diet. Nonetheless, concerns about the safety of this strategy remain. Previously, it was shown that reduced expression of the KIT receptor tyrosine kinase (whose ligand is stem cell factor) was also sufficient to allow stem cell engraftment (*13*, *14*). It would thus be interesting to examine whether the effects of valine depletion on HSCs are mediated by decreasing KIT expression.

Taya *et al.* also observed that human HSCs cultured in vitro or analyzed in xenografts depended on valine and, in contrast to mice, also on leucine. Valine and leucine are branched-chain amino acids that, when metabolized, influence α-ketoglutarate (aKG) concentrations. aKG controls embryonic stem cell pluripotency by acting as a substrate for DNA- and histone-demethylases that control the cell's epigenetic state (*15*). More studies are needed to explore such a mechanism and the molecular processes by which valine and leucine control HSC maintenance. This should help to unravel how our daily diet controls stem cell function and thus our regenerative organ systems. ■

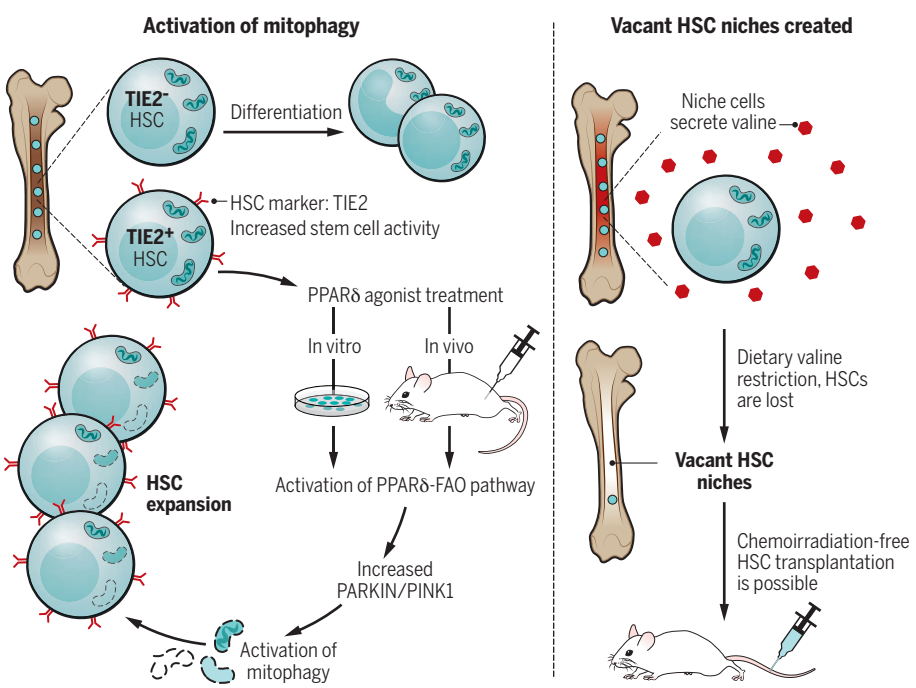
REFERENCES

1. D. Walter *et al.*, *Nature* **520**, 549 (2015).
2. K. Ito, T. Suda, *Nat. Rev. Mol. Cell Biol.* **15**, 243 (2014).
3. A. Trumpp *et al.*, *Nat. Rev. Immunol.* **10**, 201 (2010).
4. K. Ito *et al.*, *Science* **354**, 1156 (2016).
5. Y. Taya *et al.*, *Science* **354**, 1152 (2016).
6. A. Wilson *et al.*, *Cell* **135**, 1118 (2008).
7. R. Yamamoto *et al.*, *Cell* **154**, 1112 (2013).
8. F. Arai *et al.*, *Cell* **118**, 149 (2004).
9. J. Y. Chen *et al.*, *Nature* **530**, 223 (2016).
10. K. Ito *et al.*, *Nat. Med.* **18**, 1350 (2012).
11. C. Mantel *et al.*, *Cell Cycle* **9**, 2008 (2010).
12. H. M. Mehanna *et al.*, *BMJ* **336**, 1495 (2008).
13. A. Chhabra *et al.*, *Sci. Transl. Med.* **8**, 351ra105 (2016).
14. K. N. Cosgun *et al.*, *Cell Stem Cell* **15**, 227 (2014).
15. B. W. Carey *et al.*, *Nature* **518**, 413 (2015).

10.1126/science.aal3466

Metabolic characteristics of HSCs

Activation of mitophagy through a pathway that activates fatty acid oxidation (FAO) promotes TIE2⁺ HSC expansion by enhanced symmetric divisions. Valine is required for HSC maintenance, and lack of dietary valine empties bone marrow HSC niches, enabling chemoirradiation-free donor-HSC engraftment.



GRAPHIC: K. SUTLIFF/*SCIENCE*

IMMUNOLOGY

Can T cells be too exhausted to fight back?

The transcriptional network becomes less malleable in persistently activated T cells

By **Stephen J. Turner** *and* **Brendan E. Russ**

When T cells are persistently activated by antigen, such as during chronic infection or in cancer, they can become functionally incapable of performing their effector activities, a condition called T cell exhaustion. Exhaustion therefore thwarts optimal immune control of infection and tumors. There is a need to learn more about the molecular factors that drive T cell exhaustion and just how malleable T cell immunity is once exhaustion is established. On pages 1165 and 1160 of this issue, Sen *et al.* (*1*) and Pauken *et al.* (*2*), respectively, demonstrate that T cell exhaustion represents a stable differentiation state, underpinned by the apparently irreversible installation of an exhaustion-specific genetic landscape. This implies that perhaps in a majority of cases of persistent immune activation, T cells are too exhausted to fight back against cancer or pathogens.

Cancer immunotherapy aims to commandeer the T cell arm of our immune system to detect and remove tumors (*3*, *4*). This stems from observations that during an immune response, the immune system limits collateral damage to host tissues by engaging inhibitory pathways that suppress T cell function (*5*, *6*). Unfortunately, many tumors engage these same inhibitory checkpoints to evade antitumor immunity (*7*). Given that several of these immune checkpoints are mediated by receptor–ligand signaling, they have proven amenable to administration of antibodies that interrupt ligand binding (called "checkpoint blockade"), allowing T cell reactivation and tumor targeting (*3*, *4*). Indeed, checkpoint blockade has generated excitement owing to some spectacular clinical results, including complete tumor regression and remission in patients with

Department of Microbiology, Biomedical Discovery Institute, Monash University, Australia.
Email: stephen.j.turner@monash.edu

far-progressed cancers that were refractory to standard treatments (*8*–*10*). Unfortunately, checkpoint blockade fails in more instances than it works, for reasons that are as yet unclear.

One such checkpoint, programmed death–1 (PD-1), is expressed on activated T cells, and when engaged by its ligand, programmed death ligand 1 (PD-L1), limits T cell effector function (*5*). Interestingly, persistent virus infections and some cancers induce strong and stable expression of PD-1 on activated T cells, eventually resulting in permanent loss of T cell function and T cell exhaustion (*11*, *12*). Exhausted T cells are distinct from bona fide memory T cells, which are quiescent like exhausted cells but retain the capacity to be reactivated after virus or tumor reemergence (*12*). Earlier studies demonstrated that administration of monoclonal antibodies that interfere with PD-1 signaling restores exhausted T cell function in the context of chronic viral infection (*11*). Building on this finding, checkpoint blockade therapy targeting PD-1 has been curative for some cancer patients (*8*, *10*). However, a majority of patients do not respond to PD-1 checkpoint blockade, with many patients demonstrating a transient restoration of T cell function followed by disease relapse (*10*).

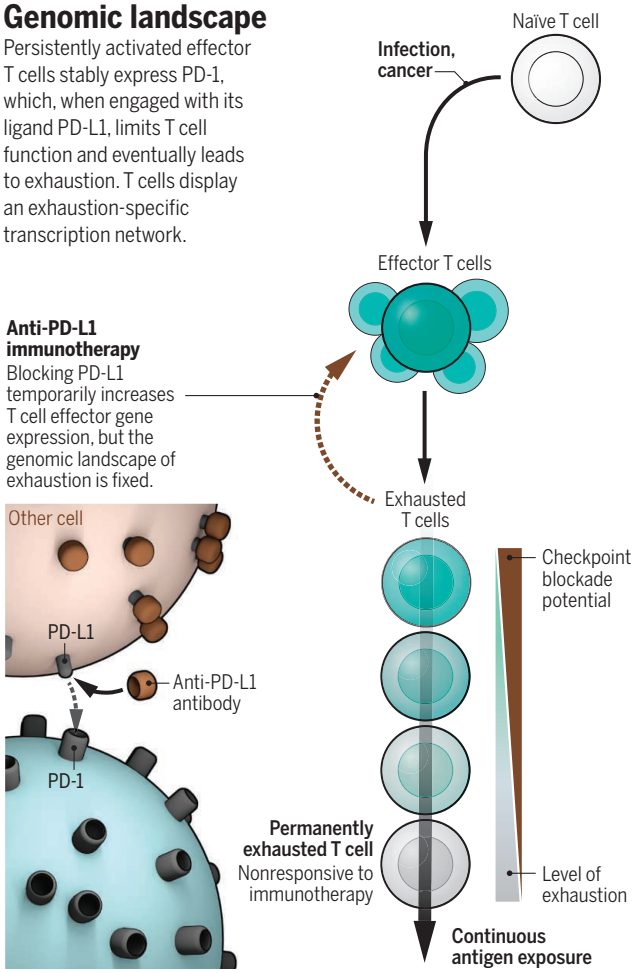
Sen *et al.* assessed the genomic characteristics of exhausted T cells induced by persistent lymphocytic choriomeningitis virus (LCMV) infection of mice. The authors used a technique called assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) to identify genomic regions that were open, or accessible, to transcription factors. These open regions [referred to by the authors as chromatin-accessible regions (ChARs)] are of interest because they contain gene regulatory elements (enhancers) that are bound by transcription factors to regulate cell type–specific gene expression (*13*). A comparison of the genomic landscape between nonexhausted and exhausted T cells demonstrated clear distinctions in the ChAR landscapes with exhausted T cells, as defined by the appearance of a unique set of accessible regulatory regions. Notably, many of these ChARs were adjacent to exhaustion-specific genes, including a ChAR that is linked to the PD-1 gene (encoded by *Pdcd1* in the mouse and by *PCDHAC1* in humans), and Sen *et al.* show that this ChAR mediates sustained PD-1 expression in exhausted T cells and that genetic deletion of this ChAR dramatically reduced PD-1 expression. Importantly, ~80% of ChARs identified as exhaustion-specific in mice had orthologs in the human genome that were engaged in HIV-specific (exhausted) human T cells, but not within nonexhausted T cells from the same donors. Taken together, these data indicate that rather than representing a state of dysfunction, T cell exhaustion results from a bona fide and deliberate differentiation program, underscored by exhaustion-specific transcriptional networks.

Pauken *et al.* explored how malleable the ChAR landscape is in exhausted T cells after blocking PD-L1. Using the same mouse model of LCMV infection, the authors found that treatment with an antibody against PD-L1 caused a temporary increase in the expression of T cell effector genes. However, withdrawal of treatment caused T cells to quickly revert back to an exhausted state, with the infection rebounding. This observation is reminiscent of the relapse seen in some cancer patients after checkpoint blockade treatment. Pauken *et al.* showed that the ChAR profile after checkpoint blockade was almost indistinguishable from that of exhausted cells. Hence, although checkpoint blockade temporarily restores T cell effector function, the underlying genetic landscape of exhausted cells is fixed and not subject to change.

The studies of Sen *et al.* and Pauken *et al.* suggest that irreversible installation of an exhaustion-specific genetic landscape is a hallmark of T cell exhaustion. This implies that perhaps in a majority of cases, T cells are too exhausted to fight back against cancer cells (see the figure). However, although these observations explain why checkpoint blockade fails, what about patients that are cured by checkpoint blockade therapy? Are their T cells in the early stages of exhaustion, and if so, can exhaustion be halted or reversed if caught early? Another question is whether the results of both studies are relevant to other checkpoint blockade strategies, such as those targeting the checkpoint protein cytotoxic T lymphocyte–associated antigen 4 (CTLA4), and in other infection or tumor settings. The answers to these questions have important implications for advancing checkpoint blockade as a cancer immunotherapy. Understanding the mechanisms that install and maintain the fixed genetic landscape of exhausted T cells may enable their reprogramming through the use of drugs targeting those proteins that modify the genetic landscape (*14*). For example, histone deacetylase inhibitors might limit the extent of T cell exhaustion (*15*). If the observations of Sen *et al.* and Pauken *et al.* are generalizable, then understanding how other strategies (such as the use of genetically engineered T cells) can combine with checkpoint blockade may lead to durable and more broadly effective therapies. ■

Genomic landscape

Persistently activated effector T cells stably express PD-1, which, when engaged with its ligand PD-L1, limits T cell function and eventually leads to exhaustion. T cells display an exhaustion-specific transcription network.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES

1. D. R. Sen *et al.*, *Science* **354**, 1165 (2016).
2. K. E. Pauken *et al.*, *Science* **354**, 1160 (2016).
3. D. M. Pardoll, *Nat. Rev. Cancer* **12**, 252 (2012).
4. D. S. Shin, A. Ribas, *Curr. Opin. Immunol.* **33**, 23 (2015).
5. H. Nishimura *et al.*, *Immunity* **11**, 141 (1999).
6. E. A. Tivol *et al.*, *Immunity* **3**, 541 (1995).
7. H. Dong *et al.*, *Nat. Med.* **8**, 1039 (2002).
8. O. Hamid *et al.*, *N. Engl. J. Med.* **369**, 134 (2013).
9. F. S. Hodi *et al.*, *N. Engl. J. Med.* **363**, 711 (2010).
10. A. Ribas *et al.*, *JAMA* **315**, 1600 (2016).
11. D. L. Barber *et al.*, *Nature* **439**, 682 (2006).
12. E. J. Wherry *et al.*, *Immunity* **27**, 670 (2007).
13. J. D. Buenrostro *et al.*, *Nat. Methods* **10**, 1213 (2013).
14. H. E. Ghoneim *et al.*, *Trends Mol. Med.* 10.1016/j.molmed.2016.10.002 (2016).
15. F. Zhang *et al.*, *Mol. Ther.* **22**, 1698 (2014).

10.1126/science.aal3204

PROTEIN SYNTHESIS

When stop makes sense

RNA sequence context matters in the termination of protein translation

By **Boris Zinshteyn** *and* **Rachel Green**

In the age of computational biology, it is easy to envision the genetic code as a set of immutable instructions that the cell follows without exception. The recent discovery of ciliate (*1*, *2*) and trypanosomatid (*3*) species in which all three stop codons, which normally act to terminate protein translation by the ribosome, encode amino acids instead is a reminder that decoding the information in messenger RNA (mRNA) depends on molecular factors that we do not entirely understand. In these organisms, stop codons specify amino acids by default, and termination of mRNA translation only occurs in close proximity to the polyadenylate [poly(A)] tail. How do these species differentiate "true" stop codons from identical ones that encode amino acids? The answers to this puzzle may provide insights into translation termination and gene regulation in all eukaryotes.

Since the discovery of the "standard" genetic code, small deviations have been discovered in various organisms and organelles (*4*). Ciliated protists have a particular propensity for reassigning one or two of the three standard stop codons (UAA, UAG, and UGA) to encode amino acids. Recent transcriptome sequencing (all the mRNA molecules expressed) revealed that in two ciliates, *Condylostoma magnum* and the unclassified *Parduzcia* sp., all three stop codons encode amino acids as well as signal translation termination (*1*, *2*). This is an altogether previously unknown paradigm containing a potentially disastrous ambiguity. The genomes of these species encode cognate transfer RNAs (tRNAs), which are complementary to the stop codons, and the balance between recruiting a tRNA to incorporate an amino acid or a release factor to terminate translation is resolved by the surrounding context of each particular stop codon. The sequence structure of ciliate transcripts provides a clue as to how this balance is accomplished. The unusually short 3′ untranslated regions (UTRs) between the "true" stop codon and the poly(A) tail (*1*, *2*) of these ciliates suggest a model in which cognate tRNAs decode stop codons by default, but in close proximity to a poly(A) tail, release factors outcompete tRNAs (see the figure). This hypothesis is supported by in vitro data showing that poly(A)-binding protein (PABP) (*5*) stimulates translation termination by recruiting release factors.

The idea of context-dependent termination sheds light on another ciliate coding oddity. More than 11% of genes in *Euplotes octocarinatus* contain an in-frame stop codon followed by a frameshift in the coding sequence (*6*, *7*). These stop codons occur in a "slippery sequence" [AAA UA(A/G), stop codon underlined], which allows the tRNA in the ribosomal P site (which is decoding the AAA codon) to pair with the codon in the +1 frame (AAU). At these positions, ribosomal frameshifting efficiently outcompetes termination (*8*), much like cognate tRNA recognition outcompetes termination in *C. magnum* and *Parducia* sp. This suggests that position-dependent stop codon recognition exists in all of these species and may be a general feature of ciliate biology.

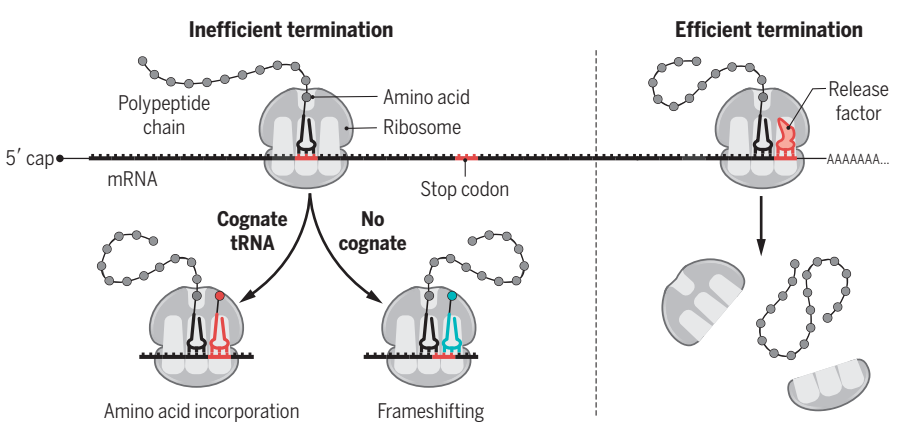
The context-dependence of stop codon recognition in ciliates bears a striking resemblance to the nonsense-mediated mRNA decay (NMD) pathway in humans and other eukaryotes. NMD degrades mRNAs with premature stop codons in early exons or long 3′ UTRs, but the mechanism by which premature stop codons are differentiated from normal ones is unclear. Several studies have argued that proximity to the poly(A) tail is one measure by which the translational machinery identifies a normal stop codon (*9*, *10*), and that decay is the default outcome of stop codon recognition in the absence of suppressing factors (*11*), much as amino acid incorporation appears to be the default for stop codons in *C. magnum*.

The distinct stop-codon recognition mechanism in these ciliates raises many questions about termination and NMD throughout the eukaryotic lineage. We still do not understand how proximity to the poly(A) tail is determined, particularly when RNA secondary structure and protein binding can affect the conformation of a mRNA in three-dimensional space. Are release factors only found in proximity to the poly(A) tail, or are they somehow activated by its presence? Is the system tightly tuned to prevent the production of truncated and extended proteins, or are ciliates constantly cleaning up the mistakes of a sloppy translational machinery? What is the role of conserved NMD factors in ciliates with ambiguous stop codon usage? Answering these questions will require both detailed biochemical studies and genome-wide analyses and will likely yield fundamental insights about gene expression in all eukaryotes. ■

Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD, USA. Email: ragreen@jhmi.edu

Context matters

In some ciliate species, termination of mRNA translation is only efficient in close proximity to the poly-A tail. *C. magnum* has cognate tRNAs for the standard stop codons, so they are translated like normal sense codons. In *E. octarinatus*, which lacks cognate tRNAs, premature stops stall the ribosome, which is resolved by +1 frameshifting.



GRAPHIC: C. BICKEL/*SCIENCE*

REFERENCES

1. E. C. Swart, V. Serra, G. Petroni, M. Nowacki, *Cell* **166**, 691 (2016).
2. S. M. Heaphy *et al.*, *Mol. Biol. Evol.* **33**, 2885 (2016).
3. K. Záhonová *et al.*, *Curr. Biol.* **26**, 2364 (2016).
4. S. Sengupta, P. G. Higgs, *J. Mol. Evol.* **80**, 229 (2015).
5. A. Ivanov *et al.*, *Nucl. Acids Res.* **44**, 7766 (2016).
6. R. Wang *et al.*, *Sci. Rep.* **6**, 21139 (2016).
7. L. A. Klobutcher, P. J. Farabaugh, *Cell* **111**, 763 (2002).
8. A. V. Lobanov *et al.*, *Nat. Struct. Mol. Biol.* 10.1038/nsmb.3330 (2016).
9. N. Amrani *et al.*, *Nature* **432**, 112 (2004).
10. A. B. Eberle *et al.*, *PLOS Biol.* **6**, e92 (2008).
11. Z. Ge, B. L. Quek, K. L. Beemon, J. R. Hogg, *eLife* **5**, e11155 (2016).

10.1126/science.aai9060

RETROSPECTIVE

Ralph J. Cicerone (1943–2016)

A leader in science and policy championed efforts to understand and tackle climate change

By **John P. Holdren**[1] *and* **Marcia K. McNutt**[2]

Ask which university in the United States does the most to improve career outcomes for low-income students, and the University of California (UC), Irvine, comes out on top, based on surveys. This is the fertile environment for teaching and research that Ralph J. Cicerone, as chancellor (1998–2005), brought to international prominence. Walk around the National Academy of Sciences in Washington, DC, and hear the echoes of more than a century and a half of lofty debate among the nation's most respected scientists. This is the sanctuary for science that Cicerone, as president (2005–2016), lovingly nurtured and enhanced.

Ralph Cicerone left his indelible fingerprints on these institutions and on the people within them. First and foremost, his emphasis was on excellence. He never compromised quality in any of the aspects of the institutions he led, whether in teaching, research, or engagement with policy-makers and the public.

His promotion of women in science had an impact. He showed, at UC Irvine, that departments with gender diversity could top the national rankings. At the National Academy of Sciences, he put into place new processes to help identify and promote worthy female candidates for election to the membership. And during his time as president, the proportion of women in the National Academy of Sciences grew from less than 10% to more than 15%; even more tellingly, the percentage of women among the new cohort of elected members skyrocketed to more than 26%.

Ralph started the transformation of the Irvine campus into a model of sustainability, such that it is now considered one of the "greenest" (most sustainable) campuses nationwide. Similarly, he oversaw the installation of solar panels on the roof of the east and west galleries of the National Academy of Sciences building. He would personally check the energy output each day, alerting the building engineer if anything looked awry.

For a scientist who lived life on the cutting edge of discovery and innovation, Cicerone was paradoxically "old school" in aspects of his professional life: paper, pen, dictation. His "filing system," at least from what could be gleaned from surveying the state of organization of his office at the National Academy, depended on stratigraphic horizons of paper in various geographic locations on a very large desk and an even larger table. Nonetheless, his system was flawless, as he could put his hand on any document as fast as a search of a hard drive might, and his recall was perfect.

PHOTO: U.S. NATIONAL ACADEMY OF SCIENCES

Always soft-spoken, Ralph Cicerone, along with his wife Carol (herself an expert in cognitive science), brought a degree of civility and a personal touch to scientific leadership, qualities that are too often lacking in the nation's capital. They were ever gracious hosts at all Academy events, whether the guests were members, staff, volunteers, or honorees. Cicerone would take extra time to remember personal bits about the background of each of his guests to make them feel at home, in his home, the home for science in America.

As a scientist, Ralph Cicerone had become deeply engaged professionally in the causes and consequences of global climate change, as well as in interdisciplinary education and advising around science, technology, and public policy. His own research in atmospheric chemistry made seminal contributions to understanding the role of chlorofluorocarbons in destroying the ozone layer. His work was cited in awarding the 1995 Nobel Prize in Chemistry to Paul J. Crutzen, Mario J. Molina, and F. Sherwood Rowland. This effort ultimately led to the adoption of the 1987 Montreal Protocol, a global treaty that banned chlorofluorocarbons and other ozone-depleting chemicals.

It was therefore no surprise that Ralph spurred the National Academies to produce the series of reports, *America's Climate Choices*, laying out possible strategies for mitigation and adaptation as well as for research investments, monitoring, and international cooperation to address the risks from changing climate. Another of Ralph's major accomplishments while president of the National Academy was securing the Gulf Research Program, a $500 million, 30-year wasting endowment funded through penalties levied against the parties responsible for the Deepwater Horizon oil spill.

Ralph's career—with its awards and recognitions far more numerous than can be detailed here—was a testament to the power of the American dream. A native of New Castle, Pennsylvania, he was the grandson of Italian immigrants and the first from his family to attend college. He graduated from the Massachusetts Institute of Technology (MIT) in 1965 with a major in electrical engineering but also made time to captain the baseball team. That love of baseball resurfaced years later when he encouraged varsity baseball at UC Irvine, where the ball field now bears his name.

After graduating from MIT, Ralph earned both masters and doctoral degrees from the University of Illinois. He then distinguished himself in a series of appointments at the University of Michigan, the Scripps Institution of Oceanography, and the National Center for Atmospheric Research, before joining UC Irvine in 1989. There he built the Earth System Science Department and rose to the positions of Dean of Physical Sciences and ultimately Chancellor, before being elected president of the National Academy.

Ralph J. Cicerone—who retired from the National Academy of Sciences presidency this past June—passed away at age 73 at his home in Short Hills, New Jersey, on 5 November 2016, surrounded by his family. ■

[1]*John P. Holdren is Assistant to the President for Science and Technology and Director of the White House Office of Science and Technology Policy.* [2]*Marcia K. McNutt is the President of the United States National Academy of Sciences.*
Email: mmcnutt@nas.edu; john_p._holdren@ostp.eop.gov

10.1126/science.aal4054

ESSAY

CELL AND MOLECULAR BIOLOGY

Trial and error

Optogenetic techniques offer insight into the dopamine circuit underlying learning

By **Neir Eshel**

We are all prediction-making machines. Granted, our predictions are often wrong—as the old saying goes, "It's tough to make predictions—especially about the future." But even wrong predictions serve a purpose: They help us learn. Each time we make a choice, we predict the outcome of that choice. When the outcome matches our prediction, there is no need to learn. When the outcome is unexpected, however, we update our predictions, hoping to do better next time.

The idea that we learn by comparing predictions to reality has been a mainstay of animal learning theory since the 1950s (*1*–*3*) and is one of the foundations of machine learning (*4*). Remarkably, the brain has evolved a simple mechanism to make precisely these comparisons. In the 1990s, Wolfram Schultz and colleagues found that dopamine neurons in the midbrains of monkeys showed a curious response to reward (*5*). When the monkeys received an unexpected reward (in this case, a squirt of juice), dopamine neurons fired a burst of action potentials. When that same reward was expected, the neurons no longer fired. And if Schultz *et al.* played a trick on the monkeys, making them expect a reward but ultimately withholding that reward, the dopamine neurons dipped below their normal firing rate (*6*). Together, these results demonstrated that dopamine neurons signal prediction error, or the difference between actual and predicted value [see the figure (A)]. If an outcome is better than predicted, dopamine neurons fire; if an outcome is the same as predicted, there is no change in firing; and if an outcome is worse than predicted, dopamine neurons dip below baseline. The level of dopamine release then informs the rest of the brain when a prediction needs to be fixed and in what direction.

This basic finding—that dopamine neurons signal errors in reward prediction—revolutionized the study of learning in the brain by supplying a powerful, mechanistic model for how reinforcement affects behavior (*7*). Despite extensive study, however, little is known about how dopamine neurons actually calculate prediction error. What inputs do dopamine neurons combine and how do they combine them? To answer these questions, we merged molecular biology, electrophysiology, and computational analysis.

We focused on the ventral tegmental area (VTA), a small brainstem nucleus that produces dopamine. Although a majority of neurons in this region are dopamine neurons, a substantial minority use the inhibitory neurotransmitter γ-aminobutyric acid (GABA) instead. A recent study from our laboratory showed that these GABA neurons do not signal prediction error; rather, they encode reward expectation (*8*).

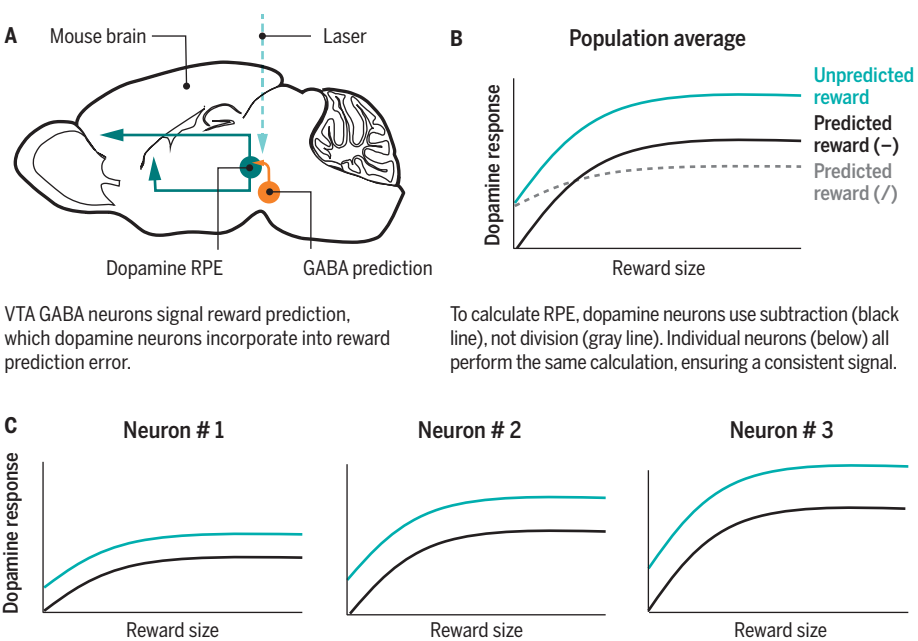
This finding raised a fascinating question: Could dopamine neurons use the GABA expectation signal to calculate prediction error? To find out, we used a virus to introduce the light-sensitive protein channelrhodopsin (ChR2) selectively in VTA GABA neurons. This enabled us to control the activity of VTA GABA neurons with light, a technique called optogenetics. We then implanted a set of electrodes surrounding a fiber optic cable into the VTA. Once the mice recovered from surgery, we recorded from the VTA and manipulated VTA GABA neuron activity, all while the mice performed simple learning tasks.

Optogenetics offers formidable precision, but there are potential pitfalls. In particular, it is easy to manipulate neural activity in ways that never occur in real life, producing results that are difficult to interpret. Our system avoided this pitfall, because we knew how VTA GABA neurons normally fire in our task. By recording during the manipulation, we made sure to mimic natural firing patterns.

When we stimulated VTA GABA neurons, dopamine neurons responded to unexpected rewards as if they were expected (*9*). Conversely, when we inhibited VTA GABA neurons, dopamine neurons responded to expected rewards as if they were unexpected. Finally, if we manipulated VTA GABA neurons simultaneously on both sides of the brain, we even changed the animals' behavior. After training mice to expect a certain size of reward, we artificially increased the expectation level by stimulating VTA GABA neurons during

Arithmetic and local circuitry of dopamine prediction errors

Dopamine neurons promote learning by conveying reward prediction error (RPE), the difference between actual and predicted reward. To probe how RPE is calculated, Eshel *et al*. recorded from dopamine neurons in mouse ventral tegmental area (VTA) while using optogenetics to manipulate nearby GABA neurons.



VTA GABA neurons signal reward prediction, which dopamine neurons incorporate into reward prediction error.

To calculate RPE, dopamine neurons use subtraction (black line), not division (gray line). Individual neurons (below) all perform the same calculation, ensuring a consistent signal.

GRAPHIC: ADAPTED FROM N. ESHEL BY G. GRULLÓN/*SCIENCE*

Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA 94305, USA. E-mail: neshel@stanford.edu

GRAND PRIZE WINNER: CELL AND MOLECULAR BIOLOGY

Neir Eshel

Neir Eshel is a psychiatry resident at Stanford University, pursuing a career at the interface of research and clinical practice. He is interested in how we learn about rewards and punishments, how we make decisions based on this knowledge, and how these systems break down in neuropsychiatric disease. He has conducted research at the National Institutes of Health, Princeton University, the World Health Organization, University College London, and Harvard University. Outside the laboratory and clinic, Neir plays clarinet in chamber groups and orchestras and is a passionate advocate for lesbian, gay, bisexual, and transgender (LGBT) health equality. **www.sciencemag.org/content/354/6316/1108**

CATEGORY WINNER: TRANSLATIONAL MEDICINE

Canan Dagdeviren

Canan Dagdeviren was born in 1985 in Istanbul, Turkey. As a Fulbright Doctoral Fellow, she received her Ph.D. degree in Material Science and Engineering at the University of Illinois at Urbana-Champaign. Dagdeviren developed a conformable, piezoelectric, energy harvester that converts mechanical energy from internal organ movements into electric energy to power medical devices. It is soft and flexible and conforms to the heart as well as other soft tissues. This technology could extend the battery life of implanted electronics or eliminate the need of battery replacement, sparing patients from repeated operations and the risk of surgical complications. Beginning in January 2017, she will assume the role of assistant professor at the MIT Media Lab, where she will direct the Conformable Decoders research group. **www.sciencemag.org/content/354/6316/1109.1**

CATEGORY WINNER: GENOMICS AND PROTEOMICS

Sam Behjati

Originally from Germany, Sam Behjati read medicine at Oxford University and pursued postgraduate clinical training in London, UK. Funded by the Wellcome Trust, he joined the Cancer Genome Project of the Sanger Institute (UK) for doctoral research. Using a "mutational postcode" that cells acquire as they divide, Behjati has shown that it is feasible to reconstruct the early developmental process in the adult mouse. Building on ideas developed during his Ph.D. preparation, he now aims to define the embryonic origin and fate of childhood cancer cells. **www.sciencemag.org/content/354/6316/1109.2**

CATEGORY WINNER: ECOLOGY AND ENVIRONMENT

David Seekell

David Seekell is an environmental scientist based in Sweden. He holds a Bachelor of Science in Natural Resources from the University of Vermont, and a Ph.D. in Environmental Sciences from the University of Virginia. In his Ph.D. research, Seekell developed statistics to provide early warning that an ecosystem is passing a tipping point and is about to undergo a regime change. He is currently an assistant professor of Ecology in the Department of Ecology and Environmental Science at Umeå University. In 2015, he became a Wallenberg Academy Fellow, and in 2016, he received a Science for Solutions Award from the American Geophysical Union. **www.sciencemag.org/content/354/6316/1109.3**

PHOTOS: (TOP TO BOTTOM) GEOFF CHESMAN; CANAN DAGDEVIREN; JET PHOTOGRAPHIC/CAMBRIDGE; ELIZABETH MURPHY

the anticipation period. The reward level, meanwhile, stayed the same. After several trials in which expectation exceeded reality, the disappointed mice stopped licking in anticipation of reward. When we turned off the laser, their behavior slowly returned to normal. We concluded that VTA GABA neurons convey to dopamine neurons how much reward to expect. In short, they put the "prediction" in "prediction error."

The VTA GABA expectation signal is only part of the puzzle. Another vital question is how do dopamine neurons actually use this input. What arithmetic do they perform? Again, we used molecular techniques to "tag" neurons with ChR2, but this time, we tagged dopamine neurons instead of GABA neurons. In each recording session, we shined pulses of light and identified neurons as dopaminergic if they responded reliably to each pulse. This ensured that the recorded neurons were indeed dopamine neurons; this eliminated the need for other, less accurate identification methods (*10*).

Using insights from the sensory literature (*11*), we designed a task to assess the input-output function of identified dopamine neurons and to determine how expectation transforms this function. We found that dopamine neurons use simple subtraction (*9*) [see the figure (B)]. Although this arithmetic is assumed in computational models, it is remarkably rare in the brain; division is much more common, as exemplified by gain control in sensory systems. However, subtraction is an ideal calculation because it allows for consistent results over a wide range of rewards. Moreover, we found that individual dopamine neurons calculated prediction error in exactly the same way (*12*). Each neuron produced an identical signal, just scaled up or down [see the figure (C)]. In fact, even on single trials, individual neurons fluctuated together around their mean activity. Such uniformity greatly simplifies information coding, allowing prediction errors to be broadcasted robustly and coherently throughout the brain—a prerequisite for any learning signal. Presumably, target neurons rely on this consistent prediction error signal to guide optimal behavior.

Our work begins to uncover both the arithmetic and the local circuitry underlying dopamine prediction errors. The method of evidence accumulation, the inputs that signal reward, and the biophysics underlying subtraction remain to be discovered—prime material for fresh predictions and unforeseen rewards. ■

REFERENCES AND NOTES

1. R. R. Bush, F. Mosteller, *Psychol. Rev.* **58**, 313 (1951).
2. L. Kamin, in *Fundamental Issues in Associative Learning: Proceedings of a Symposium on Fundamental Issues in Associative Learning*, N. J. Macintosh and W. K. Honig, Eds., Dalhousie University, Halifax, Nova Scotia, June 1968 (Dalhousie Univ. Press, Halifax, 1969), pp. 42–64.
3. R. A. Rescorla, A. R. Wagner, in *Classical Conditioning II: Current Research and Theory*, A. Black, W. Prokasy, Eds. (Appleton-Century-Crofts, New York, 1972), pp. 64–99.
4. R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction* (Cambridge Univ. Press, Cambridge, 1998), vol. 1.
5. W. Schultz, P. Dayan, P. R. Montague, *Science* **275**, 1593 (1997).
6. J. R. Hollerman, L. Tremblay, W. Schultz, *J. Neurophysiol.* **80**, 947 (1998).
7. W. Schultz, *Curr. Opin. Neurobiol.* **23**, 229 (2013).
8. J. Y. Cohen, S. Haesler, L. Vong, B. B. Lowell, N. Uchida, *Nature* **482**, 85 (2012).
9. N. Eshel *et al.*, *Nature* **525**, 243 (2015).
10. E. B. Margolis, H. Lock, G. O. Hjelmstad, H. L. Fields, *J. Physiol.* **577**, 907 (2006).
11. S. R. Olsen, V. Bhandawat, R. I. Wilson, *Neuron* **66**, 287 (2010).
12. N. Eshel, J. Tian, M. Bukwich, N. Uchida, *Nat. Neurosci.* **19**, 479 (2016).

ACKNOWLEDGMENTS

I am grateful to the many colleagues and mentors who offered criticism and encouragement over the course of this research. Particular thanks to my Ph.D. adviser, N. Uchida, and to the members of the Uchida lab, especially M. Watabe-Uchida and J. Tian.

10.1126/science.aal2190

ESSAY

ECOLOGY AND ENVIRONMENT

Passing the point of no return

Early warning signals indicate impending ecosystem regime changes

By **David Seekell**

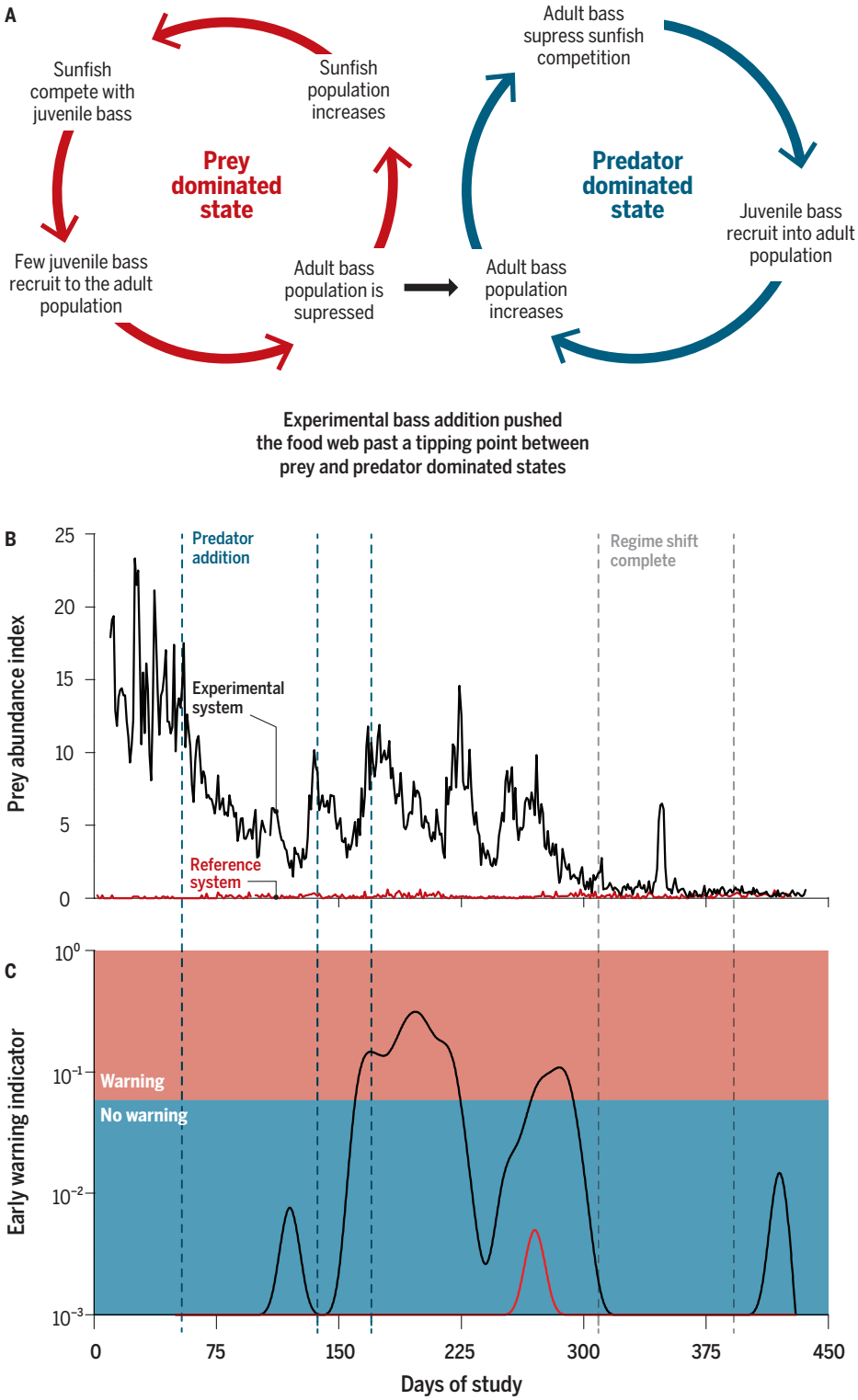
In the field of ecology, regime shifts are massive changes in function and character that occur when an ecosystem passes a tipping point. Regime shifts sometimes have severe consequences for human well-being through losses of ecosystem services, including desertification in arid regions and marine fisheries collapses (*1*, *2*). These changes are difficult to predict and sometimes impossible to reverse (*2*). For these reasons, understanding how to anticipate and prevent regime shifts is one of the most important challenges faced by environmental scientists (*1*–*3*).

Theoretical analyses have identified statistical anomalies, such as increased autocorrelation and variance in time series before regime shifts (*1*, *2*). These patterns are a manifestation of "critical slowing down"—when return rate from perturbation to equilibrium progressively declines before a tipping point (*1*). My dissertation research evaluated these anomalies as potential early warning indicators for ecosystem regime shifts.

This research was centered around a whole-lake experiment conducted with collaborators on two small lakes in the Upper Peninsula of Michigan. We manipulated the fish community of one lake to cause a trophic cascade—a type of regime shift—and made measurements throughout the food web to determine whether early warning indicators were detectable before the tipping point (*2*, *4*–*6*). An adjacent reference lake was monitored for comparison.

The experimental lake and reference lake had similar fish communities, with piscivorous largemouth bass (*Micropterus*

Department of Ecology and Environmental Science, Umeå University, SE-901 87 Umeå, Sweden. Email: david.seekell@umu.se



Lake fish regime shift. (**A**) Positive feedbacks can push either predators (largemouth bass) or prey (pumpkinseed sunfish) to dominance in the study lakes. (**B**) Average number of fish caught in minnow traps distributed around the edge of the experimental (black line) and reference (red line) lakes. This is an index of prey abundance. The study occurred during four consecutive summers, but here the data are concatenated into a continuous time series for aesthetic reasons. (**C**) Early warning of the regime shift based on moving-window conditional heteroskedasticity tests applied to chlorophyll-a concentrations from four summers concatenated into a single time series (*5*, *8*, *9*). Chlorophyll-a concentration is an index of phytoplankton biomass that strongly reflects the variability generated by the regime shift in the fish community (*2*, *5*). There was early warning for the entire summer the year before the regime shift and until the tipping point was passed during the year of the regime shift.

GRAPHIC: ADAPTED FROM D. SEEKELL BY G. GRULLÓN/*SCIENCE*

CATEGORY WINNER: ECOLOGY AND ENVIRONMENT

David Seekell

David Seekell is an environmental scientist based in Sweden. He holds a Bachelor of Science in Natural Resources from the University of Vermont, and a Ph.D. in Environmental Sciences from the University of Virginia. In his Ph.D. research, Seekell developed statistics to provide early warning that an ecosystem is passing a tipping point and is about to undergo a regime change. He is currently an assistant professor of Ecology in the Department of Ecology and Environmental Science at Umeå University. In 2015, he became a Wallenberg Academy Fellow, and in 2016, he received a Science for Solutions Award from the American Geophysical Union.

salmoides) and planktivorous pumpkinseed sunfish (*Lepomis gibbosus*) being most abundant. The competitive interactions between these species form a trophic triangle—a food web configuration where positive feedback can push either the predator (largemouth bass) or the prey (pumpkinseed sunfish) species to dominance (*4*).

Before the experiment, pumpkinseed sunfish dominated the fish community in the experimental lake, and largemouth bass dominated the fish community in the reference lake. We slowly added largemouth bass to the experimental lake over the course of 4 years in an attempt to push the food web past a tipping point where positive feedback would move the system from a sunfish-dominated regime to a bass-dominated regime [see the figure (A)].

The experimental lake passed a tipping point during the third year of the study [see the figure (B)]. Pumpkinseed sunfish competed with juvenile bass, and the tipping point occurred when the adult bass population became sufficiently large that it could suppress the sunfish, and juvenile bass could reach maturity and increase the size of the adult population (see the figure, A). This created feedback that promoted bass dominance (*2*, *4*–*6*). Variability from this shift propagated throughout the lake, and early warning indicators based on autocorrelation and variance responded strongly. Early warning signals were recorded up to a year and a half before the tipping point and were present throughout the food web including in time series of prey fish abundance, zooplankton biomass, and phytoplankton biomass (*2*, *4*–*7*). There were no early warning signals in the reference system.

This was the first ecosystem-scale proof-of-concept that early warning indicators can be detected before tipping points. These results suggest that ecosystem managers may one day be able to use adaptive management to avert unwanted regime shifts.

A specific contribution of my dissertation was to evaluate the efficacy of tests for conditional heteroskedasticity (clustered variability in time series) as early warning indicators. Conditional heteroskedasticity is widely studied by economists but almost never examined by ecologists. A key observation in my dissertation was the presence of conditional heteroskedasticity in time series from ecosystems approaching tipping points but not from stable ecosystems (*8*, *9*).

This observation has great practical importance. Most early warning indicators are interpreted by comparison with indicators from a pristine reference system. Such pristine systems are in short supply given the global nature of many environmental changes, such as climate warming. Conditional heteroskedasticity tests have thresholds that make clear distinctions between "warning" and "no warning" conditions, reducing the need for pristine reference systems (*5*, *8*) (see the figure, C). In the whole-lake experiment, conditional heteroskedasticity was the most powerful early warning indicator we tested but also the best at minimizing false-positive warnings due to the inclusion of thresholds (*5*).

Critical slowing down is a generic phenomenon that is not restricted to ecosystems. After our experiment, other research groups identified early warning indicators before regime shifts in a variety of complex systems, including in economic records before the subprime housing loan crisis (*10*), before the self-termination of epileptic seizures (*11*), and before shifts in social networking activity on Twitter (*12*). Hence, although my dissertation focused on lake ecology, the project is exemplary of how fundamental ecological research can generate tools and concepts that provide diverse benefits to society. ■

REFERENCES AND NOTES

1. M. Scheffer *et al.*, *Nature* **461**, 53 (2009).
2. S. R. Carpenter *et al.*, *Science* **332**, 1079 (2011).
3. W. J. Sutherland *et al.*, *J. Ecol.* **101**, 58 (2013).
4. D. A. Seekell, T. J. Cline, S. R. Carpenter, M. L. Pace, *Theor. Ecol.* **6**, 385 (2013).
5. D. A. Seekell, S. R. Carpenter, T. J. Cline, M. L. Pace, *Ecosystems (N. Y.)* **15**, 741 (2012).
6. T. J. Cline *et al.*, *Ecosphere* **5**, art102 (2014).
7. M. L. Pace, S. R. Carpenter, R. A. Johnson, J. T. Kurtzweil, *Limnol. Oceanogr.* **58**, 525 (2013).
8. D. A. Seekell, S. R. Carpenter, M. L. Pace, *Am. Nat.* **178**, 442 (2011).
9. V. Dakos *et al.*, *PLOS ONE* **7**, e41010 (2012).
10. J. P. L. Tan, S. S. A. Cheong, *Eur. Phys. J. B* **87**, 38 (2014).
11. M. A. Kramer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21116 (2012).
12. C. Kuehn, E. A. Martens, D. M. Romero, *J. Complex Netw.* **2**, 141 (2014).

ACKNOWLEDGMENTS

This essay is based on research supported by the NSF (DEB 0716869, DEB 0917696, and Graduate Research Fellowship Program). Whole-ecosystem experiments are logistically difficult and are therefore conducted as collaborative efforts. The following researchers contributed to the whole-ecosystem experiment described in this essay: R. Batt, W. Brock, C. Brosseau, S. Carpenter, T. Cline, J. Cole. J. Coloso, M. Dougherty, A. Farrell, J. Hodgson, R. Johnson, J. Kitchell, S. Klobucar, J. Kurtzweil, K. Lee, M. Pace, T. Matthys, K. McDonnell, H. Pack, L. Smith, T. Walsworth, B. Weidel, G. Wilkinson, C. Yang, and L. Zinn. E. Murphy gave helpful comments on the text.

10.1126/science.aal2188

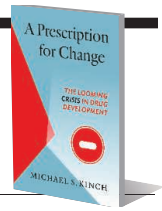
PHOTO: ELIZABETH MURPHY

HEALTH AND MEDICINE

Arrested development

Financial and regulatory pressures threaten the discovery of new medicines

A Prescription for Change
The Looming Crisis in Drug Development
Michael Kinch
UNC Press, 2016. 351 pp.



By **George Painter**

The discovery, development, and registration of new drugs is a complicated and protracted process, often taking more than 10 years. It is also quite expensive, with an estimated cost per new drug in excess of $1 billion. The development pathway has historically required input from academia, large pharmaceutical companies, and government partners, but in recent years, all of these groups have begun facing social and financial challenges that are diminishing their ability to participate in the discovery process. In *A Prescription for Change*, Michael Kinch traces the history of the pharmaceutical and biotechnology industries and describes the evolution of the scientific, regulatory, social, and marketing forces that are compromising contemporary drug development efforts.

The book starts with an introduction to the phases of the clinical development of a new drug. Kinch's descriptions of each stage are adequate, and the complications that inevitably arise throughout clinical drug development are well explained and illustrated with useful anecdotes. The process, he shows, is difficult, expensive, and fraught with unanticipated problems and setbacks.

In a chapter titled "Why regulate medicines," Kinch traces the evolution of the U.S. Food and Drug Administration (FDA) and describes how, in pursuing its responsibility to ensure the safety and efficacy of new medicines, the agency helped shape the pharmaceutical industry. Key legislation that set the FDA's trajectory (the Food, Drug, and Cosmetic Act of 1938 and the 1962 Kefauver-Harris Amendments) is introduced in historical context. But Kinch misses the chance to really emphasize how increasing public pressure to ensure safety and efficacy has resulted in larger and larger phase 3 pivotal efficacy and phase 4 postmarketing trials, which have, in turn, contributed substantially to the increased costs associated with clinical development. Daunted by both the low probability of success and the high cost of development, many venture groups have reduced funding for the development of new drugs in recent years.

The pharmaceutical industry has emerged as a formidable economic engine that has a substantial effect on the global economy. By 2017, global spending for medicines is projected to reach $1.2 trillion. In *A Prescription for Change*, Kinch focuses on the growth of two major pharmaceutical companies, Merck and Eli Lilly, both of whom rank among the dominant innovators of the 20th and 21st centuries. Their stories are told in a narrative, historical manner, starting with each company's birth as a family business. The approach is quite entertaining and provides a way to explain how innovative drugs such as morphine, penicillin, and insulin propelled these companies to world-prominent positions. Key individuals, innovative research universities, and government regulatory authorities emerge as factors that shaped the growth of the industry.

A cycle of acquisitions and downsizing is cannibalizing our pharmaceutical infrastructure, argues Michael Kinch.

In later chapters, Kinch details how, in the face of increasing competition in the marketplace and rising costs of development, large pharmaceutical companies have become financially dependent on pharmaceutical "blockbusters" (drugs that have annual sales in excess of $1 billion). However, the discovery of such drugs, at a pace necessary to maintain market expectations, is usually not possible. This has led to mergers and acquisitions as a means to increase market share and meet Wall Street growth expectations. Kinch mentions, but fails to adequately address, how these mergers inevitably led to considerable downsizing of research and development (R&D) staff and profoundly affected the ability of large companies to support productive internal R&D programs.

With the downsizing of the R&D capacity of large pharmaceutical companies, biotechnology companies (many of which arose from prominent research universities) soon became a viable source of innovative new drug candidates. Kinch describes how the relationships among research universities, biotechnology companies, and large pharmaceutical companies formed and led to a steady stream of drugs to address unmet medical needs. However, diminished financial support for joint discovery ventures has limited the productivity of these relationships in recent years. So what do we do?

Kinch believes that tackling the problems facing contemporary drug discovery efforts would require an unprecedented level of communication among academic, commercial, federal, and financial institutions. It would also require a commitment to a level and pace of change that a considerable body of evidence suggests is just not going to happen.

Yet, as pharmaceutical companies retreat and the number of biotechnology start-ups declines, a new paradigm for drug discovery based on public-private partnerships is already emerging. The drug discovery company Bridge Medicines—formed in October as a partnership between two venture capital firms, three research centers, and a major pharmaceutical company—represents a promising example of how drugs may be discovered in the future (*1*). ■

The reviewer is at the Emory Institute for Drug Discovery, Emory University, Atlanta, GA 30329, USA. Email: george.r.painter@emory.edu

REFERENCES

1. Rockefeller University, Newswire, 31 October 2016; http://newswire.rockefeller.edu/2016/10/31/pioneering-drug-discovery-company-bridge-medicines-launched-to-advance-promising-early-technologies-in-major-academic-institutions-through-human-proof-of-concept.

10.1126/science.aaj1906

LETTERS

Edited by **Jennifer Sills**

Nuclear power: Serious risks

IN THEIR POLICY Forum "China-U.S. cooperation to advance nuclear power" (5 August, p. 547), J. Cao *et al.* make the case for low-carbon energy trajectories that use "next-generation" nuclear reactors. However, they fail to address the challenges inherent in the reactors they advocate.

Cao *et al.* correctly assert that "Some studies project that a doubling or quadrupling of nuclear energy output is required in the next few decades." They neglect to add that there are also numerous peer-reviewed studies showing that 100% renewables scenarios are technologically feasible and economically competitive. Unlike some of the reactors proposed by Cao *et al.*, these could be deployed rapidly (*1*). Moreover, more scalable and commercially available energy efficiency options can displace the need for new sources of nuclear supply (*2*).

Cao *et al.* also correctly state that renewable sources of energy in Germany have been heavily subsidized. However, Germany has devoted subsidies to the nuclear industry that more than double those allotted to all renewables put together (*3*). The United States has provided the nuclear industry with at least 10 times the subsidies devoted to renewables (*4*). Despite these investments, renewables costs are falling fairly quickly, whereas nuclear costs continue to rise (*5*–*7*). In India and China, despite late starts on development, electricity production from wind has overtaken nuclear (*8*). In restructured markets that allow consumers to choose from a variety of energy options, renewables have been shown to be cheaper than nuclear power (*9*).

Cao *et al.* portray a variety of reactors as "innovative" and "next generation," yet similar reactors have been under development since the 1960s (*10*). Sodium-cooled fast reactors and liquid metal-cooled fast reactors, as well as conventional small pressurized water reactors, have a history of costly experiments. In the West, these earlier programs were abandoned, despite decades of research and development commitment and high governmental prioritization, due to economic unviability and safety issues (*11*). Even if some newer versions of these technologies prove viable and acceptable, the time scale to commercial deployment will inevitably be measured in decades.

Cao *et al.* suggest collaborative plans, especially for small modular reactors, between consortia in China and the United States as possible ways forward. However, it is unclear why these projects should have priority in terms of governmental support. Tellingly, there is no commercially operating small modular reactor anywhere in the world (*12*). There is also very limited licensing experience with small modular reactors. The cost is essentially unknown, and public acceptability completely untested.

Pouring resources into "innovative" reactor technologies could be a damaging distraction. We must give balanced consideration to a full range of alternative low-carbon energy options rather than focus uncritically on nuclear energy.

Philip Johnstone, Benjamin K. Sovacool,* Gordon MacKerron, Andy Stirling

Science Policy Research Unit, University of Sussex, Brighton, BN1 9RH, UK.

*Corresponding author. Email: b.sovacool@sussex.ac.uk

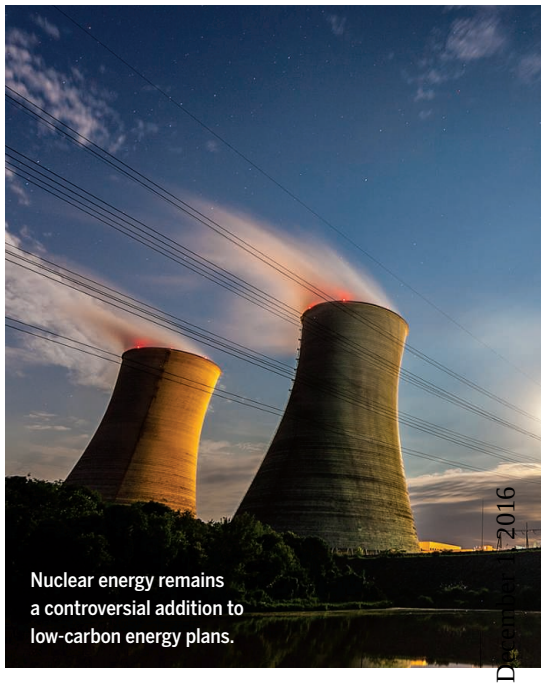
REFERENCES

1. C. Budischak *et al.*, *J. Power Sources* **225**, 60 (2013).
2. M. A. Brown, B. K. Sovacool, *Energy Efficiency* **1**, 35 (2008).
3. International Energy Agency, Research and Development Database (2014); http://wds.iea.org.
4. B. Healey, N. Pfund, *What Would Jefferson Do?* (DLB Investors, 2011).
5. M. Liebreich, *BLOOMBERG New Energy Finance Summit: In Search of the Miraculous* (Bloomberg New Energy Finance, 2016).
6. International Energy Agency, *World Energy Outlook* (International Energy Agency, 2015).
7. D. Suna, G. Resch, *Energy Pol.* **98**, 199 (2016).
8. B. Chabot, "Wind power and nuclear electricity production in China , India , Brazil and South Africa (BICS) up to 2015" (2016), pp. 1–7; http://cf01.erneuerbareenergien.schluetersche.de/files/smfiledata/5/5/8/4/6/2/141WindNucBICS2015.pdf.
9. National Audit Office, *Nuclear Power in the UK* (National Audit Office, 2016).
10. T. B. Cochran *et al.*, *Fast Breeder Reactor Programs: History and Status* (International Panel on Fissile Materials, 2010).
11. M. V. Ramana, "The Forgotten History of Small Nuclear Reactors," *IEEE Spectrum* (2015), pp. 1–9; http://spectrum.ieee.org/energy/nuclear/the-forgotten-history-of-small-nuclear-reactors.
12. World Nuclear Association, Small modular reactors (2016); www.world-nuclear.org/information-library/nuclear-fuel-cycle/nuclear-power-reactors/small-nuclear-power-reactors.aspx.

10.1126/science.aal1777

Nuclear power: Deployment speed

IN THEIR POLICY Forum "China-U.S. cooperation to advance nuclear power" (5 August, p. 547), Cao *et al.* parse data in a way that makes nuclear energy seem faster to deploy than renewable energy. A closer look at those data suggests the opposite.



Nuclear energy remains a controversial addition to low-carbon energy plans.

PHOTO: ZHONGGUO/ISTOCK

Cao *et al.* claim that nuclear power programs have historically driven the fastest growth of low-carbon power, but data in their ref. S1 (*1*) show that renewables' output (excluding hydropower) grew 6.3 times as fast as nuclear power in 2015. In Figure 2, Cao *et al.* compare seven nations' average annual increases of nuclear vs. solar-and-windpower electricity. China's world-leading (*2*) renewable increases appear to be less than a tenth as fast as Denmark's because the chart divides the energy per capita, which inflates the numbers for less populous countries. However, the population of the country producing the energy is not relevant to different technologies' relative speed in absolute carbon savings.

Cao *et al.* also disadvantage renewables' short lead times and rapid recent growth (shown in their Figure 1) by comparing output increases over the fastest decade. Recent, shorter time periods, such as the 3-year span from 2012 to 2015 according to their ref. S1 (*1*), reveal that nonhydro renewable growth dwarfs nuclear growth (in this case, by a factor of 5.3). Cao *et al.*'s Figure 1 also shows renewables' growth through 2015 [from ref. S1 (*1*)], but their Figure 2 shows 2004 to 2014 (2003 to 2013 for Spain). That truncation shrinks seven-country growth by 25% for solar power and 11% for windpower (*1*).

Cao *et al.* further deemphasize renewables by counting only solar and

windpower, omitting substantial global nonhydro renewable output. Their ref. S1 (*1*) uses gross (not net) generation, giving nuclear energy an advantage over renewables. Their national deployment-rate comparisons omit preconstruction lead times for institutional preparation, skill-building, design, siting, licensing, and financing, which typically total years for renewables but decades for nuclear energy sources. They omit end-use efficiency, which saves more global electricity than renewables or nuclear sources produce (*3*).

Cao *et al.*'s conclusion contradicts the marketplace (*2*, *4*). Fundamental differences in scale-up mechanisms, business cases, and learning curves informed China State Grid's Chairman's 2015 expectation (*5*) that 2050 global electricity will come from 14% hydro, 73% other renewables, 7% gas, 3% coal, and 3% nuclear [vs. 10.7% in 2015 (*4*)].

Amory B. Lovins

Rocky Mountain Institute, Basalt, CO 81621, USA.
Email: ablovins@rmi.org

REFERENCES

1. BP, BP Statistical Review of World Energy (BP, 2015); www.bp.com/content/dam/bp/pdf/energy-economics/statistical-review-2015/bp-statistical-review-of-world-energy-2015-full-report.pdf.
2. Renewable Energy Policy Network for the 21st Century, Renewables 2016 Global Status Report (2016); www.ren21.net/status-of-renewables/global-status-report/.
3. International Energy Agency, Energy Efficiency Market Report 2015 (www.iea.org/publications/freepublications/publication/energy-efficiency-market-report-2015-.html).
4. M. Schneider, A. Froggatt, Eds., *World Nuclear Industry Status Report* (2016); www.worldnuclearreport.org.
5. Z. Liu, *Global Energy Interconnection* (Acad. Press, 2015).

10.1126/science.aal1808

Response

WE AGREE WITH Johnstone *et al.*'s call to give "balanced consideration to a full range of alternative low-carbon energy options" to address climate change. Unlike Johnstone *et al.*, we believe that a balanced portfolio includes the investment of modest fractions of current energy research, development, and deployment (RD&D) funding into advanced nuclear energy.

Johnstone *et al.* may be confident that renewables such as wind and solar are the single fix-all bet for climate change, but the evidence provided by actual deployment rates, reported in our Policy Forum, as well as system reliability challenges, suggest that it would be good to at least hedge that bet with investments in nuclear energy RD&D. For example, wind and sun are often at low ebb for weeks or even months at a time, and current daily-cycle storage is not sufficient to fill in these gaps (*1*, *2*).

Johnstone *et al.* cite the cost of current nuclear energy as a reason not to pursue the very RD&D that might bring nuclear costs down. Indeed, it was precisely the current cost of nuclear energy, as well as too-long deployment times, that we cited as the spur for the needed RD&D. That said, even with regard to current nuclear technology, Johnstone *et al.*'s assertion that "nuclear costs continue to rise" is far from universally true. Chinese and South Korean current light water reactors built in China and the United Arab Emirates are reported to cost in the range of $2000 to $3000 per kilowatt, a 50% reduction from historical levels (*3*, *4*). Moreover, it has been shown that, with repeated construction of the same reactor design, reactor costs stabilize or decline (*3*). We are not persuaded by Johnstone *et al.*'s claim that small modular reactors should not be pursued because "there is no commercially operating small modular reactor anywhere in the world." If the same logic had been applied to utility-scale wind or solar energy technology in the 1970s, and governments had not spent billions of dollars in research and early commercial subsidies for those sources, we would likely not have the vibrant commercial wind and solar industry today that Johnstone *et al.* rightly celebrate.

Lovins points to data from 2015 and other years to show renewables exceeding nuclear in scale-up rates, but deep decarbonization requires that low-carbon energy sources sustain high rates of deployment over many decades, and results for a single year contain less useful information than the decadal trends we presented. In fact, as Figure 1 shows, our finding that nuclear power has been responsible for the fastest expansion of low-carbon electricity supplies is robust over a broad range of intervals, even single years. Worldwide, the largest 1-year increase in low-carbon electricity production was delivered by nuclear power between 1983 and 1984. In that year, almost twice as many low-carbon kilowatt hours per capita were added by nuclear power as were added by wind and solar combined in 2015—the peak year for those technologies so far (*5*, *6*). And during the peak decade for nuclear construction, 1979 to 1989, the world was adding low-carbon kilowatt hours at a faster annual rate on average than the largest amount added in a single year so far by either wind or solar (*5*). Moreover, this occurred at a time when the global economy and population were considerably smaller than they are today.

Lovins argues that it is gross additions of low-carbon energy that matter, not energy per capita. But the latter metric is the better basis for comparing the decarbonization performance of countries with different populations. By this measure, China has indeed been slower than Denmark in scaling up wind and solar over the past decade, despite its large absolute additions of these technologies. Lovins also avoids mentioning that in 2015, China, which has just begun a major nuclear scale-up program, added almost as much nuclear electricity as wind and solar combined (*5*).

More rapid deployment of solar and wind, as well as nuclear and other low-carbon energy sources, will be needed to achieve deep decarbonization goals. The future of our planet depends on it.

Junji Cao,[1] Armond Cohen,[2] James Hansen,[3]* Richard Lester,[4] Per Peterson,[5] Staffan A. Qvist,[2] Hongjie Xu[6]

[1]Key Lab of Aerosol Chemistry and Physics, Institute of Earth Environment, Xi'an, 710061, China. [2]Clean Air Task Force, Boston, MA 02108, USA. [3]Earth Institute, Columbia University, New York, NY 10025, USA. [4]Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [5]University of California, Berkeley, CA 94720, USA. [6]Shanghai Institute of Applied Physics, Shanghai, 21203, China.

*Corresponding author.
Email: jimehansen@gmail.com

REFERENCES

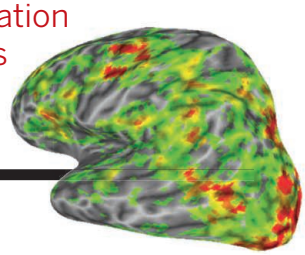
1. M. Z. Jacobson, M. A. Delucchi, M. A. Cameron, B. A. Frew, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15060 (2015).
2. A. E. MacDonald *et al.*, *Nat. Clim. Change* **6**, 526 (2016).
3. J. R. Lovering, A. Yip, T. Nordhaus, *Energy Pol.* **91**, 371 (2016).
4. Chinese Academy of Engineering, *Re-studying the Nuclear Development in China* (Tsinghua University Press, ed. 1, 2015) [in Chinese].
5. BP Statistical Review of World Energy 2016 (www.bp.com/content/dam/bp/pdf/energy-economics/statistical-review-2016/bp-statistical-review-of-world-energy-2016-full-report.pdf).
6. World Bank, World Development Indicators (2016); https://issuu.com/world.bank.publications/docs/9781464806834.

10.1126/science.aal2561

RESEARCH

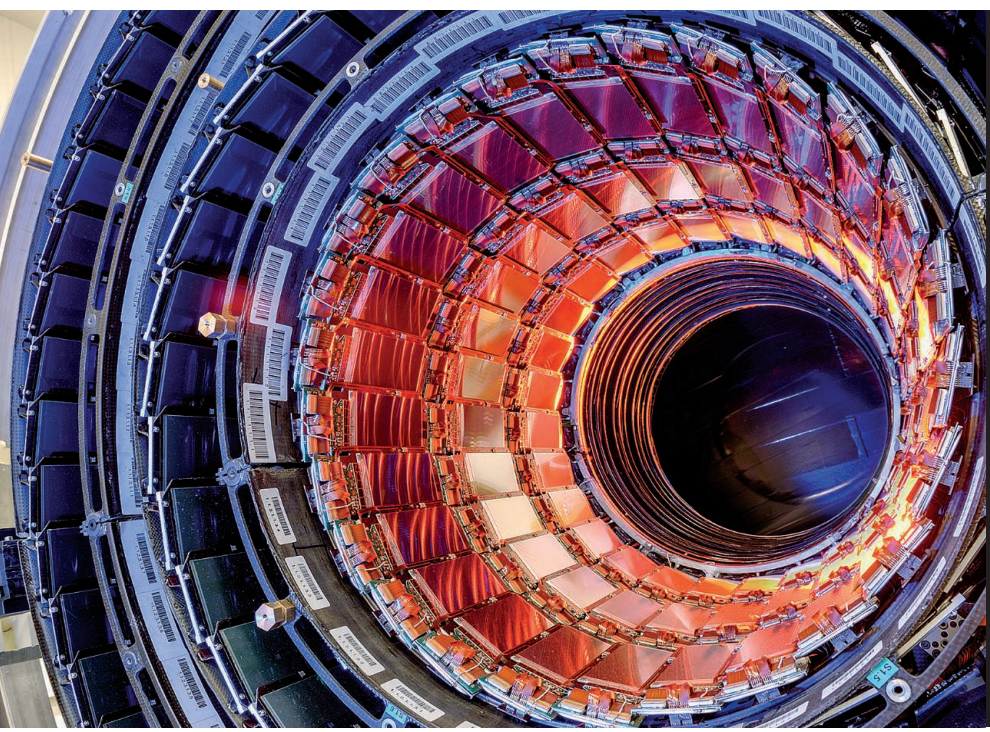
Transcranial magnetic stimulation restores misplaced memories

Rose et al., *p. 1136*



IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**



The CLOUD chamber at CERN, Switzerland

ATMOSPHERIC SCIENCE

How new particles form

New particle formation in the atmosphere produces around half of the cloud condensation nuclei that seed cloud droplets. Such particles have a pivotal role in determining the properties of clouds and the global radiation balance. Dunne *et al.* used the CLOUD (Cosmics Leaving Outdoor Droplets) chamber at CERN to construct a model of aerosol formation based on laboratory-measured nucleation rates. They found that nearly all nucleation involves either ammonia or biogenic organic compounds. Furthermore, in the present-day atmosphere, cosmic ray intensity cannot meaningfully affect climate via nucleation. —HJS

Science, this issue p. 1119

GALAXY FORMATION

A massive galaxy forming from molecular gas

The most massive galaxies gather their stars by merging with smaller galaxies and by accreting gas, which is then consumed during star formation. Emonts *et al*. investigated the Spiderweb Galaxy, a massive galaxy in the process of forming in the early universe, seen now as it was over 10 billion years ago (see the Perspective by Hatch). Radio observations of carbon monoxide revealed large quantities of molecular gas around the galaxy. The gas is not associated with the merger process but may have been recycled from earlier phases of galaxy formation. —KTS

Science, this issue p. 1128; see also p. 1102

CANCER

Running interference

Interleukin 2 (IL-2) binds to receptors on different types of T cells. $CD8^+$ T cells, which can kill tumor cells, have IL-2 receptors with two subunits. When IL-2 binds to these, it promotes T cell activation. In contrast, T regulatory cells dampen the antitumor immune response. These cells express a different type of IL-2 receptor, which contains CD25 in addition to the other two subunits. CD25 binds IL-2 tightly but does not activate a T cell response. Arenas-Ramirez *et al.* developed an antibody that can block CD25. Delivering this antibody together with IL-2 allowed IL-2 to bind specifically to the activating receptors and promote an antitumor immune response without interference from T regulatory cells. —YN

Sci. Transl. Med. **8**, 367ra166 (2016).

TOPOLOGICAL MATTER

Shining light on a peculiar coupling

One of the long-standing predictions regarding topological insulators is the magnetoelectric effect, a coupling between a material's magnetic and electric properties. Thanks to this coupling, Maxwell's equations inside topological insulators are modified, resulting in so-called axion electrodynamics. Wu *et al*. used time-domain terahertz (THz) spectroscopy to observe signatures of these unusual electrodynamics in a thin film of Be_2Se_3. They detected tiny changes to the polarization of THz light after it passed through the thin film, confirming the expected quantization of the magnetoelectric coupling. —JS

Science, this issue p. 1124

PAIN RESEARCH

Glial cells contribute to pain

Pain hypersensitivity can spread to unaffected body regions immediately surrounding the initial insult. Sometimes it can even spread to the opposite site of the body or to large body areas and cause widespread pain. Kronschläger *et al*. discovered a form of synaptic plasticity in the spinal cord that may explain the spread of pain hypersensitivity. This plasticity was induced by

CREDITS (FROM TOP): ROSE *ET AL.*; CERN

the activation of glial cells. The spread was mediated by gliotransmitters that diffuse widely, even reaching the cerebrospinal fluid at biologically relevant concentrations. —PRS

Science, this issue p. 1144

STRUCTURAL BIOLOGY

Zika virus is fit to be tied

Zika virus (ZIKV) has been associated with fetal microcephaly and Guillain-Barré syndrome. Other mosquito-born flaviviruses, such as dengue virus, encode noncoding subgenomic flavivirus RNAs (sfRNAs) in their 3′ untranslated region that accumulate during infection and cause pathology. Akiyama *et al*. now report that ZIKV also produces sfRNAs that resist degradation by host exonucleases in infected cells. The authors solved the structure of one of ZIKV's sfRNAs by x-ray crystallography and found that the multi-pseudoknot structure that it adopts underlies its exonuclease resistance. —KLM

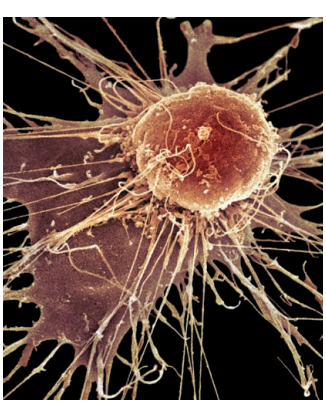
Science, this issue p. 1148

HEMATOPOIESIS

How to maintain hematopoietic stem cells

Hematopoiesis provides the body with a continuous supply of blood cells (see the Perspective by Sommerkamp and Trumpp). Taya *et al*. report that amino acid content is important for hematopoietic stem cell (HSC) maintenance in vitro and in vivo. Dietary valine restriction seems to "empty" the mouse bone marrow niche. Ito *et al*. used single-cell approaches and cell transplantation to identify a subset of HSCs at the top of the HSC hierarchy. Self-renewal relied on the induction of mitophagy, a quality-control process linked to a cell's metabolic state. Both studies may be helpful in improving clinical bone marrow transplantation. —BAP

Science, this issue p. 1103, p. 1152; see also p. 1156



Hematopoietic stem cell

PHOTOS (FROM LEFT): SPL / SCIENCE SOURCE; SCOTT CAMAZINE / ALAMY STOCK PHOTO

VACCINATION

Protecting by changing the code

Live attenuated vaccines can be very potent, but their potential to revert to their pathogenic form limits their use. In an attempt to get around this, Si *et al*. expanded the genetic code of influenza A viruses. They propagated viruses that were mutated to encode premature termination codons (PTCs) in a cell line engineered to be able to express these flu proteins. Despite not being able to replicate in conventional cells, PTC-containing viruses were highly immunogenic and protected mice, guinea pigs, and ferrets against influenza challenge. —KLM

Science, this issue p. 1170

OXIDATIVE STRESS

Overactive antiviral responses in lupus

Detection of viral RNAs causes oligomerization of mitochondrial antiviral signaling (MAVS) protein, which leads to the production of type I interferons (IFNs). Buskiewicz *et al*. found that MAVS oligomerization in the absence of virus may contribute to lupus disease severity. Mitochondrial reactive oxygen species (ROS) induced MAVS oligomerization and type I IFN production in uninfected cells. The MAVS C79F variant, which is associated with decreased lupus severity, did not oligomerize in response to ROS, and cells expressing this variant produced less type I IFN. —JFF

Sci. Signal. **9**, ra115 (2016).

IN OTHER JOURNALS

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

INSECT GENOMICS

Acquiring the genes to digest wood

The larvae of the invasive Asian longhorned beetle burrow into and kill trees. On sequencing the genome, McKenna *et al.* found that gene transfers from fungi and bacteria, followed by functional evolution and gene family expansions, appear to have conferred the ability to the beetles to find plants, digest cellulose, and nullify harmful compounds made by the plants. Interestingly, other wood-feeding beetles appear to have undergone a similar evolutionary trajectory, one that is distinct from that of wood-feeding insects such as termites. —LMZ

Genome Biol. **17**, 227 (2016).

Asian longhorned beetle larvae digest wood by using acquired genes.

PLANT BIOLOGY

Targeting tip growth

Tip growth, which characterizes cells as diverse as root hairs and brain neurons, depends on secretory vesicles to add new plasma membrane in a defined subdomain. Bloch *et al.* show that in growing *Arabidopsis* pollen tubes, the exocyst subunit SEC3a is a target for secretory vesicles at the tip. SEC3a localization defines the axis of growth and the domain where new pectin is added to the cell wall. Pollen tubes of tobacco, which are fatter than those of *Arabidopsis*, showed more complex patterns: During isotropic growth, SEC3a was distributed in a broad subapical domain, whereas during rapid elongation growth, SEC3a was localized to the apical tip. —PJH

Plant Physiol. **172**, 980 (2016).

CELL BIOLOGY

Getting an UPR hand on recovERy

The endoplasmic reticulum (ER) is an intracellular membranous labyrinth that provides the entry point to the secretory pathway. During proteotoxic stress, the ER expands to cope with the added burden of misfolded or unfolded proteins. This is termed the unfolded protein response (UPR). Fumagilli *et al.* asked how, after the stress is removed, the cell returns its ER to normal levels. In cultured mammalian cells, they uncovered a process called "recovER-phagy." During recovER-phagy, excess ER membranes are targeted for destruction. Unexpectedly, a component of the ER protein translocation machinery, Sec62, appears to provide a key autophagy receptor in this process, independently of its canonical function. —SMH

Nat. Cell Biol. **18**, 1173 (2016).

HUMAN BIOLOGY

Neuron development in human embryos

Mammalian fertility depends on the secretion of gonadotropin-releasing hormone (GnRH) from a population of specialized neurons residing in the hypothalamus. During embryogenesis, these neurons develop at the olfactory placodes, and they subsequently migrate to the brain. Very little is known about the process in humans, however. Casoni *et al*. have studied this in depth by using donated human embryonic tissue. They tracked the differentiation and migration of GnRH neurons through the first trimester of gestation by examining samples at different developmental stages and identified important differences between humans and rodents. Unexpectedly, they also found that some of these neurons migrate to extra-hypothalmic regions of the brain, suggesting that they play roles in other processes not linked to fertility. —SH

Development 10.1242/dev.139444 (2016).

Corals share gastrulation signals for mouth-anus development with vertebrates.

ANIMAL DEVELOPMENT

Keeping tissue layers separate

Gastrulation is the conversion of an embryo from a single sheet of pluripotent cells into a structure with multiple tissue layers. This process establishes the future body plan and is highly conserved among metazoans. In vertebrate embryos, the transcription factor Brachyury delineates the middle tissue layer (the mesoderm) from the outer (ectoderm) and the inner (endoderm) layers. Yasuoka *et al.* discovered that in coral embryos, which lack a mesoderm, *brachyury* is regulated by the same signaling pathway as that found in vertebrates. In corals, it demarcates ectoderm from endoderm and is essential for the development of the mouth-anus. —SH

Curr. Biol. **26**, 2885 (2016).

ROBOTICS

Autonomously eat, digest, move, repeat

Truly autonomous robots require a robust and independent way to move and a means to harvest energy from the environment. Philamore *et al.* push toward these goals by devising a soft robotic mouth for use in aquatic environments that gathers organic biomass, which is processed in a microbial fuel cell to generate useful energy. Origami-like folding of a membrane pulls particulate-laden water into the mouth; the mouth is then closed for operation of the microbial fuel cell, which can be isolated from the water and connected in series. When combined with state-of-the-art electronics, the energy output from the fuel cell was boosted to usable values for powering motors and artificial muscles within the robot. —MSL

Soft Robotics 10.1089/soro.2016.0020 (2016).

NEUROSCIENCE

Side effects for placebo poppers

Patients in clinical trials may see their conditions improve, sometimes dramatically, even when they are in the control arm of the study. Tétreault *et al.* used brain imaging to analyze the connectivity during placebo and pain relief treatment for arthritis-induced knee pain. Brain activity associated with the placebo effect was recorded in the right midfrontal gyrus circuitry of about half of the participants. In contrast, duloxetine-induced analgesia stimulated activity deep in the right parahippocampal gyrus. In some patients, duloxetine interfered adversely with the placebo effect. The responses of individuals can now be differentiated, and their exposure to ineffective therapies can be monitored. —CA

PLOS Biol. 10.1371/journal.pbio.1002570 (2016).

PHOTO: ALEXANDER SEMENOV

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

INNATE IMMUNITY

Shared logic in diverse immune systems

The innate immune systems of both plants and animals depend on the ability to recognize pathogen-derived molecules and stimulate a defense response. Jones *et al*. review how that common function is achieved in such diverse kingdoms by similar molecules. The recognition system is built for hair-trigger sensitivity and constructed in a modular manner. Understanding such features could be useful in building new pathways through synthetic biology, whether for broadening disease defenses or constructing new signal-response circuits. —PJH

Science, this issue p. 1117

RNA SPLICING

Tie me up, cut me down

Group II introns are mobile genetic elements found in all domains of life. They are large ribozymes that can excise themselves from host RNA. Costa *et al*. determined the structure of an excised group II intron in its branched conformation. This conformation is comparable to the branched "lariat" seen during the splicing of nuclear RNA transcripts. The lariat conformation helps assemble the group II active site for the reverse splicing reaction. The lariat in spliceosomal splicing may also have a similar role in the second step of messenger RNA intron removal. —GR

Science, this issue p. 1118

WATER CHEMISTRY

Frame-by-frame view of acidic transport

Protons in acidic solution constantly hop from one water molecule to the next. In between the hopping, controversy lingers over the extent to which the proton either sticks largely to one water molecule in an Eigen motif or bridges two of them in a Zundel motif. It has been hard to probe this question directly because the distinguishing vibrational bands in bulk aqueous acid spectra are so broad. Wolke *et al*. studied deuterated prototypical Eigen clusters, $D^+(D_2O)_4$, bound to an increasingly basic series of hydrogen bond acceptors (see the Perspective by Xantheas). These clusters displayed sharp bands in their vibrational spectra, highlighting a steadily evolving distortion toward a Zundel-like motif and pointing the way toward further investigations. —JSY

Science, this issue p. 1131; see also p. 1101

NEUROSCIENCE

Attention changes local brain activity

There is a well-known correlation between arousal and neuronal activity in the brain. However, it is unclear how these general effects are reflected on a local scale. Engel *et al*. recorded from higher visual areas in behaving monkeys and discovered a new principle of cortical state fluctuations. A special type of electrodes revealed that the state changes affected neuronal excitability across all layers of the neocortex. When the animals attended to a stimulus, the vigorous spiking states became longer and the faint spiking states became shorter. These states correlated with fluctuations in the local field potential. A sophisticated computational model of the state changes fitted a two-state model of neuronal responsiveness. —PRS

Science, this issue p. 1140

WORKING MEMORY

How to reactivate forgotten memories

Sophisticated techniques can decode stimulus representations for items held in a person's working memory. However, when subjects shift their attention toward something else, the neural representation of the now unattended item drops to baseline, as though the item has been forgotten. Rose *et al*. used single-pulse transcranial magnetic stimulation (TMS) to briefly reactivate the representation of an unattended item. A short pulse of TMS enhanced recognition of "forgotten" stimuli, bringing an unattended item back into focal attention. —PRS

Science, this issue p. 1136

T CELL EXHAUSTION

The epigenetics of exhaustion

During cancer or chronic infection, T cells become dysfunctional, eventually acquiring an "exhausted" phenotype. Immunotherapies aim to reverse this state. Using a mouse model of chronic infection, two studies now show that the epigenetic profile of exhausted T cells differs substantially from those of effector and memory T cells, suggesting that exhausted T cells are a distinct lineage (see the Perspective by Turner and Russ). Sen *et al*. defined specific functional modules of enhancers that are also conserved in exhausted human T cells. Pauken *et al*. examined the epigenetic profile of exhausted T cells after immunotherapy. Although there was transcriptional rewiring, the cells never acquired a memory T cell phenotype. Thus, epigenetic regulation may limit the success of immunotherapies. –KLM

Science, this issue p. 1104, p. 1165; see also p. 1160

EPIDEMIOLOGY

How bats spread viruses

Bats carry numerous viruses, such as rabies and Ebola, which they can transmit to humans. In a Perspective, Hayman highlights recent genetic studies showing that male vampire bats are key to rabies dispersal and transmission in Peru. Rabies is more often transmitted between related species than between unrelated ones. For many other bat-virus systems, little is known about how the virus is transmitted within and between species. Although challenging, further such studies of this and other bat-virus systems are needed to inform public health efforts. —JFU

Science, this issue p. 1099

ATHEROSCLEROSIS

Letting SLE-Ping plaques lie

Patients with the autoimmune disease systemic lupus erythematosis (SLE) are more likely to develop atherosclerosis than healthy individuals. Smith *et al*. hypothesized that invariant natural killer T (iNKT) cells contribute to this process because of their connection to both immune responses and lipids. The authors found that iNKT cells from SLE patients with asymptomatic plaque (SLE-P) produced more of the Th2 cytokine interleukin-4 than those from SLE patients with no plaques. These SLE-P iNKT cells were associated with changes in lipid composition and monocyte skewing to the M2 phenotype. Thus, SLE-P iNKT cells may connect changes in lipids and the immune response, contributing to the development of cardiovascular disease in SLE patients. —ACC

Sci. Immunol. **1**, eaah4081 (2016).

REVIEW SUMMARY

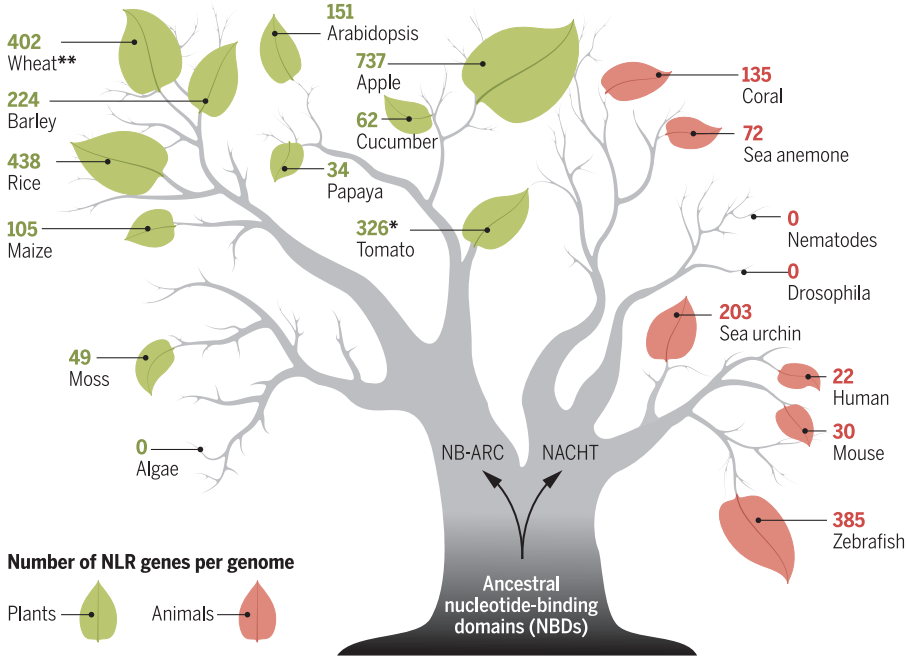
INNATE IMMUNITY

Intracellular innate immune surveillance devices in plants and animals

Jonathan D. G. Jones,*† Russell E. Vance,*† Jeffery L. Dangl*†

BACKGROUND: Pathogens cause agricultural devastation and huge economic losses. Up to 30% of our crops are lost before or after harvest to pathogens and pests, wasting water and human effort. Diseases and pests are major problems for sustainable agriculture in the face of population growth. Similarly, microbial infection remains a major cause of human mortality and morbidity, responsible for ~25% of deaths worldwide in 2012. We lack vaccines for several major infectious diseases, and antibiotic resistance is an ever-growing concern.

Plant and animal innate immune systems respond to pathogen infection and regulate beneficial interactions with commensal and symbiotic microbes. Plants and animals use intracellular proteins of the nucleotide binding domain (NBD), leucine-rich repeat (NLR) superfamily to detect many kinds of pathogens. Plant and animal NLRs evolved from distinct derivatives of a common ancestral prokaryotic adenosine triphosphatase (ATPase): the NBD shared by APAF-1, plant NLR proteins, and CED-4 (NB-ARC) domain class and that shared by apoptosis inhibitory protein (NAIP), CIITA, HET-E, TP1 (NACHT) domain class, respectively. Animals and fungi can carry both NB-ARC and NACHT domain proteins, but NACHT domain proteins are absent from plants and several animal taxa, such as *Drosophila* and nematodes. Despite the vast evolutionary distance between plants and animals, we describe trans-kingdom principles of NLR activation. We propose that NLRs evolved for pathogen-sensing in diverse organisms because the flexible protein domain architecture surrounding the NB-ARC and NACHT domains facilitates evolution of "hair trigger" switches, into which a virtually limitless number of microbial detection platforms can be integrated.



NLR tree. Evolution of NLR genes followed diverging pathways for plant and animal species. Numbers of NLR genes per genome identified computationally range widely, as shown on this stylized evolutionary tree (branches not to scale). The numbers of NLRs can vary markedly even across genomes from closely related taxa. NLRs likely derived from a common ancestor that expressed both NACHT and NB-ARC type NBDs. NACHT is found in animal NLRs, and NB-ARC in plant NLRs. Both occur in fungi. A variety of N- and C-terminal domains have been evolutionarily recruited onto NBDs, including those characteristic of NLRs. The asterisk for tomato indicates that experimental evidence exists to give this precision, as discussed in the main text. The double asterisk for wheat indicates the number of NLRs per diploid genome (wheat is hexaploid). NLR-like fungal proteins lack the LRR domain characteristic of NLRs and are thus not included here.

ADVANCES: Structural biology is beginning to shed light on pre- and postactivation NLR architectures. Various detection and activation platforms have evolved in both plant and animal NLR surveillance systems. This spectrum ranges from direct NLR activation, through binding of microbial ligands, to indirect NLR activation after the modification of host cellular targets, or decoys of those targets, by microbial virulence factors. Homo- and heterotypic dimerization and oligomerization of NLRs add complexity to signaling responses and can enable signal amplification. NLR population genomics across the plant and animal kingdoms is increasing owing to application of new capture-based sequencing methods. A more complete catalog of NLR repertoires within and across species will provide an enhanced toolbox for exploiting NLRs to develop therapeutic interventions.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaf6395

OUTLOOK: Despite breakthroughs in our molecular understanding of NLR activation, many important questions remain. Biochemical mechanisms of NLR activation remain obscure. Events downstream of plant NLR activation and outputs such as transcription of defense genes, changes in cell permeability, localized cell death, and systemic signaling remain opaque. We do not know whether activated plant NLRs oligomerize or, if they do, how this is achieved, given the diversity of subcellular sites of activation observed for various NLRs. It is not clear whether and how the different N-terminal domains of plant NLRs signal. We have increasing knowledge regarding how animal NLRs assemble and signal, although knowledge gaps remain. Therapeutic interventions in humans targeting NLRs remain on the horizon. Design of novel recognition capabilities and engineering of new or extended NLR functions to counter disease in animals and plants provides tantalizing future goals to address plant and animal health problems worldwide. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: jonathan.jones@sainsbury-laboratory.ac.uk (J.D.G.J.); rvance@berkeley.edu (R.E.V.); dangl@email.unc.edu (J.L.D.)
Cite this article as J. D. G. Jones *et al.*, *Science* **354**, aaf6395 (2016). DOI: 10.1126/science.aaf6395

REVIEW

INNATE IMMUNITY

Intracellular innate immune surveillance devices in plants and animals

Jonathan D. G. Jones,[1]*† Russell E. Vance,[2]*† Jeffery L. Dangl[3]*†

Multicellular eukaryotes coevolve with microbial pathogens, which exert strong selective pressure on the immune systems of their hosts. Plants and animals use intracellular proteins of the nucleotide-binding domain, leucine-rich repeat (NLR) superfamily to detect many types of microbial pathogens. The NLR domain architecture likely evolved independently and convergently in each kingdom, and the molecular mechanisms of pathogen detection by plant and animal NLRs have long been considered to be distinct. However, microbial recognition mechanisms overlap, and it is now possible to discern important key trans-kingdom principles of NLR-dependent immune function. Here, we attempt to articulate these principles. We propose that the NLR architecture has evolved for pathogen-sensing in diverse organisms because of its utility as a tightly folded "hair trigger" device into which a virtually limitless number of microbial detection platforms can be integrated. Recent findings suggest means to rationally design novel recognition capabilities to counter disease.

Multicellularity creates nutrient niches for microbial colonization, which in turn drives natural selection for hosts with effective innate immunity. In plants and animals, innate immunity involves both cell surface receptors (*1*) and intracellular receptors of the NLR [nucleotide binding domain (NBD) and leucine-rich repeat (LRR)] superfamily (*2*–*4*). NLRs play critical roles in organismal health in both plants and animals; absence or dysfunction can result in organ failure and death (Fig. 1). NLRs were originally referred to as "Nod-like receptors," but we do not favor this name because it arose only in the mammalian literature, and the official consensus nomenclature is that NLR stands for "NBD-LRR domain-containing" (*5*). Plant NLRs are present in angiosperms and gymnosperms, and even in bryophytes and liverworts, but not in the single-celled alga *Chlamydomonas*. NLR immune receptors are also found in diverse animals, from corals, sea urchins (*6*, *7*), and primitive chordates (*8*) to fish (*9*) and mammals (*10*). Even though chordates and plant genomes carry NLR genes, genes encoding these proteins have not been found in several animal lineages such as nematodes and arthropods. Yet, NLR-like proteins with similar core architecture, but lacking LRR domains, are present in filamentous fungi, where they can play a role in heterokaryon incompatibility (*11*, *12*). Thus, NLRs and fungal NLR-like proteins represent a protein architecture deployed across kingdoms for host defense and/or self-nonself discrimination. In this Review, we focus on an illustrative handful of the most exciting current conceptual developments in NLR biology and refer the reader to excellent recent publications for further details (*13*–*19*). We hope to convey the enthusiasm of this rapidly advancing field as an area of active basic research that is at the cusp of exploitation to address pressing plant and animal health problems worldwide.

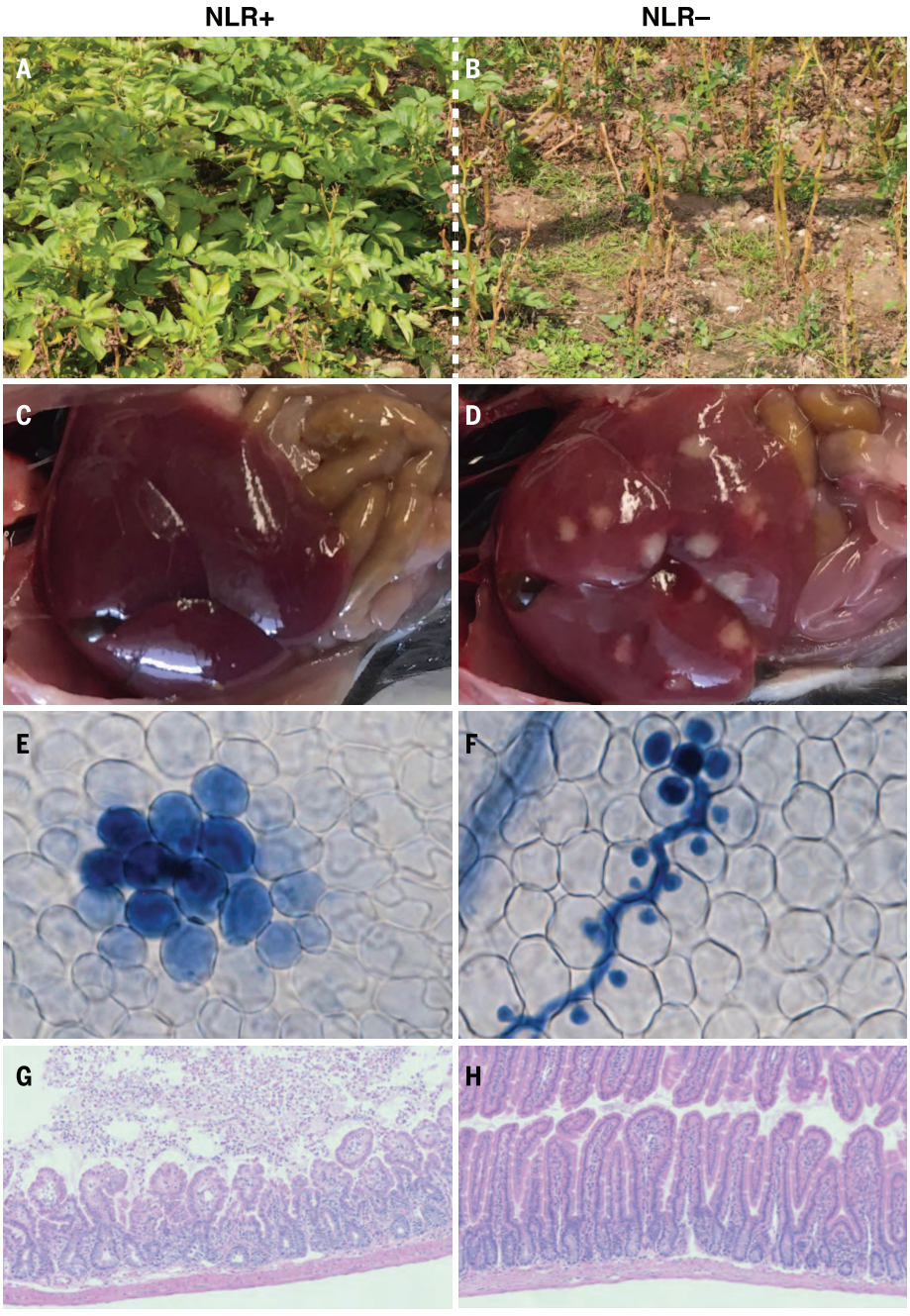
[1]Sainsbury Lab, Norwich Research Park, Colney Lane, Norwich NR4 7UH, UK. [2]Howard Hughes Medical Institute (HHMI), Department of Molecular and Cell Biology, and Cancer Research Laboratory, University of California, Berkeley, CA 94720, USA. [3]Howard Hughes Medical Institute, Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3280, USA.
*These authors contributed equally to this work. †Corresponding author. Email: jonathan.jones@sainsbury-laboratory.ac.uk (J.D.G.J.); rvance@berkeley.edu (R.E.V.); dangl@email.unc.edu (J.L.D.)

NLR architecture: Evolution sculpts sensitive switches

Plant and animal NLRs share a similar modular domain architecture, including the core NBD and LRR domain, although in both clades there is substantial diversity in N- and C-terminal accessory domains (Fig. 2). The NBD falls within the STAND [signal transduction adenosine triphosphatase (ATPases) with numerous domains] AAA^+ ATPase superfamily, which typically includes Walker A (P-loop) and Walker B motifs involved in nucleotide binding and hydrolysis (*20*, *21*). The NBD is associated with adjacent α-helical domains (*22*–*24*). Plant NLRs use a subtype of STAND NBD called the NB-ARC (nucleotide-binding, Apaf1, Resistance, CED4), associated with two α-helical domains. This domain, also known as the Apoptotic ATPase (Ap-ATPase) domain (*25*), is shared with animal proteins that lack LRRs and are involved in apoptosis such as mammalian Apaf-1, *Drosophila* DARK and nematode CED4, and likely evolved from a class of prokaryotic ATPases. Animal NLRs, in contrast, carry a distinct NBD subtype, the NACHT (NAIP, CIITA, HET-E, and TP1) domain, associated with three α-helical domains, that also likely derived from a distinct prokaryotic ancestral domain (*20*, *22*, *26*). Both NACHT and NB-ARC domains are in fungi, where they have recruited diverse N- and C-terminal domains but not LRRs (*12*). Thus, plant and animal NLRs likely evolved from distinct ancestral NBD lineages based on differential expansion from a common ancestor of these STAND AAA^+ ATPases (*27*).

Although the focus of this Review is on NLRs involved in innate pathogen-detection, some mammalian NLRs appear to have distinct functions, including transcriptional regulation in adaptive immunity (*28*, *29*). Nevertheless, despite considerable NLR diversity in sequence and function, all NLR and NLR-like proteins are presumed to involve a similar switch-like activation mechanism. Indeed, studies of Apaf1 and its homologs have established the paradigm for our current understanding of NLR activation (*30*). In this model, preactivation states of NLR proteins feature intra- and potentially intermolecular domain interactions to keep the NBD conformational equilibrium in a suppressed but not fully inactive state (*17*, *31*, *32*). In response to specific pathogen effector (virulence) proteins or other specific stimuli, the intramolecular interactions are altered, and the NBD is believed to exchange adenosine diphosphate (ADP) for adenosine triphosphate (ATP), likely driving NLR oligomerization in at least some cases. NLRs can hydrolyze ATP to ADP, but this activity does not drive oligomerization. Whether ATP hydrolysis plays an important role in NLR regulation is unclear; ATP hydrolysis may convert activated NLRs to an inactive state.

NLR oligomerization is believed to initiate signaling by the proximity-induced recruitment and activation of downstream molecules via N-terminal accessory signaling domains (*33*). These N-terminal domains vary considerably (Fig. 2) but are commonly coiled-coil or TIR domains in plant NLRs, or domains in the death-fold superfamily (such as CARD or Pyrin domains) in animal NLRs. The putative signaling molecules recruited to plant NLRs have not been identified, but several such molecules recruited to animal NLRs have been described. These include a kinase (RIPK2) that is recruited to NOD1/2, a protease (caspase-1) that is recruited directly to NLRC4 and NLRP1, and a Pyrin-CARD–containing adaptor protein (ASC) that recruits caspase-1 indirectly to several NLRs, including NLRP3. Interestingly, the pyrin domain of NLRP3 is also believed to propagate signaling by nucleating the oligomerization of ASC into polymerized filaments that coalesce into massive intracellular "specks" (*34*). The essentially irreversible conversion of ASC from a soluble to filamentous form is reminscent of the biochemical behavior of prions, and indeed, the ASC Pyrin domain exhibits prion-like properties when expressed in yeast. Mutations in ASC that disrupt its prion activities in yeast also abrogate its ability to signal in mammalian cells (*35*). Conversely, a yeast prion domain can functionally replace the N-terminal Pyrin domain in ASC (*35*). Highly cooperative polymerization that produces a "hair-trigger" all-or-none signaling output might be



CREDITS: (A) AND (B), KAMIL WITEK/(C) AND (D), VIVIEN MALTEZ AND ED MIAO/(E) AND (F), JANE PARKER/(G) AND (H), ISABELLA RAUCH

Fig. 1. NLRs make a major contribution to organismal health. (**A** and **B**) Isogenic potato plants that either (A) express or (B) lack a specific NLR (*Rpi-vnt1*) conditioning resistance to the oomycete pathogen *Phytophthora infestans*, the causal agent of potato late blight. (**C** and **D**) Livers from (C) normal NLRC4^{+} (resistant) and (D) NLRC4-deficient (susceptible) mice after lethal infection with *Chromobacterium violaceum* (white bacterial lesions are visible). (**E**) *Arabidopsis* expressing the NLR-encoding *RPP5* gene undergoes protective localized cell death (trypan blue–stained cells; debris of dead pathogen stained intensely in the center) in response to the *Arabidopsis* downy mildew pathogen (*Hyaloperonospora arabidopsidis*) strain NoCo2. (**F**) Successful downy mildew colonization of leaf tissue from isogenic susceptible *Arabidopsis* that lack *RPP5* (trypan blue–stained hyphae and intracellular pathogen haustoria). (**G**) Destruction of mouse intestinal tissue after NLRC4 activation showing sloughing of epithelial cells into the intestinal lumen. The sloughing response is believed to prevent bacterial invasion into deeper tissue. (**H**) Normal mouse intestinal tissue showing elongated villi and intact intestinal epithelium.

desirable in proteins that need to respond rapidly and sensitively to pathogen invasion. It is tempting to speculate that polymerization may be a common feature of signaling downstream of both plant and animal NLRs, although this remains to be established experimentally.

In both plant and animal NLRs, deletion of the LRR domain can result in constitutive NLR activation. Thus, a primary function of the LRR domain is likely to be negative regulation of NBD-mediated oligomerization. A structure of mouse NLRC4 suggests that autoinhibition is at least partly mediated by direct contact between the NBD and LRR (*36*), although whether this is generalizable to all NLRs remains to be seen. Once activated, NLRs often induce a characteristic cell death response termed the "hypersensitive response" in plants and "pyroptosis" in animals. Cell death is believed to restrict pathogen replication at the site of infection and, in animals, results in the release of soluble mediators that recruit and activate additional immune cells. Inappropriate spontaneous NLR activation can lead to autoimmune conditions in both plants and animals. These syndromes can be more severe at cold temperatures in both plants and animals, resulting in chilling sensitive ectopic cell death in plants and familial cold autoinflammatory syndrome (FCAS) in humans (*37*, *38*). These examples suggest that intra- or intermolecular NLR interactions required for autoinhibition can be perturbed at the nonpermissive temperature, or that ATP hydrolysis is attenuated at lower temperatures, shifting the equilibrium from the inactive toward the activated state.

Although the induced oligomerization model has underpinned investigations of animal NLR activation, it has not been demonstrated for plant NLRs. Activation of the tobacco N and *Arabidopsis* RPP1 NLRs correlates with self-association (*39*, *40*), but whether this represents the formation of oligomers, as in the case of Apaf1 and some animal NLRs, or simply homodimers remains unclear. There may be diversity in both the resting-state architectures of plant NLRs and in their modes of activation by microbial signals that are tuned by natural selection; no universal generalization may exist. For example, heterogeneity in the subcellular site of activation of a given NLR is potentially determined by the precise localization of the microbial effector target that activates it. There may be additional structural constraints imposed by requirements for interaction with partner "helper" NLRs, with microbial effector target domains, or decoys of these, or because of integration of target decoy domains into the NLR itself (*3*, *41*–*44*).

Subsequent to NLR activation, plant and animal innate immunity mechanisms likely differ greatly, although we remain quite ignorant of postactivation mechanisms in plants. Nevertheless, despite remarkable diversity in upstream and downstream signaling events, we are forced to contemplate what is so fundamentally advantageous about the NLR architecture that could explain why it arose convergently in plants and animals to play a role in pathogen detection and

defense activation (*27*, *45*). The full spectrum of mechanisms in each kingdom suggests that there is scope for more conceptual similarities than previously suspected. Given this diversity, we propose that one advantage of the NLR architecture may simply be its ability to function as a robust on-off switch in diverse signaling contexts. To illustrate the flexibility of the NLR architecture, we articulate below four distinct mechanisms of pathogen sensing ("direct," "guard," "decoy," and "integrated decoy") (Fig. 3) and discuss how these four mechanisms can be applied to individual and paired plant and animal NLRs. It remains to be seen whether these four mechanisms are the main modes of action of all NLRs or whether we are only scratching the surface.

Division of labor: Sensors and helpers

Specific NLR responses can require a pair of NLR proteins in which one senses the ligand whereas the other (the "helper NLR") is required for its downstream signaling (*46*, *47*). In mammals, the NAIP/NLRC4 inflammasomes are composed of sensor/helper NLR pairs. In this system, a NAIP sensor NLR is activated by direct binding to a specific bacterial protein ligand (such as flagellin) (*48*, *49*), leading to recruitment of NLRC4 as a helper NLR. NLRC4 does not appear to bind directly to ligands but instead functions downstream of NAIPs to recruit and activate caspase-1, a key executioner of inflammasome signaling pathways. Structural analyses demonstrate that in the absence of stimulation, NLRC4 (and presumably NAIPs) are retained in the cytosol as monomers, autoinhibited via intradomain interactions (*36*, *50*, *51*). Recognition of microbial molecules by NAIPs exposes a donor "catalytic" surface on the NAIP that binds to a "receptor" surface on NLRC4, provoking a conformational change in NLRC4 that exposes its catalytic surface. This in turn propagates the recruitment of additional NLRC4 monomers to form a ring-shaped oligomer that appears to contain one NAIP and 9 to 11 NLRC4 molecules. This striking stoichiometry suggests that NAIP activation is a hair trigger for NLRC4-mediated signal amplification.

The concept of sensors and helpers also applies to some plant NLRs. Plant genomes encode variable numbers of NLRs, with an atypical N-terminal coiled-coil domain called a CC-R (*52*) that correlates with helper NLR function. Resistance to the Tobacco Mosaic Virus requires both the sensor TIR NLR protein N and the helper CC-R NLR protein NRG1 (*53*). In *Arabidopsis*, the five CC-R–encoding genes comprise two paralogous NLR families that function as helper NLRs (*47*, *54*). One of them, a member of the ADR CC-R family, has both canonical, P-loop–dependent signaling functions in cell death control and a noncanonical P-loop–independent function as a helper NLR for several effector sensor NLRs. The noncanonical function suggested a requirement as a scaffold, much like NLRC4, which can also exhibit P-loop–independent functions in transduction of effector-activated NAIP signals (*48*). A given NLR might be operating by both of these mechanisms, depending on the activation context, as demonstrated for ADR1-L2 (*54*). Little is known about how the CC-R domain is integrated into activation mechanisms, but the fact that it is evolutionarily ancient and monophyletic suggests a generalizable function in plant NLR biology.

Natural NLR variants featuring degenerate NBD consensus sequences exist, and there is evidence to suggest that they may participate in noncanonical activation mechanisms. The rice Pb1 NLR family lacks a P-loop motif but nonetheless conditions broad spectrum resistance to rice blast, potentially by acting as a helper NLR (*55*). This is likely evolutionarily conserved, because there are Pb1 homologs in maize. The small collection of *Arabidopsis* and *Arabidopsis lyrata* proteins containing variant P-loop residues is also likely to alter or impair the canonical activation mechanism outlined above (*56*). These include NLRs with integrated decoy domains that function in NLR pairs that are encoded together and function together. For example, in the RPS4/RRS1 gene pair, RRS1 is the sensor NLR, and its P-loop is not required to activate signaling (*57*).

Guards and decoys: Getting the most from the NLR receptor repertoire

Plant NLRs were first revealed by cloning *Resistance* (*R*) genes that confer the capacity to activate defense upon detection of specific pathogen effectors. There is selective pressure for pathogens to mutate their effectors to evade NLR-dependent surveillance, which in turn selects for evolution of either new *R* gene alleles, or other *R* genes, that restore resistance. This "gene-for-gene" coevolution led to the hypothesis that NLR proteins might directly interact with the recognized effector, and in some cases this prediction was fulfilled (*58*). However, in most cases, direct interaction between a plant NLR and the "recognized" effector is not observed. Instead, many plant NLRs appear to monitor the state of self proteins, termed "guardees," whose primary function is in defense signaling and as such are frequently targeted by pathogen effectors. If a pathogen virulence protein alters the guardee's structure, then the associated NLR is activated. The "guard" strategy thus allows a relatively small repertoire of NLRs (~150 in *Arabidopsis*) to protect against diverse pathogen effectors (*3*, *59*, *60*). For example, NLR proteins RPM1 and RPS2 act at the plasma membrane to monitor the state of the plasma membrane–associated defense regulator RIN4. RPM1 detects phosphorylation of a specific threonine residue on RIN4 provoked by the pathogen effectors AvrB or AvrRpm1 (*61*, *62*) and mediated via a receptor-like cytoplasmic kinase (*63*). This effector-modulated phosphorylation interferes with both RIN4-dependent mesophyll defense responses (*62*) and stomatal immunity (*64*). In contrast, RPS2 is activated by cleavage of RIN4 by the bacterial cysteine protease effector AvrRpt2 (*65*, *66*).

Similarly, RPS5 monitors the state of protein kinase PBS1, also at the plasma membrane (*67*, *68*). PBS1 is targeted for proteolytic cleavage by plasma membrane–localized AvrPphB. Illustrating the

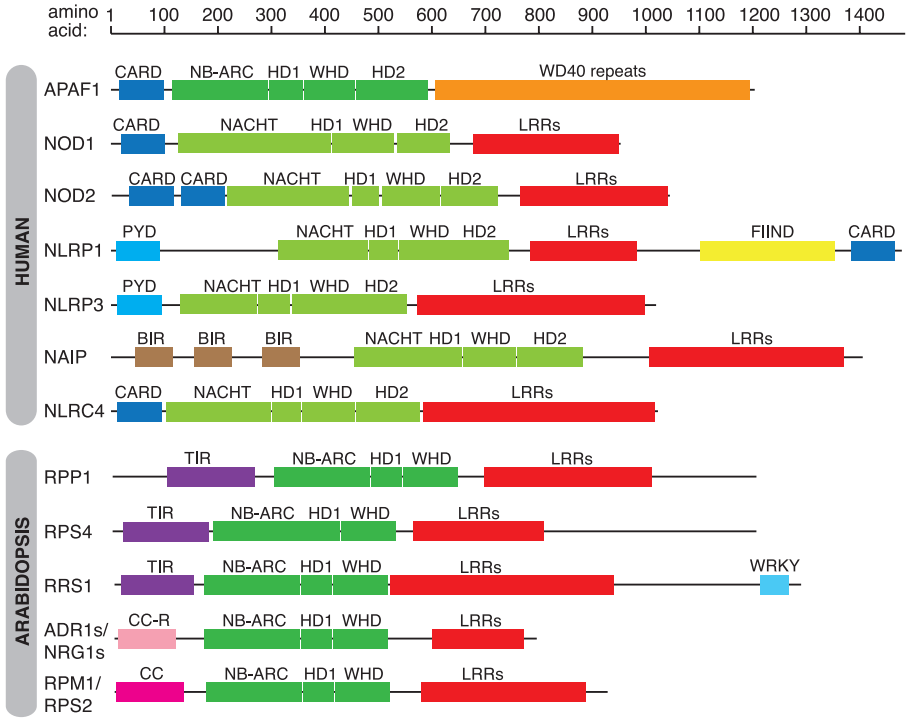


Fig. 2. Diversity of NLR and NLR-like architectures ("NLR-o-gram"). Domain structures of representative well-studied NLR proteins from human and *Arabidopsis* are shown approximately to scale. Definitions of acronyms are provided in Box 1. The NACHT and NB-ARC domains are sometimes defined as including the associated helical domains, but these domains are shown separately here for clarity. Humans contain additional NLRs not known to be directly involved in pathogen sensing.

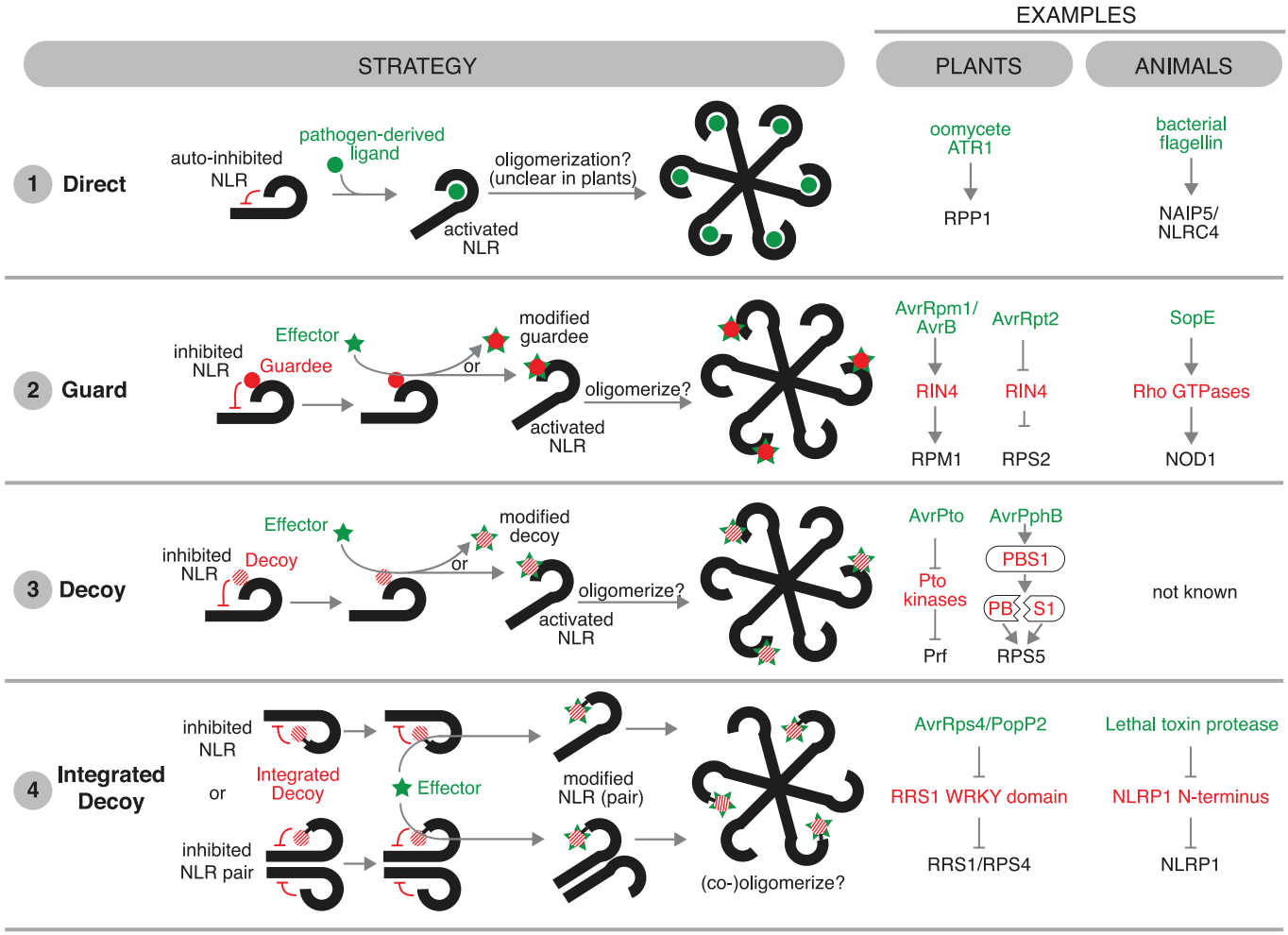


Fig. 3. Diverse strategies for NLR-mediated detection of pathogens. Four conceptually distinct strategies are illustrated. The details of how each strategy is implemented for a specific NLR example may vary. The guard and decoy strategies are analogous: In both cases, the guardee or decoy proteins are involved in maintaining the NLR in an inhibited state, and in both cases, the inhibition is relieved upon effector-mediated modification of the guardee or decoy. Guardees are distinguished from decoys by having an additional and separate function in host defense, whereas decoys are merely mimics of host defense proteins. Guardees are thus the "intended" targets of effectors, whereas decoys are inadvertently targeted by effectors.

flexibility of the guard strategy, the RPS5/PBS1 system has been engineered to expand RPS5 function to recognize other protease cleavage sites engineered into PBS1 (*69*).

Mutations of a guarded protein, such as PBS1, can result in no enhanced disease susceptibility. This suggests that paralogs of genuine targets of virulence proteins may have evolved to resemble that target, thus "luring" pathogens to reveal themselves by their action on a protein that is not required for defense. This concept was elaborated as the "decoy" model (*41*). Given the extensive functional redundancy of plant proteins, it is formally difficult to distinguish between redundant guardees and decoys; absence of evidence that a given protein is involved in host defense is not sufficient evidence to prove that it is a decoy. Nevertheless, PBS1 is likely a decoy for a class of cytosolic kinases that function during defense activated by cell surface receptors (*63*, *70*). Additionally, tomato NLR Prf has evolved an extended N-terminal domain of unknown function that acts as a detection platform with which to monitor effector-targeted protein kinases of the Pto family that are, in turn, likely decoys for the defense-relevant kinase domains of pattern recognition receptor LRR-kinases (*71*). Likewise, the *Arabidopsis* NLR ZAR1 monitors the state of multiple pseudokinases (*72*) that are likely decoys for effectors that target receptor-like cytoplasmic kinases involved in defense signaling.

In mammals, the concepts of guards and decoys have not been extensively discussed. The mammalian NOD1 and NOD2 NLRs were originally proposed to function as direct detectors of bacterial peptidoglycan fragments, which is consistent with a direct ligand-receptor model for NLR activation in animals (*73*–*75*). However, crystallographic proof that NOD1 and NOD2 bind directly to peptidoglycan-derived ligands is lacking. Instead, NOD1 and NOD2 might indirectly sense pathogens via responsiveness to disruption of host cell physiology (*76*, *77*). For example, activation of the host cytoskeletal regulators Rac1 and Cdc42 by the secreted *Salmonella* virulence factor SopE results in NOD1 activation (*76*). NOD1 and NOD2 also appear to be responsive to pathogen-induced ER stress (*77*). These results are consistent with NOD1 and NOD2 exhibiting guard-type activation, although it remains unclear how this is to be reconciled with genetic evidence that NOD1 and NOD2 can also be activated by peptidoglycan.

Mammalian NLRP3 also acts as a guard of cellular integrity because it can be activated by disruption of cellular ion gradients. The molecular mechanism of NLRP3 activation remains unresolved, but the essential role of NEK7 kinase in NLRP3 activation (*78*–*80*) is reminiscent of the role of protein kinases in plant NLR activation. Thus, the guard-type activation mechanism first described in plants may also be germane to NLR-mediated pathogen detection in animals.

Integrated decoys

An evolutionary challenge for sensor/helper, guard/guardee, or guard/decoy NLR systems is that if the corresponding *NLR* genes are unlinked and exhibit allelic variation, inappropriate allelic

Box 1. Definitions for the acronyms used in this paper.

ADR1: Activated disease resistance 1
ASC: Apoptosis-associated speck-like protein containing a CARD
BIR: Baculovirus inhibitor-of-apoptosis repeat
CARD: Caspase activation and recruitment domain
CC-R: Coiled coil domain-RPW8-like
CC: Coiled coil
FIIND: Function-to-find domain
HD1/2: Helical domain 1/helical domain 2
HMA: Heavy metal associated (same domain as RATX)
LRRs: Leucine-rich repeats
NACHT: NAIP, CIITA, HET-E, TP1 domain
NAIP: NLR family, apoptosis inhibitory protein
NB-ARC: Nucleotide binding domain shared by APAF1, R genes, CED-4
NBD: Nucleotide binding domain
NLR: Nucleotide binding domain, leucine-rich repeat–containing
NLRC4: NLR family, CARD domain–containing 4
NLRP1/3: NLR family, Pyrin domain–containing 1 and 3
NOD1/2: Nucleotide-binding oligomerization domain–containing 1 and 2
PBS1: AvrPphB susceptible 1
PYD: Pyrin domain
RIPK2: Receptor interacting protein serine/threoning kinase 2
RATX: Related to ATX1 (same domain as HMA)
RGA5: Resistance gene analog5
RIN4: RPM1 interacting protein 4
RPM1: Resistance to Pseudomonas maculicola 1
RPP1: Resistance to Peronospora parasitica 1
RPS4/5: Resistance to Pseudomonas 4 and 5
RRS1: Resistance to Ralstonia solanacearum 1
STAND: Signal transduction ATPases with numerous domains
TIR: Toll-like, interleukin-1 receptor resistance protein
WD40: tryptophan (W), aspartate (D) 40 amino acids
WHD: Winged Helix Domain
WRKY: tryptophan (W), arginine (R), lysine (K), tyrosine (Y) motif-containing domain

combinations may be generated that can result in microbe-independent autoimmunity (*81*, *82*). Such untoward consequences may facilitate selection for genetic linkage of *NLR* pairs that function together. Intriguingly, in multiple plant genomes a given *NLR* may be closely linked and divergently transcribed from another *NLR* gene that is required for its function. For example, rice RGA4 and RGA5 are two linked NLRs required for recognition of two effectors from the rice blast pathogen (*Magnaporthe oryzae*) (*83*). Similarly, *Arabidopsis* RPS4 and RRS1 are two linked NLRs that are both required to confer recognition of two bacterial effectors, AvrRps4 and the YopJ family acetyltransferase PopP2, and an unidentified fungal molecule (*84*). In both of these examples, the two NLR components appear to be preassociated (*57*) rather than associating only upon effector perception, as occurs with NAIPs and NLRC4 (*48*).

RRS1 carries a WRKY transcription factor (TF) DNA binding domain toward its C terminus. *Arabidopsis* encodes ~90 WRKY TFs, many of which are implicated in innate immunity in plants (*85*). Conceivably, then, WRKY proteins might be targets for pathogen effectors because their inactivation should result in elevated susceptibility. Recent data (*86*, *87*) fulfill this expectation: The effector PopP2 acetylates two lysines in the canonical WRKYGQK DNA-binding motif, and the GQK lysine is crucial for RPS4/RRS1–dependent defense activation. Another unrelated bacterial effector, AvrRps4, also binds to the WRKY domain. The RPS4/RRS1 complex converts the effector-dependent modification of the RRS1 WRKY domain into defense activation. Integration of the guarded decoy domain into RRS1, an indispensable and linked partner to RPS4, reduces the risk of recombination giving rise to inappropriate allelic combinations of what could be three unlinked proteins, which might result in autoimmunity. There are at least nine such linked NLR pairs in the reference *Arabidopsis* genome. Importantly, these exhibit decoy domain diversity at orthologous positions across the Brassicaceae, suggesting that the rapid shuffling of integrated decoy domains onto existing functional *NLR* pairs is a useful evolutionary strategy.

The discovery of linked paired NLRs in plants creates a new opportunity. Functional transfer of plant NLRs across species barriers has proven largely impossible. This restricted taxonomic functionality is poorly understood but, if solved, could greatly enhance prospects for crop disease control through genetics rather than chemistry. For example, RPS4 and RRS1 confer effector recognition and *Colletotrichum* fungus resistance when cotransformed into Solanaceae or Cucurbitaceae (*88*). This is consistent with the idea that restricted taxonomic functionality for one NLR arises from a requirement for the appropriate helper or partner NLR. The additional required NLRs are usually hard to identify, but in the case of paired NLR genes, comprise each other. Systematic transfer of paired NLRs between plant taxa may provide additional recognition capacities that would enable elevated crop disease resistances.

Overcoming taxonomic functionality restriction may require more than paired NLR genes. The tomato Prf/Pto guard/guardee pair confers recognition of two widespread *Pseudomonas* effectors but does not appear to function outside the Solanaceae. Prf/Pto function requires the helper NLRs NRC2a, 2b, and NRC3 (*46*). A fuller understanding of how sensors functionally integrate with helpers is required to rationally expand and transfer useful disease resistance.

The concept of integrated decoys may be widely applicable (*42*). Genome-wide analyses of plant NLR genes led to the discovery of many integrated domains in plant NLR proteins (*42*–*44*). There is an overlap between the list of integrated domains and the list of domains found to be frequent interactors of pathogen effectors in large-scale yeast 2-hybrid screens (*43*, *89*). This correlation is consistent with the view that selection favors integration into NLRs of protein domains that are targets, or decoys of targets, of pathogen effectors. For example, the rice NLR RGA5 carries a C-terminal RATX1 (related to yeast copper transporter ATX1) or HMA (heavy metal–associated) domain and likely binds metals. Effectors AVR-Pia and AVR1-CO39 from *M. oryzae* interact with this domain and trigger RGA4-dependent defense. Another rice gene pair, Pikp-1 and Pikp-2, recognizes a different effector, Avr-Pik, which exists in the fungal population as a series of alleles (AvrPikA-D). Structural studies have illustrated how effectors interact with the HMA domain that is located between the CC and the NB-ARC domain of Pikp-1 (*90*). But why do pathogens evolve effectors that interact with HMA domain proteins? A clue is provided by the observation that the recessive disease-resistance gene *Pi21* contains an HMA domain (*91*). Recessive disease resistance genes are typically interpreted as "Susceptibility" (*S*) genes encoding proteins in the host required for pathogen proliferation. Genes that encode such effector targets are promising candidates for genome editing; loss of function of the Pi21 HMA domain protein results in enhanced disease resistance. The role of HMA proteins in susceptibility remains to be established, but conceivably the metal-binding domain may influence host cell redox status, resulting in a more congenial environment for the pathogen.

Integrated decoy mechanisms have not been implicated in activation of mammalian NLRs. However, a conceptually similar mechanism appears

to underlie activation of mouse and rat NLRP1, an NLR that is activated by lethal factor, a virulence factor secreted by the anthrax bacterium *Bacillus anthracis* (*92*). Lethal factor is a protease that anthrax uses to degrade mitogen-activated protein kinases, kinases involved in host defense. As a countermeasure, certain rodent NLRP1 proteins are activated in response to direct cleavage by lethal factor. Cleavage results in removal of an N-terminal fragment of NLRP1 that normally holds NLRP1 in the "off" state, and cleavage of NLRP1 has been shown to be both necessary and sufficient to activate NLRP1B (*93*, *94*). Thus, the NLRP1 N terminus appears to behave like an integrated decoy domain, mimicking the cleavage site of the true effector targets. Unlike rodent NLRP1 proteins, human NLRP1 carries a pyrin domain at its N terminus. Interestingly, because NLRP1 signals via its C-terminal CARD, the N-terminal Pyrin domain is not required for signaling and currently is of unknown function (*95*). It is tempting to speculate that it serves as an integrated decoy to detect putative pyrin-targeting effector proteins.

Mining NLR repertoires and recognition strategies across taxa

Defining the repertoire of NLRs across species, or even across genotypes of the same species, requires an assessment of genetic diversity. We usually lack comprehensive catalogs of the diversity of NLRs present in species or strains of interest. Because NLR repertoire diversity is important for disease resistance in natural populations, definition of the pan-NLRome of any plant or animal species will catapult us beyond the limited understanding obtained by sequencing a single reference genome. In addition to the paired *NLR* genes described above, NLR-encoding genes typically exist as clustered families of closely related paralogs, or as true allelic series. Assembly of short-read whole-genome data often results in assembly errors in *NLR* loci. Sequence capture enables *NLR* gene enrichment sequencing (RenSeq) (*96*), and long-read DNA sequencing technology enables reads of complete NLRs to be obtained (*97*). Using biotinylated RNA probes designed to capture the repertoire of 450 *NLRs* predicted to be in the reference diploid potato genome, 750 NLRs were identified (*96*). These kinds of analyses on multiple accessions of plant species, combined with RenSeq applied to cDNA, will facilitate better insight into the repertoire of and genetic variation in NLRs, including those with integrated domains. Combined with mutagenesis, these methods accelerate isolation of useful resistance genes (*98*). Because integrated domains are envisaged as effector targets, a widespread understanding of their diversity will result in a broader understanding of the cellular processes usurped by pathogens (*43*). Plant genomes carry many genes encoding N- and C-terminal truncations of the canonical NLR structure, and these are also captured with RenSeq. Such "pieces" of NLRs do not fit the current mechanistic activation paradigm outlined above because many lack an active NBD. Nevertheless, their sequences are not degenerating into pseudogenes, and some are capable of signaling when overexpressed or mutated and are likely to contribute to overall NLR functional diversity (*47*, *99*).

Sequencing approaches such as RenSeq have not yet been applied to catalog animal NLR repertoires. One might anticipate that sequence capture methods would reveal considerable diversity in domain architectures as well as polymorphism, thus revealing some NLR families that are under diversifying selection and that are likely to be coevolving with pathogens in extremely large populations. Primitive vertebrates and invertebrates can carry large families of NLRs. For example, the coral *Acropora digitifera* carries ~500 NLRs, *Amphioxus* carries ~118 NLRs, and sea urchins carry ~203. There is no knowledge of the extent of diversity of NLRs in populations of these animals.

Mammals also display considerable diversity between species in their NLR repertoires. For example, unlike mice, the human reference genome appears to lack multiple NAIP paralogs, although it will be interesting to know whether this is true across all human populations. Short-read exome sequences are often difficult to assemble for paralogous and repetitive gene families; thus, RenSeq-type approaches could be valuable for assessing animal NLR diversity. Little is known about intraspecific variation in mammalian NLR repertoires, and sequence capture of NLRs could reveal diversity in innate immune surveillance capacity. Such insights could also be useful to engineer novel pathogen-resistance into animal species. For example, the NLRC4/NAIP alleles of appropriate strains of mice could enable enhanced resistance to *Salmonella* if expressed in transgenic chickens.

Toward synthetic detection platforms

Despite breakthroughs in our molecular understanding of NLR activation, knowledge of subsequent signaling steps and mechanisms remains weak. The pathways that connect NLR activation to outputs such as transcription of defense genes, changes in cell permeability, localized cell death, and systemic signaling remain poorly understood. Do activated, or dimerized, or oligomerized plant NLRs recruit new signaling proteins? How distinct are the signaling pathways controlled by the various N-terminal signaling domains recruited to the NLR chassis during evolution? Are integrated decoy domain NLRs modular? Can we engineer new or additional decoy domains into them to create or extend NLR function? As more structural and mechanistic information emerges on how plant and animal NLRs function, the engineering of novel, bespoke, and useful recognition capacities in plant and animal immune systems will become a more realistic goal.

REFERENCES AND NOTES

1. P. C. Ronald, B. Beutler, Plant and animal sensors of conserved microbial signatures. *Science* **330**, 1061–1064 (2010). doi: 10.1126/science.1189468; pmid: 21097929
2. F. Jacob, S. Vernaldi, T. Maekawa, Evolution and conservation of plant NLR functions. *Front. Immunol.* **4**, 297 (2013). doi: 10.3389/fimmu.2013.00297; pmid: 24093022
3. J. D. G. Jones, J. L. Dangl, The plant immune system. *Nature* **444**, 323–329 (2006). doi: 10.1038/nature05286; pmid: 17108957
4. E. Elinav, T. Strowig, J. Henao-Mejia, R. A. Flavell, Regulation of the antimicrobial response by NLR proteins. *Immunity* **34**, 665–679 (2011). doi: 10.1016/j.immuni.2011.05.007; pmid: 21616436
5. J. P.-Y. Ting *et al*., The NLR gene family: A standard nomenclature. *Immunity* **28**, 285–287 (2008). doi: 10.1016/j.immuni.2008.02.005; pmid: 18341998
6. M. Hamada *et al*., The complex NOD-like receptor repertoire of the coral *Acropora digitifera* includes novel domain combinations. *Mol. Biol. Evol.* **30**, 167–176 (2013). doi: 10.1093/molbev/mss213; pmid: 22936719
7. J. P. Rast, L. C. Smith, M. Loza-Coll, T. Hibino, G. W. Litman, Genomic insights into the immune system of the sea urchin. *Science* **314**, 952–956 (2006). doi: 10.1126/science.1134301; pmid: 17095692
8. S. Huang *et al*., Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Res.* **18**, 1112–1126 (2008). doi: 10.1101/gr.069674.107; pmid: 18562681
9. K. Howe *et al*., Structure and evolutionary history of a large family of NLR proteins in the zebrafish. *Open Biol.* **6**, 160009 (2016). doi: 10.1098/rsob.160009; pmid: 27248802
10. M. Lamkanfi, V. M. Dixit, Mechanisms and functions of inflammasomes. *Cell* **157**, 1013–1022 (2014). doi: 10.1016/j.cell.2014.04.007; pmid: 24855941
11. A. Daskalov *et al*., Signal transduction by a fungal NOD-like receptor based on propagation of a prion amyloid fold. *PLOS Biol.* **13**, e1002059 (2015). doi: 10.1371/journal.pbio.1002059; pmid: 25671553
12. W. Dyrka *et al*., Diversity and variability of NOD-like receptors in fungi. *Genome Biol. Evol.* **6**, 3137–3158 (2014). doi: 10.1093/gbe/evu251; pmid: 25398782
13. S. Maekawa, U. Ohto, T. Shibata, K. Miyake, T. Shimizu, Crystal structure of NOD2 and its implications in human disease. *Nat. Commun.* **7**, 11813 (2016). doi: 10.1038/ncomms11813; pmid: 27283905
14. V. A. K. Rathinam, K. A. Fitzgerald, Inflammasome complexes: Emerging mechanisms and effector functions. *Cell* **165**, 792–800 (2016). doi: 10.1016/j.cell.2016.03.046; pmid: 27153493
15. D. Sharma, T.-D. Kanneganti, The cell biology of inflammasomes: Mechanisms of inflammasome activation and regulation. *J. Cell Biol.* **213**, 617–629 (2016). doi: 10.1083/jcb.201602089; pmid: 27325789
16. R. Caruso, N. Warner, N. Inohara, G. Núñez, NOD1 and NOD2: Signaling, host defense, and inflammatory disease. *Immunity* **41**, 898–908 (2014). doi: 10.1016/j.immuni.2014.12.010; pmid: 25526305
17. Z. Duxbury *et al*., Pathogen perception by NLRs in plants and animals: Parallel worlds. *BioEssays* **38**, 769–781 (2016). doi: 10.1002/bies.201600046; pmid: 27339076
18. H. Wen, E. A. Miao, J. P.-Y. Ting, Mechanisms of NOD-like receptor-associated inflammasome activation. *Immunity* **39**, 432–441 (2013). doi: 10.1016/j.immuni.2013.08.037; pmid: 24054327
19. A. Bentham, H. Burdett, P. A. Anderson, S. J. Williams, B. Kobe, Animal NLRs provide structural insights into plant NLR function. *Ann. Bot.* mcw171 (2016). doi: 10.1093/aob/mcw171; pmid: 27562749
20. D. D. Leipe, E. V. Koonin, L. Aravind, STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: Multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J. Mol. Biol.* **343**, 1–28 (2004). doi: 10.1016/j.jmb.2004.08.023; pmid: 15381417
21. J. E. Walker, M. Saraste, M. J. Runswick, N. J. Gay, Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945–951 (1982). pmid: 6329717
22. E. V. Koonin, L. Aravind, The NACHT family—A new group of predicted NTPases implicated in apoptosis and MHC transcription activation. *Trends Biochem. Sci.* **25**, 223–224 (2000). doi: 10.1016/S0968-0004(00)01577-2; pmid: 10782090
23. E. A. van der Biezen, J. D. G. Jones, The NB-ARC domain: A novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Curr. Biol.* **8**, R226–R227 (1998). doi: 10.1016/S0960-9822(98)70145-9; pmid: 9545207

24. L. Aravind, V. M. Dixit, E. V. Koonin, The domains of death: Evolution of the apoptosis machinery. *Trends Biochem. Sci.* **24**, 47–53 (1999). doi: 10.1016/S0968-0004(98)01341-3; pmid: 10098397
25. E. V. Koonin, L. Aravind, Origin and evolution of eukaryotic apoptosis: The bacterial connection. *Cell Death Differ.* **9**, 394–404 (2002). doi: 10.1038/sj.cdd.4400991; pmid: 11965492
26. M. Albrecht, F. L. W. Takken, Update on the domain architectures of NLRs and R proteins. *Biochem. Biophys. Res. Commun.* **339**, 459–462 (2006). doi: 10.1016/j.bbrc.2005.10.074; pmid: 16271351
27. J.-X. Yue, B. C. Meyers, J.-Q. Chen, D. Tian, S. Yang, Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. *New Phytol.* **193**, 1049–1063 (2012). doi: 10.1111/j.1469-8137.2011.04006.x; pmid: 22212278
28. K. S. Kobayashi, P. J. van den Elsen, NLRC5: A key regulator of MHC class I-dependent immune responses. *Nat. Rev. Immunol.* **12**, 813–820 (2012). doi: 10.1038/nri3339; pmid: 23175229
29. J. P. Y. Ting, J. A. Duncan, Y. Lei, How the noninflammasome NLRs function in the innate immune system. *Science* **327**, 286–290 (2010). doi: 10.1126/science.1184004; pmid: 20075243
30. S. Yuan, C. W. Akey, Apoptosome structure, assembly, and procaspase activation. *Structure* **21**, 501–515 (2013). doi: 10.1016/j.str.2013.02.024; pmid: 23561633
31. J. L. Dangl, D. M. Horvath, B. J. Staskawicz, Pivoting the plant immune system from dissection to deployment. *Science* **341**, 746–751 (2013). doi: 10.1126/science.1236011; pmid: 23950531
32. M. Bernoux *et al*., Comparative analysis of the flax immune receptors L6 and L7 suggests an equilibrium-based switch activation model. *Plant Cell* **28**, 146–159 (2016). pmid: 26744216
33. O. Danot, E. Marquenet, D. Vidal-Ingigliardi, E. Richet, Wheel of Life, Wheel of death: A mechanistic insight into signaling by STAND proteins. *Structure* **17**, 172–182 (2009). doi: 10.1016/j.str.2009.01.001; pmid: 19217388
34. A. Lu *et al*., Unified polymerization mechanism for the assembly of ASC-dependent inflammasomes. *Cell* **156**, 1193–1206 (2014). doi: 10.1016/j.cell.2014.02.008; pmid: 24630722
35. X. Cai *et al*., Prion-like polymerization underlies signal transduction in antiviral immune defense and inflammasome activation. *Cell* **156**, 1207–1222 (2014). doi: 10.1016/j.cell.2014.01.063; pmid: 24630723
36. Z. Hu *et al*., Crystal structure of NLRC4 reveals its autoinhibition mechanism. *Science* **341**, 172–175 (2013). doi: 10.1126/science.1236381; pmid: 23765277
37. H. M. Hoffman, J. L. Mueller, D. H. Broide, A. A. Wanderer, R. D. Kolodner, Mutation of a new gene encoding a putative pyrin-like protein causes familial cold autoinflammatory syndrome and Muckle-Wells syndrome. *Nat. Genet.* **29**, 301–305 (2001). doi: 10.1038/ng756; pmid: 11687797
38. H. Yang *et al*., A mutant CHS3 protein with TIR-NB-LRR-LIM domains modulates growth, cell death and freezing tolerance in a temperature-dependent manner in *Arabidopsis*. *Plant J.* **63**, 283–296 (2010). doi: 10.1111/j.1365-313X.2010.04241.x; pmid: 20444230
39. P. Mestre, D. C. Baulcombe, Elicitor-mediated oligomerization of the tobacco N disease resistance protein. *Plant Cell* **18**, 491–501 (2006). doi: 10.1105/tpc.105.037234; pmid: 16387833
40. K. J. Schreiber, A. Bentham, S. J. Williams, B. Kobe, B. J. Staskawicz, Multiple domain associations within the *Arabidopsis* immune receptor RPP1 regulate the activation of programmed cell death. *PLOS Pathog.* **12**, e1005769 (2016). doi: 10.1371/journal.ppat.1005769; pmid: 27427964
41. R. A. L. van der Hoorn, S. Kamoun, From guard to decoy: A new model for perception of plant pathogen effectors. *Plant Cell* **20**, 2009–2017 (2008). doi: 10.1105/tpc.108.060194; pmid: 18723576
42. S. Césari, M. Bernoux, P. Moncuquet, T. Kroj, P. N. Dodds, A novel conserved mechanism for plant NLR protein pairs: The "integrated decoy" hypothesis. *Front. Plant Sci.* **5**, 606 (2014). doi: 10.3389/fpls.2014.00606; pmid: 25506347
43. P. F. Sarris, V. Cevik, G. Dagdas, J. D. G. Jones, K. V. Krasileva, Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biol.* **14**, 8 (2016). doi: 10.1186/s12915-016-0228-7; pmid: 26891798
44. T. Kroj, E. Chanclud, C. Michel-Romiti, X. Grand, J.-B. Morel, Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytol.* **210**, 618–626 (2016). doi: 10.1111/nph.13869; pmid: 26848538
45. F. M. Ausubel, Are innate immune signaling pathways in plants and animals conserved? *Nat. Immunol.* **6**, 973–979 (2005). doi: 10.1038/ni1253; pmid: 16177805
46. C.-H. Wu, K. Belhaj, T. O. Bozkurt, M. S. Birk, S. Kamoun, Helper NLR proteins NRC2a/b and NRC3 but not NRC1 are required for Pto-mediated cell death and resistance in Nicotiana benthamiana. *New Phytol.* **209**, 1344–1352 (2016). doi: 10.1111/nph.13764; pmid: 26592988
47. V. Bonardi, K. Cherkis, M. T. Nishimura, J. L. Dangl, A new eye on NLR proteins: Focused on clarity or diffused by complexity? *Curr. Opin. Immunol.* **24**, 41–50 (2012). doi: 10.1016/j.coi.2011.12.006; pmid: 22305607
48. E. M. Kofoed, R. E. Vance, Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity. *Nature* **477**, 592–595 (2011). doi: 10.1038/nature10394; pmid: 21874021
49. Y. Zhao *et al*., The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. *Nature* **477**, 596–600 (2011). doi: 10.1038/nature10510; pmid: 21918512
50. Z. Hu *et al*., Structural and biochemical basis for induced self-propagation of NLRC4. *Science* **350**, 399–404 (2015). doi: 10.1126/science.aac5489; pmid: 26449475
51. L. Zhang *et al*., Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization. *Science* **350**, 404–409 (2015). doi: 10.1126/science.aac5789; pmid: 26449474
52. S. M. Collier, L.-P. Hamel, P. Moffett, Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. *Mol. Plant Microbe Interact.* **24**, 918–931 (2011). doi: 10.1094/MPMI-03-11-0050; pmid: 21501087
53. J. R. Peart, P. Mestre, R. Lu, I. Malcuit, D. C. Baulcombe, NRG1, a CC-NB-LRR protein, together with N, a TIR-NB-LRR protein, mediates resistance against tobacco mosaic virus. *Curr. Biol.* **15**, 968 (2005).
54. V. Bonardi *et al*., Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16463–16468 (2011). doi: 10.1073/pnas.1113726108; pmid: 21911370
55. N. Hayashi *et al*., Durable panicle blast-resistance gene Pb1 encodes an atypical CC-NBS-LRR protein and was generated by acquiring a promoter through local genome duplication. *Plant J.* **64**, 498–510 (2010). doi: 10.1111/j.1365-313X.2010.04348.x; pmid: 20807214
56. V. Bonardi, J. L. Dangl, How complex are intracellular immune receptor signaling complexes? *Front. Plant Sci.* **3**, 237 (2012). doi: 10.3389/fpls.2012.00237; pmid: 23109935
57. S. J. Williams *et al*., Structural basis for assembly and function of a heterodimeric plant immune receptor. *Science* **344**, 299–303 (2014). doi: 10.1126/science.1247357; pmid: 24744375
58. J. G. Ellis, P. N. Dodds, G. J. Lawrence, Flax rust resistance gene specificity is based on direct resistance-avirulence protein interactions. *Annu. Rev. Phytopathol.* **45**, 289–306 (2007). doi: 10.1146/annurev.phyto.45.062806.094331; pmid: 17430087
59. E. A. Van der Biezen, J. D. Jones, Plant disease-resistance proteins and the gene-for-gene concept. *Trends Biochem. Sci.* **23**, 454–456 (1998). doi: 10.1016/S0968-0004(98)01311-5; pmid: 9868361
60. J. L. Dangl, J. D. Jones, Plant pathogens and integrated defence responses to infection. *Nature* **411**, 826–833 (2001). doi: 10.1038/35081161; pmid: 11459065
61. E.-H. Chung *et al*., Specific threonine phosphorylation of a host target by two unrelated type III effectors activates a host innate immune receptor in plants. *Cell Host Microbe* **9**, 125–136 (2011). doi: 10.1016/j.chom.2011.01.009; pmid: 21320695
62. E.-H. Chung, F. El-Kasmi, Y. He, A. Loehr, J. L. Dangl, A plant phosphoswitch platform repeatedly targeted by type III effector proteins regulates the output of both tiers of plant immune receptors. *Cell Host Microbe* **16**, 484–494 (2014). doi: 10.1016/j.chom.2014.09.004; pmid: 25299334
63. J. Liu, J. M. Elmore, Z.-J. D. Lin, G. Coaker, A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor. *Cell Host Microbe* **9**, 137–146 (2011). doi: 10.1016/j.chom.2011.01.010; pmid: 21320696
64. D. Lee, G. Bourdais, G. Yu, S. Robatzek, G. Coaker, Phosphorylation of the plant immune regulator RPM1-INTERACTING PROTEIN4 enhances plant plasma membrane H^+-ATPase activity and inhibits flagellin-triggered immune responses in *Arabidopsis*. *Plant Cell* **27**, 2042–2056 (2015). doi: 10.1105/tpc.114.132308; pmid: 26198070
65. M. J. Axtell, B. J. Staskawicz, Initiation of RPS2-specified disease resistance in Arabidopsis is coupled to the AvrRpt2-directed elimination of RIN4. *Cell* **112**, 369–377 (2003). doi: 10.1016/S0092-8674(03)00036-9; pmid: 12581526
66. D. Mackey, B. F. Holt 3rd, A. Wiig, J. L. Dangl, RIN4 interacts with Pseudomonas syringae type III effector molecules and is required for RPM1-mediated resistance in Arabidopsis. *Cell* **108**, 743–754 (2002). doi: 10.1016/S0092-8674(02)00661-X; pmid: 11955429
67. F. Shao *et al*., Cleavage of Arabidopsis PBS1 by a bacterial type III effector. *Science* **301**, 1230–1233 (2003). doi: 10.1126/science.1085671; pmid: 12947197
68. J. Ade, B. J. DeYoung, C. Golstein, R. W. Innes, Indirect activation of a plant nucleotide binding site-leucine-rich repeat protein by a bacterial protease. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2531–2536 (2007). doi: 10.1073/pnas.0608779104; pmid: 17277084
69. S. H. Kim, D. Qi, T. Ashfield, M. Helm, R. W. Innes, Using decoys to expand the recognition specificity of a plant disease resistance protein. *Science* **351**, 684–687 (2016). doi: 10.1126/science.aad3436; pmid: 26912853
70. O. Rowland *et al*., Functional analysis of Avr9/Cf-9 rapidly elicited genes identifies a protein kinase, ACIK1, that is essential for full Cf-9-dependent disease resistance in tomato. *Plant Cell* **17**, 295–310 (2005). doi: 10.1105/tpc.104.026013; pmid: 15598806
71. V. Ntoukakis, I. M. Saur, B. Conlan, J. P. Rathjen, The changing of the guard: The Pto/Prf receptor complex of tomato and pathogen recognition. *Curr. Opin. Plant Biol.* **20**, 69–74 (2014). doi: 10.1016/j.pbi.2014.04.002; pmid: 24845576
72. G. Wang *et al*., The decoy substrate of a pathogen effector and a pseudokinase specify pathogen-induced modified-self recognition and immunity in plants. *Cell Host Microbe* **18**, 285–295 (2015). doi: 10.1016/j.chom.2015.08.004; pmid: 26355215
73. N. Inohara *et al*., Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *J. Biol. Chem.* **278**, 5509–5512 (2003). doi: 10.1074/jbc.C200673200; pmid: 12514169
74. M. Chamaillard *et al*., An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. *Nat. Immunol.* **4**, 702–707 (2003). doi: 10.1038/ni945; pmid: 12796777
75. S. E. Girardin *et al*., Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan. *Science* **300**, 1584–1587 (2003). doi: 10.1126/science.1084677; pmid: 12791997
76. A. M. Keestra *et al*., Manipulation of small Rho GTPases is a pathogen-induced process detected by NOD1. *Nature* **496**, 233–237 (2013). doi: 10.1038/nature12025; pmid: 23542589
77. A. M. Keestra-Gounder *et al*., NOD1 and NOD2 signalling links ER stress with inflammation. *Nature* **532**, 394–397 (2016). doi: 10.1038/nature17631; pmid: 27007849
78. Y. He, M. Y. Zeng, D. Yang, B. Motro, G. Núñez, NEK7 is an essential mediator of NLRP3 activation downstream of potassium efflux. *Nature* **530**, 354–357 (2016). doi: 10.1038/nature16959; pmid: 26814970
79. H. Shi *et al*., NLRP3 activation and mitosis are mutually exclusive events coordinated by NEK7, a new inflammasome component. *Nat. Immunol.* **17**, 250–258 (2016). doi: 10.1038/ni.3333; pmid: 26642356
80. J. L. Schmid-Burgk *et al*., A Genome-wide CRISPR (clustered regularly interspaced short palindromic repeats) screen identifies NEK7 as an essential component of NLRP3 inflammasome activation. *J. Biol. Chem.* **291**, 103–109 (2016). doi: 10.1074/jbc.C115.700492; pmid: 26553871
81. K. Bomblies *et al*., Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLOS Biol.* **5**, e236 (2007). doi: 10.1371/journal.pbio.0050236; pmid: 17803357
82. E. Chae *et al*., Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. *Cell* **159**, 1341–1351 (2014). doi: 10.1016/j.cell.2014.10.049; pmid: 25467443
83. S. Césari *et al*., The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. *EMBO J.* **33**, 1941–1959 (2014). doi: 10.15252/embj.201487923; pmid: 25024433

84. M. Narusaka *et al*., RRS1 and RPS4 provide a dual Resistance-gene system against fungal and bacterial pathogens. *Plant J.* **60**, 218–226 (2009). doi: 10.1111/j.1365-313X.2009.03949.x; pmid: 19519800
85. N. Ishihama, H. Yoshioka, Post-translational regulation of WRKY transcription factors in plant immunity. *Curr. Opin. Plant Biol.* **15**, 431–437 (2012). doi: 10.1016/j.pbi.2012.02.003; pmid: 22425194
86. P. F. Sarris *et al*., A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell* **161**, 1089–1100 (2015). doi: 10.1016/j.cell.2015.04.024; pmid: 26000484
87. C. Le Roux *et al*., A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell* **161**, 1074–1088 (2015). doi: 10.1016/j.cell.2015.04.025; pmid: 26000483
88. M. Narusaka *et al*., Interfamily transfer of dual NB-LRR genes confers resistance to multiple pathogens. *PLOS ONE* **8**, e55954–e55958 (2013). doi: 10.1371/journal.pone.0055954; pmid: 23437080
89. R. Weßling *et al*., Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe* **16**, 364–375 (2014). doi: 10.1016/j.chom.2014.08.004; pmid: 25211078
90. A. Maqbool *et al*., Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife* **4**, 213 (2015). doi: 10.7554/eLife.08709; pmid: 26304198
91. S. Fukuoka *et al*., Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science* **325**, 998–1001 (2009). doi: 10.1126/science.1175550; pmid: 19696351
92. E. D. Boyden, W. F. Dietrich, Nalp1b controls mouse macrophage susceptibility to anthrax lethal toxin. *Nat. Genet.* **38**, 240–244 (2006). doi: 10.1038/ng1724; pmid: 16429160
93. J. Chavarría-Smith, R. E. Vance, Direct proteolytic cleavage of NLRP1B is necessary and sufficient for inflammasome activation by anthrax lethal factor. *PLOS Pathog.* **9**, e1003452 (2013). doi: 10.1371/journal.ppat.1003452; pmid: 23818853
94. J. L. Levinsohn *et al*., Anthrax lethal factor cleavage of Nlrp1 is required for activation of the inflammasome. *PLOS Pathog.* **8**, e1002638 (2012). doi: 10.1371/journal.ppat.1002638; pmid: 22479187
95. J. N. Finger *et al*., Autolytic proteolysis within the function to find domain (FIIND) is required for NLRP1 inflammasome activity. *J. Biol. Chem.* **287**, 25030–25037 (2012). doi: 10.1074/jbc.M112.378323; pmid: 22665479
96. F. Jupe *et al*., Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* **76**, 530–544 (2013). doi: 10.1111/tpj.12307; pmid: 23937694
97. K. Witek *et al*., Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. *Nat. Biotechnol.* **34**, 656–660 (2016). doi: 10.1038/nbt.3540; pmid: 27111721
98. B. Steuernagel *et al*., Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat. Biotechnol.* **34**, 652–655 (2016). doi: 10.1038/nbt.3543; pmid: 27111722
99. R. S. Nandety *et al*., The role of TIR-NBS and TIR-X proteins in plant basal defense responses. *Plant Physiol.* **162**, 1459–1472 (2013). doi: 10.1104/pp.113.219162; pmid: 23735504

ACKNOWLEDGMENTS

NLR and effector research in the laboratories of the authors is funded by the Biotechnology and Biological Sciences Research Council, European Resdearch Council, Gordon and Betty Moore Foundation (GBMF), and the Gatsby Foundation (J.D.G.J.) and the NSF, NIH (J.L.D. and R.E.V.), GMBF (J.L.D.), and HHMI (J.L.D. and R.E.V.). J.D.G.J. thanks B. Staskawicz for hosting a brief sabbatical visit, during which conversations leading to this review were initiated, and we thank A. Daskalov and O. Furzer for helpful discussion. We thank V. Maltez, E. Miao, I. Rauch, K. Witek, and J. Parker for providing unpublished images for Fig. 1 and J. Tenthorey for help with Fig. 2.

10.1126/science.aaf6395

RESEARCH ARTICLE SUMMARY

RNA SPLICING

Crystal structures of a group II intron lariat primed for reverse splicing

Maria Costa,* Hélène Walbott, Dario Monachello,† Eric Westhof, François Michel

INTRODUCTION: Self-splicing group II introns are catalytic RNAs (ribozymes) that can excise by themselves from precursor RNA molecules. These ribozymes are widespread in the bacterial world and can also be found in the bacterial-derived organelles (mitochondria and chloroplasts) of some higher organisms. Group II self-splicing is believed to have evolved into nuclear pre-mRNA splicing, a fundamental step in the expression of eukaryotic genes during which a large ribonucleoprotein machinery (the spliceosome) catalyzes the removal of introns from nascent premessenger transcripts. Both group II and pre-mRNA splicing proceed via two sequential phosphoryl transfer reactions. First, a 2′-5′ phosphodiester bond is created between a conserved intron adenosine and the first intron nucleotide. The resulting splicing intermediate with a branched conformation is called a "lariat." In a second step, completion of splicing leads to the ligation of the flanking 5′ and 3′ exons and the release of the intron lariat. Bacterial group II introns are composite elements that, in addition to their ribozyme core, carry an open reading frame encoding a reverse transcriptase (RT) enzyme. In association with their RT, freed group II intron lariats behave as mobile genetic elements that colonize genomes through retrotransposition. The group II mobility pathway is initiated by "reverse splicing" of the intron lariat into a DNA target, followed by synthesis of a DNA copy of the integrated intron by the RT enzyme.

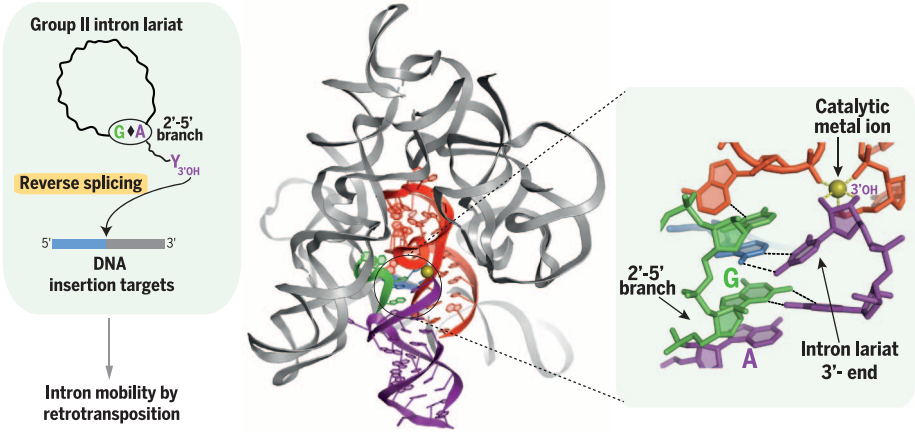
RATIONALE: Eukaryotic pre-mRNA splicing relies entirely on formation of the 2′-5′ branch structure. In group II introns, the same branch structure is required for efficient and faithful catalysis of the second step of splicing and for complete, accurate reverse splicing of the intron into DNA targets during intron mobility. Moreover, in both splicing systems, the branched nucleotides must translocate within the active site between the two steps of splicing. To understand the molecular mechanisms at play, we crystallized the lariat form of a group II intron either alone or bound to a nonreactive analog of the 5′ exon.

RESULTS: Our crystal structures at 3.4 and 3.5 Å resolution reveal that the 2′-5′ branched nucleotides are part of a network of hydrogen bonds and stacking interactions that involve highly conserved nucleotides at the intron core and boundaries. The resulting architecture organizes the second-step active site by juxtaposing the intron 5′ and 3′ ends and promotes positioning of the last intron nucleotide into the catalytic center. After the ligated exons have been released, the terminal ribose of the lariat intron remains docked in the reaction center, with its 3′-hydroxyl group activated by a highly coordinated metal ion and poised for catalysis of the reverse-splicing reaction. Stable docking of the 2′-5′ branch structure into the active site is promoted by a rearrangement of the base-pairing pattern within the helix that contains the adenosine branchpoint. This rearrangement operates between the two steps of splicing and is essential for recognition of the proper 3′ splice site. Comparison of lariat structures in the presence and absence of the 5′ exon reveals that substrate binding results in an induced fit that extends into the catalytic center and contributes to coordination of a second catalytic metal ion. This "exon-sensing" device could ensure that catalysis of reverse splicing is dependent on the accuracy of intron-exon pairings.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaf9258

CONCLUSION: The present crystal structures bring to light the crucial role of the 2′-5′ branch in organizing the lariat intron catalytic site for efficient and accurate ligation of the flanking exons during the last stage of splicing. Making use of the branch structure to build the second-step active site results in coupling the two steps of splicing and contributes decisively to the fidelity of the overall process. Moreover, the presence of the 2′-5′ branch locks the active site into a near-transition-state configuration for catalysis of reverse splicing, which must have contributed to selection of the lariat conformation during the evolution of mobile group II introns. As all nucleotides involved in the catalytic center have potential homologs in the spliceosomal system, a group II-based model in which the 2′-5′ branch fulfills the same organizational role is proposed for the spliceosome second-step active site. As in group II introns, the postulated architecture implies a notable conformational rearrangement of the spliceosome active center between the two steps of pre-mRNA splicing. Our findings rationalize the extreme conservation of the branched conformation both during the diversification of group II introns and along the evolutionary path that gave rise to the nuclear pre-mRNA splicing apparatus of eukaryotes. ■



Structural basis for reverse splicing by group II introns. Self-splicing group II introns and their evolutionary descendants, the spliceosomal introns of eukaryotes, are excised as branched molecules (lariats) with a 2′-5′ phosphodiester bond. The 2′-5′ branch organizes the lariat active site, priming it for catalysis of splicing and reverse splicing. Reverse splicing into DNA is used to initiate group II intron mobility in bacteria.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: maria.costa@i2bc.paris-saclay.fr
Cite this article as M. Costa et al., Science 354, aaf9258 (2016). DOI: 10.1126/science.aaf9258

RESEARCH ARTICLE

RNA SPLICING

Crystal structures of a group II intron lariat primed for reverse splicing

Maria Costa,[1]* Hélène Walbott,[2] Dario Monachello,[1]† Eric Westhof,[3] François Michel[1]

The 2′-5′ branch of nuclear premessenger introns is believed to have been inherited from self-splicing group II introns, which are retrotransposons of bacterial origin. Our crystal structures at 3.4 and 3.5 angstrom of an excised group II intron in branched ("lariat") form show that the 2′-5′ branch organizes a network of active-site tertiary interactions that position the intron terminal 3′-hydroxyl group into a configuration poised to initiate reverse splicing, the first step in retrotransposition. Moreover, the branchpoint and flanking helices must undergo a base-pairing switch after branch formation. A group II–based model of the active site of the nuclear splicing machinery (the spliceosome) is proposed. The crucial role of the lariat conformation in active-site assembly and catalysis explains its prevalence in modern splicing.

Group II introns are large ribozymes (catalytic RNAs) with the capacity to self-excise from their host precursor RNAs in vitro (*1*). The group II ribozyme, which is composed of six structural domains (DI to DVI; fig. S1), self-splices through two consecutive transesterification reactions (Fig. 1). First, a 2′-5′ phosphodiester bond is formed between the 2′-hydroxyl group of an adenosine located in DVI and the first intron nucleotide. The resulting splicing intermediate with a branched conformation is called "lariat." Subsequently, the terminal 3′-hydroxyl group of the 5′ exon attacks the 3′ splice site, generating the ligated exons and freeing the intron lariat. Although splicing may occasionally be initiated by hydrolysis, which leads to the release of the excised intron in linear instead of branched form, the use of a branched splicing intermediate is the hallmark of group II introns. Because the spliceosome—a large and dynamic ribonucleoprotein machinery—uses the same reactional mechanism to remove introns from pre-mRNA transcripts in the nucleus of eukaryotes, group II and spliceosome-catalyzed splicing are widely believed to share a common origin [reviewed in (*2*)].

Group II introns are widespread in the bacterial world. Most bacterial group II introns encode a multifunctional reverse transcriptase that associates with the intron to promote its genomic mobility through retrotransposition, a pathway initiated by "reverse splicing" (*3*). The latter process, by which the freed intron lariat catalyzes its own insertion into a DNA target (Fig. 1), rests on the chemical reversibility of the two transesterification reactions of splicing. Introns in linear form, which lack the 2′-5′ bond, require host functions to complete reverse splicing and insert into DNA: They transpose much less efficiently and in a predominantly imprecise manner when tested in vivo (*4*).

Aside from its contribution to retrotransposition, the 2′-5′ branch is also important for the second step of splicing, because branched intron–3′ exon reaction intermediates are much more efficient than their linear counterparts at carrying out exon ligation (*5*). Despite the central role of the lariat conformation, the exact arrangement of nucleotides that participate in and surround the group II 2′-5′ branch has nonetheless remained elusive. Available x-ray structures of the *Oceanobacillus iheyensis* group II ribozyme (*6*, *7*) correspond to a linear form of the intron truncated at its 3′ end; they lack the small helical DVI that carries the branchpoint adenosine (bpA). Structures of lariat introns were more recently generated by crystallography (*8*) and cryogenic electron microscopy (cryo-EM) (*9*), but in both of them, the resolution of the intron active site is insufficient to identify some essential nucleotides and ascertain that they are correctly assembled.

To understand the structural basis for the supremacy of a 2′-5′ branched structure in group II splicing, we crystallized the lariat form of a chimeric group II ribozyme derived from the *O. iheyensis* intron. Like most members of structural subclass IIC (*3*), the wild-type *O. iheyensis* ribozyme self-splices exclusively through 5′ splice-site hydrolysis in vitro (*6*), resulting in linear molecules. However, we recently succeeded in activating the branching pathway for *O. iheyensis* intron constructs in which DVI and its first-step, ι (iota) binding site (*10*) in substructure IC1, had been replaced by their counterparts in *Azotobacter vinelandii* intron I2 (construct Oc19; fig. S1B) (*11*). The intron lariat excised from the Oc19 chimeric precursor by in vitro splicing was purified under denaturing conditions and subsequently refolded and crystallized, either alone or in the presence of a nonreactive 5′-exon analog RNA (*12*). The two structures were solved at 3.4 and 3.5 Å resolution, respectively, by molecular replacement (*12*), which allowed us to locate DVI in both electron density maps and unambiguously model the active-site region.

The 2′-5′ branch organizes the reactive site

The Oc19 lariat crystallized alone or in the presence of an unreactive 5′ exon exhibits the same overall folding of intron subdomains around catalytic DV (Fig. 2A) as the previously published structures of the *O. iheyensis* ribozyme that lacked DVI (*6*). In particular, the "catalytic triplex," formed in the major groove of DV by the "catalytic" components of this domain (C358, G359, and C377; fig. S1) and G288 and C289 of the J2/3 strand, displays the same configuration as in the absence of DVI.

In both our lariat structures, DVI is not bound to its ι receptor (fig. S1A). Instead, its basal section (the one immediately next to the secondary structure central wheel; fig. S1) is only slightly shifted with respect to the basal stem of DV, with whose axis it forms an angle of ~160° (Fig. 2B). This spatial arrangement suggested that the lariat was crystallized in its second-step conformation, which was confirmed by fitting the intron 3′ end into the density map. The last intron nucleotide (U419, position γ′) forms a Watson-Crick base pair with A287(γ) of J2/3, the highly conserved linker between DII and DIII (Fig. 2, C and D, and fig. S2). The γ-γ′ interaction, which positions the 3′ splice site for exon ligation (*13*), holds the terminal ribose in the catalytic center even after the ligated exons have been released. The 2′- and 3′-oxygens of this terminal ribose coordinate a metal ion (M1; Fig. 2C and figs. S2 and S3E), which is also bound to catalytically important oxygens in DV (*6*). The position of the essential A287(γ) base (Fig. 2C) differs markedly compared to previous structures (*6*–*9*). The particular backbone path adopted by the catalytically essential J2/3 linker is instrumental in positioning the A287(γ) and G288 bases sufficiently far apart to allow the second-step active site to form at the side of the catalytic triplex (fig. S3D).

Stabilization of γ-γ′ is achieved by A287(γ) being directly stacked on top of the 2′-5′ branch structure, which is formed by base-stacking of the first intron nucleotide (G1) on the bpA (Fig. 2, C and D, and figs. S2 and S3, B and C). Stable docking of the branched nucleotides, which involves hydrogen bonding between the donor groups of the G1 base and the O1P (pro-Sp) oxygen of C357 in DV, places G1 in the proper register to base pair with the penultimate intron nucleotide through a non–Watson-Crick interaction (Fig. 2, C and D) specific to the second step of splicing (*14*). The

[1]Group II introns as ribozymes and retrotransposons, Institute for Integrative Biology of the Cell (I2BC), UMR 9198 CNRS, Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), University Paris-Sud, University Paris-Saclay, 1 Avenue de la Terrasse, Bâtiment 26, 91198 Gif-sur-Yvette cedex, France. [2]Structure and Dynamics of RNA, I2BC, UMR 9198 CNRS, CEA, University Paris-Sud, University Paris-Saclay, 91198 Gif-sur-Yvette cedex, France. [3]Architecture and Reactivity of RNA, Institute of Molecular and Cellular Biology of the CNRS, University of Strasbourg, 15 rue René Descartes, 67084 Strasbourg, France.
***Corresponding author. Email: maria.costa@i2bc.paris-saclay.fr**
†Present address: Institut des Sciences des Plantes de Paris-Saclay (IPS2)–Institut National de la Recherche Agronomique, Bâtiment 630, rue de Noetzlin, Plateau du Moulon, 91405 Orsay, France.

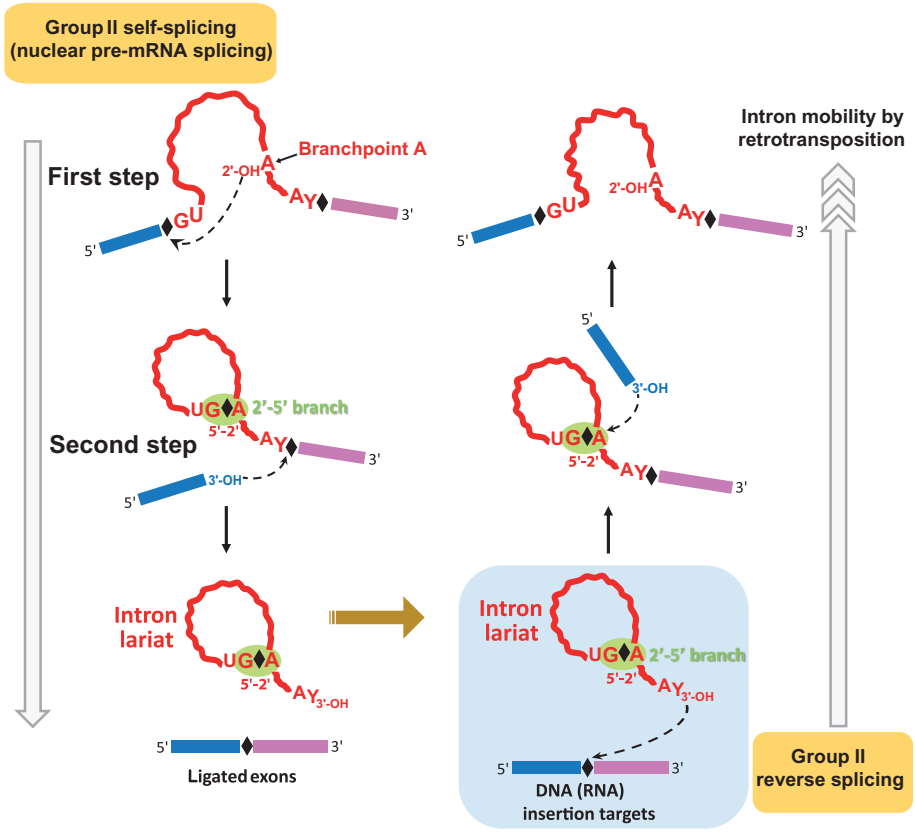


Fig. 1. Splicing and reverse-splicing pathways catalyzed by group II introns. Red wavy line, group II intron; blue and pink boxes, 5′ and 3′ exons (for splicing) or DNA sequences flanking the target site (for reverse splicing). Conserved nucleotides at the branchpoint and intron boundaries are shown. Splicing proceeds through two consecutive transesterification reactions. Dashed lines and arrowheads indicate nucleophilic attack at each reactional step, diamonds at intron ends stand for reactive phosphate groups, and the 2′-5′ branch is highlighted by a green background. The group II splicing pathway as depicted here is valid for nuclear pre-mRNA splicing catalyzed by the major spliceosome subtype, except that the latter's intron substrates end with a G. The freed group II lariat intron is ready to undergo reverse splicing, in which the two steps of splicing are performed in the opposite direction. The intron lariat in our crystal structures is in the right conformation to carry out the first step of reverse splicing, here shown on a blue background. For clarity, the intron-encoded reverse transcriptase that assists splicing and reverse splicing in vivo by stabilizing the ribozyme is not depicted.

A412(bpA)-G1-A287(γ) purine stack is topped by the second intron nucleotide (U2), whose stretched backbone allows its base moiety to stack onto A287(γ) (Fig. 2C and figs. S2 and S3, B and C). The U2 base also interacts with the sugar edge of A376 in the catalytically critical DV bulge. This constrains the U2 nucleobase into a *syn* conformation, which is unusual for a pyrimidine (Fig. 2D) (*15*). Finally, two metal ions (M3 and M4) participate in organizing the active site by stabilizing the contorted backbone path between nucleotides 1 and 5 (Fig. 2C and fig. S3E).

By revealing the network of stacking and hydrogen bonding interactions that connects the highly conserved nucleotides of the J2/3 linker and intron boundaries to the 2′-5′ branch, our structures make visible the crucial role played by the branched nucleotides in assembling a functional active site for the second step of splicing. This role had been anticipated by biochemical experiments: A linear intron–3′ exon splicing intermediate carries out the second step of splicing ~800 times slower than its branched counterpart (*5*), and disruption of the interaction between G1 and the penultimate intron nucleotide affects splicing only of the lariat form of the intron, not its linear form (*14*). Having the 2′-5′ branch as an essential component of the second-step active site ensures coupling of the two steps of splicing and makes it possible to verify at the exon ligation stage that the proper 5′ splice site was selected during branch formation: Incorrectly assembled splicing intermediates will either get "debranched" through reversal of the first step reaction or eventually dissociate from the 5′ exon.

Alternate branchpoint conformations

On the basis of extensive comparative analyses (*16*, *17*), the bpA of group II introns has generally been assumed to form a single-nucleotide bulge on the 3′ side of DVI (Fig. 3, top left). However, such a secondary structure is not supported by our electron density maps. These show instead an alternate conformation in which the DVI basal helix comprises only 3 base pairs (bp), and the branchpoint A412 is part of a 2-nucleotide (nt) bulge (Fig. 3, top right, and fig. S3F). A 2-nt branchpoint bulge was present in the crystal structure of a small construct intended to mimic DV and DVI of a group II intron (*18*). This observation and the estimated thermodynamic stabilities in Fig. 3 suggest that the conformation with a 2-nt bulge could constitute the DVI ground state in many group II introns.

Because our crystal structures imply that exon ligation makes use of a 2-nt bulge, the conformation with a 1-nt bulge (Fig. 3, top left) could be specific to the first (branching) step of splicing. We verified this by introducing base substitutions that selectively destabilize either one of the two DVI conformations. In constructs G395C:C410G and G396C:C409G, the postulated first-step conformation should be almost as stable as in Oc1δ, whereas a 2-nt bulge should be highly unfavorable (Fig. 3). As expected, neither of these substitutions interfered with the initiation of splicing: Branched products form rapidly under splicing conditions (figs. S4 and S5). However, both constructs fail to make efficient use of the 3′ splice site, as shown by the nearly complete absence of ligated exons and the transient accumulation of both the intron–3′ exon lariat intermediate and branched products of intermediate lengths, which are eventually converted into a molecule that migrates like the Oc1δ intron lariat [fig. S4; the same defects were seen with an *O. iheyensis*–derived DVI whose sequence lacks the capacity for strand shifting (*11*)]. In contrast, construct G395C:C409G, which should overwhelmingly favor the conformation with a 2-nt bulge, is inefficient at initiating splicing by branching, as seen from the predominance of linear products generated by 5′ splice-site hydrolysis (fig. S4), but yields a single lariat band in a second-step reaction whose rate constant is at least equal to that estimated for Oc1δ (fig. S5).

In the Oc1δ crystal structure, the last six nucleotides of the intron form a continuous helical stack (Fig. 2B). Adding U413 to this stack, by pairing it with A393 in a 4-bp DVI basal helix, would drag not only A412 but also G1 out of place, which explains why, in a branched molecule, the 1-nt bulge conformation fails to use the correct 3′ splice junction efficiently. The crucial role played by the branched G1 in binding the intron 3′-terminal segment is confirmed by the behavior of linear intron–3′ exon molecules, in which the absence of the 2′-5′ bond and consequent destabilization of the intron ends can be invoked to explain why intron excision is both slow (*5*) and imprecise (fig. S4; G395C:C409G lanes).

As opposed to exon ligation, efficient branch formation was shown to require a 4-bp, rather than 3-bp, DVI basal helix (*11*, *19*). This explains why the 2-nt bulge conformation reacts preferentially by 5′ splice-site hydrolysis and accounts for the need to reorganize the DVI bulge between the two steps of splicing. Other group II introns with a 7-nt spacer between the branchpoint and 3′ splice site should share with *A. vinelandii* I2—whose DVI was part of the Oc1δ construct (fig.

S1B)—the ability to toggle between conformations with a 4- and 3-bp DVI basal helix. This is true for all but one of the intron lineages with a 7-nt spacer (fig. S6). As in *A. vinelandii* I2, interconversion between the 1-nt bulge, first-step conformation and the 2-nt bulge, second-step conformation is achieved by reciprocal shifting of the two strands that constitute the middle part of DVI (in red in fig. S6) by one nucleotide. Such a mechanism requires that in a segment that undergoes rearrangement of its base-pairing pattern, one strand contains only G's (except at its extremities) and the other one contains only pyrimidines. This explains the peculiar base distribution and the slow evolution of sequences surrounding the DVI branchpoint in these intron subgroups, which constitute about half of recognized group II lineages.

Conformational reorganization of DVI could be helped in vivo by proteins—either the intron-encoded reverse transcriptase or cellular helicases like yeast Mss116 (*20*) that would take advantage of the accessibility of this domain in the three-dimensional structure of the intron (Fig. 2 and fig. S3A) to promote rearrangement of its secondary structure. Although the recent cryo-EM structure of a group II intron bound to its reverse transcriptase (*9*) lacks any contact between the intron-encoded protein and DVI, footprinting and cross-linking data (*21*) do suggest a direct, possibly transient, physical interaction between these components.

Future work will investigate how the branchpoint rearrangement we identified here is mechanistically and structurally coupled to the formation of tertiary contacts between DVI and other intron components. These contacts include ι-ι′ (fig. S1A), which is specific to the branching step (*10*), and two interactions between DII and the basal and distal sections of DVI (*8*, *22*). The latter interactions, which are specific to the exon ligation step (*22*), were removed in our Oc19 construct for the sake of efficient crystal packing.

Exon-driven induced fit and implications

Group II introns bind their 5′ and 3′ exons for splicing, or their DNA target site for retrotransposition, by base pairing to exon binding site (EBS) sequences located in DI (fig. S1) (*3*, *23*, *24*). The linear *O. iheyensis* ribozyme (DI to DV) had been reported to retain essentially the same structure when crystallized in the presence or absence of exons, except in the EBS1 segment, which is complementary to the last nucleotides of the 5′ exon (fig. S1B) and was found to be slightly disordered in ligand-free structures (*6*, *7*). However, here, we report that whereas the 5′ exon–bound Oc19 lariat (*12*) differs little from previously published structures over DI to DV, our exon-free lariat structure reveals substantial local rearrangements.

The expected disorganization of the middle section of EBS1 in the absence of the 5′ exon results in the neighboring GAAC terminal loop of DV (ζ′) to adopt an alternate conformation in which all four loop bases are stacked (Fig. 4A and fig. S7A). The reorganized loop cannot interact with its receptor motif (ζ), which becomes partly disordered. Concomitantly, the symmetrical internal loop in the basal section of subdomain IC (fig. S1) adopts a fold that completely extrudes nucleotide A72 (Fig. 4A and fig. S7A). Stacking of the flipped A72 base on the highly conserved A106(λ) is made possible by a slight rotation of A106 away from its position in the exon-bound structure: This chain of rearrangements is ultimately triggered by the loss of the hydrogen bond between the N1 atom of A106 and the 2′-OH group at position −2 of the exon (Fig. 4, B and C, and fig. S7B).

These structural rearrangements are functionally relevant, as illustrated by data on the binding of the 5′ exon to a subgroup IIB intron lariat (*25*). Two binding modes that differ by a ~100-fold difference in K_d (dissociation constant) were identified, with the high-affinity configuration being dependent on the distal section of DV. Moreover, footprinting experiments revealed that the N1 position of the nucleotide homologous to A106(λ) becomes protected from modification only when the 5′ exon is tightly bound to the lariat, as expected from our crystal structures.

Exon-driven rearrangements extend to the active site, because the nonbridging oxygen (O1P) of U375, which participates in catalysis as a metal ion ligand, is properly positioned to fulfill this role only when the 5′-exon substrate is stably bound to the intron (fig. S7C). This observation explains why removal of the 2′-OH group at exon position −2 decreases the rate of catalysis by as much as 500-fold (*25*) and illustrates the crucial role of the 5′ exon in helping to bend DV into its catalytically competent conformation.

Just like protein enzymes and the ribosome (*26*), therefore, group II introns bind their 5′ exon through an induced-fit mechanism, in which initial substrate binding induces structural rearrangements in the catalytic core that, ultimately, strengthen enzyme-substrate contacts and trigger catalysis. This device allows the system to "verify" the quality of intron-substrate pairings before engaging in splicing or reverse splicing. The

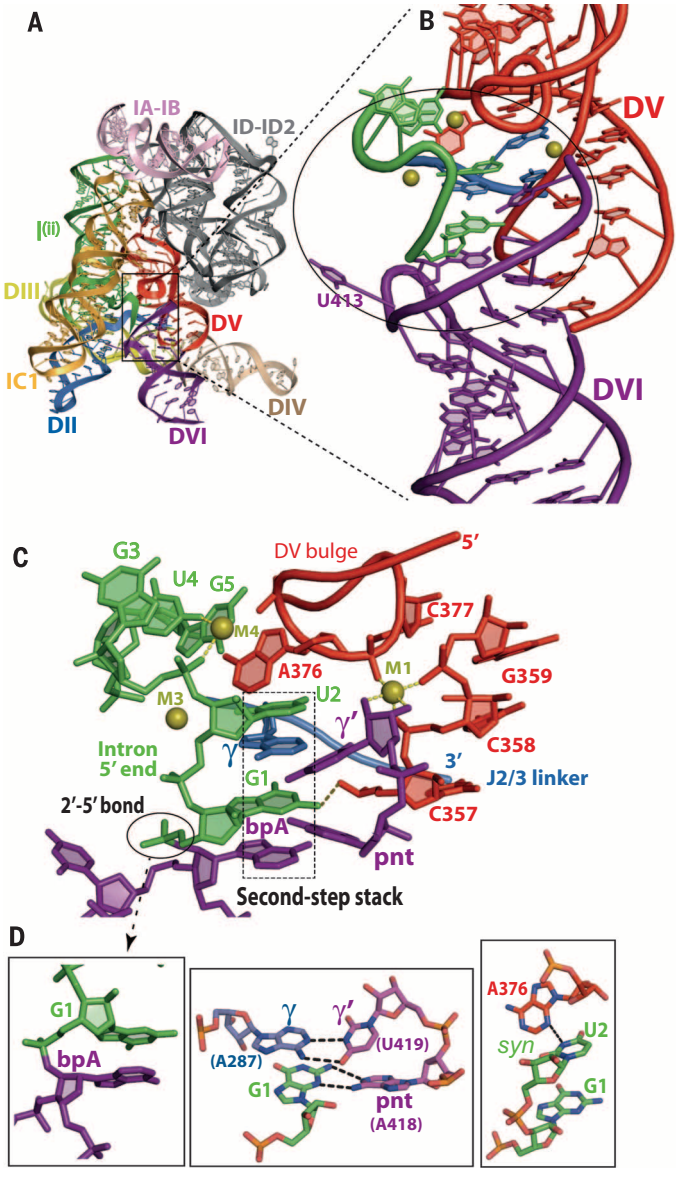


Fig. 2. Location of DVI and tertiary structure of the active site of the crystallized lariat. (**A**) Overall three-dimensional structure of the Oc19 lariat showing the position of DVI relative to other intron components (color coding as in fig. S1A). (**B**) Close-up view of DV (red), DVI (purple), the intron 5′-terminal segment (green), and J2/3 linker (blue). (**C**) Base-stacking array that connects the branched nucleotides to the reaction center. M1, M3, and M4 are metal ions (fig. S3E); dashed yellow and brown lines indicate direct metal ion coordination and hydrogen bonding, respectively. pnt, penultimate intron nucleotide. (**D**) Base-base interactions that organize the active site. Dotted black lines are hydrogen bonds. Coloring by atom is superimposed onto coloring by domains in the last two panels.

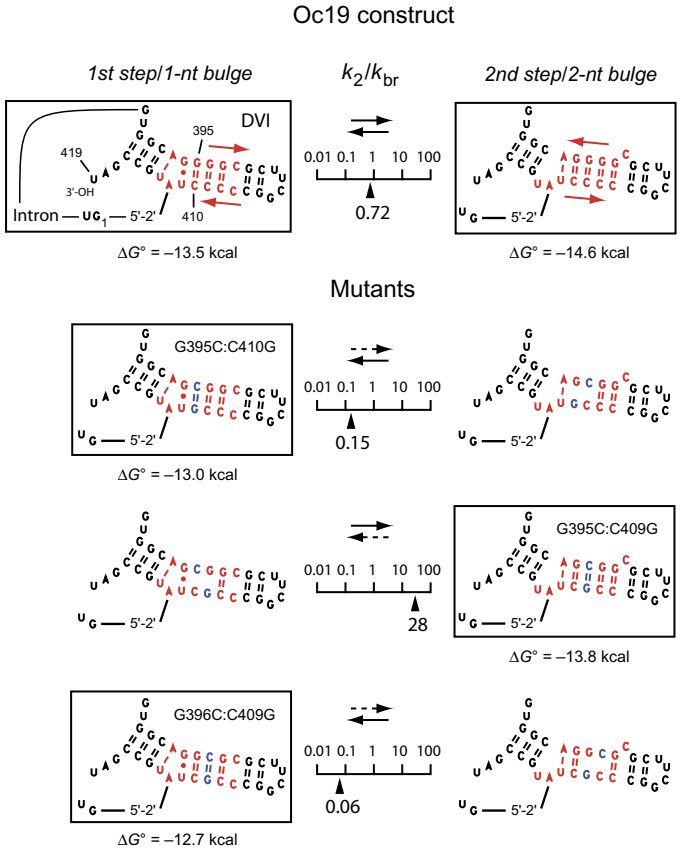


Fig. 3. Experimental evidence for a rearrangement of DVI between the first (branching) and second (exon ligation) steps of splicing. The section of DVI that is proposed to undergo conformational rearrangement by reciprocal strand shifting (red arrows) is in red; base substitutions in Oc19 mutants are in blue. Favored conformations are boxed; dotted arrows indicate unfavorable rearrangements; values of ΔG°_{37} were calculated with RNAstructure 5.6 (*59*). k_2 and k_{br} are experimentally determined rate constants for steps 2 and 1, respectively (fig. S5B) (*12*); their ratio was proposed to depend on the thermodynamic equilibrium between step 1 and step 2 conformations (*22*).

interaction with the 2′-hydroxyl group at position –2 of the 5′ exon is of particular biological interest because it enables group II ribozymes to discriminate between RNA and DNA substrates (*25*). Because exon-intron pairings are directly recognized by the intron-encoded reverse transcriptase (*9*), the latter could ensure that DNA is favored over RNA as a target for reverse splicing during intron mobility.

Metal ions and catalysis of reverse splicing

The architecture of the second-step active site is not altered by binding of the 5′-exon substrate, and in both Oc19 lariat structures, the native electron density map revealed a large peak in the reaction center (Fig. 5A). This peak was modeled as metal ion M1 and further assigned to a Mg^{2+} ion based on ytterbium (Yb^{3+}) soaks (Y1 site in Fig. 5A) (*12*). Coordination of M1 involves a remarkably high number of inner-sphere contacts. Besides the three coordinations with nonbridging phosphate oxygens in DV [(*6*) and references therein], two other inner-sphere contacts with the 2′- and 3′-oxygen atoms of the terminal U419(γ′) ribose are visualized for the first time (Fig. 5B). These newly seen ribose ligands had been predicted, based on biochemical evidence, to bind essential divalent metal ions in the transition state for the second step of splicing (*27*, *28*). Our crystal structures support and extend these findings by demonstrating that a single ion, which is already observable in the Oc19 lariat ground state, is simultaneously bound to the 2′- and 3′-oxygen atoms and that this unusual coordination (*28*) is made possible by the C2′-endo conformation of the terminal ribose (Fig. 5B).

Because the unreactive 5′-exon analog we used to prevent lariat debranching during cocrystallization (*12*) lacks a terminal 3′-OH group (Fig. 5A), neither of the Oc19 native maps exhibits electron density that could be interpreted as the second divalent metal ion, which was shown biochemically (*29*) and crystallographically (*30*) to be bound to that 3′-oxygen. However, our anomalous difference map obtained in Yb^{3+}, which binds with higher affinity than Mg^{2+} (*31*), does reveal a second strong Yb^{3+} binding site (Y2; Fig. 5A), which lies at the appropriate location to coordinate a terminal 3′-oxygen and stands 4.0 Å apart from Y1/M1 (Fig. 5A). The latter distance is typical for catalysis of phosphoryl transfer reactions by the two–metal ion mechanism (*32*).

When a ligated exon substrate is docked (*12*) into our 5′ exon–bound Oc19 structure (Fig. 5B), the scissile phosphate of that substrate falls into the reaction center, right between M1 and Y2, with its O1P (pro-Sp) oxygen properly positioned to make inner-sphere contacts with Y2 and the apical position of catalytic metal ion M1 (Fig. 5, B and C). This configuration of the scissile linkage is the one determined biochemically for reverse splicing by group II introns (*33*). Moreover, the 3′-oxygen of the terminal U419(γ′) ribose, which is the nucleophile of the reverse-splicing reaction, is correctly prepositioned in our crystal for inline nucleophilic attack (*34*) on the scissile phosphate (Fig. 5, B and C). The entire arrangement of metal ions and ligands is fully consistent with the two–metal ion mechanism of catalysis also in use for polymerases and group I introns (*32*). That essential features of this mechanism should directly be seen in, or readily deduced from, our ground-state crystal structure brings to light the efficiency with which the 2′-5′ branch structure enforces a near–transition state configuration to the reaction center. This ability to prime the intron for reverse splicing must have contributed decisively to the selection of the lariat over linear form during the emergence of mobile group II introns.

Implications for the spliceosome active site

Nuclear pre-mRNA splicing is catalyzed within the spliceosome, a highly dynamic ribonucleoprotein particle whose active core is composed of three small nuclear RNAs (snRNAs) named U2, U5, and U6 and numerous protein factors (*35*). In both group II and spliceosomal introns, the 2′-5′ linkage of the lariat results from attack of the 5′ splice site by a conserved adenosine that bulges out from a helical stem. This shared peculiarity and the presence of similar consensus sequences at intron boundaries (5′-GU…AG-3′ for the most common subtype of spliceosomal introns, 5′-GU…AY-3′ for group II introns) led to the hypothesis that the two systems have a common evolutionary origin (*36*). There is compelling biochemical evidence that part of the highly conserved U6 snRNA is homologous to DV of group II introns: The terminal **GA** of the invariant ACAGA**GA** sequence of U6 interacts with the major groove of U2-U6 helix Ib to form a group II intron–like catalytic triplex (*37*). The latter promotes binding of two catalytic metal ions to conserved, nonbridging phosphate oxygens of U6 in a manner similar to DV (Fig. 5) (*38*, *39*).

We now propose that the architecture of the second-step active site of the spliceosome rests on a network of RNA-RNA interactions (Fig. 6) similar to the one our structures have revealed for the Oc19 lariat. As in group II, the 2′-5′ branch and the terminal nucleotides of spliceosomal introns are both required for the second step of pre-mRNA splicing (*40*). Moreover, the conserved guanosines at the boundaries of nuclear introns form a non–Watson-Crick base pair that is specifically required for exon ligation (*41*, *42*) and probably fulfills the same function as the pair between G1 and the penultimate nucleotide of group II introns. Although the two pairings are not isomorphic, it is nevertheless possible to fit the spliceosomal G1:Gn interaction into the Oc19 active site in such a way that the Gn terminal ribose remains positioned in the catalytic center (Fig. 6A) (*12*). In the resulting model, G1 no longer stacks on the bpA (Fig. 6B). However, the stacking interaction between nucleotides U2 and A287(γ)

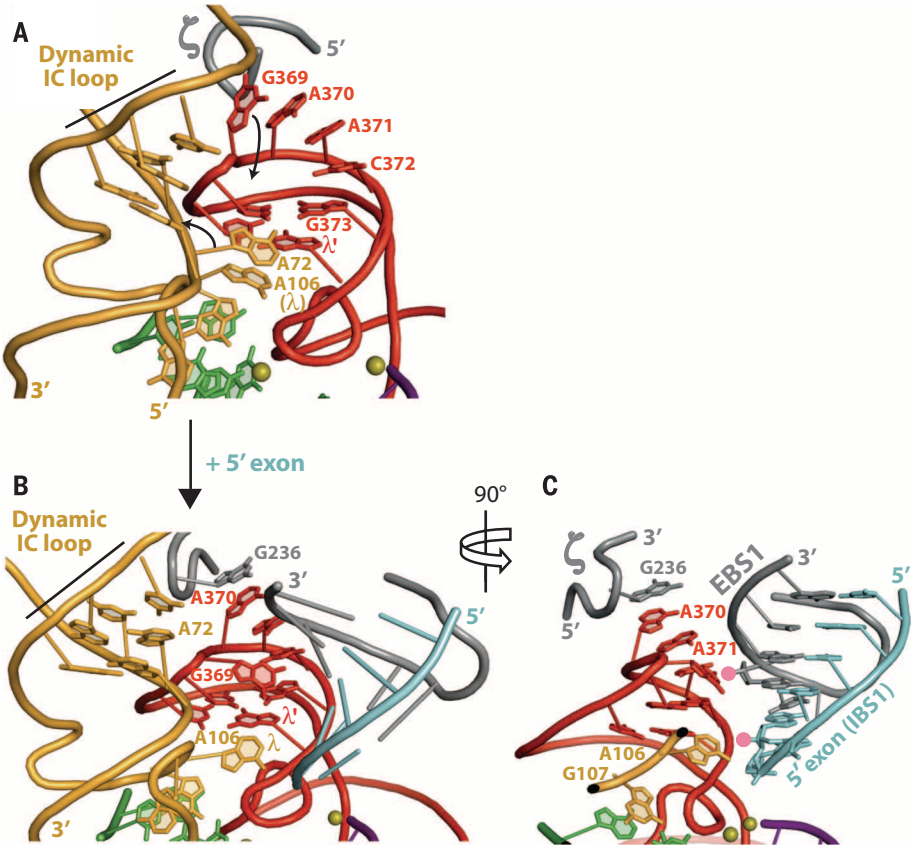


Fig. 4. Conformational rearrangements induced by the 5′ exon. (**A**) Conformation of the distal section of DV (red) and basal section of domain IC (bright orange) in the absence of the 5′ exon. The ζ motif (gray tube) in subdomain ID and the middle section of the EBS1 loop (not shown) are disordered. Black arrows indicate the wide movements undergone by A72 and G369 (in the GAAC tetraloop) upon 5′-exon binding. (**B**) Base-pairing of the 5′-exon substrate (cyan) to EBS1 (gray) promotes integration of A72 into the dynamic IC loop and formation of tertiary interaction between ζ and the GAAC tetraloop. (**C**) Rotated view from (B) highlighting the 2′-hydroxyl groups (pink dots) of EBS1 and the 5′ exon that are directly recognized by the distal section of DV and nucleotide A106(λ), respectively. See fig. S7 for the complete networks of stacking and hydrogen bonding interactions that stabilize these two structural states. IBS1, intron binding site 1.

is preserved, and we specifically propose that, in the spliceosome, the third invariant A (A45 in mammals, A51 in yeast) of the ACAGAGA motif of U6 snRNA (Fig. 6) is the counterpart of group II A287. This assignment is supported by several lines of evidence. First, just like A287(γ) in group II introns, A45/51 lies immediately 5′ to the dinucleotide engaged in the catalytic triplex. Second, substitution of A51 totally blocks the second step of pre-mRNA splicing in yeast (*43*). Third, A45/51 and nucleotide U2 of spliceosomal introns cross-link with each other specifically during the second step of pre-mRNA splicing in both mammals (*44*) and yeast (*45*), and the chemical mechanism of the cross-link is suggestive of a stacking interaction between these nucleotides.

Genetic suppression data in yeast (*46*) led to the suggestion that the highly conserved Prp8 spliceosomal protein, which lies at the heart of the splicing machinery and is evolutionarily related to group II–encoded reverse transcriptases (*47*), recognizes a second-step active-site RNA structure composed of intron U2, A51 of U6 snRNA, and the AG dinucleotide at the 3′ intron boundary. These four nucleotides are tightly clustered in our model (Fig. 6) and could be recognized by the "catalytic cavity" of Prp8 (*48*). Moreover, the adenosine N1 atoms at the branchpoint and pnt, which are both known to be important for the exon ligation step (*49*, *50*), yet without an identified RNA partner, constitute additional candidates for interaction with Prp8 (Fig. 6C).

Available cryo-EM structures of the catalytically activated spliceosome pertain to the first step of pre-mRNA splicing. They reveal the positioning of the 2′-5′ branch structure either immediately (*51*) or soon (*52*) after its formation. Comparison with our second-step spliceosomal model leads to the conclusion that, as already proposed for group II introns (*10*), the 2′-5′ branch needs to undergo a major translocation between

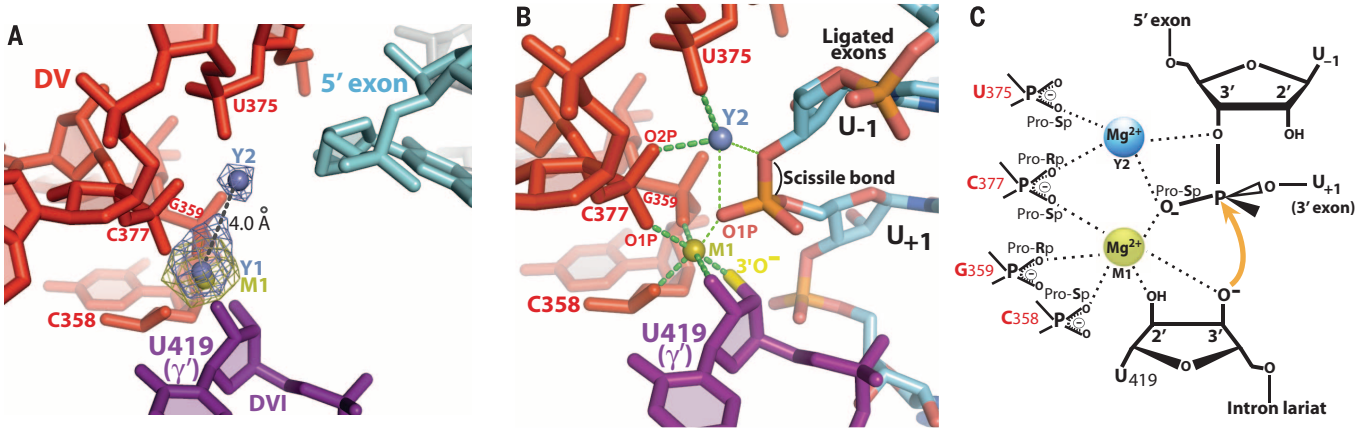


Fig. 5. The catalytic mechanism of group II reverse splicing. (**A**) Ytterbium (Yb^{3+}; an anomalous scatterer that mimics Mg^{2+}) anomalous difference map (violet, contoured at 21σ) reveals two ions (Y1 and Y2) at the catalytic center in the structure of the Oc19 lariat complexed with an unreactive 5′-exon analog. The native density $F_o - F_c$ omit map (olive, contoured at 7σ) for metal M1 perfectly superimposes with the anomalous difference map for Y1. (**B**) Inner-sphere coordinations directly observed in our structure for metals M1 and Y2 are shown as thick green dashed lines, whereas thin dotted lines are inner-sphere contacts inferred after docking [without further modeling (*12*)] a ligated exon substrate (from PDB entry 4E8K, colored by atoms). The nucleophile of the reverse-splicing reaction (the intron terminal 3′-oxyanion) is in yellow. (**C**) Crystallographically derived model of the transition state according to the two–metal ion mechanism for group II reverse splicing (O1P and O2P are pro-Sp and pro-Rp oxygens, respectively).

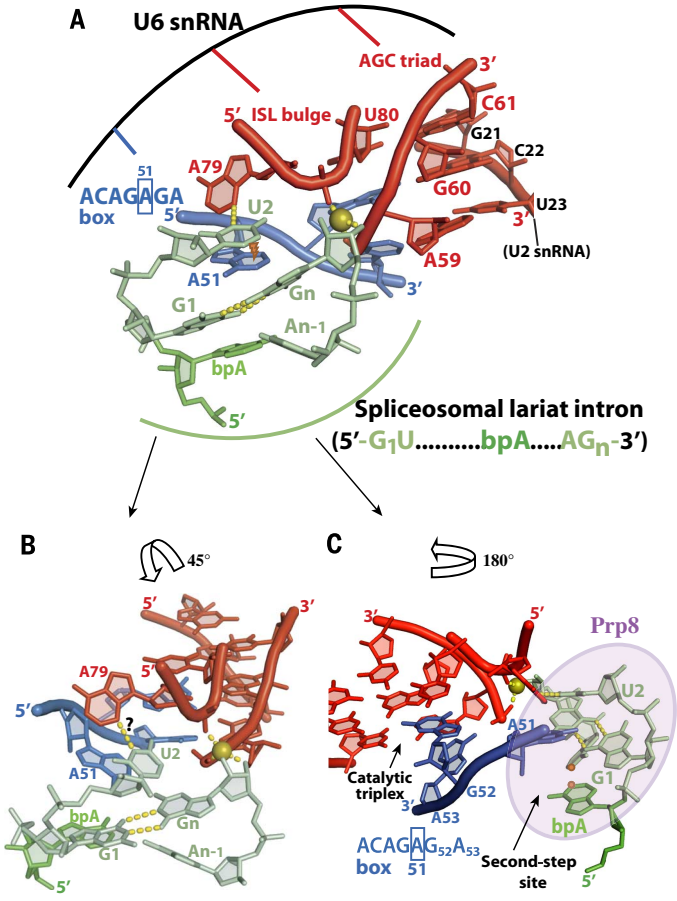


Fig. 6. Second-step active site of the spliceosome modeled after its group II counterpart. (**A**) The second step–specific interaction between the terminal G's of spliceosomal introns (light green) was modeled into the group II intron active site (*12*); numbering of snRNA sequences is according to the yeast spliceosome. The catalytically essential components of U6 snRNA (AGC triad, ISL bulge, and ACAGAGA box) are colored as their group II counterparts in DV and the J2/3 linker (Fig. 2). The orange lightning bolt stands for a second step–specific cross-link between nucleotide A51/A45 of U6 snRNA and the U at intron position 2 (see text). The location of group II metal M1 is potentially preserved in the spliceosome (*12*). (**B**) In this view, G1:Gn hydrogen bonds are shown as dashed yellow lines. A potential interaction between U6 snRNA A79 (homologous to Oc19 A376; Fig. 2) and intron nucleotide U2 is suggested. (**C**) Back view of the spliceosomal active site and the potential lariat recognition site for Prp8. The nucleotide groups highlighted as orange dots are required for the second step of pre-mRNA splicing (see text).

the two transesterification reactions. Such a large-scale movement could be part of the extensive conformational rearrangement that is believed to occur between the two catalytic steps of pre-mRNA splicing (*35*), and although the structure of the spliceosome active site for exon ligation remains to be established, our model specifically predicts that the 2′-5′ branch structure will prove essential to its assembly. Just as in group II introns, such structural coupling of the two chemical steps of splicing provides an RNA-based "proofreading" device that could complement protein-dependent proofreading (*53*) in ensuring the fidelity of spliceosomal splicing.

On the other hand, implications for nuclear pre-mRNA splicing of the strand-shifting mechanism that we have uncovered in the group II system are not immediately obvious because, in spliceosomal introns, the presence of a "spacer" of variable length and sequence that generally separates the branchpoint from the 3′ splice site would seem to make it superfluous to rearrange the base-pairing pattern around the branchpoint. Nevertheless, it is possible that in some unicellular eukaryotes, rearrangement of the branch site takes place through an RNA-based mechanism similar to the one used to remodel DVI of group II introns when seven nucleotides separate the branchpoint from the 3′ splice site (fig. S8).

In conclusion, our crystal structures have brought to light key structural features that explain the prominent role of the 2′-5′ branch in the assembly of the intron active site for exon ligation and the initiation of reverse splicing. In doing so, they elucidate the reasons why the lariat bond was so stubbornly conserved not only during the diversification of group II introns but also all the way along the evolutionary path that led to the emergence of nuclear premessenger introns and their splicing machinery from a group II ancestor.

Methods summary

RNA preparation and crystallography

The Oc19 chimeric construct, in which DVI and part of the IC1 subdomain of the Oceanobacillus intron have been replaced by their counterparts in intron Av.I.2 (fig. S1) (*12*), was selected for crystallography based on its ability to generate abundant intron lariat during self-splicing (fig. S4). For crystallization purposes, RNA synthesis from linearized Oc19 plasmid DNA was performed at 37°C with home-prepared T7 RNA polymerase. To generate Oc19 lariat, posttranscription samples were desalted prior to incubation under self-splicing conditions. The Oc19 lariat RNA was purified by denaturing polyacrylamide gel electrophoresis and stored at −20°C. Prior to crystallization trials, the purified lariat was refolded and subsequently concentrated to 0.7 μg/μl. Crystals were grown in sitting drops by vapor diffusion at 28°C. Crystallization drops of the lariat alone contained 1 μl of reservoir solution [50 mM Na-cacodylate, pH 6.5, 225 mM NH_4Cl, 95 mM $MgCl_2$, and 22% 2-methyl-2,4-pentanediol (MPD)] and 1.5 μl of renatured RNA. For cocrystallization of the Oc19 lariat with the 5′ exon, an unreactive RNA analog of the latter was added to the drops at a final concentration of 12 μM. Ytterbium derivative crystals were obtained by soaking native crystals in the crystallization solution supplemented with 0.5 mM Yb^{3+} chloride. X-ray diffraction data (table S1) were collected on the PROXIMA 1 beamline at the synchrotron SOLEIL (Saint-Aubin, France), and both structures were solved by molecular replacement with MOLREP (*54*) using as a search model, PDB entry 4FAW (7), which corresponds to a linear form of the O.i.I1 ribozyme lacking domain VI. Model building was done with COOT (*55*), and structure refinement was achieved with BUSTER (*56*) and PHENIX (*57*). The positions of ytterbium atoms were identified by MR-SAD (*12*) with PHASER (*58*).

Kinetic analyses

Kinetic analyses of Oc19 and derived mutant constructs (figs. S4 and S5) were performed with ^{32}P-labeled precursors generated from in vitro transcriptions. Self-splicing assays were carried out in 50 mM Tris-HCl, pH 7.5 (37°C), 2 M NH_4Cl, 10 mM $MnCl_2$, and 0.01% SDS, and reaction products were analyzed on 4% polyacrylamide–8 M urea gels (fig. S4). Rate constants for branching (kbr) and hydrolysis (khy), two parallel reactions at the 5′ splice site, were calculated from simple exponential fits (e.g., fig. S5A) (*12*).

REFERENCES AND NOTES

1. F. Michel, J.-L. Ferat, Structure and activities of group II introns. *Annu. Rev. Biochem.* **64**, 435–461 (1995). doi: 10.1146/annurev.bi.64.070195.002251; pmid: 7574489
2. A. M. Lambowitz, M. Belfort, Mobile bacterial group II introns at the crux of eukaryotic evolution. *Microbiol. Spectrum* **3**, MDNA3-0050-2014 (2015). doi: 10.1128/microbiolspec.MDNA3-0050-2014; pmid: 26104554
3. A. M. Lambowitz, S. Zimmerly, Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* **3**, a003616 (2011). doi: 10.1101/cshperspect.a003616; pmid: 20463000
4. F. Zhuang, M. Mastroianni, T. B. White, A. M. Lambowitz, Linear group II intron RNAs can retrohome in eukaryotes and may use nonhomologous end-joining for cDNA ligation. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18189–18194 (2009). doi: 10.1073/pnas.0910277106; pmid: 19833873
5. E. Dème, A. Nolte, A. Jacquier, Unexpected metal ion requirements specific for catalysis of the branching reaction in a group II intron. *Biochemistry* **38**, 3157–3167 (1999). doi: 10.1021/bi982462j; pmid: 10074371
6. N. Toor, K. S. Keating, S. D. Taylor, A. M. Pyle, Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82 (2008). doi: 10.1126/science.1153803; pmid: 18388288
7. M. Marcia, A. M. Pyle, Visualizing group II intron catalysis through the stages of splicing. *Cell* **151**, 497–507 (2012). doi: 10.1016/j.cell.2012.09.033; pmid: 23101623
8. A. R. Robart, R. T. Chan, J. K. Peters, K. R. Rajashankar, N. Toor, Crystal structure of a eukaryotic group II intron lariat. *Nature* **514**, 193–197 (2014). doi: 10.1038/nature13790; pmid: 25252982

9. G. Qu *et al*., Structure of a group II intron in complex with its reverse transcriptase. *Nat. Struct. Mol. Biol.* **23**, 549–557 (2016). doi: 10.1038/nsmb.3220; pmid: 27136327
10. C.-F. Li, M. Costa, F. Michel, Linking the branchpoint helix to a newly found receptor allows lariat formation by a group II intron. *EMBO J.* **30**, 3040–3051 (2011). doi: 10.1038/emboj.2011.214; pmid: 21712813
11. D. Monachello, F. Michel, M. Costa, Activating the branch-forming splicing pathway by reengineering the ribozyme component of a natural group II intron. *RNA* **22**, 443–455 (2016). doi: 10.1261/rna.054643.115; pmid: 26769855
12. Materials and methods are available as supporting material on *Science* Online.
13. A. Jacquier, F. Michel, Base-pairing interactions involving the 5′ and 3′-terminal nucleotides of group II self-splicing introns. *J. Mol. Biol.* **213**, 437–447 (1990). doi: 10.1016/S0022-2836(05)80206-2; pmid: 2191139
14. G. Chanfreau, A. Jacquier, Interaction of intronic boundaries is required for the second splicing step efficiency of a group II intron. *EMBO J.* **12**, 5173–5180 (1993). pmid: 8262060
15. J. E. Sokoloski, S. A. Godfrey, S. E. Dombrowski, P. C. Bevilacqua, Prevalence of *syn* nucleobases in the active sites of functional RNAs. *RNA* **17**, 1775–1787 (2011). doi: 10.1261/rna.2759911; pmid: 21873463
16. F. Michel, K. Umesono, H. Ozeki, Comparative and functional anatomy of group II catalytic introns—A review. *Gene* **82**, 5–30 (1989). doi: 10.1016/0378-1119(89)90026-7; pmid: 2684776
17. M. A. Candales *et al*., Database for bacterial group II introns. *Nucleic Acids Res.* **40**, D187–D190 (2012). doi: 10.1093/nar/gkr1043; pmid: 22080509
18. L. Zhang, J. A. Doudna, Structural insights into group II intron catalysis and branch-site selection. *Science* **295**, 2084–2088 (2002). doi: 10.1126/science.1069268; pmid: 11859154
19. V. T. Chu, C. Adamidi, Q. Liu, P. S. Perlman, A. M. Pyle, Control of branch-site choice by a group II intron. *EMBO J.* **20**, 6866–6876 (2001). doi: 10.1093/emboj/20.23.6866; pmid: 11726522
20. R. Russell, I. Jarmoskaite, A. M. Lambowitz, Toward a molecular understanding of RNA remodeling by DEAD-box proteins. *RNA Biol.* **10**, 44–55 (2013). doi: 10.4161/rna.22210; pmid: 22995827
21. L. Dai *et al*., A three-dimensional model of a group II intron RNA and its interaction with the intron-encoded reverse transcriptase. *Mol. Cell* **30**, 472–485 (2008). doi: 10.1016/j.molcel.2008.04.001; pmid: 18424209
22. G. Chanfreau, A. Jacquier, An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO J.* **15**, 3466–3476 (1996). pmid: 8670849
23. A. Jacquier, F. Michel, Multiple exon-binding sites in class II self-splicing introns. *Cell* **50**, 17–29 (1987). doi: 10.1016/0092-8674(87)90658-1; pmid: 3297351
24. M. Costa, F. Michel, E. Westhof, A three-dimensional perspective on exon binding by a group II self-splicing intron. *EMBO J.* **19**, 5007–5018 (2000). doi: 10.1093/emboj/19.18.5007; pmid: 10990464
25. M. Costa, F. Michel, Tight binding of the 5′ exon to domain I of a group II self-splicing intron requires completion of the intron active site. *EMBO J.* **18**, 1025–1037 (1999). doi: 10.1093/emboj/18.4.1025; pmid: 10022844
26. K. A. Johnson, Role of induced fit in enzyme specificity: A molecular forward/reverse switch. *J. Biol. Chem.* **283**, 26297–26301 (2008). doi: 10.1074/jbc.R800034200; pmid: 18544537
27. E. J. Sontheimer, P. M. Gordon, J. A. Piccirilli, Metal ion catalysis during group II intron self-splicing: Parallels with the spliceosome. *Genes Dev.* **13**, 1729–1741 (1999). doi: 10.1101/gad.13.13.1729; pmid: 10398685
28. P. M. Gordon, E. J. Sontheimer, J. A. Piccirilli, Kinetic characterization of the second step of group II intron splicing: Role of metal ions and the cleavage site 2′-OH in catalysis. *Biochemistry* **39**, 12939–12952 (2000). doi: 10.1021/bi001089o; pmid: 11041859
29. P. M. Gordon, R. Fong, J. A. Piccirilli, A second divalent metal ion in the group II intron reaction center. *Chem. Biol.* **14**, 607–612 (2007). doi: 10.1016/j.chembiol.2007.05.008; pmid: 17584608
30. N. Toor, K. Rajashankar, K. S. Keating, A. M. Pyle, Structural basis for exon recognition by a group II intron. *Nat. Struct. Mol. Biol.* **15**, 1221–1222 (2008). doi: 10.1038/nsmb.1509; pmid: 18953333
31. A. V. Kazantsev, A. A. Krivenko, N. R. Pace, Mapping metal-binding sites in the catalytic domain of bacterial RNase P RNA. *RNA* **15**, 266–276 (2009). doi: 10.1261/rna.1331809; pmid: 19095619
32. T. A. Steitz, J. A. Steitz, A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6498–6502 (1993). doi: 10.1073/pnas.90.14.6498; pmid: 8341661
33. M. Podar, P. S. Perlman, R. A. Padgett, Stereochemical selectivity of group II intron splicing, reverse splicing, and hydrolysis reactions. *Mol. Cell. Biol.* **15**, 4466–4478 (1995). doi: 10.1128/MCB.15.8.4466; pmid: 7542746
34. R. A. Padgett, M. Podar, S. C. Boulanger, P. S. Perlman, The stereochemical course of group II intron self-splicing. *Science* **266**, 1685–1688 (1994). doi: 10.1126/science.7527587; pmid: 7527587
35. M. C. Wahl, C. L. Will, R. Lührmann, The spliceosome: Design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009). doi: 10.1016/j.cell.2009.02.009; pmid: 19239890
36. T. R. Cech, The generality of self-splicing RNA: Relationship to nuclear mRNA splicing. *Cell* **44**, 207–210 (1986). doi: 10.1016/0092-8674(86)90751-8; pmid: 2417724
37. S. M. Fica, M. A. Mefford, J. A. Piccirilli, J. P. Staley, Evidence for a group II intron–like catalytic triplex in the spliceosome. *Nat. Struct. Mol. Biol.* **21**, 464–471 (2014). doi: 10.1038/nsmb.2815; pmid: 24747940
38. S. M. Fica *et al*., RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**, 229–234 (2013). doi: 10.1038/nature12734; pmid: 24196718
39. M. Anokhina *et al*., RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO J.* **32**, 2804–2818 (2013). doi: 10.1038/emboj.2013.198; pmid: 24002212
40. J. G. Umen, C. Guthrie, The second catalytic step of pre-mRNA splicing. *RNA* **1**, 869–885 (1995). pmid: 8548652
41. R. Parker, P. G. Siliciano, Evidence for an essential non-Watson–Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature* **361**, 660–662 (1993). doi: 10.1038/361660a0; pmid: 8437627
42. A. Deirdre, J. Scadden, C. W. Smith, Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs. *EMBO J.* **14**, 3236–3246 (1995). pmid: 7621835
43. P. Fabrizio, J. Abelson, Two domains of yeast U6 small nuclear RNA required for both steps of nuclear precursor messenger RNA splicing. *Science* **250**, 404–409 (1990). doi: 10.1126/science.2145630; pmid: 2145630
44. E. J. Sontheimer, J. A. Steitz, The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* **262**, 1989–1996 (1993). doi: 10.1126/science.8266094; pmid: 8266094
45. C. H. Kim, J. Abelson, Site-specific crosslinks of yeast U6 snRNA to the pre-mRNA near the 5′ splice site. *RNA* **2**, 995–1010 (1996). pmid: 8849776
46. C. A. Collins, C. Guthrie, Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. *Genes Dev.* **13**, 1970–1982 (1999). doi: 10.1101/gad.13.15.1970; pmid: 10444595
47. M. Dlakić, A. Mushegian, Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. *RNA* **17**, 799–808 (2011). doi: 10.1261/rna.2396011; pmid: 21441348
48. C. Yan *et al*., Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349**, 1182–1191 (2015). doi: 10.1126/science.aac7629; pmid: 26292707
49. C. C. Query, S. A. Strobel, P. A. Sharp, Three recognition events at the branch-site adenine. *EMBO J.* **15**, 1392–1402 (1996). pmid: 8635472
50. R. K. Gaur, L. Beigelman, P. Haeberli, T. Maniatis, Role of adenine functional groups in the recognition of the 3′-splice-site AG during the second step of pre-mRNA splicing. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 115–120 (2000). doi: 10.1073/pnas.97.1.115; pmid: 10618380
51. W. P. Galej *et al*., Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537**, 197–201 (2016). doi: 10.1038/nature19316; pmid: 27459055
52. R. Wan, C. Yan, R. Bai, G. Huang, Y. Shi, Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* **353**, 895–904 (2016). doi: 10.1126/science.aag2235; pmid: 27445308
53. D. R. Semlow, J. P. Staley, Staying on message: Ensuring fidelity in pre-mRNA splicing. *Trends Biochem. Sci.* **37**, 263–273 (2012). doi: 10.1016/j.tibs.2012.04.001
54. A. Vagin, A. Teplyakov, Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25 (2010). doi: 10.1107/S0907444909042589; pmid: 20057045
55. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004). doi: 10.1107/S0907444904019158; pmid: 15572765
56. E. Blanc *et al*., Refinement of severely incomplete structures with maximum likelihood in *BUSTER-TNT*. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2210–2221 (2004). doi: 10.1107/S0907444904016427; pmid: 15572774
57. P. D. Adams *et al*., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010). doi: 10.1107/S0907444909052925; pmid: 20124702
58. A. J. McCoy *et al*., Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007). doi: 10.1107/S0021889807021206; pmid: 19461840
59. J. S. Reuter, D. H. Mathews, RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010). doi: 10.1186/1471-2105-11-129; pmid: 20230624

ACKNOWLEDGMENTS

We thank the SOLEIL Synchrotron (Saint-Aubin, France) for beamtime allocation and the scientists at the PROXIMA 1 beamline, especially P. Legrand, for advice and help in data collection and initial processing. We thank L. Sperling and D. Fourmy for comments on the manuscript. This work was supported by the French Agence Nationale de la Recherche (grant ANR-10-BLAN-1502 to F.M. and E.W.). Coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession codes 5J01 (lariat intron) and 5J02 (lariat intron bound to the 5′-exon analog). The authors will provide coordinates of the model in Fig. 6 upon request.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/aaf9258/suppl/DC1
Materials and Methods
Figs. S1 to S8
Table S1
References (*60*–*67*)

21 April 2016; accepted 27 October 2016
10.1126/science.aaf9258

RESEARCH ARTICLE

ATMOSPHERIC SCIENCE

Global atmospheric particle formation from CERN CLOUD measurements

Eimear M. Dunne,[1]*† Hamish Gordon,[2]*‡ Andreas Kürten,[3] João Almeida,[2,3] Jonathan Duplissy,[4] Christina Williamson,[3]§ Ismael K. Ortega,[5]|| Kirsty J. Pringle,[1] Alexey Adamov,[6] Urs Baltensperger,[7] Peter Barmet,[7] Francois Benduhn,[8] Federico Bianchi,[6,7] Martin Breitenlechner,[9]¶ Antony Clarke,[10] Joachim Curtius,[3] Josef Dommen,[7] Neil M. Donahue,[11,6] Sebastian Ehrhart,[2,3] Richard C. Flagan,[12] Alessandro Franchin,[6] Roberto Guida,[2] Jani Hakala,[6] Armin Hansel,[9,13] Martin Heinritzi,[3] Tuija Jokinen,[6]# Juha Kangasluoma,[6] Jasper Kirkby,[2,3] Markku Kulmala,[6] Agnieszka Kupc,[14]§ Michael J. Lawler,[15]# Katrianne Lehtipalo,[6,7] Vladimir Makhmutov,[16] Graham Mann,[1] Serge Mathot,[2] Joonas Merikanto,[6] Pasi Miettinen,[15] Athanasios Nenes,[17,18,19] Antti Onnela,[2] Alexandru Rap,[1] Carly L. S. Reddington,[1] Francesco Riccobono,[7] Nigel A. D. Richards,[1] Matti P. Rissanen,[6] Linda Rondo,[3] Nina Sarnela,[6] Siegfried Schobesberger,[6] Kamalika Sengupta,[1] Mario Simon,[3] Mikko Sipilä,[6] James N. Smith,[15]# Yuri Stozkhov,[16] Antonio Tomé,[20] Jasmin Tröstl,[7] Paul E. Wagner,[14] Daniela Wimmer,[3,6] Paul M. Winkler,[14] Douglas R. Worsnop,[6,21] Kenneth S. Carslaw[1]‡**

Fundamental questions remain about the origin of newly formed atmospheric aerosol particles because data from laboratory measurements have been insufficient to build global models. In contrast, gas-phase chemistry models have been based on laboratory kinetics measurements for decades. We built a global model of aerosol formation by using extensive laboratory measurements of rates of nucleation involving sulfuric acid, ammonia, ions, and organic compounds conducted in the CERN CLOUD (Cosmics Leaving Outdoor Droplets) chamber. The simulations and a comparison with atmospheric observations show that nearly all nucleation throughout the present-day atmosphere involves ammonia or biogenic organic compounds, in addition to sulfuric acid. A considerable fraction of nucleation involves ions, but the relatively weak dependence on ion concentrations indicates that for the processes studied, variations in cosmic ray intensity do not appreciably affect climate through nucleation in the present-day atmosphere.

Nucleation of particles occurs throughout Earth's atmosphere by condensation of trace vapors (*1*–*3*). Around 40 to 70% of global cloud condensation nuclei (CCN) (*4*–*6*) are thought to originate as nucleated particles, so the process has a major influence on the microphysical properties of clouds and the radiative balance of the global climate system. However, laboratory measurements are needed to disentangle and quantify the processes that contribute to particle formation, and very few laboratory measurements exist under atmospheric conditions (*7*–*10*). This leaves open fundamental questions concerning the origin of particles on a global scale. First, it is not known whether nucleation is predominantly a neutral process, as assumed in most models (*11*–*13*), or whether atmospheric ions are important (*6*, *14*–*16*). This relates to the question of whether solar-modulated galactic cosmic rays (GCRs) affect aerosols, clouds, and climate (*17*–*21*). Second, the lack of measurements of nucleation rates at low temperatures means that the origin of new particles in the vast regions of the cold free troposphere has not yet been experimentally established. Third, whereas it has been shown that nucleation of sulfuric acid (H_2SO_4)–water particles in the boundary layer requires stabilizing molecules such as ammonia (NH_3), amines, or oxidized organic compounds (*7*, *8*, *22*–*24*), it is not yet known from existing experimental data over how much of the troposphere these molecules are important for nucleation. Robust atmospheric models to answer these questions need to be founded on direct measurements of nucleation rates. At present, to simulate nucleation over a very wide range of atmospheric conditions, global models must use theoretical nucleation models (*25*, *26*), which can require adjustments to the nucleation rates of several orders of magnitude to obtain reasonable agreement with ambient observations (*27*, *28*).

The lack of an experimentally based model of global particle nucleation is in stark contrast to global models of atmospheric gas-phase chemistry, which have been based on laboratory kinetics measurements since the 1970s (*29*, *30*). We built a model of the global aerosol system based on laboratory nucleation-rate measurements that is able to explain global particle concentrations without any adjustment to the rates. We derive some selected implications for CCN, cloud albedo, and hence Earth's radiative forcing.

Chamber measurements of inorganic nucleation

In Fig. 1, we present ~350 measurements of ion-induced and neutral particle formation from H_2SO_4 and NH_3 vapors, conducted in the CLOUD (Cosmics Leaving Outdoor Droplets) chamber at the CERN Proton Synchrotron. To enable globally applicable nucleation rate expressions to be developed, we combined ~250 earlier measurements (*7*, *23*, *31*) with ~100 new measurements covering a much wider range of vapor concentrations than was achieved in the previous

[1]School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK. [2]European Organization for Nuclear Research (CERN), 1211 Geneva, Switzerland. [3]Goethe-University Frankfurt am Main, Institute for Atmospheric and Environmental Sciences, Altenhöferallee 1, 60438 Frankfurt am Main, Germany. [4]Helsinki Institute of Physics, FI-00014 Helsinki, Finland. [5]Laboratoire de Physique des Lasers, Atomes et Molécules (PhLAM), Université Lille 1, UMR 8523 CNRS, 59655 Villeneuve d'Ascq, France. [6]Department of Physics, University of Helsinki, Post Office Box 64, FI-00014 Helsinki, Finland. [7]Laboratory of Atmospheric Chemistry, Paul Scherrer Institute, 5232 Villigen, Switzerland. [8]Institute for Advanced Sustainability Studies, Berliner Straße 130, D-14467 Potsdam, Germany. [9]Institute of Ion Physics and Applied Physics, Leopold-Franzens University, Technikerstraße 25, 6020 Innsbruck, Austria. [10]Department of Oceanography, University of Hawaii, 1000 Pope Road, Honolulu, HI 96822, USA. [11]Center for Atmospheric Particle Studies, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA. [12]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA. [13]Ionicon, 6020 Innsbruck, Austria. [14]Faculty of Physics, University of Vienna, Boltzmanngaße 5, 1090 Vienna, Austria. [15]University of Eastern Finland, Post Office Box 1627, 70211 Kuopio, Finland. [16]Lebedev Physical Institute, Solar and Cosmic Ray Research Laboratory, 119991 Moscow, Russia. [17]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA. [18]Institute of Chemical Engineering Sciences (ICE-HT), Foundation for Research and Technology, Hellas, 26504 Patras, Greece. [19]Institute for Environmental Research and Sustainable Development, National Observatory of Athens, I. Metaxa & Vas. Pavlou, 15236 Palea Penteli, Greece. [20]CENTRA-SIM, University of Lisbon and University of Beira Interior, 1749-016 Lisbon, Portugal. [21]Aerodyne Research, Billerica, MA 01821, USA.

*These authors contributed equally to this work. †Present address: Finnish Meteorological Institute, Atmospheric Research Centre of Eastern Finland, PL 1627, 70211 Kuopio, Finland. **‡Corresponding author. Email: hamish.gordon@cern.ch (H.G.); leckscc@ds.leeds.ac.uk (K.S.C.)** §Present address: Chemical Sciences Division, Earth Systems Research Laboratory, 325 Broadway, National Oceanic and Atmospheric Administration, Boulder, CO 80305, USA, and Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA. ||Present address: ONERA–The French Aerospace Lab, F-91123 Palaiseau, France. ¶Present address: School of Engineering and Applied Sciences, Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. #Present address: Department of Chemistry, University of California, 1102 Natural Science II, Irvine, CA 92697, USA. **Present address: Department of Atmospheric Sciences, University of Washington, Seattle, WA 98195, USA.

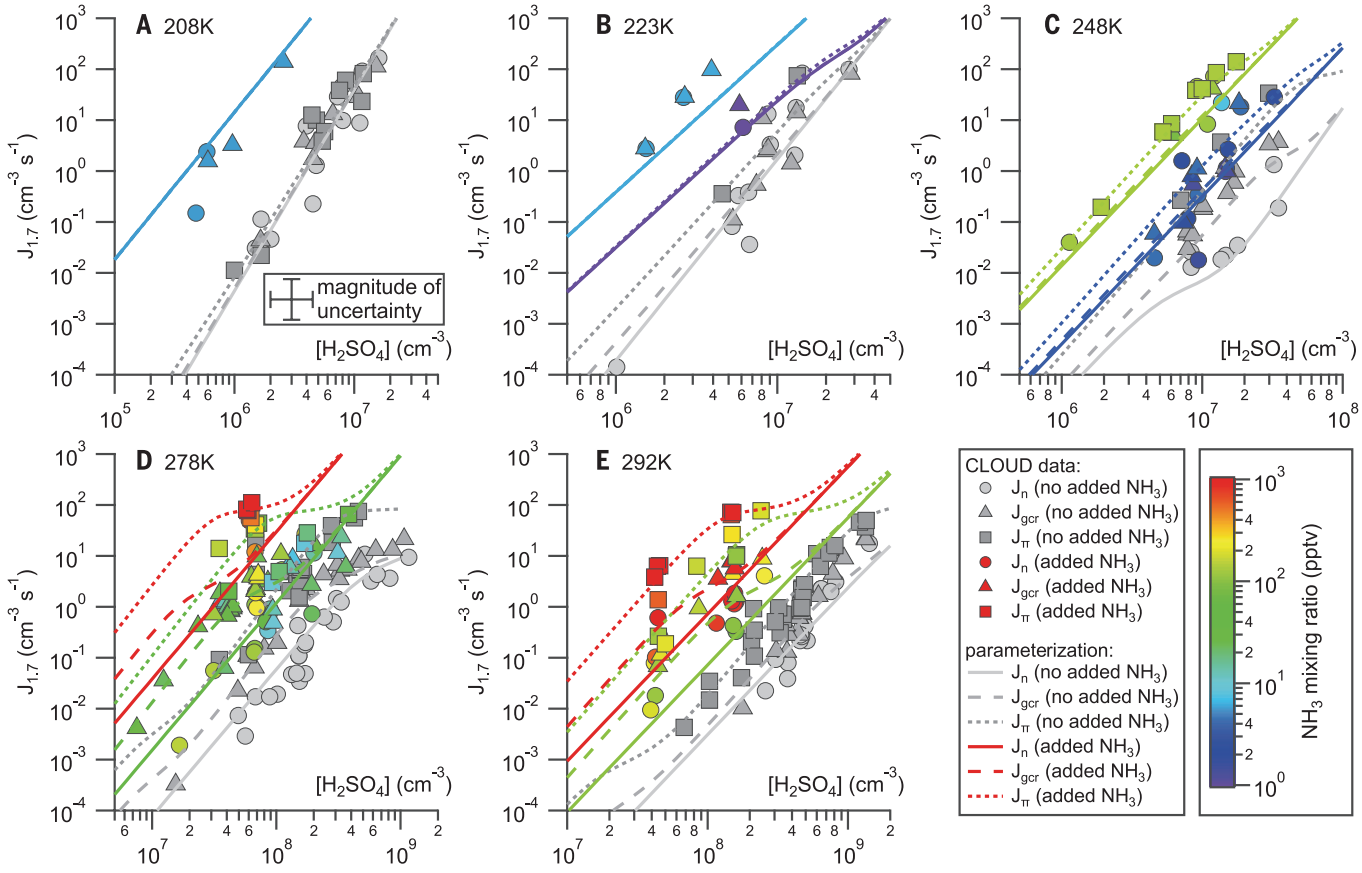


Fig. 1. Measured and parameterized nucleation rates. Neutral, GCR, and pion-beam nucleation rates (J) are shown at 1.7 nm mobility diameter as a function of sulfuric acid concentration. Rates are shown at (**A**) 208, (**B**) 223, (**C**) 248, (**D**) 278, and (**E**) 292 K. The symbols show measured values of nucleation rates: circles for neutral (n) rates (ion-pair production rate q = 0 cm^{-3} s^{-1}), triangles for GCR rates (q = 2 cm^{-3} s^{-1}), and squares for pion beam (π) rates (q ~ 75 cm^{-3} s^{-1}). The lines show parameterized nucleation rates (supplementary materials, section 8): solid lines for neutral rates, dashed lines for GCR rates, and dotted lines for pion beam rates. Gray symbols and lines indicate contaminant concentrations of NH_3 below the detection limit of the instruments (supplementary materials, section 6), whereas colored symbols and lines represent measurements at NH_3 concentrations indicated by the color scale. For clarity, the uncertainties on each data point are not shown, but the overall uncertainty of a factor of 2.5 on nucleation rate and a factor of 1.5 on $[H_2SO_4]$ is shown separately from the real data in (A). The contaminant level of ammonia increases as temperature increases. This explains why the ionization effect without added ammonia at 292 K is smaller than that at 278 K and why the nucleation rates without added ammonia are similar at these temperatures.

experiments. We also studied ternary nucleation at temperatures as low as 208 K, typical of the upper troposphere. We combined these ~350 inorganic measurements with data on organic-mediated nucleation (*24*) to quantify nucleation rates throughout the troposphere. The experiments were performed under neutral, natural GCR and charged pion beam conditions (supplementary materials) (*32*). GCRs create ion pairs in the chamber at a rate of about 2 cm^{-3} s^{-1}, characteristic of the lower atmosphere, and the controllable pion beam is able to reproduce equilibrium ion-pair concentrations between ground level and the upper troposphere (*33*). Neutral conditions are achieved by removing ions from the chamber with an electric field.

Figure 1 shows how our full data set of inorganic nucleation rates depends on H_2SO_4, NH_3, temperature, and ionization rates. At 208 K, the nucleation rates are up to 10^4 times as high as at 248 K, the lowest temperature investigated in any previous study (*7*). The enhancement of the neutral nucleation rate caused by natural GCR ionization reaches about a factor of 15 at temperatures found in the lower troposphere. However, we measured no appreciable enhancement due to ions at lower temperatures representative of the upper troposphere, indicating that evaporation of the corresponding neutral clusters is strongly suppressed. Ammonia mixing ratios of a few parts per trillion by volume (pptv) greatly enhance the nucleation rate. For example, at 223 K, the neutral rate rises by more than a factor of 1000 when NH_3 is increased from an estimated 0.05 pptv contaminant level to 6 pptv (Fig. 1B). The negative ion cluster composition (fig. S1) (*34*) confirms that NH_3 molecules are participating in and enhancing ion-induced nucleation, although binary nucleation of sulfuric acid and water without ammonia cannot be neglected.

Although most of our measurements were conducted at 38% relative humidity (RH), the dependence of nucleation rates on RH was also measured (supplementary materials, section 9) and was found to be stronger than the dependence on ion concentrations but weaker than that on other factors. At 223 and 208 K, temperatures typical of the upper free troposphere, a change in RH between 20 and 100% induces at most a factor of 5, and typically a 50 to 100%, change in the nucleation rate, whereas ambient atmospheric concentrations of H_2SO_4 and NH_3 vary over many orders of magnitude. At 298 K, the nucleation rate increases by about a factor of 10 between 40 and 80% RH, but this does not strongly affect CCN concentrations (see the model results below).

The inorganic nucleation rates are parameterized in four dimensions—temperature, $[H_2SO_4]$, $[NH_3]$, and ion concentrations—and fitted to our full data set of ~350 inorganic measurements. The RH dependence was not included in the fit because of insufficient data, although we tested its effect in separate model sensitivity studies, described below. Because one of our objectives was to determine the relative importance of binary and ternary nucleation in the global atmosphere, we used the molecular composition of the charged nucleating clusters from APi-TOF (atmospheric pressure interface time-of-flight)

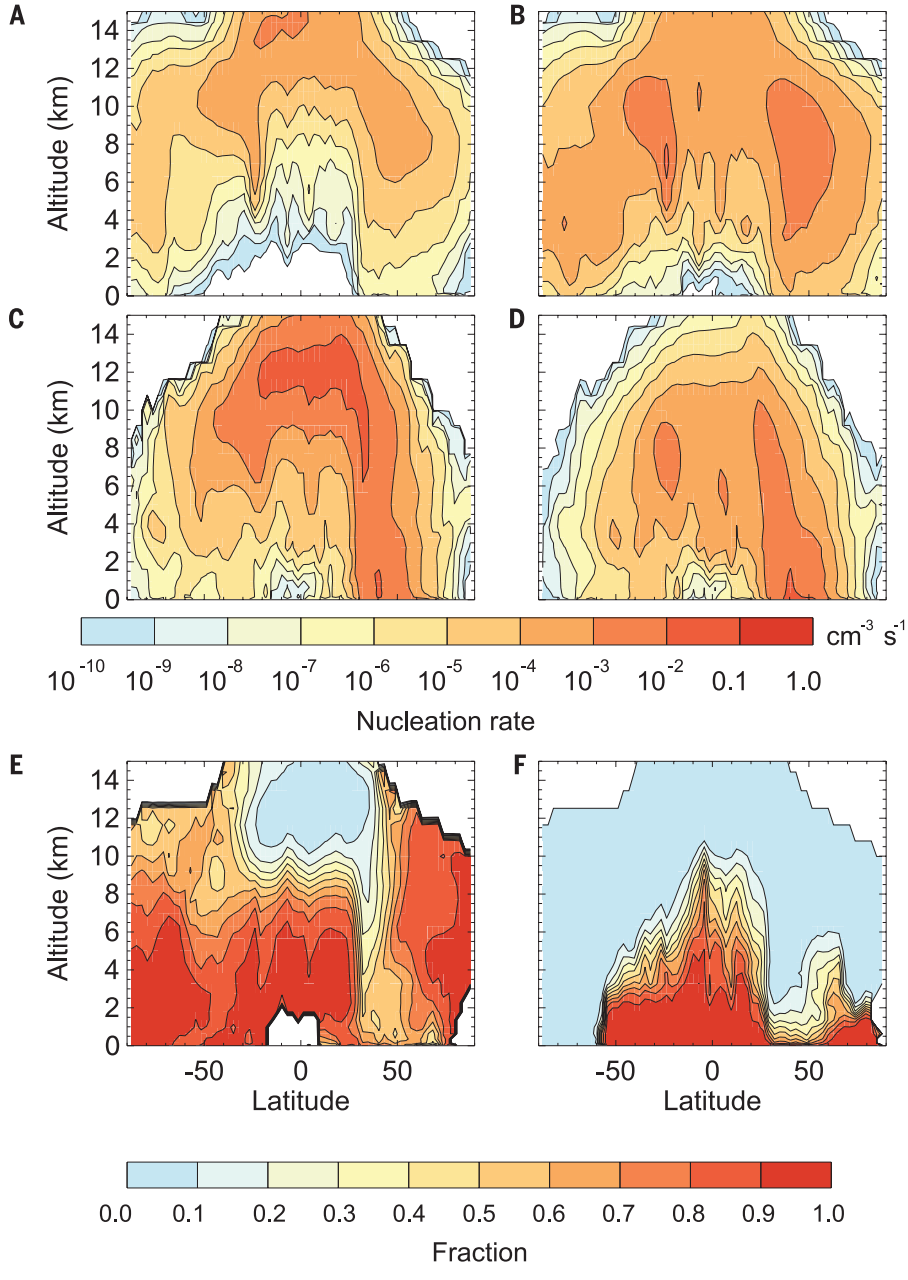


Fig. 2. Modeled zonal and annual mean particle formation rates, per cubic centimeter per second, at 3 nm diameter. (**A**) Binary (H_2SO_4-H_2O) neutral nucleation rate, (**B**) binary ion-induced nucleation rate, (**C**) ternary (H_2SO_4-NH_3-H_2O) neutral nucleation rate, (**D**) ternary ion-induced nucleation rate, (**E**) ion-induced fraction of inorganic nucleation, and (**F**) fraction of all nucleation from ternary organic nucleation (H_2SO_4-BioOxOrg-H_2O). In (E) and (F), the model data are shown only where the overall mean nucleation rate exceeds 10^{-6} cm^{-3} s^{-1}.

mass spectrometry measurements to unambiguously verify the amount of ammonia or organic species in, or their absence from, the charged nucleating clusters (*34*, *35*) (supplementary materials, section 6). Guided by these mass spectra, the four-dimensional global fit enables us to determine the dependence on trace gas concentrations and ions, even though the data are sparse in any one dimension. Over almost the full range of the measurements (supplementary materials, section 8), the nucleation rate varies approximately as $[H_2SO_4]^3$, linearly with $[NH_3]$, and linearly with ion concentration.

Global particle formation pathways

The CLOUD nucleation-rate measurements allow us, for the first time, to evaluate the global importance of competing particle sources on the basis of experimental data. The model is described in the supplementary materials, sections 11 to 16. The total nucleation rate is determined by adding the inorganic rates (previous section) to the neutral and ion-induced ternary organic H_2SO_4-BioOxOrg-H_2O nucleation rates from our earlier CLOUD chamber measurements reported in (*24*). Here BioOxOrg is a proxy for α-pinene oxidation products, and we can use the term "ternary" by treating them as a single class of vapors.

Figure 2 shows that the binary nucleation rates peak in the upper troposphere, consistent with earlier models that considered only binary neutral or ion-induced H_2SO_4-H_2O nucleation above the boundary layer (*6*, *11*–*13*). However, we find that the fractional contributions to the production rate of 3-nm-diameter particles below 15 km altitude are 15% binary (2.6% neutral and 12% ion-induced), 65% ternary inorganic with ammonia (54% neutral and 11% ion-induced), and 21% ternary organic. We are unable to quantify the fraction of ternary organic nucleation that is ion-induced as accurately as for inorganic nucleation (supplementary materials, section 15), but we estimate that 28% of all new particles are formed by ion-induced nucleation. Overall, ion-induced nucleation is the dominant process over large regions of the troposphere where particle formation rates are low. Consequently, it produces more particles than neutral nucleation in 67% of the troposphere below 15 km. Thus, almost all new particle formation over the entire troposphere involves NH_3 or organic compounds, and much of this is ion-induced.

Although our model of global nucleation rates accounts for the most important tropospheric variables, there remain some missing pieces. The most important are that we are unable to fully account for the variation of nucleation rates with RH, we do not include the contribution of amines to nucleation, and we assume that the organic nucleation rate is independent of temperature. The third of these possible sources of variation is not yet constrained by CLOUD laboratory measurements. If we assume a ternary organic rate that increases with decreasing temperature according to theoretical estimates (supplementary materials, sections 10 and 20), the ternary organic fraction of nucleation increases to 69%. This temperature dependence is likely an extreme estimate, given that terpene oxidation products are less oxidized at lower temperatures because isomerization rates are lower (*36*, *37*). This offsets the increased ease with which highly oxidized organic molecules condense at lower temperatures. An estimate of a more likely temperature dependence (supplementary materials, section 20) results in a fraction of organic nucleation of 43%. Further numerical studies of the uncertainties in the fractions of nucleation from different pathways are detailed in table S7. Because the neutral and ion-induced organic nucleation pathways have different, but unknown, dependencies on temperature (supplementary materials, section 20), the overall fraction of ion-induced nucleation is affected by this shortcoming. In the sensitivity tests in table S6, the highest fraction of ion-induced nucleation is around 63%, and the lowest is 9%.

Amines can also nucleate with sulfuric acid (*23*, *38*–*40*), but they are unlikely to influence

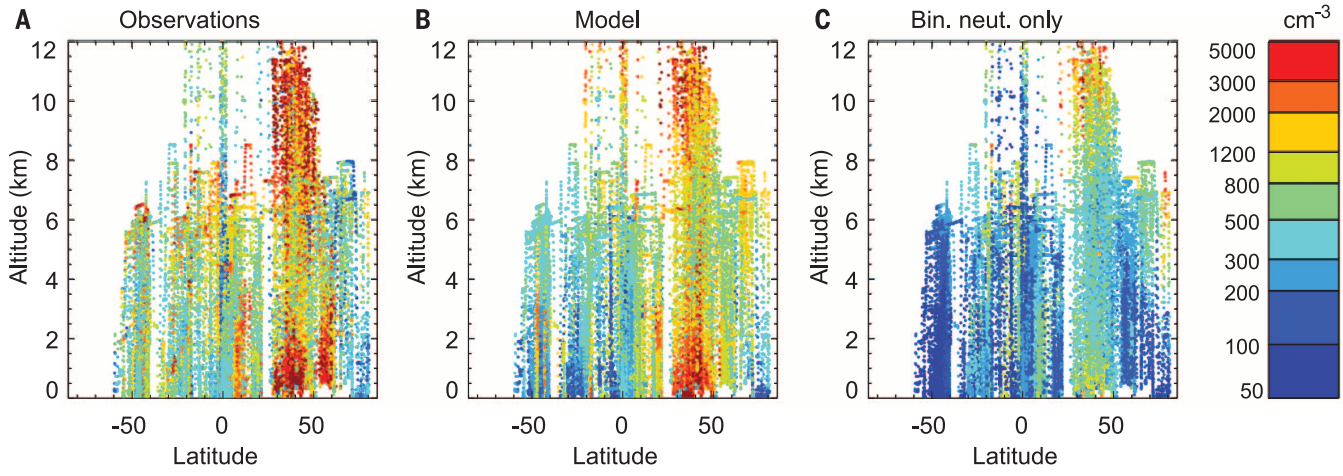


Fig. 3. Comparison of measured and modeled particle concentrations by latitude and altitude. (**A**) Measured 3-nm-diameter particle concentrations (fig. S17) (*44*). (**B**) Modeled particle concentrations (all processes). (**C**) Modeled particle concentrations including only primary particle emissions and binary neutral nucleation of sulfuric acid and water. Modeled particle concentrations in (C) are much higher than the concentrations that result from the binary-only pathway in the full model because the losses due to the condensation sink for these particles in the full model are higher than those in the binary-only model.

nucleation in the free troposphere owing to their short atmospheric lifetimes (*41*) and low fluxes. However, they are important in polluted areas of the boundary layer (*42*). A preliminary calculation (supplementary materials, section 17, and fig. S14) with a slightly different global model suggests that 6% of new particles below 500 m altitude are formed by an amine-driven mechanism, with a large uncertainty range of 3 to 27%. Amine-driven nucleation is almost certainly negligible above 500 m because of the short lifetime of amines in the atmosphere. The amine-driven nucleation has only a minimal effect on CCN concentrations (fig. S14C) because the highest amine emissions are in polluted areas with high condensation sinks, which suppress nucleation.

When we included a temperature-dependent factor to model the RH dependence of the binary nucleation rate with a polynomial function (supplementary materials, section 9), we found a 4.5% change in the concentration of 3-nm particles and a 0.3% change in the concentration of soluble 70-nm particles (approximately representative of CCN) in the troposphere up to 15 km altitude. If we assume that the ternary inorganic and organic nucleation rates depend on RH in the same way as the binary rate, we find that tropospheric 3-nm particle concentrations increase by 14%, and 70-nm particle concentrations increase by 6.5%. At cloud base level, the concentration of soluble 70-nm particles increases by 6.0%. Ternary nucleation should be less affected by RH than binary nucleation, so this change represents an upper bound. All of these numbers are comparable to typical differences between the model and observations.

Evidence from global aerosol measurements

Comparison of the global model results with atmospheric observations helps to establish how different nucleation pathways contribute to global particle concentrations. Using the simplest inorganic pathway (binary neutral nucleation) in the model systematically underestimates particle concentrations measured in the lower atmosphere and fails to explain their seasonal variation (supplementary materials, section 18). Across 35 global surface sites (*43*), including mountaintops, this binary simulation, together with primary particle emissions, can account for only 31% of the particles observed in winter in the Northern Hemisphere and 25% in summer (fig. S15). When we include ternary H_2SO_4-NH_3-H_2O nucleation, 65% of particles in winter are explained but only 36% in summer. By further including ternary H_2SO_4-BioOxOrg-H_2O nucleation, the modeled seasonal cycle of particle concentrations represents the observations well, explaining 71% of particles observed in winter and also 71% in summer. These comparisons suggest that ternary inorganic nucleation is a major source of particles in winter in the Northern Hemisphere, whereas ternary organic nucleation is a major source in summer.

We compared the predictions of the model with aircraft measurements collected at higher altitudes, a region of the atmosphere in which new particle formation is frequently modeled by binary nucleation of sulfuric acid. The aircraft campaigns (supplementary materials, section 18 and references therein) aimed to determine the most favorable chemical and meteorological conditions for nucleation in the troposphere (*44*). Relatively low model resolution prevents us from fully simulating the effects of meteorology on nucleation, especially humidity variations near clouds (*45*, *46*). Nevertheless, the full model shows good agreement with observations (Fig. 3B). In contrast, as is the case at the surface, a model with binary neutral nucleation alone can account for only 25% of the observed particle concentrations (Fig. 3C), with especially large biases of up to a factor of 5 in the lower troposphere below about 6 km altitude.

In addition to our imperfect coverage of the full parameter space of tropospheric nucleation rates, uncertainties in the model might also affect the confidence in our conclusions about the causes of global nucleation. Tables S6 and S7 summarize the principal sources of uncertainty in the CLOUD measurements and parameterization, as well as the implementation of the parameterization in the global model. The dominant sources of uncertainty, we estimate, are vapor concentrations in the model itself, rather than the parameterized nucleation rates. These and other important sources of uncertainty, such as the aerosol microphysical processes, precursor gas and primary aerosol emissions, and removal processes, have been studied comprehensively in a different configuration of this model for CCN (*47*) and 3-nm particle concentrations (*48*) by perturbing 28 model parameters in a way that allows combined uncertainties to be quantified. If we assume that the relative effect of these uncertainties on particle concentrations would be similar in the model configuration used for this study, then we can assess where we can be confident that free-tropospheric nucleation is mainly binary and where it is mainly ternary. For the southernmost remote observations in Fig. 3 [ACE-1 (Atmospheric Chemistry Experiment 1) measurements collected south of Australia; supplementary materials, section 18], the standard deviation of 3-nm particle concentrations from the 28 uncertainties is about 60% of the mean at 850 m altitude. So the binary neutral model plus one standard deviation would still lie about a factor of 5 below the observations. Although the binary model is closer to observations in some areas, such as over the Pacific Ocean, overall, its low bias relative to observations suggests that ternary nucleation (H_2SO_4 with organics or NH_3) is the primary source of

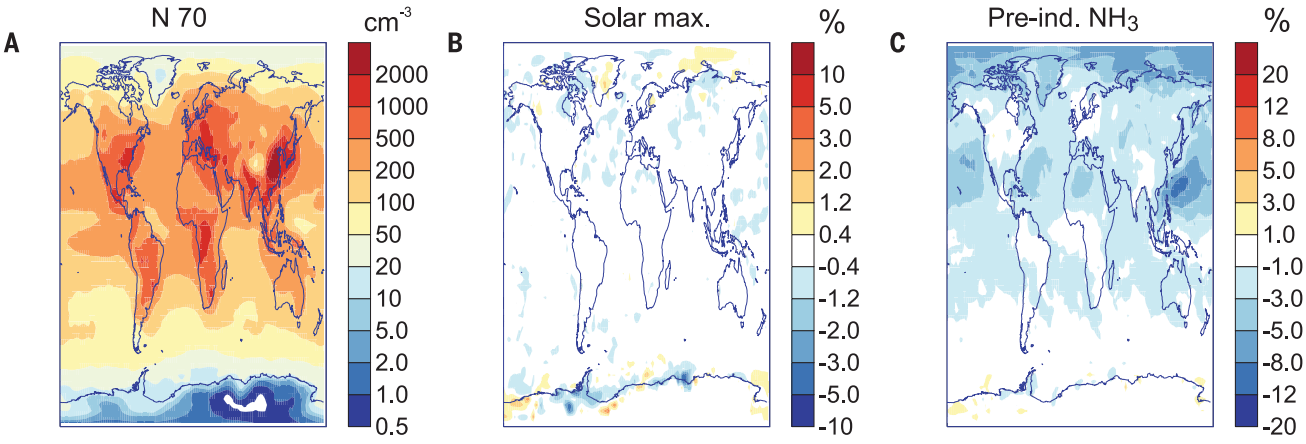


Fig. 4. Modeled present-day CCN concentrations and the effect of perturbations. Here hygroscopic particles above 70 nm diameter are used as a proxy for CCN. (**A**) Annual mean CCN concentrations at about cloud base altitude (915 hPa) (N 70, soluble particles with diameters greater than 70 nm). (**B**) Effect of changing the heliospheric modulation potential from solar minimum to maximum. (**C**) Effect of reducing ammonia concentrations to preindustrial levels. Perturbations are shown as percentage changes from the baseline shown in (A) where concentrations are higher than 5 cm^{-3}.

particles below about 6 km altitude in the environments that we have analyzed. We therefore conclude that binary nucleation becomes important only at the highest altitudes in the troposphere, and at lower altitudes, ternary nucleation dominates.

Implications for the atmosphere

The quantified effects of NH_3, oxidized organic compounds, and ions on global particle formation rates enabled us to estimate the effect on climate of changes in nucleation rates due to changing environmental conditions. We tested the effect of changes in the GCR ionization rate that occur between the solar maximum and minimum (supplementary materials, section 13) (*17*). Over the solar cycle, the global mean change in CCN at cloud base altitude (915 hPa, usually around 850 m above the surface) is only 0.1% (Fig. 4B), with local changes of no more than 1%. This is expected from the experimentally derived sublinear dependence of the inorganic nucleation rate on the ionization rate (supplementary materials, section 8) and consistent with previous assessments (*49*, *50*). The results in (*24*) suggest, with a large uncertainty, that organic nucleation is less sensitive to ionization rate than inorganic nucleation, so it would be unlikely to substantially increase the effect.

We also studied the effect of the estimated 80% increase in NH_3 emissions over the industrial period (*51*). To calculate the baseline aerosol-cloud albedo radiative forcing, we simulated preindustrial aerosols by removing anthropogenic emissions and keeping other model parameters constant, and we compared it with our present-day simulation. We then simulated present-day aerosols, keeping ammonia concentrations at preindustrial levels. In this simulation, the present-day global mean CCN concentration at cloud base level is 1.7% lower than in the usual present-day simulation (and locally up to 10 to 20% lower; Fig. 4C). Comparing the present-day simulation using preindustrial ammonia with the preindustrial simulation allows us to calculate the aerosol-cloud albedo forcing without the effect of ammonia. We can infer that ammonia has led to a strengthening of the anthropogenic aerosol-cloud radiative forcing from –0.62 to –0.66 Wm^{-2}. The 0.04 Wm^{-2} change in global mean forcing is within the uncertainty of forcing previously calculated for this model (*52*). However, the effect of ammonia on nucleation is a new process in the model, so the entire probability distribution of forcing reported in (*52*) would be shifted to lower values.

Global aerosol concentrations may be affected by future temperature changes through the temperature dependence of the formation rates. When we increase the temperatures used to calculate the inorganic nucleation rate by 2.2 K [the projected global mean change by 2100 (*53*)], mean CCN concentrations decrease by 1.0% at cloud base (locally by 10%) and cause a radiative effect of 0.02 Wm^{-2}. Therefore, a temperature-driven climate feedback related to changes in inorganic nucleation (*54*) is likely to be small compared with the large greenhouse gas forcings that are projected to occur by 2100. This result also shows that global inorganic aerosol nucleation provides a pervasive source of CCN that is relatively insensitive to environmental perturbation. The effect of rising global temperatures on organic ternary nucleation and CCN is less straightforward to calculate because there is probably compensation between decreasing nucleation rates (thus far not measured in the laboratory) and rising biogenic vapor emissions. Available observations suggest that the net effect could be to increase particle concentrations (*55*).

Conclusions

Atmospheric aerosol nucleation has been studied for over 20 years, but the difficulty of performing laboratory nucleation-rate measurements close to atmospheric conditions means that global model simulations have not been directly based on experimental data. This contrasts with the case of chemical transport modeling, which is well founded on reaction rate constants measured under controlled laboratory conditions over the past few decades (*56*). The multicomponent inorganic and organic chemical system is highly complex and is likely to be impossible to adequately represent in classical nucleation theories, just as ab initio prediction of reaction rate constants remains largely out of reach. This highlights the importance of replacing theoretical calculations with laboratory measurements, as we have done here. The CERN CLOUD measurements are the most comprehensive laboratory measurements of aerosol nucleation rates so far achieved, and the only measurements under conditions equivalent to the free and upper troposphere.

This work offers a new understanding of global particle formation as based almost entirely on ternary rather than binary nucleation, with ions playing a major but subdominant role. Our results suggest that about 43% of cloud-forming aerosol particles in the present-day atmosphere originate from nucleation, which is similar to a previous estimate of 45% obtained using the same chemical transport model and nonexperimental nucleation rates (*4*) (supplementary materials, section 16) and broadly consistent with other studies (*5*, *6*). An experimentally based model of global nucleation provides a basis for understanding how this complex system of inorganic and organic molecules responds to changes in trace gas emissions and environmental factors and, therefore, how these factors affect past and future climate.

REFERENCES AND NOTES

1. M. Kulmala *et al.*, *Atmos. Chem. Phys.* **4**, 2553–2560 (2004).
2. C. A. Brock, P. Hamill, J. C. Wilson, H. H. Jonsson, K. R. Chan, *Science* **270**, 1650–1653 (1995).
3. A. D. Clarke, *J. Atmos. Chem.* **14**, 479–488 (1992).
4. J. Merikanto, D. V. Spracklen, G. W. Mann, S. J. Pickering, K. S. Carslaw, *Atmos. Chem. Phys.* **9**, 8601–8616 (2009).
5. M. Wang, J. E. Penner, *Atmos. Chem. Phys.* **9**, 239–260 (2009).
6. F. Yu, G. Luo, *Atmos. Chem. Phys.* **9**, 7691–7710 (2009).

7. J. Kirkby *et al.*, *Nature* **476**, 429–433 (2011).
8. R. Zhang *et al.*, *Science* **304**, 1487–1490 (2004).
9. S. M. Ball, D. R. Hanson, F. L. Eisele, P. H. McMurry, *J. Geophys. Res. Atmos.* **104**, 23709–23718 (1999).
10. D. R. Benson, J. H. Yu, A. Markovich, S.-H. Lee, *Atmos. Chem. Phys.* **11**, 4755–4766 (2011).
11. R. Makkonen *et al.*, *Atmos. Chem. Phys.* **9**, 1747–1766 (2009).
12. P. J. Adams, J. H. Seinfeld, *J. Geophys. Res. Atmos.* **107**, AAC 4-1–AAC 4-23 (2002).
13. D. V. Spracklen, K. J. Pringle, K. S. Carslaw, M. P. Chipperfield, G. W. Mann, *Atmos. Chem. Phys.* **5**, 2227–2252 (2005).
14. S.-H. Lee *et al.*, *Science* **301**, 1886–1889 (2003).
15. J. Kazil *et al.*, *Atmos. Chem. Phys.* **10**, 10733–10752 (2010).
16. M. Kulmala *et al.*, *J. Aerosol Sci.* **35**, 143–176 (2004).
17. H. Svensmark, E. Friis-Christensen, *J. Atmos. Sol. Terr. Phys.* **59**, 1225–1232 (1997).
18. G. Bond *et al.*, *Science* **294**, 2130–2136 (2001).
19. U. Neff *et al.*, *Nature* **411**, 290–293 (2001).
20. K. S. Carslaw, R. G. Harrison, J. Kirkby, *Science* **298**, 1732–1737 (2002).
21. J. Kirkby, *Surv. Geophys.* **28**, 333–375 (2007).
22. M. Chen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18713–18718 (2012).
23. J. Almeida *et al.*, *Nature* **502**, 359–363 (2013).
24. F. Riccobono *et al.*, *Science* **344**, 717–721 (2014).
25. H. Vehkamäki *et al.*, *J. Geophys. Res. Atmos.* **107**, AAC 3-1–AAC 3-10 (2002).
26. I. Napari, M. Noppel, H. Vehkamäki, M. Kulmala, *J. Geophys. Res.* **107**, AAC 6-1–AAC 6-6 (2002).
27. S. D. D'Andrea *et al.*, *Atmos. Chem. Phys.* **13**, 11519–11534 (2013).
28. J. Jung, C. Fountoukis, P. J. Adams, S. N. Pandis, *J. Geophys. Res. Atmos.* **115**, D03203 (2010).
29. L. K. Peters, A. A. Jouvanis, *Atmos. Environ.* **13**, 1443–1462 (1979).
30. H. Rodhe, I. Isaksen, *J. Geophys. Res. Oceans* **85**, 7401–7409 (1980).
31. J. Duplissy *et al.*, *J. Geophys. Res. Atmos.* **121**, 1752–1775 (2016).
32. A. Kürten *et al.*, *J. Geophys. Res. Atmos.* 10.1002/2015JD023908 (2016).
33. A. Franchin *et al.*, *Atmos. Chem. Phys.* **15**, 7203–7216 (2015).
34. S. Schobesberger *et al.*, *Atmos. Chem. Phys.* **15**, 55–78 (2015).
35. S. Schobesberger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17223–17228 (2013).
36. J. D. Crounse, L. B. Nielsen, S. Jørgensen, H. G. Kjaergaard, P. O. Wennberg, *J. Phys. Chem. Lett.* **4**, 3513–3520 (2013).
37. M. Ehn *et al.*, *Nature* **506**, 476–479 (2014).
38. M. E. Erupe, A. A. Viggiano, S.-H. Lee, *Atmos. Chem. Phys.* **11**, 4767–4775 (2011).
39. C. N. Jen, P. H. McMurry, D. R. Hanson, *J. Geophys. Res. Atmos.* **119**, 7502–7514 (2014).
40. T. Bergman *et al.*, *J. Geophys. Res. Atmos.* **120**, 9606–9624 (2015).
41. X. Ge, A. S. Wexler, S. L. Clegg, *Atmos. Environ.* **45**, 561–577 (2011).
42. J. Zhao *et al.*, *Atmos. Chem. Phys.* **11**, 10823–10836 (2011).
43. D. V. Spracklen *et al.*, *Atmos. Chem. Phys.* **10**, 4775–4793 (2010).
44. A. D. Clarke, V. N. Kapustin, *J. Atmos. Sci.* **59**, 363–382 (2002).
45. A. D. Clarke *et al.*, *J. Geophys. Res. Atmos.* **103**, 16397–16409 (1998).
46. A. D. Clarke *et al.*, *J. Geophys. Res. Atmos.* **104**, 5735–5744 (1999).
47. L. A. Lee *et al.*, *Atmos. Chem. Phys.* **13**, 8879–8914 (2013).
48. K. S. Carslaw, L. A. Lee, C. L. Reddington, G. W. Mann, K. J. Pringle, *Faraday Discuss.* **165**, 495–512 (2013).
49. J. Kazil *et al.*, *Geophys. Res. Lett.* **39**, L02805 (2012).
50. J. R. Pierce, P. J. Adams, *Geophys. Res. Lett.* **36**, L09820 (2009).
51. A. F. Bouwman *et al.*, *Global Biogeochem. Cycles* **11**, 561–587 (1997).
52. K. S. Carslaw *et al.*, *Nature* **503**, 67–71 (2013).
53. U. Cubasch *et al.*, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the IPCC*, T. F. Stocker *et al.*, Eds. (Cambridge Univ. Press, 2013), pp. 119–158.
54. F. Yu, G. Luo, R. P. Turco, J. A. Ogren, R. M. Yantosca, *Atmos. Chem. Phys.* **12**, 2399–2408 (2012).
55. P. Paasonen *et al.*, *Nat. Geosci.* **6**, 438–442 (2013).
56. R. Atkinson *et al.*, *Atmos. Chem. Phys.* **4**, 1461–1738 (2004).

ACKNOWLEDGMENTS

We thank CERN for supporting CLOUD with important technical and financial resources and for providing a particle beam from the CERN Proton Synchrotron. We also thank P. Carrie, L.-P. De Menezes, J. Dumollard, K. Ivanova, F. Josa, I. Krasin, R. Kristic, A. Laassiri, O. S. Maksumov, B. Marichy, H. Martinati, S. V. Mizin, R. Sitals, H. U. Walther, A. Wasem, and M. Wilhelmsson for their important contributions to the experiment. The computer modeling simulations were performed on ARC1 and ARC2, part of the high-performance computing facilities at the University of Leeds, UK. This work also made use of the POLARIS facility of the N8 High Performance Computing Centre of Excellence, provided and funded by the N8 consortium and the Engineering and Physical Sciences Research Council (grant no. EP/K000225/1). The Centre is coordinated by the Universities of Leeds and Manchester. This research has received funding from the European Commission Seventh Framework Programme [Marie Curie Initial Training Networks CLOUD- ITN (no. 215072) and CLOUD-TRAIN (no. 316662)]; European Research Council (ERC) Starting Grant no. 5736 [MOCAPAF (Role of Molecular Clusters in Atmospheric Particle Formation)] and ERC Advanced grant no. 227463 [ATMNUCLE (Atmospheric Nucleation: From Molecular to Global Scale)]; the German Federal Ministry of Education and Research (project nos. 01LK0902A and 01LK1222A); the Swiss National Science Foundation (project nos. 200020 135307 and 206620 141278); the Academy of Finland (Center of Excellence project no. 1118615 and other projects 135054, 133872, 251427, 139656, 139995, 137749, 141217, 141451, and 138951); the Finnish Funding Agency for Technology and Innovation; the V.is.l. Foundation; the Nessling Foundation; the Austrian Science Fund (FWF) (project no. J3198-N21); the Portuguese Foundation for Science and Technology (project no. CERN/FP/116387/2010); the Swedish Research Council; Vetenskapsrådet (grant 2011-5120); the Presidium of the Russian Academy of Sciences and the Russian Foundation for Basic Research (grants 08-02-91006-CERN and 12-02-91522-CERN); the U.S. National Science Foundation (grants AGS1136479, AGS1447056, AGC1439551, and CHE1012293); the U.S. Department of Energy (grant DE-SC0014469); the PEGASOS (Pan-European Gas-Aerosol-Climate Interaction Study) project funded by the European Commission under Framework Programme 7 (FP7-ENV- 2010-265148); the Davidow Foundation; and the Natural Environment Research Council project GASSP (Global Aerosol Synthesis and Science Project) under grant NE/J024252/1. We acknowledge financial support from the Royal Society Wolfson Merit Award. The nucleation rates used in our manuscript are available in the supplementary materials as a CSV file.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1119/suppl/DC1
Materials and Methods
Figs. S1 to S21
Tables S1 to S7
References (*57*–*135*)
Data S1

18 January 2016; accepted 12 October 2016
Published online 27 October 2016
10.1126/science.aaf2649

REPORTS

TOPOLOGICAL MATTER

Quantized Faraday and Kerr rotation and axion electrodynamics of a 3D topological insulator

Liang Wu,[1]*† M. Salehi,[2] N. Koirala,[3] J. Moon,[3] S. Oh,[3] N. P. Armitage[1]*

Topological insulators have been proposed to be best characterized as bulk magnetoelectric materials that show response functions quantized in terms of fundamental physical constants. Here, we lower the chemical potential of three-dimensional (3D) Bi_2Se_3 films to ~30 meV above the Dirac point and probe their low-energy electrodynamic response in the presence of magnetic fields with high-precision time-domain terahertz polarimetry. For fields higher than 5 tesla, we observed quantized Faraday and Kerr rotations, whereas the dc transport is still semiclassical. A nontrivial Berry's phase offset to these values gives evidence for axion electrodynamics and the topological magnetoelectric effect. The time structure used in these measurements allows a direct measure of the fine-structure constant based on a topological invariant of a solid-state system.

Topological phenomena in condensed matter physics provide some of the most precise measurements of fundamental physical constants. The measurement of the quantum conductance $G_{xy} = e^2/h$ from the quantum Hall effect (*1*) and the flux quantum from the Josephson effect (*2*, *3*) provide the most precise value for Planck's constant h. More recently, topological insulators have been discovered (*4*–*6*), in which topological properties of the bulk wave functions give rise to a topologically protected surface metal with a massless Dirac spectrum. It has been proposed that topological insulators are best characterized not as surface conductors but as bulk magnetoelectrics (*7*, *8*) with a quantized magnetoelectric response coefficient whose size is set by the fine-structure constant $\alpha = e^2/2\varepsilon_0 hc$. Such a measurement could provide precise values for three fundamental physical constants: the electric charge e, Planck's constant h, and the vacuum impedance $Z_0 = \sqrt{\mu_0/\varepsilon_0}$ in a solid-state context.

Magnetoelectrics (ME) are materials in which a polarization can be created by an applied magnetic field or a magnetization can be created by an applied electric field (*9*); representative examples are Cr_2O_3 (*10*) with an ME coupling of the $\boldsymbol{E}\cdot\boldsymbol{B}$ form and multiferroic $BiFeO_3$ (*11*), where the ME coupling can be expressed (in part) in a $\boldsymbol{E} \times \boldsymbol{B}$

[1]Institute for Quantum Matter, Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, MD 21218, USA. [2]Department of Materials Science and Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA. [3]Department of Physics and Astronomy, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA.
***Corresponding author. Email: liangwu@berkeley.edu (L.W.); npa@jhu.edu (N.P.A.)** †Present address: Department of Physics, University of California, Berkeley, Berkeley, CA 94720, USA.

form. Topological insulators (TIs) can be characterized as special $\boldsymbol{E}\cdot\boldsymbol{B}$ magnetoelectrics (*7*, *8*), which in the topological field theory can be shown to be a consequence of an additional term $\mathcal{L}_\theta = -2\alpha\sqrt{\frac{\varepsilon_0}{\mu_0}}\frac{\theta}{2\pi}\boldsymbol{E}\cdot\boldsymbol{B}$ added to the usual Maxwell Lagrangian (*7*). Here, α is the fine-structure constant, and ε_0 and μ_0 are the permittivity and permeability of free space.

Although $\mathcal{L}_\theta$ is a generic expression that can be applied even to Cr_2O_3 [with $\theta \approx \frac{\pi}{36}$ at low temperature (*12*)], its form merits additional discussion when applied to TIs. Although it is usually said that one must break both time-reversal symmetry (TRS) and inversion symmetry (IS) to define a magnetoelectric coefficient, this is not formally true. The Lagrangian defines the action $\mathcal{S} = \int dt dx^3 \mathcal{L}$, and because all physical observables depend on $\exp(i\mathcal{S}/\hbar)$, they are invariant to global changes to θ of 2π. Therefore, due to the transformation properties of $\boldsymbol{E}$ and $\boldsymbol{B}$, if either TRS or IS are present, θ is constrained to be not just zero (as it is in a nonmagnetoelectric conventional material) but can take on integer multiples of π. Three-dimensional (3D) insulators in which either TRS or IS is preserved can be divided into two classes depending on whether θ is $2\pi(N+\frac{1}{2})$ (topological) or $2\pi(N)$ (conventional) (*7*). Here, N is an integer that indicates the highest fully filled Landau level (LL) of the surface if TRS is broken. In either case, θ can be formulated as a bulk quantity modulo a quantum (here, 2π) in much the same way as the electric polarization $\mathbf{P}$ in a ferroelectric can be defined only as a bulk quantity modulo a dipole quantum that depends on the surface charge (*13*). It is important to note that to support a macroscopic magnetic/electric moment of the sample from an applied electric/magnetic field, macroscopic TRS and IS must both be broken (as they are in conventional magnetoelectrics), but a finite magnetoelectric $\boldsymbol{E}\cdot\boldsymbol{B}$ term is more general than the capacity to support a moment. Because inversion-symmetric Bi_2Se_3 in magnetic field breaks only TRS, such a sample cannot exhibit a net macroscopic moment from magnetoelectricity unless IS is broken macroscopically through some other means. In the case relevant for our experiment, IS constrains the crystal's bulk θ term to be $2\pi(N+\frac{1}{2})$. A net macroscopic moment cannot be generated, but the sample is still magnetoelectric in the sense that $\mathcal{L}_\theta$ still applies. The topological magnetoelectric effect (TME) of this kind has been called "axion electrodynamics" because of an analogy that can be made to the physics of the hypothetical axion particle that was proposed to explain charge conjugation parity symmetry violation (CP violation) in the strong interaction (*14*).

In the limit where a TRS breaking field is small and the surface chemical potential is tuned near the Dirac point, modified Maxwell's equations can be derived (section 1 of supplementary text) from the full Lagrangian. The modified Gauss's and Ampère's laws read

$$\nabla\cdot\mathbf{E} = \frac{\rho}{\varepsilon_0} - 2c\alpha\nabla\left(\frac{\theta}{2\pi}\right)\cdot\mathbf{B} \qquad (1)$$

$$\nabla\times\mathbf{B} = \mu_0\mathbf{J} + \frac{1}{c^2}\frac{\partial\mathbf{E}}{\partial t} + \frac{2\alpha}{c}\left[\mathbf{B}\frac{\partial}{\partial t}\left(\frac{\theta}{2\pi}\right) + \nabla\left(\frac{\theta}{2\pi}\right)\times\mathbf{E}\right] \qquad (2)$$

The consequences of axion electrodynamics are the additional source and current terms in Eqs. 1 and 2 (*7*, *15*). The additional current term gives a half-integer quantum Hall effect (QHE) on the TI surface (*7*). Although there has been some evidence for half-integer QHE effects in gated TI $BiSbTeSe_2$ exfoliated flakes (*16*), gated $(Bi_{1-x}Sb_x)_2Te_3$ thin films (*17*), and surface charge-transfer doped pure Bi_2Se_3 films (*18*) at very high magnetic fields, it is generally not straightforward to observe the QHE in a conventional dc transport-style experiment with leads connected to sample edges, as TIs have a closed surface with no boundaries (*19*). It is desirable then to use noncontact probes such as Faraday and Kerr rotations (*7*, *20*, *21*), which have been predicted to be quantized with a scale set by the fine-structure constant. One can proceed from the modified Ampère's law Eq. 2 in conjunction with the usual Faraday's law to derive the reflection and transmission coefficients for a traveling wave incident on a TI surface (section 2 of supplementary text). In an applied magnetic field, one finds that for a TI film on a simple dielectric substrate, the Faraday rotation in the quantum regime is

$$\tan(\phi_F) = \frac{2\alpha}{1+n}\left(N_t + \frac{1}{2} + N_b + \frac{1}{2}\right) \qquad (3)$$

where $n \sim 3.1$ is the THz range index of refraction of the substrate (sapphire) and N_t, N_b are the highest fully filled LL of the top and bottom surfaces of the film, which depend on the chemical potential and size of the TRS breaking field.

There have been a number of interrelated challenges in realizing the TME experimentally. First, one must have a negligible level of bulk carriers and a low chemical potential at the surface, but most known topological insulators suffer from inadvertent bulk doping; a metallic gate cannot be used easily in an optical experiment to gate away charge carriers because it would have its own Faraday effect in field. Second, as the topological field theory is derived for the translationally invariant case, one may expect that it will apply only when the TRS breaking perturbation is strong enough to overcome disorder and establish a surface QHE. Third, to reveal the TME, the probe frequencies and temperatures must be well below the Landau level spacing of the surface states, which are given by $E = v_F\sqrt{2Ne\mathbf{B}\hbar}$ (where v_F is the Fermi velocity). This puts the

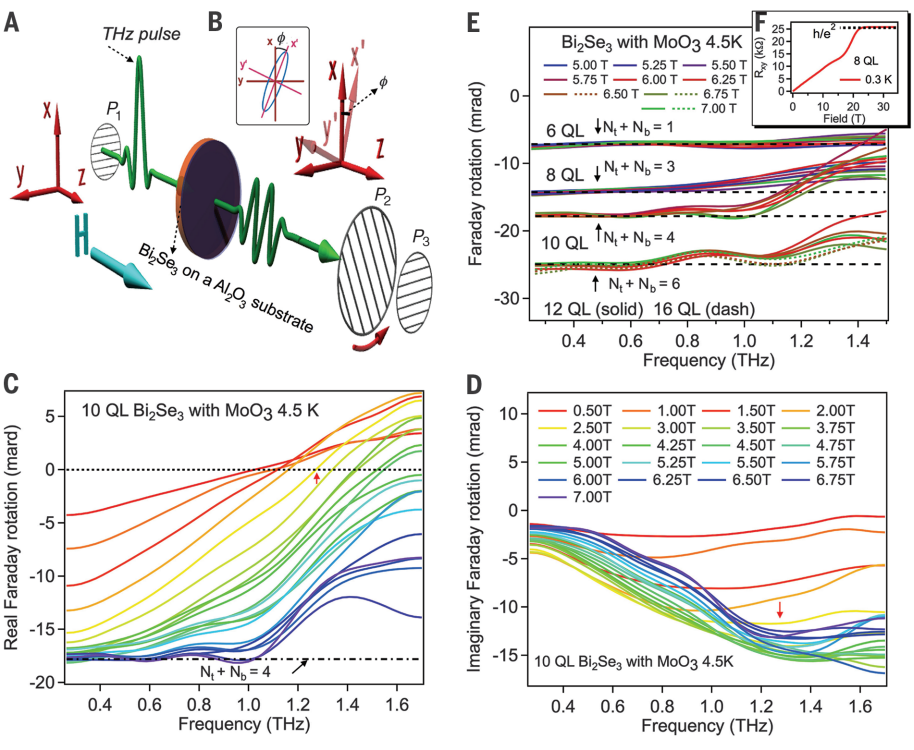


Fig. 1. Quantized Faraday rotation of the topological surface states. (**A**) Diagram of the Faraday rotation experiment. P_1, P_2, and P_3 are polarizers. The polarization plane (xz) of the linearly polarized incoming THz beam is rotated by the Faraday angle ϕ_F (into the $x'z$ plane) after passing through Bi_2Se_3 on a sapphire substrate in a perpendicular magnetic field (z direction). The polarization acquires an ellipticity simultaneously, as shown in (**B**). (**C**) Real part of Faraday rotation of 10-QL Bi_2Se_3 films with MoO_3 at 4.5 K for various values of the magnetic field [color coded as in (D)]. The dash-dot line is the expectation from Eq. 3. (**D**) Imaginary part of Faraday rotation. A representative cyclotron frequency is marked by a red arrow for data at 2.5 T. (**E**) Quantized Faraday rotation for different samples. Dashed black lines are theoretical expectation values assuming certain values for the filling factor of the surface states. (**F**) dc transport Hall resistance of a representative 8-QL sample.

relevant frequency in the traditionally challenging sub-THz part of the electromagnetic spectrum. Fourth, THz range experiments with their long wavelengths require large uniform samples of at least a few mm in spatial extent. Fifth, as the size of the effect is set by the fine-structure constant, the rotations are expected to be very small and much smaller than the capacity of conventional THz range polarimetry.

Here, we overcome these challenges by using recently developed low-density and high-mobility Bi_2Se_3 molecular beam epitaxy (MBE) films (*18*), in conjunction with a high-precision polarimetry technique (*22*). Time-domain terahertz spectroscopy (TDTS) is a powerful tool to study the low-energy electrodynamics of topological insulators. Samples are thin films of Bi_2Se_3 grown by MBE with a recently developed recipe (*18*) that results in true bulk-insulating TIs with low surface chemical potential. These films were further treated in situ by a thin charge-transfer layer of deposited MoO_3 that further decreases the carrier density and puts the chemical potential close to the Dirac point. MoO_3 is a semiconductor with a gap of ~3 eV (*23*) and does not contribute to Faraday rotation. Details of the film growth can be found in (*18*).

First, as topological surface states have been shown to be the only conducting channel in these films (*18*), we further verify the low Fermi energy of these films through measurements of their optical conductivity and cyclotron resonance at low field in the semiclassical transport regime (figs. S1 and S2 and section 3 of supplementary text). Next, we explore their low-frequency Faraday rotation (Fig. 1A). The complex THz-range Faraday rotation was measured with the polarization modulation technique (*22*, *24*). The Faraday rotation is a complex quantity in which the real part is the rotation of the major axis of the ellipse and the imaginary part is related to the ellipticity (Fig. 1B). The full-field data of a 10-QL sample is shown in Fig. 1, C and D. At low fields (<4 T), the Faraday rotation shows semiclassical cyclotron resonance, as demonstrated by the shifting of the inflection point (close to the zero value) in the real part and the shifting of the minimum in the imaginary part with fields (*24*). For the 10-QL sample, above 5 T, the inflection point in the real part of the Faraday rotation moves above our frequency range, and the low-frequency tail becomes flat and overlaps with higher field data. In our TDTS measurements, top and bottom states are measured simultaneously (*18*, *24*, *25*), so the quantized Faraday rotation is given by Eq. 3.

These data are well described by the prediction for the $5\frac{2\alpha}{1+n}$ plateau. Because the resolution of our THz polarimetry is within 1 mrad, we conclude that the 10-QL sample enters the quantized regime when the field is above 5.75 T, with the low-frequency tails falling on the expected value. Similarly, for the 6-, 8-, 12-, and 16-QL samples, the low-frequency Faraday rotations fall on the $2\frac{2\alpha}{1+n}$, $4\frac{2\alpha}{1+n}$, $7\frac{2\alpha}{1+n}$, and $7\frac{2\alpha}{1+n}$ plateaus, respectively (Fig. 1E). Aside from the filling factor differences, the only qualitative differences between samples is that thicker samples have a narrower magnetic field range where the Faraday rotation

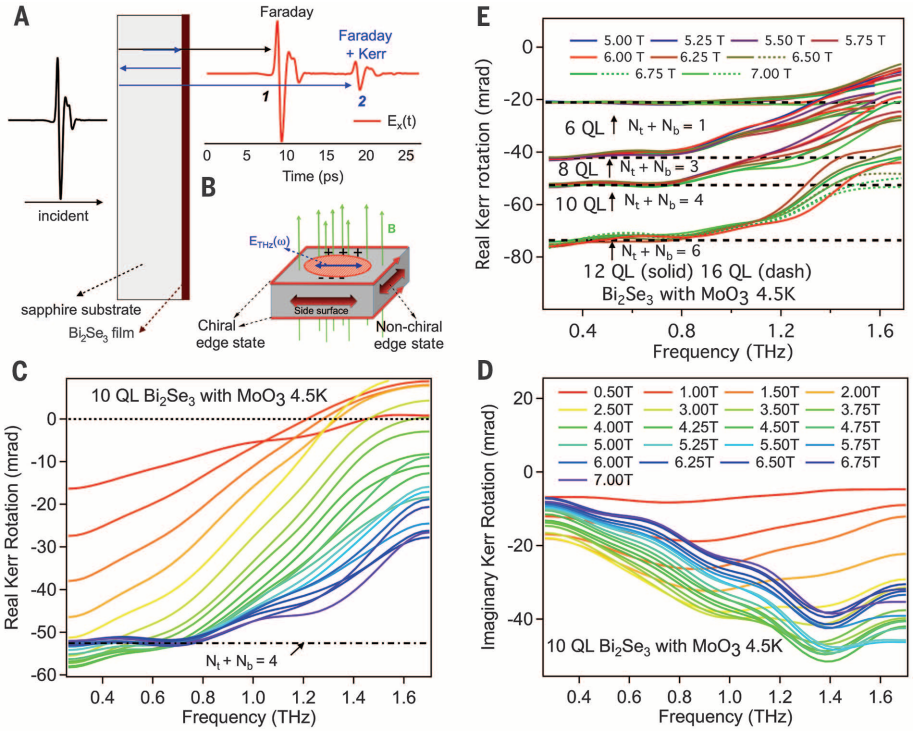


Fig. 2. Quantized Kerr rotation of the surface states. (**A**) Diagram of the Kerr rotation experiment. The black and blue arrows show the optical path for the first and second pulses in the time trace. (**B**) Bi_2Se_3 thin film in a magnetic field (substrate not shown). Nonchiral edge states from the side surface states are shown by bidirectional arrows. The circle on the top surface indicates the THz spot where transverse oscillating charge density is built up by the THz electric field. (**C**) Real and (**D**) imaginary part of Kerr rotation of 10-QL new Bi_2Se_3 with MoO_3 at 4.5 K. (**E**) Quantized Kerr rotation for different samples.

is quantized because they have a slightly higher carrier density and filling factor at the same magnetic fields (figs. S3 and S4 and section 3 of supplementary text). Because we measure the top and bottom surfaces of the thin film simultaneously, essential for our interpretation in terms of axion electrodynamics is that we can treat the top and bottom surfaces independently. Previous angle-resolved photoemission spectroscopy work (*26*) and theory (*27*) showed that the hybridization gap from top and bottom surfaces was negligible for film thicknesses more than 6 QL (see further discussion in section 4 of supplementary text).

It is important to point out that this effect is not just the conventional dc quantum Hall effect. We can contrast the quantized optical response with dc transport that has shown quantum Hall resistivity plateaus in these films only above ~24 T, as shown for a typical sample 8-QL film in Fig. 1F. When an external magnetic field is applied perpendicular to the films, top and bottom surface states are gapped because of LL formation, whereas the side surfaces parallel to the magnetic field remain gapless because a small in-plane field will cause only a shift of the Dirac points (*28*). The dc QHE in conventional 2D electron gas (2DEG) is usually regarded as occurring through ballistic 1D chiral states formed at the edge of the sample. In the present case, the dc QHE is corrupted at low fields by the nonchiral side states (Fig. 2B) . These side surface states can be gapped by an amount $\hbar v_F/d$ through finite size effects (where d is the film thickness), but in order that they do not contribute to dc transport, this gap must be larger than the LL spacing $E = v_F\sqrt{2Ne\mathbf{B}\hbar}$ (*19*). This condition is hard to fulfill with films thick enough to be effectively 3D and with fields large enough to establish a surface QHE. We believe that quantized dc transport is achieved in high fields because the nonchiral side states localize in high magnetic field in the highly disordered edges. In the present experiment, THz radiation is focused onto a local spot far from the edges of the film, so irrespective of their properties they cannot contribute to the spectral response. The Hall response measured here originates in the "bulk" of the sample (topological surface states), and the edge state picture does not apply. Section 5 of the supplementary text provides further discussion on the ac QHE and on how the incompressible bulk responds to an oscillating charge density in an ac experiment.

Data in Fig. 1 gives evidence for a Faraday rotation set by the fine-structure constant. However, such measurements by themselves are limited, as Eq. 3 shows that the Faraday rotation still depends nonuniversally on the index of refraction of the substrate n, and our ability to measure the fine-structure constant to high precision is limited by our knowledge of n. However, by using the explicit time structure of TDTS we can define and measure a quantity that depends only on the

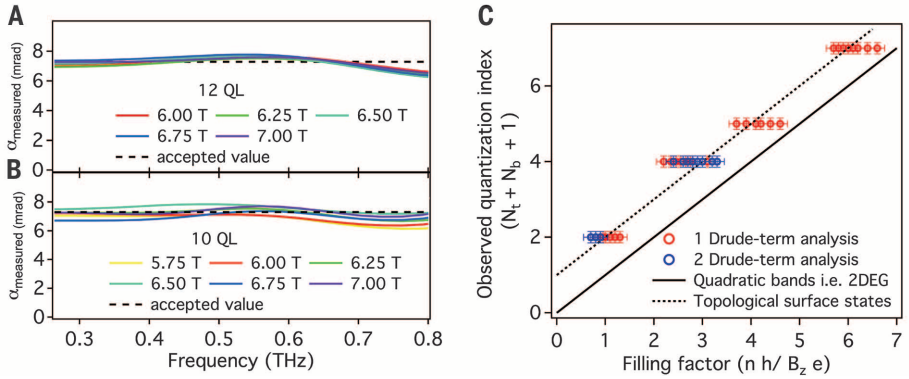


Fig. 3. A measure of the fine-structure constant and evidence for axion electrodynamics and the TME. The measured fine-structure constant of (**A**) 12-QL and (**B**) 10-QL new Bi_2Se_3 with MoO_3 at different fields at 4.5 K. (**C**) Measured quantization index versus filling factor. The solid line is the expectation for quadratic bands, and the dashed line is for two topological surface states.

fine-structure constant (and surface filling factors). When THz light is transmitted through a film and substrate, the substrate itself can be used as an optical resonator (*29*, *30*), resulting in a series of pulses that each have different histories of interaction with the film (Fig. 2A). The first peak that is transmitted through the film undergoes a Faraday rotation, whereas the second peak undergoes an additional reflection and Kerr rotation ϕ_K. By subtracting the Faraday rotation, we can measure the Kerr rotation separately. In the quantized regime, one can show (section 2 of supplementary text) that the Kerr rotation (up to factors of the third order in α) is

$$\tan(\phi_K) = \frac{4n\alpha}{n^2-1}\left(N_t + \frac{1}{2} + N_b + \frac{1}{2}\right) \quad (4)$$

Representative data for the 10-QL sample for the Kerr rotation is shown in Fig. 2, C and D. Similar to the Faraday rotation, the signatures of cyclotron resonance are inflection points in the real part and dips in the imaginary part. Above 5.75 T, the Kerr rotation of a 10-QL sample is quantized as $5\frac{4n\alpha}{n^2-1}$ to within our experimental resolution at frequencies below 0.8 THz. The prefactor of 5 is the same as arrived at in the Faraday rotation. We measured Kerr rotation on samples with different thickness—6, 8, 12, and 16 QL—and in all cases the rotation is given by $\frac{4n\alpha}{n^2-1}$ times the filling factor found in the Faraday rotation experiments (Fig. 2E). Combining Eqs. 3 and 4, one can eliminate the dependence on the index of the substrate and measure the fine-structure constant directly.

$$\alpha_{measured} = \frac{1}{N_t + N_b + 1/2 + 1/2} \frac{\tan(\phi_F)^2 - \tan(\phi_F)\tan(\phi_K)}{\tan(\phi_K) - 2\tan(\phi_F)} \quad (5)$$

Measuring these quantities in a single scan and taking ratios also serves to minimize the systematic noise in the output for $\alpha_{measured}$. Using $N_t + N_b + 1/2 + 1/2 = 5$ and 7 for the 10- and 12-QL samples, respectively, we plot the results of Eq. 5 for two samples in Fig. 3, A and B, and find for both that the measured value is close to 1/137 (~7.3) mrad. Averaging over the frequency range that quantized rotation is observed (0.2 to 0.8 THz) for all samples measured, we find a best measured value for $\alpha_{measured}$ of 1/137.9, which is close to the accepted value 1/137.04. This represents a direct measurement of the fine-structure constant based on a topological invariant in a solid-state system. Although the level of precision that we have achieved for α is far less than, for instance, its determination via the anomalous magnetic moment of the electron (*31*), the quantization should be considered quite good. Its deviation from the accepted value is ~0.5%, which can be compared favorably to the quantization seen in the quantum spin Hall effect, which was quantized only to the 10% level (*32*). Moreover, the observed quantization is far better than that observed previously in the ac QHE of 2DEG-like GaAs heterostructures and graphene (*33*, *34*). If this measurement could be further refined, it could, along with measures of the Josephson effect and quantized Hall resistances in 2DEG, provide a purely solid-state measure in a redefined conventional electrical unit scheme for the impedance of free space $Z_0 = \sqrt{\mu_0/\varepsilon_0}$, which would in turn allow c to become a measured quantity in a condensed matter experiment.

It is important to distinguish our results from a conventional QHE, as may be observed in a 2DEG. Do we truly probe axion electrodynamics and the TME effect? As discussed above, the TME is characterized by a θ angle that is $2\pi(N + \frac{1}{2})$ or equivalent a half-integer QHE effect. In Fig. 3C, we plot the observed quantization index versus the total filling factor (which can be measured independently, as discussed above and in sections 3 and 6 of the supplementary text). There is a systematic offset of 1 in the position of the plateaus that originates from the Berry's phase. With our previous results establishing surface state transport from two independent surfaces (*18*, *24*), one must associate a contribution to this offset of 1/2 for each surface by itself. This establishes a π value of the axion angle of the topological insulators and the TME.

The quantized response that we find here should not be viewed as a simple manifestation of the quantized quantum Hall transport seen in usual 2DEGs, because TI surface states live on a closed 2D manifold embedded in 3D space. In a formalism in which the TIs are described as bulk magnetoelectrics, this response can be described in the context of a topological magnetoelectric effect and axion electrodynamics. Going forward, the technique may prove to be an essential tool in the discovery of theoretically anticipated states of matter such as fractional topological insulators in the form of a fractional magnetoelectric effect (*35*, *36*).

REFERENCES AND NOTES

1. K. Klitzing, G. Dorda, M. Pepper, *Phys. Rev. Lett.* **45**, 494–497 (1980).
2. B. D. Josephson, *Phys. Lett.* **1**, 251–253 (1962).
3. P. W. Anderson, J. M. Rowell, *Phys. Rev. Lett.* **10**, 230–232 (1963).
4. J. E. Moore, *Nature* **464**, 194–198 (2010).
5. M. Z. Hasan, C. L. Kane, *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
6. X.-L. Qi, S.-C. Zhang, *Rev. Mod. Phys.* **83**, 1057–1110 (2011).
7. X.-L. Qi, T. L. Hughes, S.-C. Zhang, *Phys. Rev. B* **78**, 195424 (2008).
8. A. M. Essin, J. E. Moore, D. Vanderbilt, *Phys. Rev. Lett.* **102**, 146805 (2009).
9. M. Fiebig, *J. Phys. D Appl. Phys.* **38**, R123–R152 (2005).
10. I. Dzyaloshinskii, *Soviet Physics Jetp-Ussr* **10**, 628 (1959).
11. J. C. Wojdeł, J. Iñiguez, *Phys. Rev. Lett.* **103**, 267205 (2009).
12. F. W. Hehl, Y. N. Obukhov, J.-P. Rivera, H. Schmid, *Eur. Phys. J. B* **71**, 321–329 (2009).
13. D. Vanderbilt, R. D. King-Smith, *Phys. Rev. B* **48**, 4442–4455 (1993).
14. F. Wilczek, *Phys. Rev. Lett.* **58**, 1799–1802 (1987).
15. X.-L. Qi, R. Li, J. Zang, S.-C. Zhang, *Science* **323**, 1184–1187 (2009).
16. Y. Xu *et al.*, *Nat. Phys.* **10**, 956–963 (2014).
17. R. Yoshimi *et al.*, *Nat. Commun.* **6**, 6627 (2015).
18. N. Koirala *et al.*, *Nano Lett.* **15**, 8245–8249 (2015).
19. O. Vafek, *Phys. Rev. B* **84**, 245417 (2011).
20. J. Maciejko, X.-L. Qi, H. D. Drew, S.-C. Zhang, *Phys. Rev. Lett.* **105**, 166803 (2010).
21. W.-K. Tse, A. H. MacDonald, *Phys. Rev. Lett.* **105**, 057401 (2010).
22. C. M. Morris, R. V. Aguilar, A. V. Stier, N. P. Armitage, *Opt. Express* **20**, 12303–12317 (2012).
23. Z. Hussain, *J. Mater. Res.* **16**, 2695–2708 (2011).
24. L. Wu *et al.*, *Phys. Rev. Lett.* **115**, 217602 (2015).
25. L. Wu *et al.*, *Nat. Phys.* **9**, 410–414 (2013).
26. Y. Zhang *et al.*, *Nat. Phys.* **6**, 584–588 (2010).
27. J. Linder, T. Yokoyama, A. Sudbø, *Phys. Rev. B* **80**, 205401 (2009).
28. S. S. Pershoguba, V. M. Yakovenko, *Phys. Rev. B* **86**, 165404 (2012).
29. R. Valdés Aguilar *et al.*, *Phys. Rev. Lett.* **108**, 087403 (2012).
30. J. N. Hancock *et al.*, *Phys. Rev. Lett.* **107**, 136803 (2011).
31. P. J. Mohr, B. N. Taylor, *Rev. Mod. Phys.* **72**, 351–495 (2000).
32. M. König *et al.*, *Science* **318**, 766–770 (2007).
33. Y. Ikebe *et al.*, *Phys. Rev. Lett.* **104**, 256802 (2010).
34. R. Shimano *et al.*, *Nat. Commun.* **4**, 1841 (2013).
35. J. Maciejko, X.-L. Qi, A. Karch, S.-C. Zhang, *Phys. Rev. Lett.* **105**, 246809 (2010).
36. B. Swingle, M. Barkeshli, J. McGreevy, T. Senthil, *Phys. Rev. B* **83**, 195139 (2011).

ACKNOWLEDGMENTS

We thank M. Franz, T. Hughes, A. MacDonald, J. Maciejko, J. Moore, M. Orlita, V. Oganesyan, W.-K. Tse, A. Turner, R. Valdés Aguilar, X. L. Qi, and S.C. Zhang for helpful discussions. Experiments were supported by the Army Research Office Grant W911NF-15-1-0560, with additional support by the Gordon and Betty Moore Foundation through grant GBMF2628 to N.P.A. at Johns Hopkins University. Film growth for this work was supported by the NSF DMR-1308142, EFMA-1542798, and the Gordon and Betty Moore Foundation EPiQS Initiative Grant GBMF4418 to S.O. at Rutgers.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1124/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S4
Movies S1 and S2
References (*37*–*45*)

26 February 2016; resubmitted 7 July 2016
Accepted 8 November 2016
10.1126/science.aaf5541

GALAXY FORMATION

Molecular gas in the halo fuels the growth of a massive cluster galaxy at high redshift

B. H. C. Emonts,[1]* M. D. Lehnert,[2] M. Villar-Martín,[1,3] R. P. Norris,[4,5] R. D. Ekers,[4] G. A. van Moorsel,[6] H. Dannerbauer,[7,8,9] L. Pentericci,[10] G. K. Miley,[11] J. R. Allison,[4] E. M. Sadler,[12] P. Guillard,[2] C. L. Carilli,[6,13] M. Y. Mao,[14,15] H. J. A. Röttgering,[11] C. De Breuck,[16] N. Seymour,[17] B. Gullberg,[18] D. Ceverino,[1,19] P. Jagannathan,[6] J. Vernet,[16] B. T. Indermuehle[4]

The largest galaxies in the universe reside in galaxy clusters. Using sensitive observations of carbon monoxide, we show that the Spiderweb galaxy—a massive galaxy in a distant protocluster—is forming from a large reservoir of molecular gas. Most of this molecular gas lies between the protocluster galaxies and has low velocity dispersion, indicating that it is part of an enriched intergalactic medium. This may constitute the reservoir of gas that fuels the widespread star formation seen in earlier ultraviolet observations of the Spiderweb galaxy. Our results support the notion that giant galaxies in clusters formed from extended regions of recycled gas at high redshift.

The formation of the largest galaxies in the universe is thought to be a two-stage process. For the past 10 billion years, these giant galaxies have grown mostly by cannibalizing smaller galaxies (*1*, *2*). However, computer simulations predict that in an earlier phase, lasting a few billion years, their stars condensed directly out of large reservoirs of accreted gas (*3*, *4*).

We present observational evidence for an extended gas reservoir fueling star formation in the massive Spiderweb galaxy, MRC 1138-262, which is located in a protocluster at a redshift of z = 2.161 (*5*–*9*). The Spiderweb galaxy is not a single galaxy, but rather an aggregation of protocluster galaxies. They are embedded in a giant halo of atomic (neutral and ionized) hydrogen gas, which radiates Lyα emission across a region of ~200 kpc (*6*). The central protocluster galaxy has a supermassive black hole at its core, which emits jets of relativistic particles visible in radio observations (*5*). Observations suggest that the protocluster galaxies will eventually merge and evolve into a single giant elliptical galaxy in the center of the cluster (*10*). We therefore refer to the Spiderweb galaxy as the entire region encompassed by the Lyα halo, and we refer to the gas between the protocluster galaxies as the intergalactic medium (IGM).

Earlier observations of line emission by carbon monoxide revealed the presence of large amounts of molecular gas in the Spiderweb galaxy (*11*). Because molecular gas is the raw fuel for the formation of stars, observations of molecular gas give us insight into the processes driving the evolution of the distant Spiderweb galaxy. We have obtained sensitive observations of the transition between the lowest two rotational levels of carbon monoxide, ^{12}CO (J = 1 → 0), using the Australia Telescope Compact Array (ATCA; 90-hour exposure time) and the Karl G. Jansky Very Large Array (VLA; 8-hour exposure time) (*12*). The ATCA observations were optimized for detecting low-surface-brightness emission from broadly distributed CO, with a 4.8″ × 3.5″ resolution. The VLA observations complement the ATCA data with a higher 0.7″ × 0.6″ resolution, making the VLA data sensitive to small-scale features but not to large-scale ones. Sampling these different spatial scales allows us to obtain a complete picture of the CO distribution, from the gas in the individual protocluster galaxies to that across the IGM.

Figure 1 shows that the CO emission in the ATCA data covers a region of ~70 kpc around the central radio galaxy. The CO (J = 1 → 0) luminosity is $L'_{\mathrm{CO,ATCA}}$ ~ 5.6 (±1.7) × 10^{10} K km s^{-1} pc^2. The high-resolution VLA data detect CO only on an arcsecond scale within the central radio galaxy itself. Figure 2 shows that this central emission accounts for only one-third of the CO luminosity observed with the ATCA, or $L'_{\mathrm{CO,VLA}}$ ~ 1.9 (±0.6) × 10^{10} K km s^{-1} pc^2. Thus, two-thirds of the CO detected with the ATCA, 3.7 (±1.1) × 10^{10} K km s^{-1} pc^2, originates from gas outside the central radio galaxy. Figure S1 shows that the CO-emitting gas outside the central radio galaxy is spread on scales larger than the individual protocluster galaxies. This large-scale molecular gas cannot be imaged with our VLA data, which are far less sensitive to low-surface-brightness emission than the ATCA data (fig. S1). Figure 3 shows that the bulk of this large-scale CO is not peaking cospatially with the central radio galaxy or any of its brightest surrounding protocluster galaxies; the CO-emitting gas instead appears to be concentrated between the galaxies (*12*). The velocity dispersion of the CO-emitting gas, σ_{CO} ~ 220 km s^{-1}, is also much lower than that of the protocluster galaxies, σ_{gal} > 1000 km s^{-1} (Fig. 2) (*7*). This agrees with the fact that our VLA data do not detect CO down to a 3σ limit of L'_{CO} < 9 × 10^9 K km s^{-1} pc^2 in any of the surrounding protocluster galaxies (*12*).

On the basis of Figs. 1 to 3 and fig. S1, we argue that our data show a molecular phase of the IGM embedded in the giant Lyα-emitting halo of the Spiderweb galaxy. There is no evidence that this large-scale molecular gas is rotating. Instead, the most blueshifted CO lies toward the southwest, while gas at increasingly higher velocities is distributed increasingly counterclockwise around the radio galaxy (Fig. 3). The brightest CO peak lies in the region between two bright spots in the radio jet (Fig. 3, E and F). This is consistent with recent detections of H_2O emission along the jets in the Spiderweb galaxy, which imply rapid dissipation of the kinetic energy supplied by the relativistic jets (*13*). Following recent models (*14*, *15*), the peak in molecular CO emission along the jet also likely indicates that jet-induced gas cooling occurs within the IGM (*12*).

Hubble Space Telescope (HST) imaging of extended rest-frame ultraviolet (UV) light in the IGM of the Spiderweb galaxy has previously revealed ongoing star formation across a region of ~70 kpc,

[1]Centro de Astrobiología (Consejo Superior de Investigaciones Científicas–Instituto Nacional de Técnica Aerospacial), Carretera de Torrejón a Ajalvir, km 4, 28850 Torrejón de Ardoz, Madrid, Spain. [2]Sorbonne Universités, Université Pierre et Marie Curie, Université Paris 6 et Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7095, Institut d'Astrophysique de Paris, 75014 Paris, France. [3]Facultad de Ciencias, Unidad Asociada Consejo Superior de Investigaciones Científicas, Astro-Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid, Spain. [4]Commonwealth Scientific and Industrial Research Organisation, Astronomy and Space Science, Australia Telescope National Facility, Epping, NSW 1710, Australia. [5]Western Sydney University, Penrith South, NSW 1797, Australia. [6]National Radio Astronomy Observatory, Socorro, NM 87801, USA. [7]Instituto de Astrofísica de Canarias, E-38205 La Laguna, Tenerife, Spain. [8]Departamento Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain. [9]Institut für Astrophysik, Universität Wien, 1180 Wien, Austria. [10]Istituto Nazionale di Astrofisica, Osservatorio Astronomico di Roma, 00040 Monteporzio (RM), Italy. [11]Leiden Observatory, University of Leiden, 2300 RA Leiden, Netherlands. [12]Sydney Institute for Astronomy, School of Physics, University of Sydney, NSW 2006, Australia. [13]Cavendish Laboratory, Cambridge CB3 0HE, UK. [14]Joint Institute for Very Long Baseline Interferometry, European Research Infrastructure Consortium, 7990 AA Dwingeloo, Netherlands. [15]Jodrell Bank Observatory, University of Manchester, Macclesfield, Cheshire SK11 9DL, UK. [16]European Southern Observatory, 85748 Garching, Germany. [17]International Centre for Radio Astronomy Research, Curtin University, Perth, Australia. [18]Centre for Extragalactic Astronomy, Department of Physics, Durham University, Durham DH1 3LE, UK. [19]Institut für Theoretische Astrophysik, Zentrum für Astronomie der Universität Heidelberg, 69120 Heidelberg, Germany.
*Corresponding author. Email: bjornemonts@gmail.com

with a star formation rate $SFR_{IGM} \sim 142\ M_\odot\ year^{-1}$ (where $M_\odot$ is the mass of the Sun) (*16*). The morphology of this UV-emitting region is similar to that of the CO emission (Fig. 3). The extended UV light is most likely produced by young O- and B-type stars that formed in situ within the IGM, because its rest-frame UV color (as measured through the HST Advanced Camera for Surveys F475W and F814W filters) is bluer than that of the protocluster galaxies, and the extended UV light is also not easily explained by nebular continuum, scattered light, or synchrotron emission (*16*). The estimated mass of molecular H_2 gas that is available to sustain this in situ star formation is $M_{H_2IGM} \sim 1.5\ (\pm 0.4) \times 10^{11} \times (\alpha_{CO}/4) M_\odot$, where $\alpha_{CO} \equiv M_{H_2}/L'_{CO}$ is the conversion factor between the CO luminosity and the molecular gas mass in the IGM (*17*). We assume a value of $\alpha_{CO} = 4 M_\odot$ (K km s^{-1} $pc^2)^{-1}$, because the IGM of the Spiderweb galaxy is likely to have a metallicity well below the solar value (*17*) and because an independent estimate of the H_2 mass based on the dust emission (*18*) gives the same value (*12*). Figure 4 shows that with this available molecular gas mass, the in situ star formation in the IGM of the Spiderweb galaxy follows the relation between the surface density of the SFR and the gas surface density observed in galaxies, also known as the Kennicutt-Schmidt relation (*12*, *19*). We therefore conclude that there is sufficient molecular gas available to fuel the in situ star formation within the IGM.

The molecular gas in the IGM of the Spiderweb galaxy could sustain the current rate of in situ star formation for a time scale $t_{depletion} \equiv M_{H_2}/SFR_{IGM} \sim 1.1 \times (\alpha_{CO}/4)$ billion years. With our assumed value of α_{CO}, this could fuel the star formation until $z \sim 1.6$. Therefore, even if the gas is replenished for another billion years, our results are consistent with earlier predictions that the halo must stop forming stars by $z = 1$, so that the stellar population has at least 5 billion years' time to age and reach the colors seen across the stellar halos of local central-cluster ellipticals (*10*).

Our study of the Spiderweb galaxy demonstrates that giant cluster galaxies can grow their stellar mass in situ out of very extended reservoirs of molecular gas, early in their formation process. The carbon and oxygen required to form the observed CO were made in the cores of stars. Therefore, the gaseous halo must have been polluted with recycled material that has been processed by previous episodes of star formation and was subsequently expelled back into the IGM. Our CO results complement absorption-line studies of distant quasars, which infer the presence of large (≥100 kpc) halos of warm ($T \sim 10^4$ K), metal-enriched gas (*20*). Our results also support recent computer models of gas infall into massive dark-matter halos of submillimeter-bright galaxies, which predict that stellar feedback both from galaxies and from in situ star formation can enrich gas across ~200 kpc with dust and metals (*21*). In addition to enrichment by star formation, powerful radio jets may drag metals out of the host galaxy and far into the halo environment (*22*). We therefore conclude that our observations

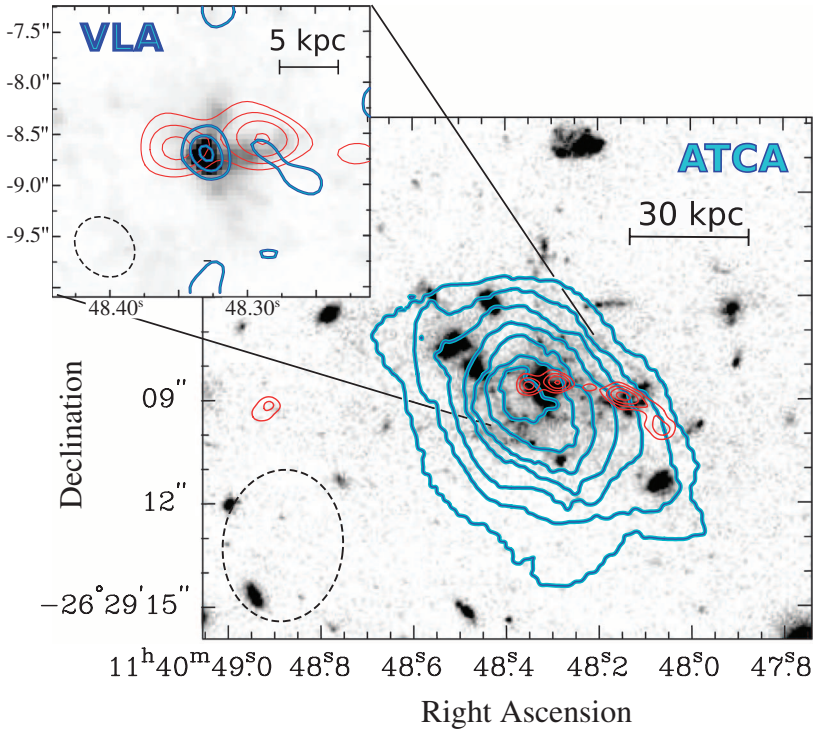


Fig. 1. Molecular gas on multiple scales in the Spiderweb galaxy. ^{12}CO ($J = 1 \rightarrow 0$) total-intensity contours from the ATCA (light blue) overlaid onto a negative grayscale HST image taken with the Advanced Camera for Surveys through the combined F475W and F814W filters [HST image copyright American Astronomical Society, reproduced with permission (*6*)]. Contour levels are 0.020, 0.038, 0.056, 0.074, 0.092, 0.110, and 0.128 Jy beam^{-1} × km s^{-1}. Red contours show the 36-GHz radio continuum from our VLA data at 0.20, 0.43, 0.80, and 1.40 mJy beam^{-1}. The inset shows the CO ($J = 1 \rightarrow 0$) total-intensity contours from the VLA (dark blue) at 2.8σ, 3.5σ, and 4.2σ, with σ = 0.019 Jy beam^{-1} × km s^{-1}. No negative contours are visible at this level around the central radio galaxy in the VLA data, likely because underlying large-scale flux skews the noise to slightly more positive values (fig. S1). The uncertainty in the astrometry of the HST image is ~0.3". The two inner radio-continuum components (red contours in the inset) likely trace the two-sided base of the radio jets. The dashed ellipses in the lower left corners visualize the beam size at full width of half the maximum intensity (FWHM). Coordinates are given in epoch J2000.

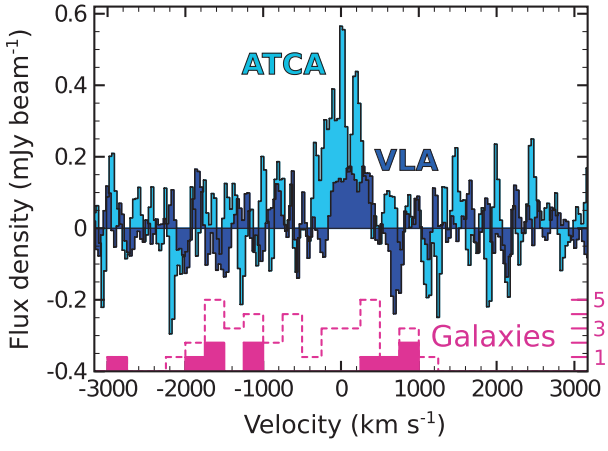


Fig. 2. Spectra of the CO-emitting gas. ^{12}CO ($J = 1 \rightarrow 0$) spectra at the location of the peak emission in Fig. 1. The light and dark blue spectra show the data taken with the ATCA and the VLA, respectively. The ATCA spectrum was obtained by tapering the data to a projected baseline length of ~200 m to ensure that we recover all the CO emission (fig. S1) (*12*). The VLA spectrum shows the CO emission at the highest spatial resolution. The histogram outlined with a dashed magenta line shows the velocity distribution of the protocluster galaxies (*7*). The solid magenta histogram includes only galaxies that lie within the CO-emitting region, with velocities derived from the rest-frame optical emission lines [O II], [O III], and Hα (*7*). The corresponding number of galaxies per bin is indicated in magenta on the right axis. Velocities are with respect to $z = 2.161$ (*12*).

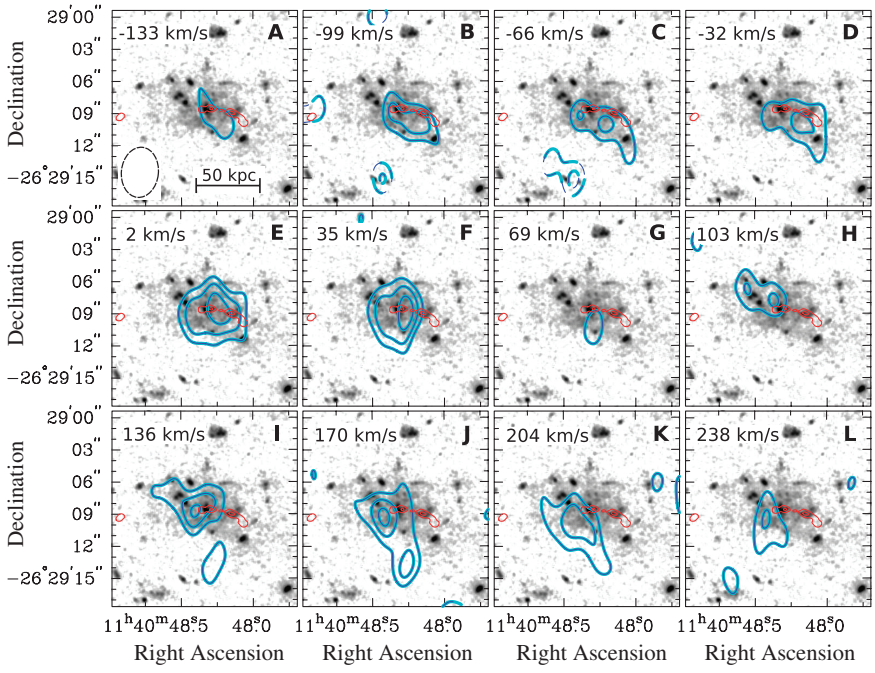


Fig. 3. Distribution and kinematics of the CO-emitting gas. (**A** to **L**) Maps of ^{12}CO ($J = 1 \rightarrow 0$) detected with the ATCA at different velocities, increasing in sequence from (A) to (L). Blue CO contour levels are at −3.5σ, −2.5σ (both dashed), 2.5σ, 3.5σ, and 4.5σ (all solid), where σ = 0.085 mJy beam^{-1}. The background grayscale image was taken with the HST Advanced Camera for Surveys through the F475W filter. This HST image is placed on a logarithmic scale to show the extended rest-frame UV light from in situ star formation discovered in earlier work (*16*). We downloaded these data from the HST Legacy Archive and reprocessed them by applying a 5 × 5–pixel boxcar smoothing to highlight the low-surface-brightness UV emission. Red contours show the radio-continuum source from Fig. 1. The beam size of the CO data is visualized with the dashed ellipse in (A). The positional accuracy of the CO peaks is ~0.5″ for a 4.5σ signal (*12*). The CO signal is not entirely independent across adjacent panels because of the applied Hanning smoothing (*12*). Velocities are with respect to $z = 2.161$ (*12*). Coordinates are given in epoch J2000.

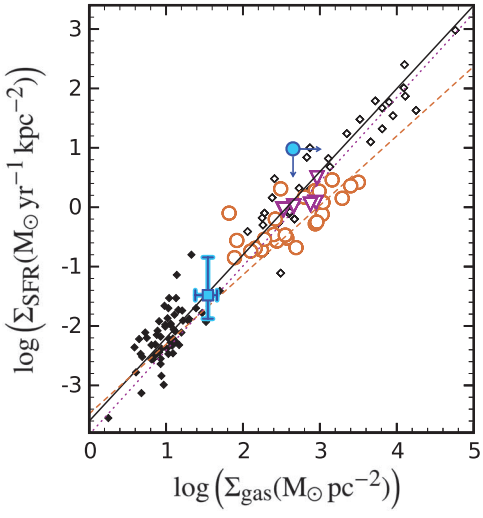


Fig. 4. The molecular star-forming gas in the IGM on the Kennicutt-Schmidt relation. The surface density of the star formation rate (Σ_{SFR}) is plotted as a function of the molecular gas mass surface density (Σ_{gas}). The solid blue square represents the in situ star formation across the IGM of the Spiderweb galaxy. The solid blue circle shows the limits for the central radio galaxy from our high-resolution VLA data for $\alpha_{CO} \geq 0.8M$ (K km s^{-1} pc^2)$^{-1}$. For limits and error calculations, see (*12*). Solid diamonds represent local spiral galaxies; open diamonds represent circumnuclear starbursts (*19*). Brown circles and purple triangles represent star-forming galaxies at $1.0 \leq z \leq 3.5$ (*23*) and $z \sim 1.5$ (*24*), respectively. The lines represent the best fit to the different samples [black solid (*19*), brown dashed (*23*), and purple dotted (*24*)].

have identified the predicted cold baryon cycle that governs the early growth of massive cluster galaxies (*14*, *21*).

REFERENCES AND NOTES

1. C. Lidman *et al.*, *Mon. Not. R. Astron. Soc.* **427**, 550–568 (2012).
2. G. De Lucia, V. Springel, S. D. M. White, D. Croton, G. Kauffmann, *Mon. Not. R. Astron. Soc.* **366**, 499–509 (2006).
3. L. Oser, J. P. Ostriker, T. Naab, P. H. Johansson, A. Burkert, *Astrophys. J.* **725**, 2312–2323 (2010).
4. C. A. Collins *et al.*, *Nature* **458**, 603–606 (2009).
5. L. Pentericci, H. J. A. Roettgering, G. K. Miley, C. L. Carilli, P. McCarthy, *Astron. Astrophys.* **326**, 580 (1997).
6. G. K. Miley *et al.*, *Astrophys. J.* **650**, L29 (2006).
7. E. Kuiper *et al.*, *Mon. Not. R. Astron. Soc.* **415**, 2245–2256 (2011).
8. Y. Koyama *et al.*, *Mon. Not. R. Astron. Soc.* **428**, 1551–1564 (2013).
9. H. Dannerbauer *et al.*, *Astron. Astrophys.* **570**, A55 (2014).
10. N. A. Hatch *et al.*, *Mon. Not. R. Astron. Soc.* **395**, 114–125 (2009).
11. B. H. C. Emonts *et al.*, *Mon. Not. R. Astron. Soc.* **430**, 3465–3471 (2013).
12. See supplementary materials on *Science* Online.
13. B. Gullberg *et al.*, *Astron. Astrophys.* **591**, A73 (2016).
14. G. M. Voit, M. Donahue, G. L. Bryan, M. McDonald, *Nature* **519**, 203–206 (2015).
15. Y. Li, G. L. Bryan, *Astrophys. J.* **789**, 54 (2014).
16. N. A. Hatch, R. A. Overzier, H. J. A. Röttgering, J. D. Kurk, G. K. Miley, *Mon. Not. R. Astron. Soc.* **383**, 931–942 (2008).
17. A. D. Bolatto, M. Wolfire, A. K. Leroy, *Annu. Rev. Astron. Astrophys.* **51**, 207–268 (2013).
18. J. A. Stevens *et al.*, *Nature* **425**, 264–267 (2003).
19. R. C. Kennicutt Jr., *Astrophys. J.* **498**, 541–552 (1998).
20. J. X. Prochaska, M. W. Lau, J. F. Hennawi, *Astrophys. J.* **796**, 140 (2014).
21. D. Narayanan *et al.*, *Nature* **525**, 496–499 (2015).
22. C. C. Kirkpatrick, B. R. McNamara, K. W. Cavagnolo, *Astrophys. J.* **731**, L23 (2011).
23. R. Genzel *et al.*, *Mon. Not. R. Astron. Soc.* **407**, 2091–2108 (2010).
24. E. Daddi *et al.*, *Astrophys. J.* **714**, L118 (2010).

ACKNOWLEDGMENTS

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013 under REA grant agreement 624351. We thank L. Colina and S. Arribas for useful discussions. Supported by Spanish Ministerio de Economía y Competitividad (MINECO) grant AYA2012-32295 (B.H.C.E. and M.V.-M.), MINECO under Ramón y Cajal program RYC-2014-15686 (H.D.), European Community grant H2020-MSCA-IF-2014 660432 (M.Y.M.), and European Research Council advanced grants DUSTYGAL 321334 and STARLIGHT 33177 (B.G. and D.C.). The Australia Telescope is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities Inc. The data reported in this paper are tabulated in the supplementary materials and archived at http://atoa.atnf.csiro.au under project codes C2052 and C2717 (ATCA) and at https://archive.nrao.edu/archive/advquery.jsp under project code 14B-160 (VLA).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1128/suppl/DC1
Materials and Methods
Supplementary Text
Fig. S1
Tables S1 to S3
References (*25*–*61*)

17 May 2016; accepted 21 October 2016
10.1126/science.aag0512

WATER CHEMISTRY

Spectroscopic snapshots of the proton-transfer mechanism in water

Conrad T. Wolke,[1] Joseph A. Fournier,[1,2] Laura C. Dzugan,[3] Matias R. Fagiani,[4,5] Tuguldur T. Odbadrakh,[6] Harald Knorke,[4] Kenneth D. Jordan,[6]* Anne B. McCoy,[3,7]* Knut R. Asmis,[4]* Mark A. Johnson[1]*

The Grotthuss mechanism explains the anomalously high proton mobility in water as a sequence of proton transfers along a hydrogen-bonded (H-bonded) network. However, the vibrational spectroscopic signatures of this process are masked by the diffuse nature of the key bands in bulk water. Here we report how the much simpler vibrational spectra of cold, composition-selected heavy water clusters, $D^+(D_2O)_n$, can be exploited to capture clear markers that encode the collective reaction coordinate along the proton-transfer event. By complexing the solvated hydronium "Eigen" cluster $[D_3O^+(D_2O)_3]$ with increasingly strong H-bond acceptor molecules (D_2, N_2, CO, and D_2O), we are able to track the frequency of every O-D stretch vibration in the complex as the transferring hydron is incrementally pulled from the central hydronium to a neighboring water molecule.

The translocation of positive charge through water is widely understood to occur through intermolecular transfer of a proton between oxygen atoms linked by a strong ionic hydrogen bond (*1*, *2*). This transport requires thermal fluctuations in the water H-bond network and occurs in a cooperative fashion such that H-bonds distant from the center of charge control the barrier for proton transfer (*3*–*5*). Many models of this process have been advanced, ranging from molecular-level pictures emphasizing the key role of the second solvation shell (*3*) to more holistic treatments where the entire medium is treated as a supermolecule in an electronic structure calculation (*4*, *6*–*10*). At the heart of this discussion is the relative importance of the two limiting forms of proton accommodation in water: the charge-delocalized, tricoordinated hydronium ion $[H_3O^+(\cdots OH_2)_3$ (or Eigen form)] (*11*) and the more charge-localized configuration where the proton is located at the midpoint of two closely separated oxygen atoms $[H_2O\cdots H^+\cdots OH_2$ (or Zundel form)] (*12*). The extremely diffuse vibrational signature of the aqueous proton (*11*–*13*) is a clear manifestation of the wide-ranging intramolecular distortions intrinsic to its molecular speciation, but this breadth also precludes unambiguous structural characterization of the proton defect by analysis of these bands.

Recently, Thämer *et al.* (*5*) reported ultrafast two-dimensional (2D) vibrational spectra of an aqueous 4 M HCl solution, which they interpreted to indicate a surprisingly large population of Zundel-type configurations. The ultrafast transients associated with the $H_5O_2^+$ Zundel motif were subsequently explored by Dahms *et al.* (*14*) in a scheme where this species was isolated in a room temperature solution of $HClO_4$ in acetonitrile. An important aspect of that study was the analysis of the ultrafast spectral diffusion displayed by the excess proton absorptions. In particular, Dahms *et al.* emphasized how the local electric field fluctuations in the liquid drive large structural changes in the embedded $H_5O_2^+$ moiety due to the contribution of the bridging proton to its polarizability. This, in turn, leads to the diffuse character of the vibrational spectra of weak acids, an idea first introduced by Zundel in 1972 (*13*).

The interpretation of the ultrafast infrared (IR) spectroscopy results rests on understanding how the local distortions in the charge accommodation motif are manifested in the vibrational frequencies associated with the displacement of the transferring proton parallel to the O-O axis, as well as the bending modes of the donor and acceptor water molecules. Analyses of the spectra displayed by $H^+(H_2O)_n$ clusters at low temperature and over a wide spectral range have played an important role in providing this information (*15*–*19*). A long-standing goal for the cluster work is to capture specific configurations along the pathway for proton transfer as locally stable structures and then identify the spectral signatures of all key protons involved in the transfer event. Arrangements that are transient in the bulk are often metastable or even global minima in clusters (*17*, *20*, *21*). Our strategy exploits the fact that H-bond acceptors attach preferentially to the OH groups of one of the three solvent water molecules that bind to the H_3O^+ core ion in the Eigen form of the $n = 4$ cluster, denoted **E4**. The scheme in Fig. 1 illustrates how, by complexing molecules with increasingly larger proton affinities (represented as "A"; A = H_2, N_2, CO, and H_2O) to the **E4** scaffold, we can follow the concerted distortions of the hydronium core and of the solvent water molecules as one of the hydronium protons is incrementally pulled closer to the solvated water molecule. From this perspective, the critical configuration for proton transfer corresponds to a cluster in which the proton is trapped at the midpoint between the donor and acceptor oxygen atoms. These measurements are carried out on cold (~20 K), size-selected protonated water clusters using cryogenic ion trap photofragmentation mass spectrometry. The multiple photon action spectrum of the bare $D^+(D_2O)_4$ cluster was obtained using the instrument at Yale, along with the linear action spectra of the solvated clusters, $D^+(D_2O)_4$-A (with A = D_2 and N_2). Spectra of the A = CO and D_2O clusters were obtained on a similar instrument in Leipzig. Schematic diagrams of these instruments are included in fig. S1.

Until very recently, the use of cluster vibrational spectroscopy to reveal the spectral reporters of the collective proton-transfer reaction coordinate has been hampered by the strongly anharmonic potentials that govern this process. As a result, cold-cluster vibrational spectra often display broad features that mask the small splittings expected to occur at the start of the proton-transfer path. Moreover, the observed patterns routinely display more bands than are expected for the fundamentals of a particular structural isomer. Such extra bands in the spectra of the crucial $n = 4$ and 5 clusters have been attributed, alternatively, to anharmonic features arising from a single isomer (*19*, *22*, *23*) or to overlapping bands from two isomers identified using a classical ab initio molecular dynamics approach (*24*). For $n = 4$, one of these isomers corresponds to the Eigen-based structure (**E4**), which was originally proposed (*25*). The other corresponds to a more open arrangement with an equally shared proton between central water molecules, as occurs in the Zundel ion, and is hereafter denoted **Z4**.

The difficulty in disentangling the roles of isomers and anharmonic effects can be traced to the fact that these systems are floppy and may not be reliably treated with the commonly applied theoretical and computational tools for predicting vibrational spectra. For example, calculating the simple band pattern displayed by the $H_5O_2^+$ cluster required combining very accurate potential energy and dipole surfaces over a large configuration space with a full (15D) quantum mechanical treatment of the nuclear motion (*26*). Such demanding methods have not yet been extended to larger cluster sizes due to their extreme computational overhead. As such, this work is being carried out in a regime where experiment leads theory, and our approach here is to combine analysis of empirical trends with several theoretical approaches to provide compelling assignments for the structures in play and the OH stretching fundamentals associated with them.

Implementation of the scheme outlined in Fig. 1 requires reliable isolation of the Eigen form of the $n = 4$ cluster, a structure that has

[1]Department of Chemistry, Yale University, New Haven, CT 06520, USA. [2]James Frank Institute and Department of Chemistry, University of Chicago, Chicago, IL 60637, USA. [3]Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH 43210, USA. [4]Wilhelm-Ostwald-Institut für Physikalische und Theoretische Chemie, Universität Leipzig, 04103 Leipzig, Germany. [5]Fritz-Haber-Institut der Max-Planck-Gesellschaft, 14195 Berlin, Germany. [6]Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15620, USA. [7]Department of Chemistry, University of Washington, Seattle, WA 98195, USA.
*Coresponding author. Email: jordan@pitt.edu (K.D.J.); abmccoy@uw.edu (A.B.M.); knut.asmis@uni-leipzig.de (K.R.A.); mark.johnson@yale.edu (M.A.J.)

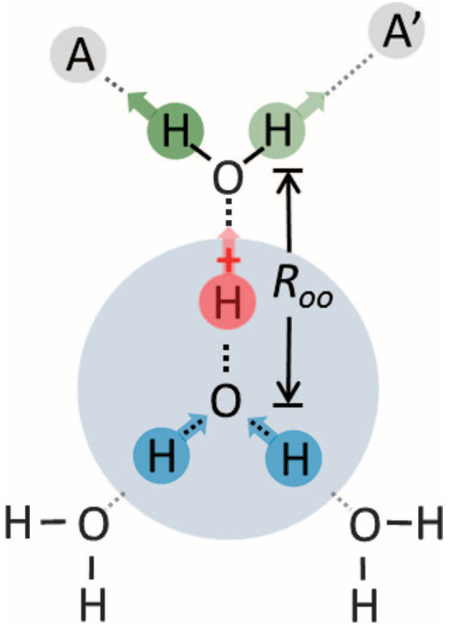


Fig. 1. Schematic of the proton-relay mechanism. The symmetric $H_9O_4^+$ Eigen ion is distorted upon the addition of proton acceptors A and A′ (A = H_2, N_2, CO, H_2O; A′ = H_2O) to one of the water molecules. Formation of these complexes induces the attraction of a proton in the H_3O^+ core toward the solvated H_2O molecule and reduces the corresponding O-O distance, R_{OO}.

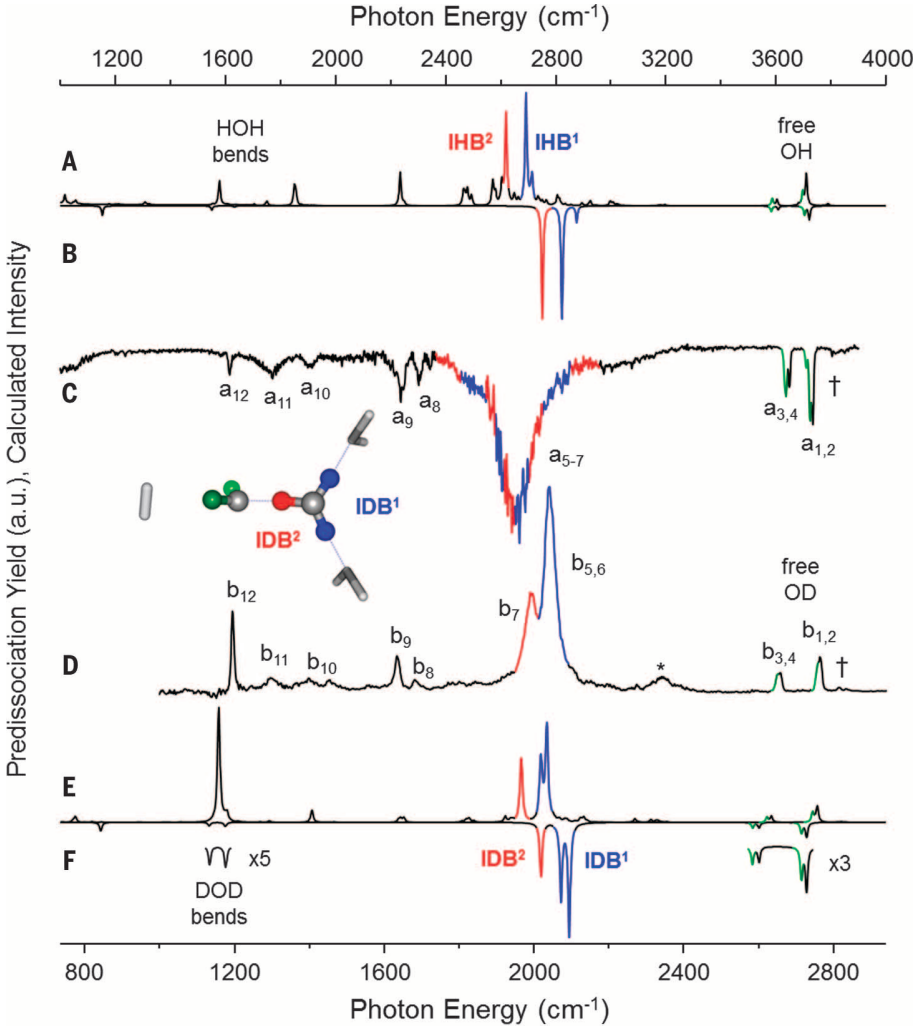


Fig. 2. Comparison of the experimental and calculated vibrational spectra of the $H^+(H_2O)_4$-D_2 and $D^+(D_2O)_4$-D_2 clusters. (**A**) Calculated anharmonic (VPT2) and (**B**) harmonic spectra of $H^+(H_2O)_4$-D_2, compared with the experimental vibrational predissociation spectrum in (**C**). (**D**) Experimental spectrum of $D^+(D_2O)_4$-D_2 compared with (**E**) its calculated anharmonic (VPT2) and (**F**) harmonic spectra. Calculations were performed at the MP2/aug-cc-pVDZ level, with the harmonic frequencies scaled by 0.9538 and the VPT2 frequencies left unscaled. Bands indicated by IH(D)B^1 (blue) and IH(D)B^2 (red) refer to symmetric and antisymmetric OH(D) stretches of the core D_3O^+ ion, which are slightly perturbed by complexation with D_2. Band labels (a_1 to a_{12}, b_1 to b_{12}) aid referencing in the text and in table S1. a.u., arbitrary units.

been invoked since the earliest reports of the $H^+(H_2O)_4$ vibrational spectrum by Okumura *et al.* (*25*). The D_2-tagged spectrum, $H^+(H_2O)_4$-D_2, is presented in Fig. 2C. (*19*) This assignment has recently been challenged by Kulig and Agmon (*24*), however, on the basis of cluster spectra calculated using classical molecular dynamics methods. The role of isomers in the n = 4 spectra was clarified by the recent determination [using an isomer-selective, IR-IR hole-burning technique] (*19*) that the $H^+(H_2O)_4$ spectrum (Fig. 2C), reproduced in many laboratories (*17*, *18*, *23*, *25*), is homogeneous. Application of vibrational perturbation theory (VPT2) (*27*), using the harmonic approximation for the unperturbed normal modes and frequencies to the **E4** structure, provides compelling assignments for several of the key features in question (Fig. 2A), but it does not account for the doubling of the peaks $a_{8,9}$ and $a_{10,11}$. As a result, it is presently unclear whether the diffuse features (a_{10} and a_{11}) above the intramolecular HOH bend are due to fundamentals of the H_3O^+ bend (*22*) or combination bands involving lower-frequency modes (*19*), and their assignments are still uncertain. Of most concern for our study, however, is the observation of the predicted splitting in OH stretching features [ionic H-bonded OHs labeled IHB^1 (blue) and IHB^2 (red) in Fig. 2A], calculated (at both the harmonic and VPT2 levels) to signal the initial distortion of the embedded H_3O^+ ion in the **E4** structure by the action of the weakly bound D_2 molecule. The three OH stretches in the hydronium are calculated to evolve into two distinct features that split apart as a proton is transferred; therefore, these features are denoted IHB^1 and IHB^2. This predicted splitting is completely obscured in the experimental $H^+(H_2O)_4$-D_2 spectrum, however, by broadening that is not anticipated at these levels of theory, the origin of which is not presently known.

A useful empirical tool in the assignment of anharmonic spectra involving hydrogen bonds is to follow the evolution of the band pattern with H/D substitution (*18*, *28*). Fundamentals primarily involving displacements of the hydrogen atoms are expected (at the harmonic level) to appear lower in energy by a factor of ~1.36 derived from the reduced mass change of the OH(D) system, and this scaling relation is closely followed by non-bonded OH stretching and HOH intramolecular bending fundamentals in many systems (*29*). Figure 2D presents the D_2-tagged n = 4 spectrum [$D^+(D_2O)_4$-D_2] with the energy axis scaled by 1.36. An unexpected dividend of this scheme is that some of the bands in the $H^+(H_2O)_4$-D_2 spectrum, whose assignments have been in question ($a_{8\text{-}11}$ in Fig. 2C), are suppressed relative to bend fundamentals (a_{12} and b_{12}) in the $D^+(D_2O)_4$-D_2 spectrum. Such selective suppression is anticipated for features that arise from coupling between the high-frequency OH stretches and soft modes of the scaffold (*23*). The persistent bands in the $D^+(D_2O)_4$-D_2 spectrum are close to features in the scaled $H^+(H_2O)_4$-D_2 spectrum, thus revealing that these transitions primarily involve displacements of the hydrons. Moreover, the pattern of

sharp bands in the $D^+(D_2O)_4$-D_2 spectrum can be readily assigned to OD stretching and bending fundamentals expected for the Eigen structure **E4**, as indicated by comparison with the calculated VPT2 spectrum in Fig. 2E and table S1. The prominent sharp feature (b_9) not assigned to a fundamental has been analyzed in detail for the H_3O^+-Y_3 series (Y = Ar, N_2, CH_4) and traced to a very strong enhancement of the transition moment of the H_3O^+ bending mode (not evident in Fig. 2D) upon rotation of the hydronium moiety about its symmetry axis within the tricoordinated solvent cage (*22*).

The fact that the fundamentals in the $D^+(D_2O)_4$-D_2 spectrum are close to strong absorptions in the scaled $H^+(H_2O)_4$-D_2 spectrum also provides compelling evidence that the Eigen structure is adopted by both isotopologues ($\mathbf{E4_H}$ and $\mathbf{E4_D}$). A comparison of the predictions obtained by application of VPT2 to both isotopologues is included in table S1, which includes assignments of the transitions in the calculated (VPT2) spectra for $\mathbf{E4_H}$-D_2 and $\mathbf{E4_D}$-D_2. The underlying causes for the extra features and broadening in the $\mathbf{E4_H}$-D_2 spectrum are of great interest but cannot be resolved with currently available theoretical tools. Consequently, we now focus on the simpler perdeutero isotopologues to identify the spectroscopic markers of intracluster proton transfer.

Key features that were obscured by broadening in the $\mathbf{E4_H}$-D_2 spectrum sharpen notably in that of $\mathbf{E4_D}$-D_2. As a result, the subtle splittings (IHB^1 and IHB^2) predicted to signal symmetry breaking of the hydronium core in the D_2-tagged **E4** structure are clearly revealed in bands $b_{5,6}$ and b_7. In the C_3 symmetric $\mathbf{E4_D}$ structure, two OD stretching fundamentals of the hydronium ion are predicted to be separated by only 38 cm^{-1} at the harmonic level. The lower-energy transition corresponds to the nondegenerate, symmetric OD stretch, whereas the higher-energy feature is due to the doubly-degenerate antisymmetric OD stretch. This near-degeneracy of the three OD stretches in the bare $\mathbf{E4_D}$ species is quickly broken as a central hydron is displaced away from D_3O^+, causing one of the OD stretch vibrations of the D_3O^+ to become localized on this OD group and to be displaced energetically below the nearly degenerate symmetric and antisymmetric stretching bands associated with the $-OD_2$ moiety. For reference, we recorded the infrared multiple photon dissociation (IRMPD) spectrum of $\mathbf{E4_D}$ (at a temperature of ~20 K) over the limited range afforded by that technique, with the spectrum displayed in Fig. 3A. A single feature is observed close to the centroid of the $IDB^{1,2}$ doublet in $\mathbf{E4_D}$-D_2.

The experimental results for a series of H-bond acceptors (A and A′ in Fig. 1) are presented in Fig. 3. For clarity, we color code the features traced primarily to the motion of the active hydrons engaged in the transfer process (red and blue for the hydronium, light and dark green for the relay or AD water molecule). The calculated harmonic and VPT2 spectra are presented in figs. S2 to S4, along with assignments of the calculated (VPT2) spectra in tables S2 to S4. The collective spectral

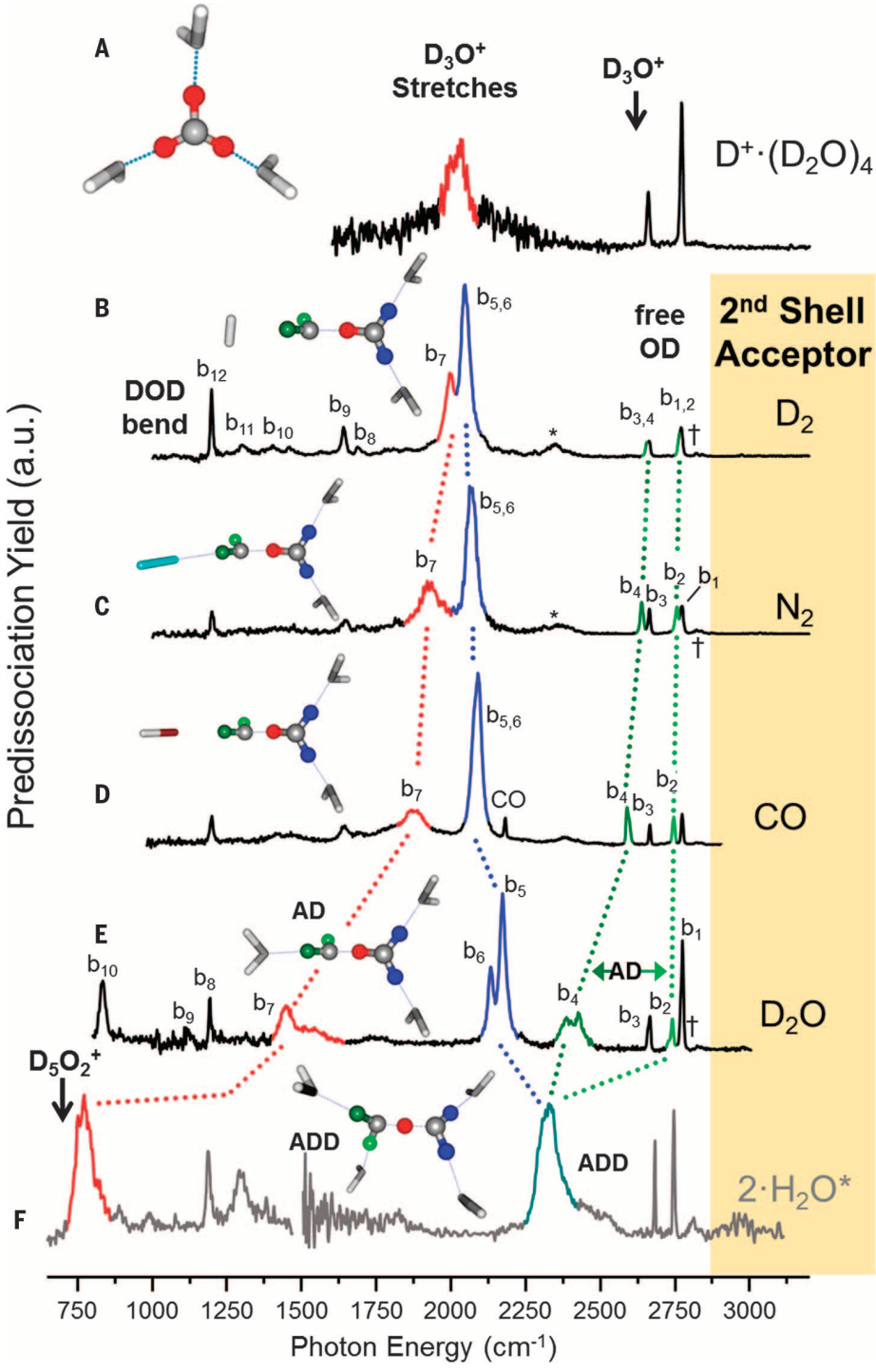


Fig. 3. Vibrational predissociation spectra of $D^+(D_2O)_4$-A clusters. These spectra reveal the local frequencies of all five OD oscillators involved in the transfer of a deuteron from one water molecule to another upon addition of incrementally stronger H-bonding molecules, A (see Fig. 1), to one of the water molecules in the primary hydration shell around the hydronium ion. (**A**) IRMPD spectrum of the bare ion. (**B** to **D**) Spectra of A = D_2, N_2, and CO. (**E**) D_2-tagged spectrum for A = D_2O [i.e., $D^+(D_2O)_5$-D_2]. (**F**) H_2-tagged spectrum of the Zundel isomer of the $H^+(H_2O)_6$ cluster, with the structure indicated in the inset (*21*), scaled by 1/1.36 to estimate the locations of key bands in the heavy isotopologue with this geometry. The arrow indicates the position of the bridging proton stretch in the $D_5O_2^+$ Zundel ion (*28*). This scaling is required because the $D^+(D_2O)_6$-D_2 cluster does not occur in this structure, as is evident by its vibrational spectrum (fig. S5). Band colors are correlated with atoms in the structure insets that are most associated with these transitions (red represents the transferred proton, dark green the OD group attached to the exterior acceptor, light green the other OD group in this DOD unit, and blue the $-OD_2$ motif that will reform a neutral water molecule after the transfer is complete).

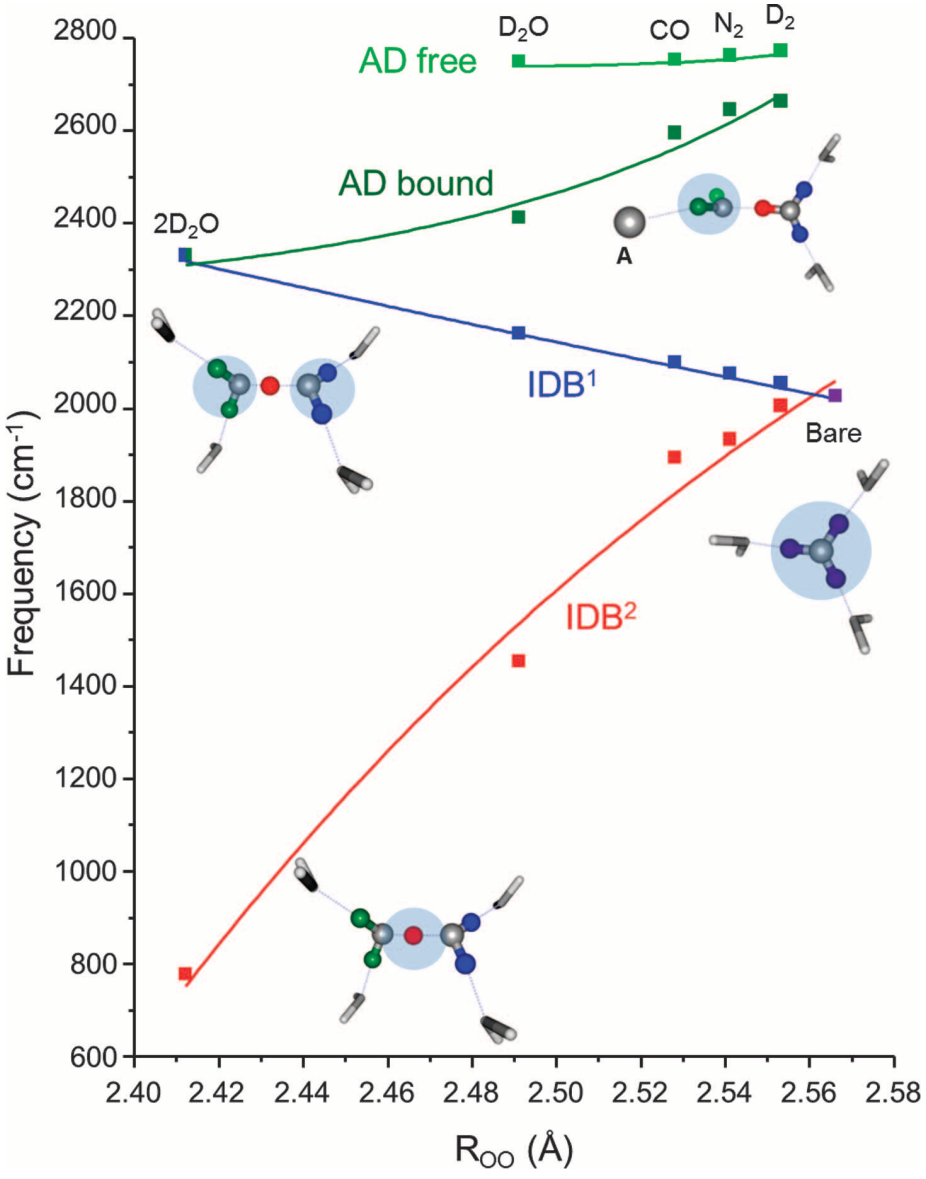


Fig. 4. Experimentally observed frequencies of the five OD stretches associated with the transfer of the hydron are plotted against the computed O-O distance between the acceptor and donor waters (MP2/aug-cc-pVDZ) for the series of $D^+(D_2O)_4$-A complexes. Red, transferring proton (IDB^2); blue, OD stretch of the donating water (IDB^1); dark green, OD of the water molecule bound to the H-bond acceptor, A; light green, free OD of the accepting water. The trends for the five OD stretches are fit to exponentials to guide the eye. The centroid frequencies and computed O-O distances are given in table S5.

response to complexation with increasingly basic (and, hence, more strongly H-bonded) molecules can be qualitatively understood as the spectral manifestation of partial charge accumulation on one of the central hydrons as it is pulled closer to the AD water molecule. Thus, when D_2 is replaced by N_2 and CO (with gas-phase basicities of 424, 495, and 594 kJ/mol, respectively) in the second shell, the highest-energy OD stretches are now clearly split apart such that two bands (b_1 and b_3) remain unaffected, whereas new bands [light green (b_2) and dark green (b_4)] appear lower in energy (Fig. 3, C and D). At the same time, the doublet associated with the hydronium ion [red (b_7) and blue ($b_{5,6}$)] splits farther apart ($\Delta\nu$ = 213 cm^{-1} for CO) in such a fashion that b_7 red-shifts to a greater extent than $b_{5,6}$ shifts to the blue. The CO stretching band is also evident in Fig. 3D and occurs 39 cm^{-1} above the stretch of the isolated CO molecule (2143 cm^{-1}), reflecting its response to the formation of the H-bond from a water molecule in the primary hydration shell.

The evolution of the OD_{sym}^{free} and OD_{asym}^{free} bands ($b_{3,4}$ and $b_{1,2}$) upon complexation warrants comment. Although the first H-bond acceptor molecule, A, is calculated to attach to only one of the OD groups, both OD_{asym} and OD_{sym} bands (b_2 and b_4) shift to the red relative to the two fundamentals (b_1 and b_3) associated with the spectator water molecules. This effect results from the relatively strong (~60 cm^{-1}) coupling between the two OD oscillators, which maintain their collective character so long as the perturbation of one of the OD groups upon accepting the H-bond is small relative to their intrinsic coupling. The phenomenon is accurately recovered at the harmonic and VPT2 levels, and a detailed analysis of the tag-induced decoupling is presented in the supplementary text (section SIV, part C) and fig. S7.

Formally, replacement of the more weakly perturbing molecules with a water molecule generates the Eigen-based form of the protonated water pentamer, **E5**, which is the traditionally accepted structure of this cluster (*16*, *17*, *30*, *31*). As in the situation regarding the assignments of the $H^+(H_2O)_4$ spectrum discussed above, the band pattern displayed by the $H^+(H_2O)_5$ cluster has recently come into question (*24*). The resolution of this issue and assignment of the fundamentals in the $D^+(D_2O)_5$-D_2 spectrum exclusively to the $\mathbf{E5_D}$ isomer follows the same protocol described above for the n = 4 system and is presented in a separate publication (*23*). Figure S4 shows both VPT2 and more computationally demanding vibrational self-consistent field results on both n = 5 isotopologues to address the location and fine structure associated with the key IDB^2 fundamental, which presents the greatest challenge for theory. The assignments of the OD stretches in the $\mathbf{E5_D}$-D_2 spectrum are collected in table S4 and are color-coded in Fig. 3E to highlight the key bands, which fall in line with the overall trend set by the weaker perturbers (Fig. 3, A to D). Because the bands associated with the hydrons most involved in the transfer (red and dark green) are increasingly red-shifted, the H-bond donor band (b_4, dark green) approaches the blue-shifted $-OD_2$ bands ($b_{5,6}$, blue) due to the increasingly distorted hydronium core. The symmetric (b_6) and antisymmetric (b_5) OD stretches of the $-OD_2$ group are now clearly split apart as a result of the lower symmetry of the **E5** scaffold, a feature that is accurately recovered in the VPT2 calculations (fig. S4).

The observation that the n = 5 cluster is approaching the tipping point for transfer of the bridging proton to the adjacent water molecule is related to the fact that the addition of one more water molecule to form the $H^+(H_2O)_6$ cluster results in the formation of at least two isomers, one with a Zundel arrangement (**Z6**) and the other best described as a distorted Eigen accommodation motif (**E6**) (*20*, *21*). Their spectra have been isolated through the application of isomer-specific IR-IR double-resonance spectroscopy (*21*), and a particularly important structure in this regard is that in which the proton is stabilized at the midpoint between the oxygen atoms of a proton-bound water dimer while the remaining four water molecules form the first hydration shell around a $H_5O_2^+$ Zundel core (*16*). The spectrum of the D_2-tagged n = 6 perdeutero cluster, included in fig. S5, does not clearly reflect either of the structures found in the light, H_2-tagged isotopologue (*21*), likely emphasizing the role of nuclear quantum effects when different arrangements are close in energy (*23*).

To estimate the band locations in the $n = 6$ perdeutero cluster with the Zundel core, we return to the isotopic scaling approach exploited earlier in our assignments of the **E4** and **E5** structures. Figure 3F presents the similarly scaled (1/1.36) spectrum for the Zundel form of the $H^+(H_2O)_6$ cluster ($\mathbf{Z6_H}$, isolated by double resonance) (*21*) to gauge how this arrangement would be encoded in the $D^+(D_2O)_6$ spectrum. The measured location of the bridging deuteron stretch in the isolated $D_5O_2^+$ ion is indicated by the arrow at the left of Fig. 3F, which, falling only 62 cm^{-1} below the location obtained by scaling in $\mathbf{Z6_D}$ (*15*, *28*), indicates that the errors introduced by this procedure are small compared with the overall shifts in the bands. The comparison over the entire spectral region can be found in fig. S6.

The band positions in the $\mathbf{Z6_D}$ spectrum (Fig. 3F) generated by scaling are readily understood as a continuation of the trend displayed by the series shown above in Fig. 3, B to E. The parallel stretch of the shared deuteron is estimated by the scaling procedure to occur at ~750 cm^{-1} (red), whereas the four H-bonded OD stretches of the water molecules in the first solvation shell around the Zundel core (light green, dark green, and blue) now appear as a single broadened feature at 2321 cm^{-1}, labeled "ADD" in Fig. 3F. In effect, the **Z6** structure can be viewed as a pair of coupled Eigen ions, each with a maximal splitting between the three OD oscillators.

The observation of clear spectroscopic signatures that encode progress along the proton-transfer reaction provides an opportunity to quantify the relationship between a hydron's stretching frequency and the environment in which the hydron is located. A useful index to describe the extent of proton transfer is the shortest distance between two oxygen atoms that surround the hydron defect, R_{OO} in Fig. 1, which ranges from 2.57 to 2.41 Å in the transition from the Eigen to the Zundel accommodation motifs. The experimental (and, for $n = 6$, estimated by scaling from $\mathbf{Z6_H}$) band positions of the five most active OD groups for six R_{OO} values calculated for the various complexes are presented in Fig. 4. The points along this curve correspond to the centroids of the key bands in Fig. 3 (given in table S5). These are colored to highlight the D atom between the nearest oxygen atoms (red), the remaining $-OD_2$ group on the Eigen core (blue), the OD group bound to A (dark green), and the free OD stretch in that water molecule (light green). Specifically, the three closely spaced OD stretching bands in the Eigen cation (red and blue at R_{OO} ~ 2.57 Å) quickly split apart such that the one corresponding to the transferred deuteron falls in frequency in accordance with the usual correlation with the distance between the heavy atoms (*32*–*34*). The most important result of this work is therefore the determination of the parametric dependence of the four other OD stretches on R_{OO}, which encodes the collective response of the surrounding H-bonding network throughout the course of the hydron transfer. Overall, attachment of an H-bond acceptor (A) to an OD group on a water molecule in the primary hydration shell of hydronium acts to break the near degeneracy of the three OD groups, whereas the newly formed H-bond decouples the two OD oscillators on the AD water molecule. This effect can be driven to the point that the deuteron is equally shared with addition of two water molecules to yield an ADD arrangement for the deuteron acceptor, at which point the stretches of the outer four OD groups are quasi-degenerate and most widely separated from the very low frequency of the deuteron trapped at the midpoint.

The key to understanding the large solvatochromic response of the excess proton lies in the nature of the Zundel ion, **Z2**. In Zundel's model for the breadth of the hydrated proton spectrum (*13*), the potential that governs the parallel vibration of the excess proton is strongly perturbed by the electric field of the solvent surrounding the $H_5O_2^+$ moiety. This effect arises because of the very large mechanical contribution to the polarizability due to field-induced displacement of the central proton (*12*). Because this potential is softest for **Z2**, low-energy displacements of the shared proton lead to large changes in the effective dipole moment, as evidenced by the very intense vibrational fundamental associated with this mode (*15*, *28*). Dahms *et al.* (*14*) have very recently revisited this calculation using modern quantum chemical methods in the context of understanding the ultrafast relaxation of the Zundel ion. The spectroscopic behavior of the cold-cluster model systems now allows a more quantitative exploration of this proton polarizability model. In fig. S8, we illustrate how the effect of H-bond acceptors on the potential-energy curve for the bridging proton can be recovered by considering the electric field they impart to an isolated $H_5O_2^+$.

It is clear from the distance dependence of the OD stretching frequency in Fig. 4 that there is a continuum in the spectral response of the system to hydron transfer that covers the entire infrared spectral region. A particularly important aspect of this work is the determination of the correlations between the hydron stretching absorptions across a very large spectral range, with the lower range due to the transferring hydron and the upper range due to the flanking OD groups. The geometries represented in a spectroscopic measurement in bulk water will therefore reflect the convolution between the population of R_{OO} values and the oscillator strengths for the bands correlated with this distance. Such information should be accessible with two-color 2D IR methods and could affect the present interpretation of correlated transients based on intramolecular HOH bending and excess proton stretching assignments (*5*). In the extrapolation of cluster behavior to the liquid, however, the R_{OO} values operative in solution are likely to be substantially different from those in small gas-phase clusters. This effect was recently emphasized in our work on the surface-embedded D_3O^+, where the OD stretches manifest a red shift of ~300 cm^{-1} as the cluster size increased from $n = 4$ to 21 (*18*, *19*). The extension of the cluster studies to follow how symmetry breaking is manifested in larger networks is thus a challenging but important direction for future work.

REFERENCES AND NOTES

1. C. J. D. von Grotthuss, *Ann. Chim.* **LVIII**, 54–74 (1806).
2. D. Marx, *ChemPhysChem* **7**, 1848–1870 (2006).
3. T. C. Berkelbach, H. S. Lee, M. E. Tuckerman, *Phys. Rev. Lett.* **103**, 238302 (2009).
4. D. Marx, M. E. Tuckerman, J. Hutter, M. Parrinello, *Nature* **397**, 601–604 (1999).
5. M. Thämer, L. De Marco, K. Ramasesha, A. Mandal, A. Tokmakoff, *Science* **350**, 78–82 (2015).
6. M. Tuckerman, K. Laasonen, M. Sprik, M. Parrinello, *J. Chem. Phys.* **103**, 150–161 (1995).
7. M. E. Tuckerman, D. Marx, M. L. Klein, M. Parrinello, *Science* **275**, 817–820 (1997).
8. M. Heyden *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12068–12073 (2010).
9. H. T. Bian *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4737–4742 (2011).
10. O. Markovitch *et al.*, *J. Phys. Chem. B* **112**, 9456–9466 (2008).
11. M. Eigen, L. de Maeyer, *Proc. R. Soc. London* **247**, 505–533 (1958).
12. G. Zundel, H. Metzger, *Z. Phys. Chem.* **58**, 225–245 (1968).
13. R. Janoschek, E. G. Weidemann, H. Pfeiffer, G. Zundel, *J. Am. Chem. Soc.* **94**, 2387–2396 (1972).
14. F. Dahms *et al.*, *Angew. Chem. Int. Ed.* **55**, 10600–10605 (2016).
15. K. R. Asmis *et al.*, *Science* **299**, 1375–1377 (2003).
16. J. M. Headrick *et al.*, *Science* **308**, 1765–1769 (2005).
17. K. Mizuse, A. Fujii, *J. Phys. Chem. A* **116**, 4868–4877 (2012).
18. J. A. Fournier *et al.*, *Science* **344**, 1009–1012 (2014).
19. J. A. Fournier *et al.*, *J. Phys. Chem. A* **119**, 9425–9440 (2015).
20. K. Mizuse, A. Fujii, *Phys. Chem. Chem. Phys.* **13**, 7129–7135 (2011).
21. N. Heine *et al.*, *J. Am. Chem. Soc.* **135**, 8266–8273 (2013).
22. A. B. McCoy, T. L. Guasco, C. M. Leavitt, S. G. Olesen, M. A. Johnson, *Phys. Chem. Chem. Phys.* **14**, 7205–7214 (2012).
23. M. R. Fagiani *et al.*, *Phys. Chem. Chem. Phys.* **18**, 26743–26754 (2016).
24. W. Kulig, N. Agmon, *J. Phys. Chem. B* **118**, 278–286 (2014).
25. M. Okumura, L. I. Yeh, J. D. Myers, Y. T. Lee, *J. Chem. Phys.* **85**, 2328–2329 (1986).
26. O. Vendrell, F. Gatti, H. D. Meyer, *J. Chem. Phys.* **127**, 184303 (2007).
27. V. Barone, *J. Chem. Phys.* **122**, 14108 (2005).
28. L. R. McCunn, J. R. Roscioli, M. A. Johnson, A. B. McCoy, *J. Phys. Chem. B* **112**, 321–327 (2008).
29. S. Horvath *et al.*, *J. Phys. Chem. A* **114**, 1556–1568 (2010).
30. G. E. Douberly, R. S. Walters, J. Cui, K. D. Jordan, M. A. Duncan, *J. Phys. Chem. A* **114**, 4570–4579 (2010).
31. C. Chaudhuri *et al.*, *Mol. Phys.* **99**, 1161–1173 (2001).
32. A. Novak, in *Large Molecules* (Springer, 1974), pp. 177–216.
33. W. Mikenda, *J. Mol. Struct.* **147**, 1–15 (1986).
34. E. Libowitzky, *Monatsh. Chem.* **130**, 1047–1059 (1999).

ACKNOWLEDGMENTS

M.A.J. and K.D.J. acknowledge financial support from the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award Numbers DE-FG02-06ER15800 and DE-FG02-06ER15066. K.D.J. acknowledges the use of resources in the University of Pittsburgh's Center for Simulation and Modeling. A.B.M. thanks the U.S. NSF (grant CHE-1619660) and the Ohio Supercomputing Center for resources on the Oakley Cluster. M.R.F. and K.R.A. acknowledge financial support from Collaborative Research Center 1109 of the German Research Foundation (Deutsche Forschungsgemeinschaft). Additional data supporting the conclusions are available in supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1131/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 to S5
References (*35*–*43*)

8 April 2016; resubmitted 11 August 2016
Accepted 29 September 2016
10.1126/science.aaf8425

WORKING MEMORY

Reactivation of latent working memories with transcranial magnetic stimulation

Nathan S. Rose,[1,2]* Joshua J. LaRocque,[1,3] Adam C. Riggall,[1,5] Olivia Gosseries,[1,4] Michael J. Starrett,[1] Emma E. Meyering,[1] Bradley R. Postle[1,5]*

The ability to hold information in working memory is fundamental for cognition. Contrary to the long-standing view that working memory depends on sustained, elevated activity, we present evidence suggesting that humans can hold information in working memory via "activity-silent" synaptic mechanisms. Using multivariate pattern analyses to decode brain activity patterns, we found that the active representation of an item in working memory drops to baseline when attention shifts away. A targeted pulse of transcranial magnetic stimulation produced a brief reemergence of the item in concurrently measured brain activity. This reactivation effect occurred and influenced memory performance only when the item was potentially relevant later in the trial, which suggests that the representation is dynamic and modifiable via cognitive control. The results support a synaptic theory of working memory.

The ability to mentally retain information in an accessible state, to manipulate it, and to use it to guide behavior is a critical building block for cognition. It has long been assumed that the neural basis for this working memory (WM) ability is elevated and persistent neuronal firing (*1*). This assumption has been called into question by recent proposals that information can be held in WM via synaptic mechanisms that do not require sustained, elevated brain activity (*2*–*4*).

Building on theoretical frameworks that information can be held in WM in one of several states of activation (*5*, *6*), we recorded neural activity while participants performed a multistep task in which two items were presented as memoranda for each trial. A cue indicated which item would be tested by the impending recognition memory probe, followed by the probe, then by a second cue, and then a second probe (Fig. 1). There was equal probability following the first cue, but not the second, that the uncued item might be needed for an ensuing memory judgment. This procedure moves the uncued item into a different state than the cued item, which, by definition, is in the focus of attention (*7*). Cognitive theories refer to the intermediate state of this unattended memory item (UMI) as "activated long-term memory" (LTM) (*5*, *6*).

For experiment 1, multivariate pattern analysis (MVPA) showed evidence for an active representation of the UMI that dropped to baseline levels (Fig. 2) (*7*–*9*). This suggests that information in WM (but outside of focal attention) can be maintained in a latent state via mechanisms other than sustained, elevated activity. Although a similar drop-to-baseline pattern is observed when participants are instructed to drop information

[1]Department of Psychiatry, University of Wisconsin, Madison, WI 53706, USA. [2]Department of Psychology, University of Notre Dame, Notre Dame, IN 46556, USA. [3]Neuroscience Training Program, University of Wisconsin, Madison, WI 53706, USA. [4]Coma Science Group, University of Liège, 4000 Liège, Belgium. [5]Department of Psychology, University of Wisconsin, Madison, WI 53706, USA.
*Corresponding author. Email: nrose1@nd.edu (N.S.R.); postle@wisc.edu (B.R.P.)

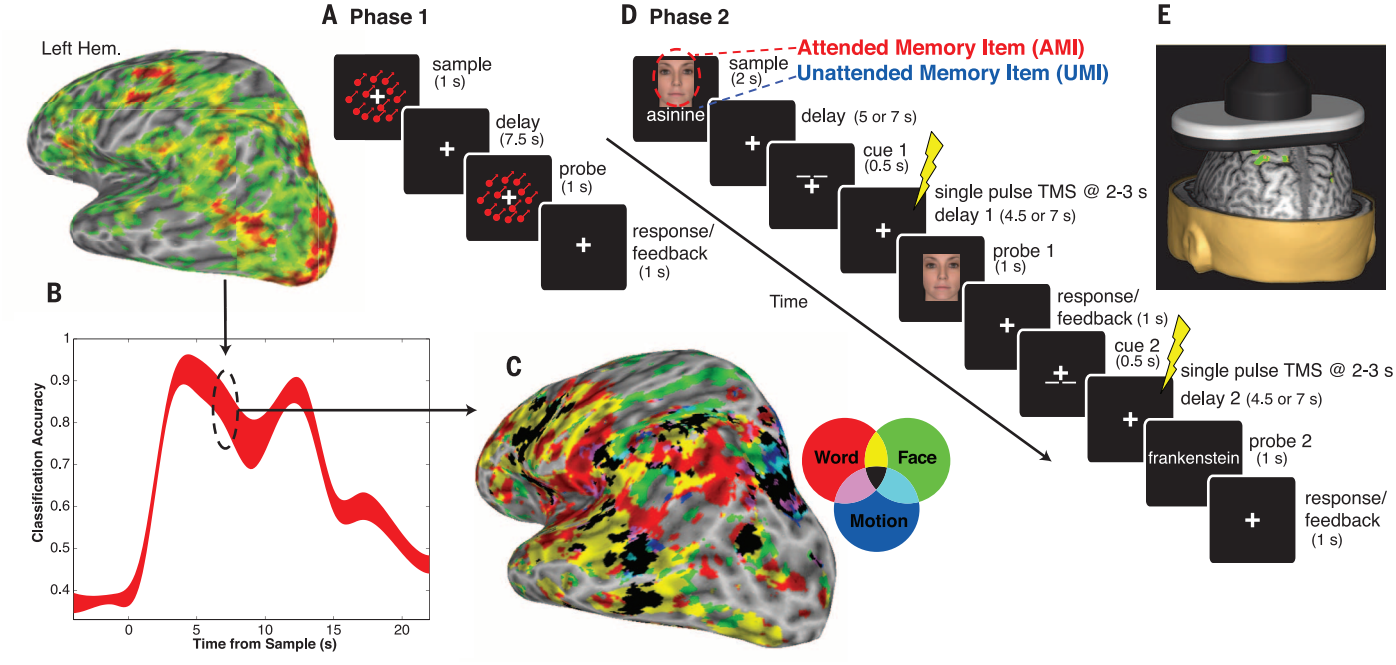


Fig. 1. General procedure. (**A**) In phase 1, functional magnetic resonance imaging (fMRI) data were acquired while participants performed a one-item delayed-recognition task for words, faces, or directions of motion; these data were used for multivariate pattern analysis (MVPA). (**B**) Classifiers trained on the delay period were used for subsequent analyses. For experiment 1, these classifiers were used to decode fMRI activity from phase 2 (Fig. 2). (**C** and **D**) For experiments 2 and 3, they were used in a whole-brain searchlight conjunction analysis to generate participant-specific maps of category-sensitive areas (C); nonoverlapping areas were used for transcranial magnetic stimulation (TMS) targeting in phase 2 (D). (**E**) In phase 2, single pulses of TMS were delivered during the postcue delay periods.

from WM (*10*, *11*), here the UMI remained in WM because, when so instructed by the second cue, participants accurately reactivated it and used it to evaluate the final probe (Fig. 2B).

In three additional experiments, we tested the hypothesis that if a UMI is encoded in a distributed pattern of synaptic weights and held in a state that is more accessible than trial-irrelevant information, the readout from a nonspecific burst of activity filtered through this network might reveal this latent representation (*2*) (fig. S1). This would be consistent with the idea that networks in the posterior cortex can be dynamically configured as matched filters to encode behaviorally relevant information (*3*, *4*, *12*, *13*).

For experiments 2 and 3, participants performed the phase 2 WM task (Fig. 1) while we recorded electroencephalography (EEG) and applied single-pulse transcranial magnetic stimulation (TMS) 2 to 3 s after the cue. For experiment 2, we targeted brain regions identified from the phase 1 MRI task as preferentially supporting MVPA decoding for one category, but not the other two. MVPA of the spectrally transformed EEG data from only the phase 2 task detected reliable evidence for an active representation of both memory items across the initial portion of the trial, until the onset of the first cue, at which point decoding accuracy remained elevated for the attended memory item (AMI) but dropped to the baseline for the UMI (*14*).

After a single pulse of TMS, there was a brief recovery of MVPA decoding of the UMI—a "reactivation effect"—before it returned to baseline and remained there while the cued item was tested [P = 0.01; Bayes factor (BF) = 3.64 against the null] (Fig. 3A). TMS affected neither broadband decoding of the AMI nor recognition memory judgments (fig. S4). When we analyzed bandpass-filtered data, the TMS reactivation effect was isolated to signal from the beta band (fig. S5) and was associated with a transient period of above-chance decoding performance for both the UMI and the AMI. The TMS reactivation effect was specific for information that was in WM on that trial, because above-chance MVPA performance, as assessed with the AUC (area under the curve) analysis, necessarily means that TMS did not activate a representation of the category that was irrelevant on that trial.

In experiment 2, we administered blocks of trials with TMS targeting one of the category-selective regions, but we varied, on a trial-by-trial basis, which category was the AMI and which was the UMI. Each block included trials for which the UMI belonged to the targeted region's preferred category, and trials for which it did not. A TMS reactivation effect was observed (Fig. 3B) whether or not TMS targeted the UMI's category-preferred region, although the effect was larger and more prolonged when it did (BF = 4.02 for targeted sites, 1.72 for nontargeted sites). This finding suggests that WM is supported by heightened connectivity between cortical networks that represent all trial-relevant information (AMI and UMI) relative to trial-irrelevant information (*15*, *16*).

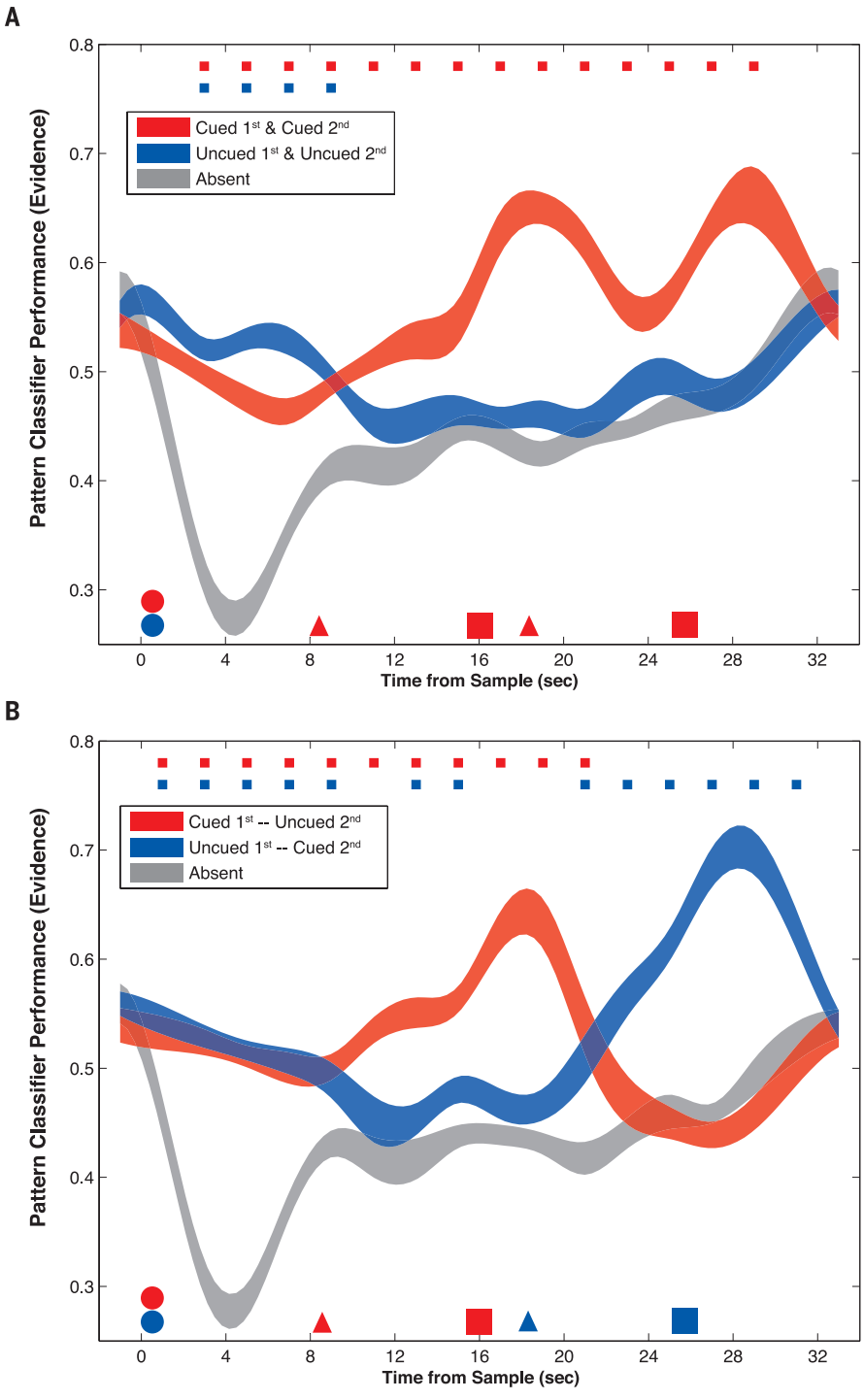


Fig. 2. Experiment 1 fMRI decoding (train phase 1, test phase 2): Classifier evidence as a function of an item's status, collapsed across stimulus category. After stimulus presentation (red and blue circles), delay-period classifier evidence for both items was elevated relative to the empirical baseline of evidence for the category that was not presented on that trial ("absent," gray). Upon presentation of the first cue (red triangle), evidence for the cued category (red) remained elevated, but for the uncued category (blue) dropped to baseline. (**A**) After the first probe (red square), on half the trials the second cue designated that the same item would be tested by the second probe, and evidence for the two categories remained the same relative to baseline. (**B**) When the second cue designated the previously uncued item, evidence for the two categories reversed for the remainder of the trial. Color-coded small squares at the top of each plot indicate $P < 0.01$; line width reflects SEM.

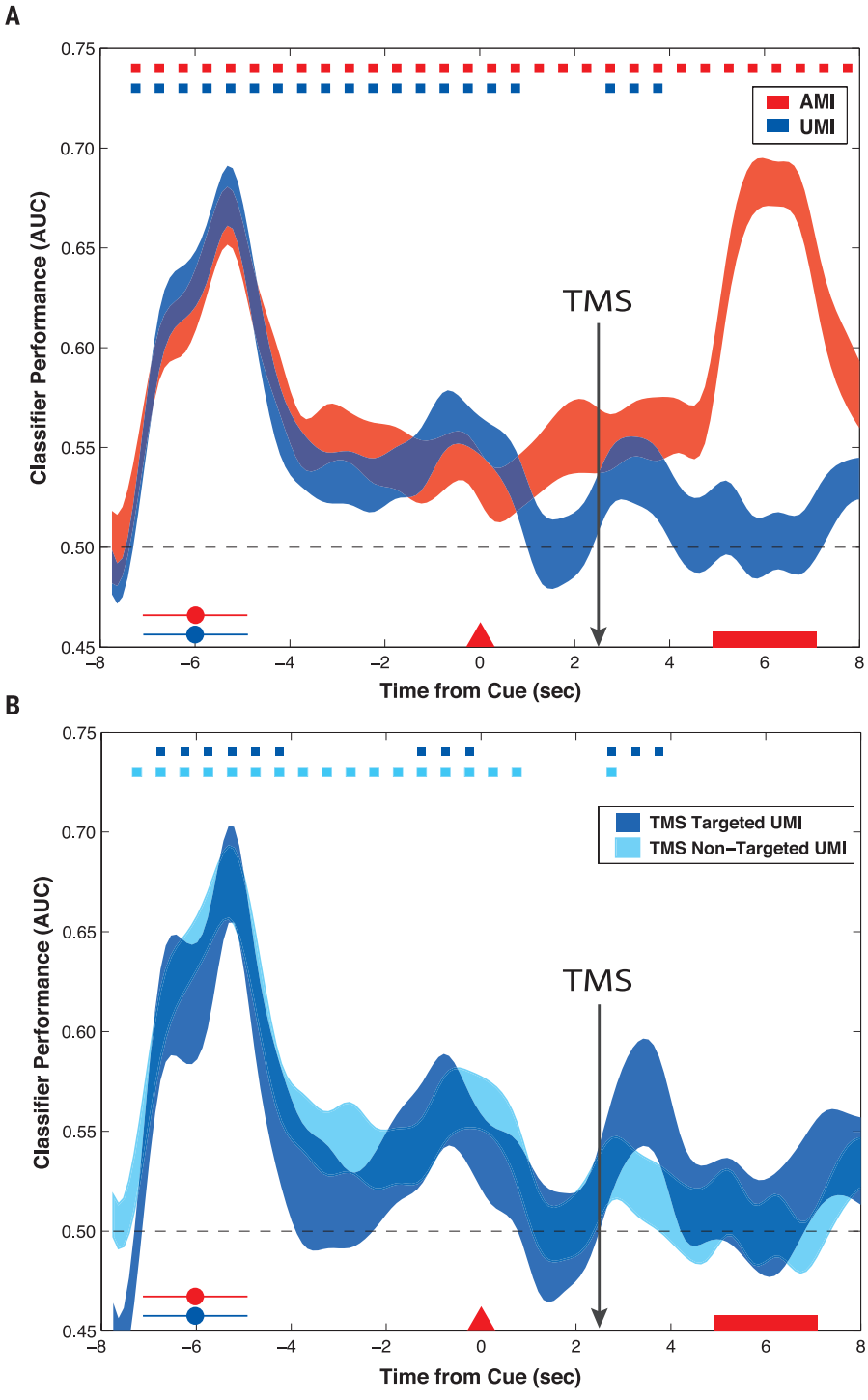


Fig. 3. Experiment 2 EEG decoding (train and test on phase 2 data): Classifier accuracy (area under curve, AUC) as a function of an item's status at the time of the first cue, collapsed across stimulus category. AUC reflects classifier sensitivity to discriminating between evidence for the AMI or UMI relative to the absent category. (**A**) Classification time series of the AMI and UMI upon stimulus presentation (red and blue circles), the first cue (red triangle), TMS, and first probe (red rectangle), averaged over $N = 18$ sessions, 2952 trials (decoding ends where the AMI and UMI switched on 50% of the trials). (**B**) Decoding UMIs as a function of whether TMS targeted that item's phase 1–defined region or a different category's region. Color-coded small squares at the top of each plot indicate $P < 0.05$; line width reflects SEM.

Retrocues that inform subjects that they can drop an item from memory result in a rapid loss of multivariate evidence for the no longer relevant item (*11*, *17*). Nonetheless, proactive interference from stimuli presented on previous trials indicates that the brain retains a residual trace of such recent, but no longer relevant, information (*18*). An important test of state-based models of WM is whether there is a functional distinction between UMIs (putatively held in a state of activated LTM) and dropped information (no longer in WM). In experiment 3, with a different group of participants, we also administered TMS after the second cue, after which the uncued item would no longer be relevant on the trial, and at which point it should have the same status as an irrelevant item. If the TMS reactivation effect is a consequence of an item being maintained in a privileged state, it should only be observed when that item is still potentially relevant for the trial. We also jittered the onset of TMS between 2 and 3 s after the cues (*14*) and standardized TMS by targeting the same region on every trial for all participants—an MVPA-defined region in the right precuneus known to be critical for the top-down control of visual attention (*19*) (Fig. 4A).

For the first half of the trial, the results from experiment 3 replicated those from experiment 2 (Fig. 4B), with a robust TMS reactivation effect for the UMI (BF = 9.8 against the null). For the delay period following the second cue, however, there was no evidence for significant decoding of the uncued item following the TMS pulse (BF = 3.4 in favor of the null). These results suggest that UMIs are maintained in a different state than are items that have been dropped from WM, and that the mechanisms that maintain latent representations in WM are dynamic and modifiable via cognitive control (*20*).

Because our design entails decoding at the category level, it does not rule out the possibility that the TMS reactivation effect reflects a general reinstatement of category context (*21*), rather than the temporary activation of the UMI itself. The idea that the representation of the UMI itself drives this effect would be strengthened by demonstrating that TMS can influence recognition memory decisions on this task. If the TMS reactivation effect reflects a temporary reinstatement of the UMI back into the focus of attention, participants should have more difficulty rejecting the UMI as a lure when probing their memory of the AMI.

In experiment 4, we presented recognition memory probes that matched the AMI on 50% of trials; of the 50% of nonmatch probes, 30% were drawn from the same category as the AMI, and a critical 20% matched the UMI (*14*). Participants were instructed to reject memory probes that did not match the AMI. Critically, only for the first probe was there an increased proportion of false alarms to the UMI for TMS relative to no-TMS trials (Fig. 3C, $P = 0.01$, BF = 3.48) (*14*).

Our results have important implications for the understanding of WM at many levels. They

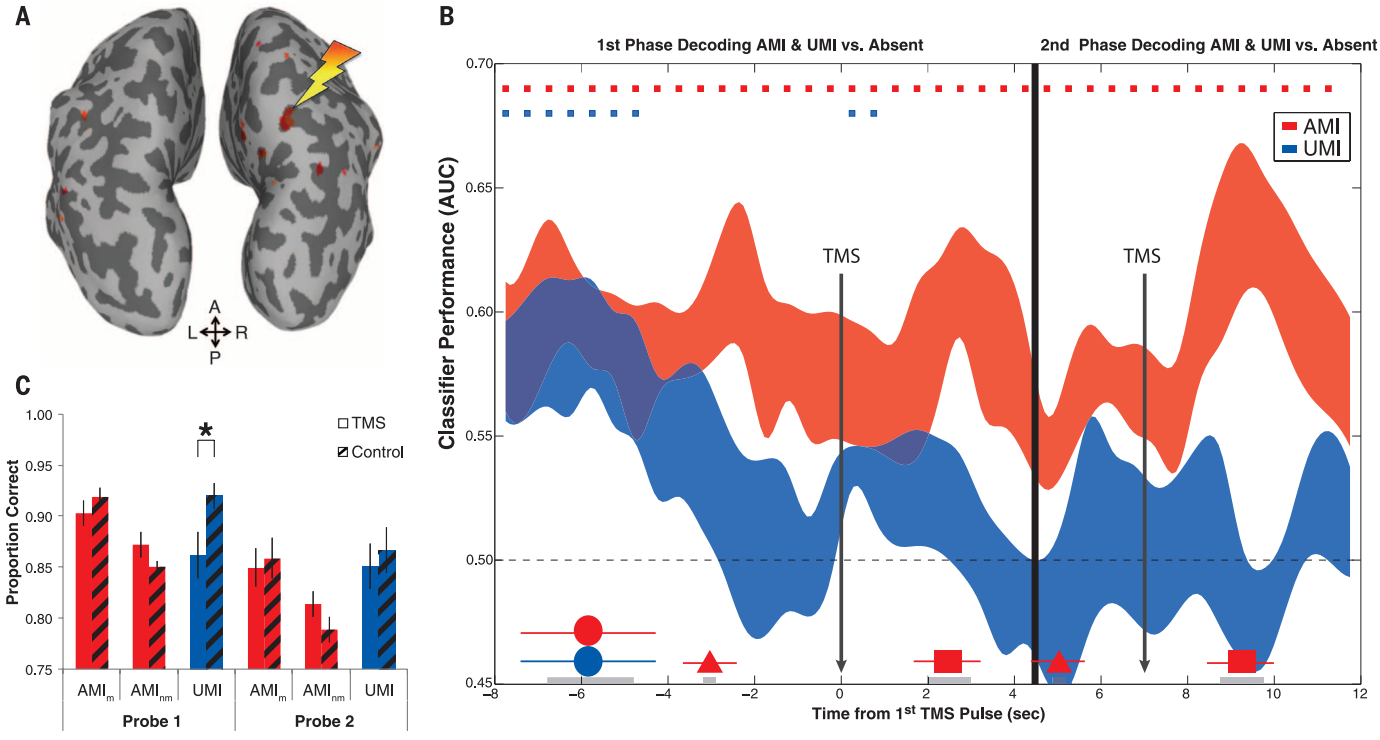


Fig. 4. Results from experiments 3 and 4. (**A**) The MVPA-defined TMS target for experiments 3 and 4 (right precuneus). A, anterior; L, left; R, right; P, posterior. (**B**) Classification time series from experiment 3 showing TMS reactivation of the UMI after the first cue, when the UMI was still relevant (left), but not after the second cue, when the UMI was no longer relevant on the trial (right) averaged over 1152 trials. Color-coded small squares at the top of each plot indicate $P < 0.05$; line width reflects SEM. (**C**) Experiment 4 recognition memory for AMI match probes (AMI_m), AMI nonmatch probes (AMI_{nm}), and UMI (nonmatch) probes. *$P < = 0.01$; error bars denote SEM.

provide neural evidence for at least two levels of WM that are distinct from the default state of LTM representations (*5*, *6*). They are inconsistent with models positing just one level of WM storage (*22*, *23*). They also suggest that instead of "activated LTM," a more apt label for the second level of WM would be "prioritized LTM." Information can be held in WM in latent "activity-silent" traces (*11*, *20*). What might be the physiological bases of such representations? Computational models of WM have proposed that short-term synaptic plasticity could be the basis for the transient formation of weight-based networks that can represent information over short time periods (*2*, *24*).

Our results provide empirical evidence for the existence of a short-term plasticity mechanism that is likely to be fundamental to a wide range of cognitive functions involving attentional selection (*25*) and may provide the building blocks for long-term potentiation mechanisms that support LTM (*26*). Therefore, our findings introduce a potential avenue for reactivating and strengthening representations that underlie many classes of high-level cognition.

REFERENCES AND NOTES

1. P. S. Goldman-Rakic, *Neuron* **14**, 477–485 (1995).
2. G. Mongillo, O. Barak, M. Tsodyks, *Science* **319**, 1543–1546 (2008).
3. Y. Sugase-Miyamoto, Z. Liu, M. C. Wiener, L. M. Optican, B. J. Richmond, *PLOS Comput. Biol.* **4**, e1000073 (2008).
4. M. J. Wolff, J. Ding, N. E. Myers, M. G. Stokes, *Front. Syst. Neurosci.* **9**, 123 (2015).
5. N. Cowan, in *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, A. Miyake, P. Shah, Eds. (Cambridge Univ. Press, 1999), pp. 62–101.
6. K. Oberauer, *Psychol. Learn. Motiv.* **51**, 45–100 (2009).
7. J. J. LaRocque, J. A. Lewis-Peacock, A. T. Drysdale, K. Oberauer, B. R. Postle, *J. Cogn. Neurosci.* **25**, 127–142 (2013).
8. J. A. Lewis-Peacock, A. T. Drysdale, K. Oberauer, B. R. Postle, *J. Cogn. Neurosci.* **24**, 61–79 (2012).
9. J. J. LaRocque, A. C. Riggall, S. M. Emrich, B. R. Postle, *Cereb. Cortex* 10.1093/cercor/bhw283 (2016).
10. E. F. Ester, D. E. Anderson, J. T. Serences, E. Awh, *J. Cogn. Neurosci.* **25**, 754–761 (2013).
11. A. C. Riggall, B. R. Postle, *J. Neurosci.* **32**, 12990–12998 (2012).
12. B. Y. Hayden, J. L. Gallant, *Front Neurosci* **7**, 18 (2013).
13. N. E. Myers *et al.*, *eLife* **4**, e09000 (2015).
14. See supplementary materials on *Science* Online.
15. B. Kundu, J. Y. Chang, B. R. Postle, B. D. Van Veen, *Neuroimage* **114**, 320–327 (2015).
16. T. G. Lee, M. D'Esposito, *J. Neurosci.* **32**, 15458–15466 (2012).
17. T. B. Christophel, M. N. Hebart, J. D. Haynes, *J. Neurosci.* **32**, 12983–12989 (2012).
18. S. Monsell, *Cognit. Psychol.* **10**, 465–501 (1978).
19. D. M. Beck, N. Muggleton, V. Walsh, N. Lavie, *Cereb. Cortex* **16**, 712–717 (2006).
20. M. G. Stokes, *Trends Cogn. Sci.* **19**, 394–405 (2015).
21. S. M. Polyn, V. S. Natu, J. D. Cohen, K. A. Norman, *Science* **310**, 1963–1966 (2005).
22. B. McElree, *Psychol. Learn. Motiv.* **46**, 155–200 (2006).
23. A. Baddeley, *Annu. Rev. Psychol.* **63**, 1–29 (2012).
24. V. Itskov, D. Hansel, M. Tsodyks, *Front. Comput. Neurosci.* **5**, 40 (2011).
25. R. Desimone, J. Duncan, *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
26. Y. Dudai, *Annu. Rev. Psychol.* **55**, 51–86 (2004).

ACKNOWLEDGMENTS

We thank J. Samaha, A. Sheldon, B. Kundu, J. Lewis-Peacock, and J. Johnson for assistance and helpful discussions. Supported by NIH grant MH095984 (B.R.P.). The data are stored at curate.nd.edu. Author contributions: N.S.R., B.R.P., J.J.L., and A.C.R. designed the research; N.S.R., O.G., M.J.S., and E.M.M. conducted the research; N.S.R., J.J.L., and A.C.R. analyzed the data; and N.S.R., B.R.P., and J.J.L. wrote the manuscript. The authors declare no conflict of interest.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1136/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Table S1
Movie S1
References (*27*–*38*)

3 August 2016; accepted 31 October 2016
10.1126/science.aah7011

NEUROSCIENCE

Selective modulation of cortical state during spatial attention

Tatiana A. Engel,[1,2]*† Nicholas A. Steinmetz,[3]* Marc A. Gieselmann,[4] Alexander Thiele,[4] Tirin Moore,[2,3] Kwabena Boahen[1]

Neocortical activity is permeated with endogenously generated fluctuations, but how these dynamics affect goal-directed behavior remains a mystery. We found that ensemble neural activity in primate visual cortex spontaneously fluctuated between phases of vigorous (On) and faint (Off) spiking synchronously across cortical layers. These On-Off dynamics, reflecting global changes in cortical state, were also modulated at a local scale during selective attention. Moreover, the momentary phase of local ensemble activity predicted behavioral performance. Our results show that cortical state is controlled locally within a cortical map according to cognitive demands and reveal the impact of these local changes in cortical state on goal-directed behavior.

Endogenous fluctuations in neocortical spiking activity vary on a continuum between synchronized and desynchronized states, and the level of synchrony has been associated with the overall level of arousal (*1*, *2*). During slow-wave sleep and anesthesia, ensemble neural activity exhibits slow synchronous transitions between periods of high activity and quiescence. In individual neurons, these transitions manifest as alternating Down (hyperpolarized) and Up (depolarized) phases of the membrane potential due to, respectively, an ebb and flow of synaptic activity (*3*–*7*). In awake animals, these slow synchronous transitions are less frequent, and thus, ensemble neural activity appears less synchronized than during anesthesia or slow-wave sleep (*8*–*12*). This relationship between arousal and cortical synchrony suggests that mechanisms controlling cortical state are brain-wide and unrelated to neural circuits involving the selective recruitment of local populations during goal-directed behavior. In particular, changes in cortical state should be orthogonal to the modulations of spiking activity observed locally within cortical maps during selective attention.

We discovered that spontaneous transitions between episodes of vigorous (On) and faint (Off) spiking occur synchronously across cortical layers in the visual cortex of behaving monkeys. We recorded ensemble spiking activity in area V4 of two rhesus macaques (G and B) with 16-channel linear array microelectrodes (Fig. 1C, left) arranged so that receptive fields (RFs) on all channels largely overlapped (Fig. 1A, fig. S1, and materials and methods 1). The On-Off transitions occurred synchronously throughout the cortical depth during fixation and in the absence of visual stimulation (Fig. 1, B and C). The On and Off episodes resembled the Up and Down phases commonly observed during anesthesia and slow-wave sleep (*1*–*3*) and were consistent with the large fluctuations in cortical membrane potentials recorded intracellularly in behaving monkeys (*13*).

To examine whether these On-Off fluctuations also occur during more demanding cognitive behaviors, we trained monkeys to perform a selective attention task. Monkeys were rewarded for detecting changes in a visual stimulus and indicating those changes with an antisaccade response (Fig. 1D). During each trial, a small central cue indicated the stimulus that was most likely to change orientation. The cued stimulus was thus the target of covert attention, whereas because of anticipation of antisaccadic response, the stimulus opposite to the cue was the target of overt attention (*14*). In spite of the difficulty of the task, monkeys performed well above chance, with 69 and 67% correct responses for monkeys G and B, respectively. While monkeys performed this task, we recorded from area V4 in 46 sessions (25 in monkey G and 21 in monkey B). As in the fixation task, we observed prominent On-Off transitions occurring synchronously across the cortical depth in both spontaneous and stimulus-driven activity, before and after the attention cue was presented, and evident in both single- and multiunit activity (Fig. 1E, fig. S2, supplementary text 3.1). On episodes reliably followed stimulus onset on a majority of trials. However, subsequent On-Off transitions occurred irregularly within and across trials during the sustained response to the stable RF stimulus. Transitions were also irregular with respect to the attention-cue's onset.

To characterize the On-Off dynamics, we counted spikes in 10-ms time bins and used a Hidden Markov Model (HMM) as a statistically principled way to segment spike-count data into On and Off episodes (Fig. 2A and materials and methods 2.2) (*15*, *16*). An HMM was fitted to 16-channel multiunit activity, but all analyses based on the fitted HMM were performed on both single- and multiunit activities, yielding highly consistent results. The HMM has a one-dimensional, latent variable representing an unobserved population state that switches between two phases, On and Off. Spikes on 16 recorded channels are assumed to be generated by inhomogeneous Poisson processes, with different mean rates during the On and Off phases. When an HMM is fitted to the spiking data, 34 parameters are estimated: firing rates in the On and Off phases for each of 16 channels and transition probabilities p_{on} and p_{off} for the entire ensemble (Fig. 2B). Using these parameters, we can then infer the most likely sequence of On and Off episodes that underlie the observed spike trains on a trial-by-trial basis (Fig. 2C). From visual inspection, On and Off episodes inferred by the HMM were closely aligned to the periods of vigorous and faint spiking. These On-Off transitions in spike rates were also phase-locked to low-frequency fluctuations in the local field potential (LFP) (fig. S3 and supplementary text 3.2). On average, the HMM captured about half of the maximal explainable variance in the data (fig. S4 and supplementary text 3.3). For most recording sessions (31 total, 67%), the two-phase HMM was the most parsimonious model among HMMs, with 1 or up to 8 possible phases (fig. S5). These 31 sessions were therefore used in subsequent analyses of On and Off episode durations. For the remaining 15 (33%) sessions, a one-phase HMM was the most parsimonious model. Consistent with HMM assumptions, the durations of On and Off episodes were distributed exponentially, with the decay time constants τ_{on} and τ_{off} given by the average On and Off episode durations (Fig. 2D and fig. S6A). The averages across sessions of these average On and Off episode durations were, respectively, τ_{on} = 149 ± 77 ms and τ_{off} = 102 ± 33 ms for stimulus-driven activity and τ_{on} = 97 ± 36 ms and τ_{off} = 118 ± 47 ms for spontaneous activity (mean ± SD across 31 sessions) (fig. S6, B and C). We also analyzed laminar recordings from area V4 performed with a different type of linear electrode array, in two different behaving monkeys, and in a different laboratory. In this additional data set, On-Off transitions also occurred synchronously across cortical layers during spontaneous and stimulus-driven activity and were equally well described by the two-phase HMM (Fig. 2, E to G, fig. S7, and supplementary text 3.4).

To investigate the extent to which On-Off dynamics reflect arousal or selective attention, we analyzed data from the two monkeys performing the attention task. In rodents, cortical state dynamics closely covary with global arousal, as measured by pupil size, with dilation characterized by desynchronization of neural activity and constriction by an increase in low-frequency fluctuations (*17*, *18*). Similarly, the pupil size was positively correlated with the average duration of On episodes on a trial-by-trial basis, during fixation and attention (fig. S8). Thus, the On-Off dynamics indeed reflected global changes of cortical state associated with arousal. Spatial attention, on the other hand, involves the selective

[1]Departments of Bioengineering and Electrical Engineering, Stanford University, Stanford, CA, USA. [2]Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA. [3]Department of Neurobiology, Stanford University, Stanford, CA, USA. [4]Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK.
*These authors contributed equally to this work. †Corresponding author Email: tatiana.engel@stanford.edu

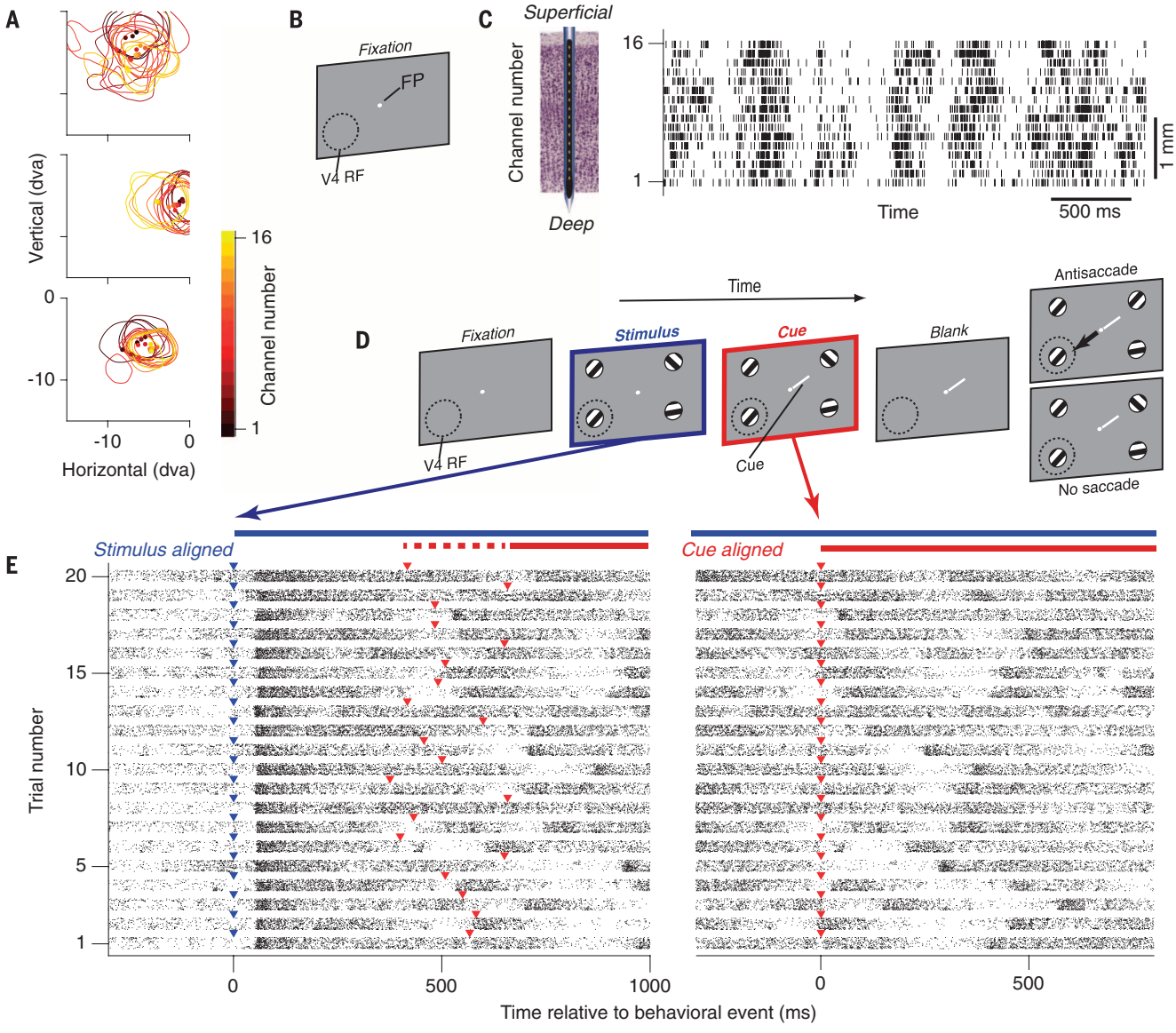


Fig. 1. Spontaneous On and Off transitions in spiking activity during fixation and attention tasks. (**A**) Overlap of V4 receptive fields measured simultaneously across the cortical depth on 16 channels (lines, RF contours; dots, RF centers) for three example recordings (dva, degrees of visual angle). (**B**) Fixation task. Monkeys fixated a central dot (FP) on a blank screen for 3 s on each trial; the dashed circle outlines approximate V4 receptive field locations (V4 RF). (**C**) An example trial showing spontaneous transitions between episodes of vigorous (On) and faint (Off) spiking in multiunit activity simultaneously recorded with 16-channel electrodes (left). Spikes are marked by vertical ticks. (**D**) Attention task. Monkeys reported orientation changes with an antisaccade. A cue indicated which stimulus was likely to change. Monkeys initiated a trial by fixating a central dot (Fixation). After a brief delay (333 ms and 170 ms in monkeys G and B, respectively), four peripheral oriented-grating stimuli appeared, one in each of the screen's quadrants (Stimulus). After a variable delay (200 to 2 700 ms), stimuli briefly disappeared (Blank, < 270 ms) then reappeared either with or without one of them changing orientation. Monkeys reported an orientation change by executing a saccade to the stimulus diametrically opposite to the change location (Antisaccade; arrow indicates saccade direction). If no change happened, monkeys had to maintain fixation (No saccade). A small, central cue (white line; illustrated larger than actual size) appeared shortly (200 to 500 ms) after stimulus onset (Cue), pointing toward the stimulus that was most likely to change. (**E**) On and Off transitions in multiunit spiking activity on 16 simultaneously recorded channels (each horizontal band) for 20 example trials. Activity is aligned to the stimulus' (left; blue triangles) and attention cue's onset times (right; red triangles).

recruitment of local neuronal populations encoding behaviorally relevant stimuli at one retinotopic location and the simultaneous suppression of populations encoding irrelevant stimuli in other retinotopic locations (*19*–*21*). We used three behavioral conditions to measure the local effects of both covert and overt attention on neuronal activity. In the covert attention condition, the cue directed the animal's attention to the RF stimulus. In the overt attention condition, the cue directed attention to the stimulus opposite the RF and indicated that a saccade to the RF stimulus was likely to be required. In the control condition, the cue directed covert and overt attention to stimuli in directions orthogonal to the RF direction. The overall mean firing rate of V4 neurons was greater in covert and overt attention conditions relative to the control condition (Fig. 3A and fig. S9B), as has been previously reported (*14*). If these local effects of selective attention are indeed orthogonal to

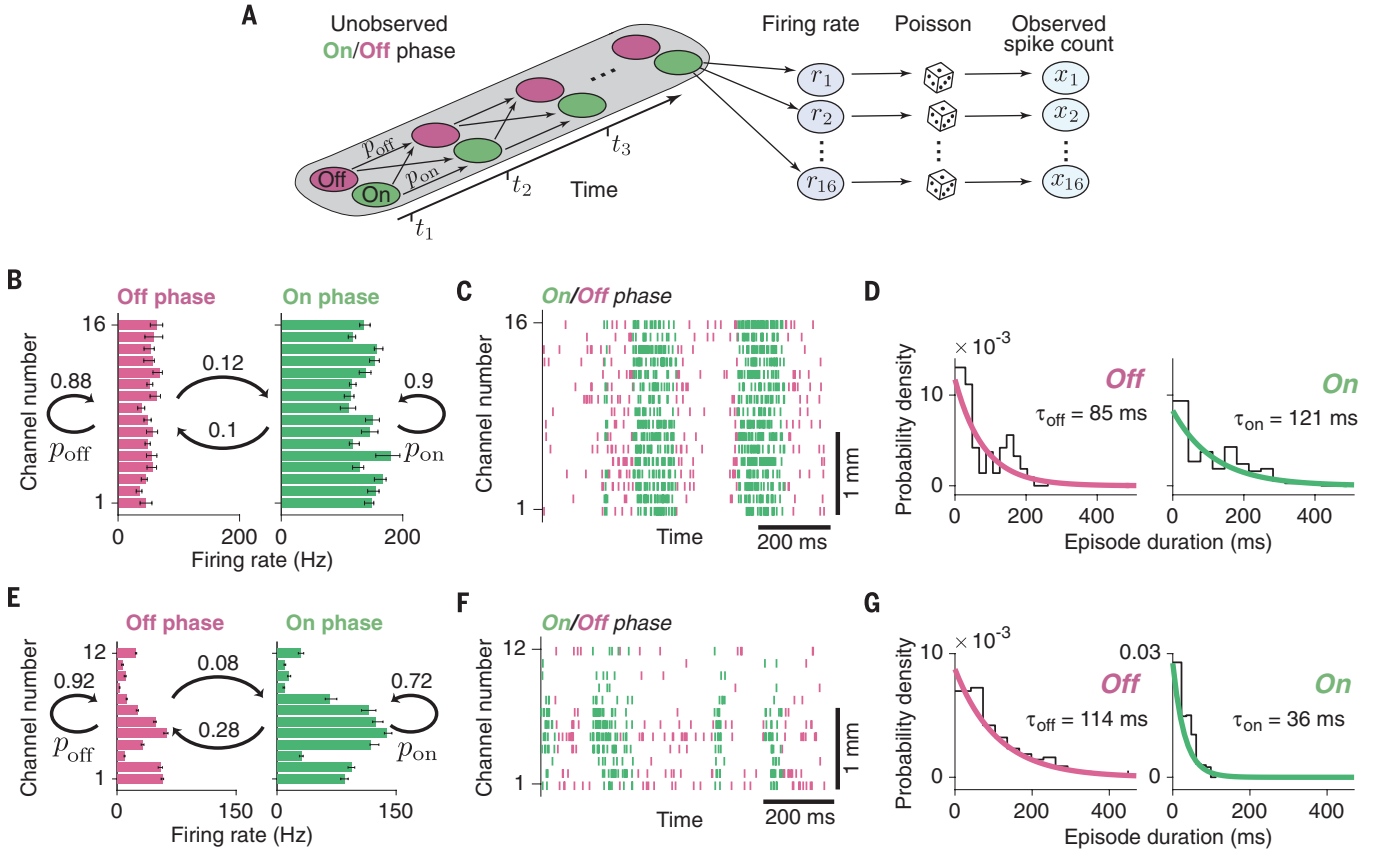


Fig. 2. HMM of On-Off dynamics. (**A**) HMM schematic. In each time bin (t_1, t_2, t_3, ...), spike counts x_1, ...x_{16} on 16 channels are generated through inhomogeneous Poisson processes with mean rates r_1, ...r_{16} that are different between the On and Off phase. Transitions between unobserved On and Off phases are governed by probabilities p_{on} and p_{off}. (**B**) Example HMM fit. Firing rates of 16 channels in the On and Off phase and transition probabilities (numbers above the curved arrows) estimated by the model. Error bars are 5th and 95th percentiles over 10 bootstrap samples. The HMM was fitted to multiunit activity within the time-window beginning 400 ms after the attention-cue onset and ending at the start of the blank period (Fig. 1D). (**C**) Example epoch of spiking activity segmented into On (green) and Off (pink) episodes by the HMM. (**D**) Distributions (black lines) of On (right) and Off (left) episode durations overlaid by exponential distributions (green and pink lines), with the decay time-constants set by HMM transition probabilities. (**E** to **G**) Same as (B)to (D), respectively, but for the additional laminar data set.

global state changes, then we should not expect On-Off dynamics to be modulated by attention.

We considered three ways in which an attention-induced increase in mean firing rate could co-occur with the On-Off dynamics (Fig. 3B). First, mean firing rate could be enhanced during On phase, Off phase, or both, without any difference in the transition dynamics across attention conditions. This outcome would suggest that the On-Off dynamics are not selective for local neuronal populations but reflect a global state. Second, the enhancement in the mean firing rate could arise entirely from an increase in the duration of On episodes, or a decrease in the duration of Off episodes, or both, but without any change in the firing rates during On or Off phases. Third, a combination of the first two scenarios is also possible: The firing rates and durations of the On and Off episodes could both be modulated. The last two outcomes would both indicate that On-Off dynamics are locally and selectively modulated within confined retinotopic regions and do not solely reflect a global arousal state.

The On-Off dynamics were modulated by attention consistent with the third scenario. Firing rates during the On and Off phases were slightly, but significantly, enhanced during both types of attention [Wilcoxon signed rank test; covert, On-phase median modulation index (MI) = 0.008, P = 0.002, Off-phase median MI = 0.005, P = 0.046; overt, On-phase median MI = 0.014, $P < 10^{-5}$, Off-phase median MI = 0.029, $P < 10^{-10}$] (Fig. 3C and materials and methods 2.2.5). The average duration of Off episodes was significantly longer in the covert, but not the overt, attention condition compared with control conditions (Wilcoxon signed rank test; covert, median change in duration 4 ms, P = 0.004; overt, median change in duration −2 ms, P = 0.652) (Fig. 3D). However, the average duration of On episodes was significantly longer during both covert and overt attention as compared with controls (Wilcoxon signed rank test; covert, median change in duration 7 ms, $P < 10^{-3}$; overt, median change in duration 13 ms, $P < 10^{-4}$) (Fig. 3D). Correspondingly, the On-Off transition frequency was significantly lower during covert and overt attention compared with control conditions (Wilcoxon signed rank test; covert, median reduction in frequency 0.2 Hz, P = 0.001; overt, median reduction in frequency 0.2 Hz, $P < 10^{-3}$) (fig. S9A). In separate control analyses, we confirmed that the changes in On-Off dynamics were not an artifact of the attention-related increase in firing rates, or of the HMM's assumption of discrete On-Off phases (figs. S9 and S10 and supplementary text 3.5 and 3.6). We also considered the influence of microsaccades. Unlike stimulus onset, which was reliably followed by On-episodes, only a small fraction of transitions was preceded by a microsaccade (fig. S11, A to D). Nonetheless, changes in frequency or direction of microsaccades could not account for observed increases in On-episode durations (fig. S11, E to G). Moreover, increases in On-episode durations were also observed on trials without microsaccades (fig. S11H and supplementary text 3.7).

Last, we asked whether On-Off dynamics, in addition to being modulated by attention, predicted behavioral performance. In our task, the probability of detecting a change was greater at

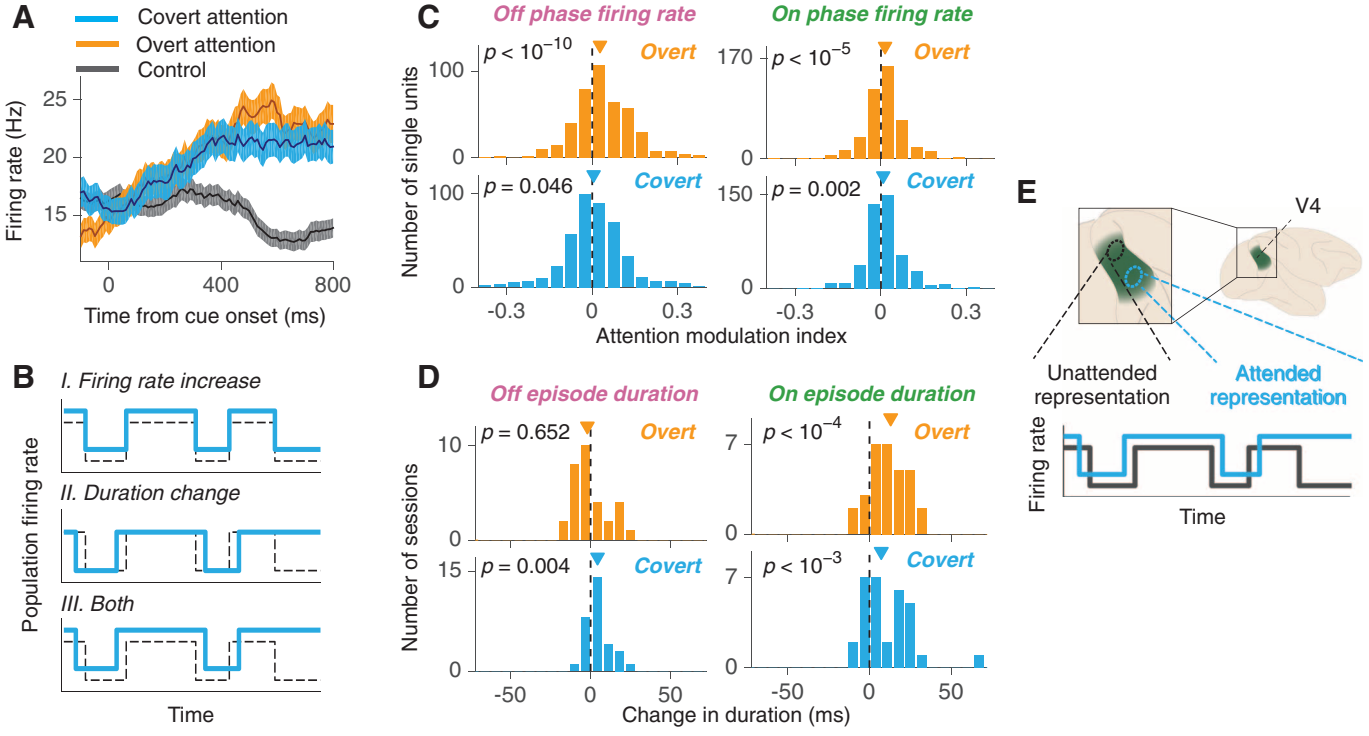


Fig. 3. Effects of covert and overt attention on dynamics of On-Off transitions. (**A**) Average firing rates of an example single-unit for covert (blue), overt (orange), and control trials (gray). Shaded area represents ± SEM. (**B**) Three hypothetical ways On-Off dynamics can interact with attention-related increase in average firing rate (blue line, attention; black dashed line, control). (**C**) Distribution of attention modulation index of single-unit firing rates in the On (right) and Off (left) phases during covert (blue) and overt attention (orange). (**D**) Distribution across recordings of the difference in average durations of On (right) and Off (left) episodes between covert attention and control (blue), and overt attention and control (orange) conditions. In (C) and (D), triangles indicate medians of the distributions; *P* values are for Wilcoxon signed rank test. (**E**) Cartoon of two sites in area V4 corresponding to attended (blue ellipse) and unattended (black ellipse) retinotopic locations (top) with corresponding On-Off dynamics (bottom).

the cued location as compared with uncued locations (*14*). We investigated how the increase in detection probability at the cued location was related to On-Off dynamics (materials and methods 2.2.6). When the cued orientation change occurred in the RFs of recorded neurons, the probability of detecting that change was significantly greater when it occurred during an On-phase than during an Off-phase (median detection probability 64.8% during Off-phase, 78.3% during On-phase, difference 13.5%, $P < 10^{-3}$, Wilcoxon signed rank test) (Fig. 4A). This difference in detection probability was evident beginning ~150 ms before the stimulus change (Fig. 4B), which is consistent with the average duration of On episodes. This effect was spatially selective; we found no difference in detection probability between On and Off phases when the cued change occurred outside the RFs of recorded neurons (median detection probability difference 0.4%, $P = 0.943$, Wilcoxon signed rank test) (Fig. 4A).

Spontaneous On-Off transitions, occurring synchronously throughout the cortical depth, were modulated locally within a cortical map during selective attention and predicted behavioral performance. These On-Off dynamics represent a substantial source of correlated variability classically observed in cortical responses (*22*), and many features of this correlated variability,

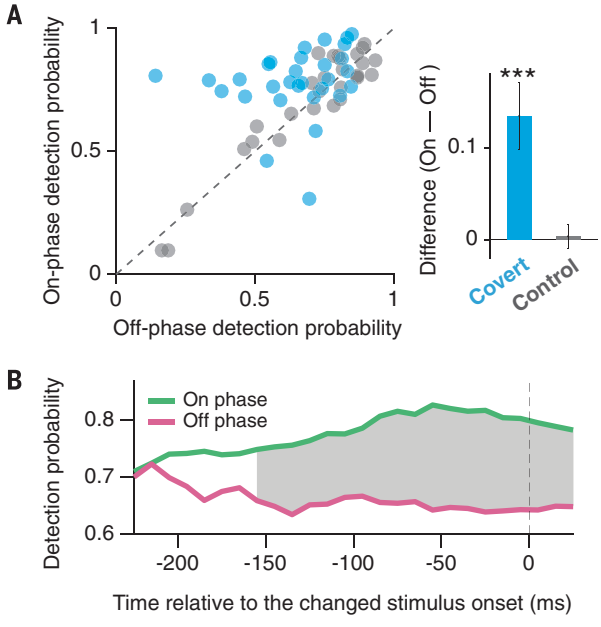


Fig. 4. On-Off population state predicts behavioral performance. (**A**) Probability to detect an orientation change that occurred during the Off phase (*x* axis) versus a change that occurred during the On phase (*y* axis). Each point represents one recording session (blue, covert attention; gray, control condition). (Inset) Average difference between the On and Off detection probability. Error bars represent SD across recordings; asterisks indicate $P < 10^{-3}$ for Wilcoxon signed rank test. (**B**) Time course of the detection probability. At each time bin, the detection probability was calculated separately for trials on which the instantaneous population state was in the On phase (green line) and on which it was in the Off phase (pink line) at the corresponding time bin. Gray shading indicates significant difference in detection probability (two-sided paired *t* test, $P < 0.05$ corrected for multiple comparisons across all time bins).

such as spike-count correlations (*23*), can be understood as arising from the On-Off dynamics (fig. S12 and supplementary text 3.8). Correlated variability can be affected by cognitive factors (*24*–*26*). In particular, spike-count correlations can increase or decrease during selective attention (*27*–*30*), and changes in the On-Off dynamics account for changes in spike-count correlations during attention in our data (fig. S13 and supplementary text 3.8.4). Recent models parsimoniously attribute changes in spike-count correlations during attention to fluctuations in shared modulatory signals (*31*), with smaller spike-count correlations accounted for by reduced fluctuations in these modulatory signals (*32*). The On-Off dynamics observed here could underlie the apparent trial-to-trial fluctuations in shared modulatory signals (*32*, *33*) but can account for within-trial fluctuations as well (fig. S12 and supplementary text 3.8.5).

What mechanisms underlie the spatially and temporally precise control of cortical state during selective attention? Our results suggest that global mechanisms governing cortical states may themselves also operate on a local scale or, alternatively, may interact with separate attentional control mechanisms operating locally. Indeed, neuromodulators known to act on a brain-wide scale (*1*, *34*, *35*) also mediate the effects of selective attention (*36*) and influence circuits that control selective attention (*37*). On the other hand, cortico-cortical inputs appear to influence state changes in a spatially targeted manner (*38*, *39*). Because diffuse neuromodulatory signals are interspersed with topographically precise projections throughout the cortex, local modulation of cortical state is likely to be widespread, extending to modalities beyond vision and serving many cognitive functions.

REFERENCES AND NOTES

1. K. D. Harris, A. Thiele, *Nat. Rev. Neurosci.* **12**, 509–523 (2011).
2. S.-H. Lee, Y. Dan, *Neuron* **76**, 209–222 (2012).
3. M. Steriade, D. A. McCormick, T. J. Sejnowski, *Science* **262**, 679–685 (1993).
4. M. Steriade, I. Timofeev, F. Grenier, *J. Neurophysiol.* **85**, 1969–1985 (2001).
5. B. Haider, A. Duque, A. R. Hasenstaub, Y. Yu, D. A. McCormick, *J. Neurophysiol.* **97**, 4186–4202 (2007).
6. A. Hasenstaub, R. N. S. Sachdev, D. A. McCormick, *J. Neurosci.* **27**, 9607–9622 (2007).
7. A. Renart *et al.*, *Science* **327**, 587–590 (2010).
8. S. Crochet, C. C. H. Petersen, *Nat. Neurosci.* **9**, 608–610 (2006).
9. J. F. A. Poulet, C. C. H. Petersen, *Nature* **454**, 881–885 (2008).
10. C. M. Niell, M. P. Stryker, *Neuron* **65**, 472–479 (2010).
11. M. Okun, A. Naim, I. Lampl, *J. Neurosci.* **30**, 4440–4448 (2010).
12. A. Luczak, P. Bartho, K. D. Harris, *J. Neurosci.* **33**, 1684–1695 (2013).
13. A. Y. Y. Tan, Y. Chen, B. Scholl, E. Seidemann, N. J. Priebe, *Nature* **509**, 226–229 (2014).
14. N. A. Steinmetz, T. Moore, *Neuron* **83**, 496–506 (2014).
15. E. Seidemann, I. Meilijson, M. Abeles, H. Bergman, E. Vaadia, *J. Neurosci.* **16**, 752–768 (1996).
16. G. Rainer, E. K. Miller, *Neurocomputing* **32-33**, 961–966 (2000).
17. M. Vinck, R. Batista-Brito, U. Knoblich, J. A. Cardin, *Neuron* **86**, 740–754 (2015).
18. J. Reimer *et al.*, *Neuron* **84**, 355–362 (2014).
19. J. Moran, R. Desimone, *Science* **229**, 782–784 (1985).
20. S. Treue, J. H. Maunsell, *Nature* **382**, 539–541 (1996).
21. P. Fries, J. H. Reynolds, A. E. Rorie, R. Desimone, *Science* **291**, 1560–1563 (2001).
22. A. Renart, C. K. Machens, *Curr. Opin. Neurobiol.* **25**, 211–220 (2014).
23. M. R. Cohen, A. Kohn, *Nat. Neurosci.* **14**, 811–819 (2011).
24. M. R. Cohen, W. T. Newsome, *Neuron* **60**, 162–173 (2008).
25. Y. Gu *et al.*, *Neuron* **71**, 750–761 (2011).
26. A. S. Ecker *et al.*, *Neuron* **82**, 235–248 (2014).
27. J. F. Mitchell, K. A. Sundberg, J. H. Reynolds, *Neuron* **63**, 879–888 (2009).
28. M. R. Cohen, J. H. R. Maunsell, *Nat. Neurosci.* **12**, 1594–1600 (2009).
29. D. A. Ruff, M. R. Cohen, *Nat. Neurosci.* **17**, 1591–1597 (2014).
30. D. A. Ruff, M. R. Cohen, *J. Neurosci.* **36**, 7523–7534 (2016).
31. R. L. T. Goris, J. A. Movshon, E. P. Simoncelli, *Nat. Neurosci.* **17**, 858–865 (2014).
32. N. C. Rabinowitz, R. L. Goris, M. Cohen, E. P. Simoncelli, *eLife* **4**, e08998 (2015).
33. M. R. Cohen, J. H. R. Maunsell, *J. Neurosci.* **30**, 15241–15253 (2010).
34. G. Aston-Jones, J. D. Cohen, *Annu. Rev. Neurosci.* **28**, 403–450 (2005).
35. T. W. Robbins, A. F. T. Arnsten, *Annu. Rev. Neurosci.* **32**, 267–287 (2009).
36. J. L. Herrero *et al.*, *Nature* **454**, 1110–1114 (2008).
37. B. Noudoost, T. Moore, *Nature* **474**, 372–375 (2011).
38. E. Zagha, A. E. Casale, R. N. S. Sachdev, M. J. McGinley, D. A. McCormick, *Neuron* **79**, 567–578 (2013).
39. G. G. Gregoriou, S. J. Gotts, H. Zhou, R. Desimone, *Science* **324**, 1207–1210 (2009).

ACKNOWLEDGMENTS

This work was supported by NIH grants EY014924 and NS076460, a Stanford NeuroVentures grant, Medical Research Council (MRC) grant MR/K013785/1, and Wellcome Trust grant 093104. We thank E. I. Knudsen, K. Harris, S. Ganguli, R. N. S. Sachdev, and M. Zirnsak for their comments on the manuscript. We thank D. S. Aldrich for technical assistance. All behavioral and electrophysiological data are presented in (*14*) and are archived at the Stanford Neuroscience Institute server at Stanford University. T.A.E., N.A.S., T.M., and K.B. designed the study. N.A.S. and T.M. designed the experiments. N.A.S. performed experiments, spike sorting, microsaccade detection, and RF measurements. M.A.G. and A.T. performed experiments for the additional laminar data set. T.A.E. analyzed and modeled the data. T.A.E., N.A.S., T.M., and K.B. discussed the findings and wrote the paper.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1140/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S13
Table S1
References (*40*–*58*)

16 May 2016; accepted 31 October 2016
10.1126/science.aag1420

PAIN RESEARCH

Gliogenic LTP spreads widely in nociceptive pathways

M. T. Kronschläger,* R. Drdla-Schutting,* M. Gassner, S. D. Honsek, H. L. Teuchmann, J. Sandkühler†

Learning and memory formation involve long-term potentiation (LTP) of synaptic strength. A fundamental feature of LTP induction in the brain is the need for coincident pre- and postsynaptic activity. This restricts LTP expression to activated synapses only (homosynaptic LTP) and leads to its input specificity. In the spinal cord, we discovered a fundamentally different form of LTP that is induced by glial cell activation and mediated by diffusible, extracellular messengers, including D-serine and tumor necrosis factor (TNF), and that travel long distances via the cerebrospinal fluid, thereby affecting susceptible synapses at remote sites. The properties of this gliogenic LTP resolve unexplained findings of memory traces in nociceptive pathways and may underlie forms of widespread pain hypersensitivity.

Activity-dependent, homosynaptic long-term potentiation (LTP) (*1*) at synapses in nociceptive pathways contributes to pain amplification (hyperalgesia) at the site of an injury or inflammation (*2*–*5*). Homosynaptic LTP can, however, not account for pain amplification at areas surrounding (secondary hyperalgesia) or remote from (widespread hyperalgesia) an injury. It also fails to explain hyperalgesia that is induced independently of neuronal activity in primary afferents—e.g., by the application of or the withdrawal from opioids (opioid-induced hyperalgesia) (*6*). Glial cells are believed to contribute to these forms of hyperalgesia and to LTP in nociceptive pathways (*7*–*10*). Induction of homosynaptic LTP can be accompanied by LTP in adjacent, inactive synapses converging onto the same neuron, especially early in development. The respective molecular signals for this heterosynaptic form of LTP are thought to be confined within the cytoplasm of the activated neuron, spreading tens of micrometers only (*11*). We have now tested the hypothesis that, in contrast to current beliefs, activation of glial cells is causative for the induction of LTP at spinal C-fiber synapses and that this gliogenic LTP constitutes a common denominator of homo- and heterosynaptic LTP in the spinal cord.

Our previous study revealed that selective activation of spinal microglia by fractalkine induces

Department of Neurophysiology, Center for Brain Research, Medical University of Vienna, Spitalgasse 4, 1090 Vienna, Austria.
*These authors contributed equally to this work. **†Corresponding author. Email: juergen.sandkuehler@meduniwien.ac.at**

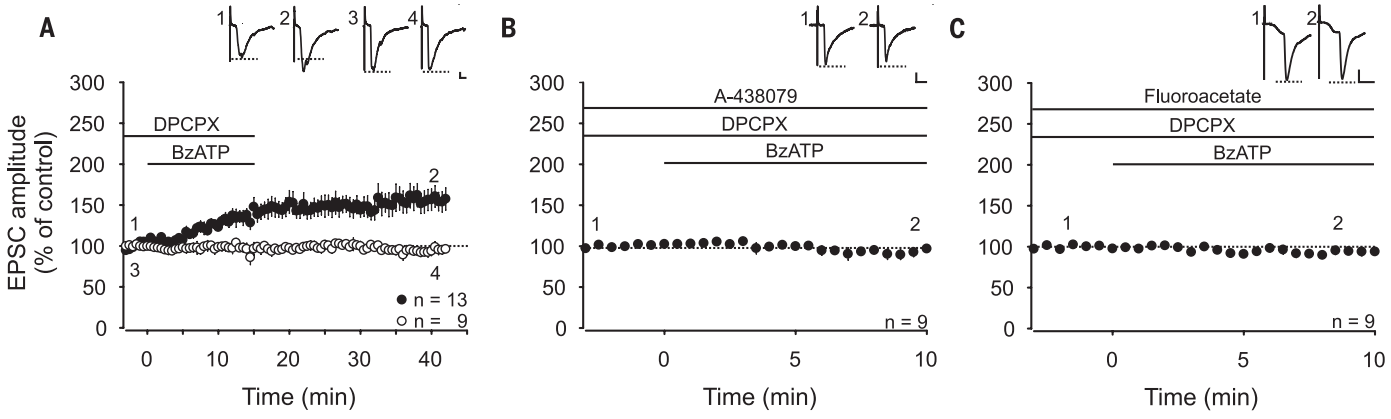


Fig. 1. Activation of spinal P2X$_7$ receptors induces gliogenic LTP at C-fiber synapses. Recordings were performed on lamina I neurons with independent monosynaptic C-fiber inputs from two dorsal root halves. Amplitudes of EPSCs were normalized to six baseline values, and the mean (±1 SEM) was plotted against time (min). Horizontal bars indicate drug application. (**A**) DPCPX (1 μM) application started at time point −3 min. Bath application of BzATP (100 μM) started at time point 0 min and induced LTP at 13 out of 22 C-fiber inputs (filled circles) ($P < 0.001$, at 30 min of wash-out compared with control values). At 9 out of 22 C-fiber inputs, BzATP did not influence EPSC amplitudes (open circles) ($P = 0.650$, at 30 min of wash-out compared with control values). (**B**) Bath application of the P2X$_7$R antagonist A-438079 (10 μM) 13 min before BzATP prevented the BzATP-induced LTP at all C-fiber inputs tested ($n = 9$, $P = 0.054$, at 10 min compared with baseline). (**C**) In the presence of fluoroacetate (10 μM), BzATP had no effect on synaptic transmission ($n = 9$, $P = 0.114$ at 10 min compared to baseline). Insets show individual EPSCs at indicated time points. Calibration bars indicate 50 pA and 10 ms. Statistical significance was determined by using repeated measures analysis of variance (RM ANOVA) followed by Bonferroni *t* test. Paired *t* test was used for control recordings.

transient facilitation, but no LTP, at C-fiber synapses (*12*). Here, we recorded monosynaptic C-fiber–evoked excitatory postsynaptic currents (EPSCs) from lamina I neurons in rat lumbar spinal cord slices. To investigate whether selective activation of spinal astrocytes is sufficient for the induction of synaptic plasticity in the absence of any other conditioning stimulus, we used ultraviolet (UV)–flash photolysis of caged inositol 1,4,5-trisphosphate (IP$_3$) in astrocytic networks (fig. S1 and movie S1). This induced a robust long-term depression at C-fiber synapses (gliogenic LTD; to 69 ± 9%, $n = 7$, $P < 0.001$) (fig. S1C) but no LTP. UV flashes were without any effect on synaptic strength when applied in the absence of caged IP$_3$ (fig. S1D) or in the presence of the glial cell toxin fluoroacetate (fig. S1E). To coactivate microglia and astrocytes, we next applied the purinergic P2X$_7$ receptor (P2X$_7$R) agonist benzoyl-benzoyl adenosine triphosphate (BzATP). This never affected holding currents or membrane potentials in any of the spinal neurons tested (fig. S2), supporting the observation that, in the spinal dorsal horn, and unlike other P2X receptors (*13*), P2X$_7$Rs are expressed exclusively on glial cells (*14*–*18*). ATP is finally hydrolyzed to adenosine. We therefore applied the adenosine 1 receptor antagonist DPCPX to block adenosine-mediated presynaptic inhibition (fig. S3). Combined activation of microglia and astrocytes by BzATP induced LTP in 13 out of 22 C-fiber inputs (to 156 ± 13%, $P < 0.001$) (Fig. 1A). BzATP-induced LTP was abolished by the selective P2X$_7$R antagonist A-438079 (Fig. 1B) and by fluoroacetate (Fig. 1C). This demonstrates that selective activation of P2X$_7$R on spinal glial cells caused gliogenic LTP at synapses between C fibers and lamina I neurons.

High-frequency stimulation (HFS) of primary afferent C fibers triggers the release of ATP from primary afferent neurons (*19*, *20*), activates glial cells (*21*, *22*), and induces LTP (*2*, *3*), leading to the intriguing hypothesis that HFS-induced LTP at spinal C-fiber synapses might be a variety of gliogenic LTP. If true, one would predict that HFS induces LTP not only at conditioned but also at unconditioned C-fiber synapses and that, in striking contrast to current beliefs, homo- and heterosynaptic LTP could be expressed independently of each other. To directly test these predictions, we used transverse lumbar spinal cord slices with long dorsal roots attached that were separated into halves. We recorded from 22 dorsal horn lamina I neurons that received independent monosynaptic C-fiber input from each dorsal root half. HFS applied to one dorsal root half induced LTP in the conditioned pathway in 12 of these neurons ("homosynaptic LTP"; to 134 ± 9%, $P < 0.001$) (Fig. 2Aa, red filled circles). Out of these 12 neurons, where homosynaptic LTP was induced, 6 also showed LTP at the unconditioned pathway ("heterosynaptic LTP"). In total, heterosynaptic LTP was induced in 11 out of 22 neurons (to 174 ± 19%, $P < 0.001$) (Fig. 2Ba, blue filled circles) because, importantly, in 5 of these neurons, heterosynaptic LTP was induced in the absence of homosynaptic LTP (to 161 ± 9%, $P < 0.005$) (Fig. 2C), a finding that cannot be explained by current models of synaptic plasticity.

We investigated whether HFS-induced homo- and heterosynaptic LTP require activation of glial cells via P2X$_7$R. Blockade of glial P2X$_7$R by A-438079 fully blocked LTP induction at the conditioned and at the unconditioned sites (Fig. 2, Ab and Bb). This was also achieved by blocking glial cell metabolism with fluoroacetate [Fig. 2, Ac and Bc and (*21*)]. Both homo- and heterosynaptic LTP were abolished by blocking postsynaptic *N*-methyl-D-aspartate receptors (NMDARs) (Fig. 2, Ad and Bd). D-serine is a coagonist at NMDARs that is released from astrocytes (*23*). Here, preincubation of slices with the D-serine–degrading enzyme D-amino acid oxidase (DAAO) abolished both homo- and heterosynaptic LTP (Fig. 2, Ae and Be). We then investigated whether D-serine alone is sufficient to enhance synaptic strength at C-fiber synapses. Bath application of D-serine facilitated synaptic strength at C-fiber synapses (to 120 ± 2% in 13 out of 32 cells; $P < 0.001$) (fig. S4A). This amplification was abolished by blockade of NMDARs (in 12 out of 13 cells; $P = 0.094$) (fig. S4B). Taken together, our data demonstrate that the combined activation of microglia and astrocytes, either via P2X$_7$R or by HFS, was sufficient to induce gliogenic LTP. When gliogenic LTP is induced by conditioning HFS, it may appear as homo- and/or heterosynaptic LTP that can be elicited independently of each other.

We next asked whether gliogenic LTP also exists in vivo. HFS applied to the sciatic nerve induced LTP of spinal C-fiber–evoked field potentials in deeply anesthetized rats (to 211 ± 16% at 220 to 240 min; $n = 49$, $P < 0.001$) (Fig. 3A). HFS-induced LTP was blocked by spinal application of either fluoroacetate (Fig. 3B) or DAAO (Fig. 3C), indicating that it required the activation of spinal glial cells and D-serine signaling. Application of fluoroacetate or DAAO after the induction of LTP had no effects on LTP maintenance (to 192 ± 23% and to 181 ± 30%, respectively, at 220 to 240 min; $n = 6$; $P = 0.433$ and 0.546, respectively) (fig. S5), indicating that once LTP was induced, glial cells were no longer required. Thus, the gliogenic nature refers to the induction but not to the maintenance phase of LTP.

We then investigated whether HFS leads to the release of diffusible mediators that spread over long distances to trigger LTP. We induced LTP by HFS, collected the spinal superfusate from the respective lumbar segments, and transferred

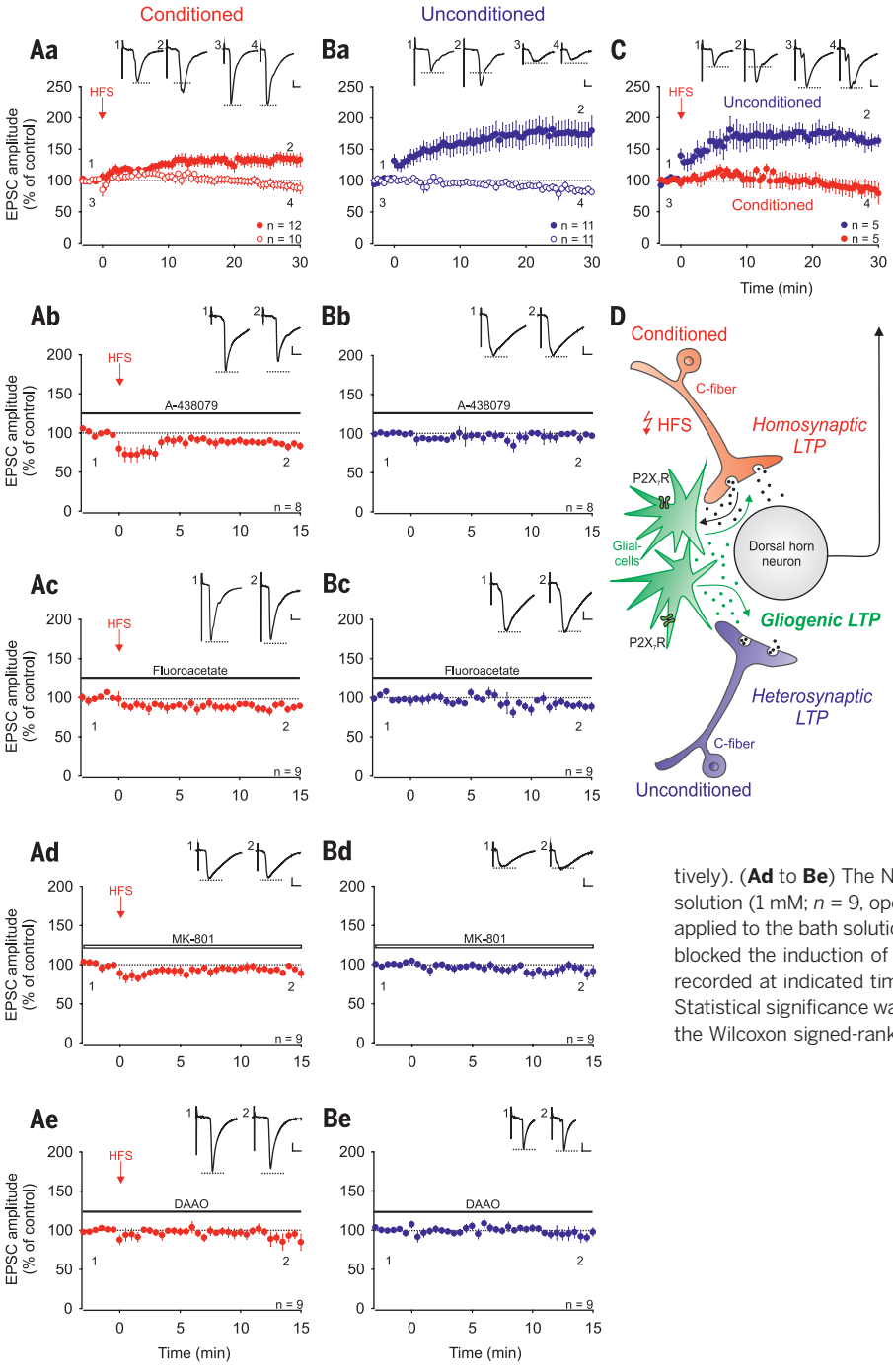


Fig. 2. Homo- and heterosynaptic forms of LTP are induced independently of each other at C-fiber synapses by conditioning HFS. Recordings were performed on lamina I neurons with independent monosynaptic C-fiber inputs from two dorsal root halves. Amplitudes of EPSCs were normalized to six baseline values and the mean (±1 SEM) was plotted against time (min). HFS was applied to one dorsal root (arrow; conditioned site in red) at time point 0 min. Horizontal bars indicate drug application. (**Aa**) HFS induced LTP at conditioned synapses in 12 out of 22 neurons (homosynaptic LTP in red, filled circles; $P < 0.001$, at 30 min compared with control values). In 10 of these neurons, no homosynaptic LTP was induced (open circles; $P = 0.105$). (**Ba**) HFS induced LTP at unconditioned synapses in 11 out of the same 22 neurons tested (heterosynaptic LTP in blue, filled circles; $P < 0.001$, at 30 min compared with control values). In 11 of these neurons, no heterosynaptic LTP was observed (open circles; $P = 0.003$). (**C**) In 5 out of these 22 neurons tested, HFS induced LTP at unconditioned (filled circles in blue; 161 ± 10%, $P = 0.005$) but not at conditioned synapses (filled circles in red; $P = 0.313$). (**D**) Schematic illustration of homo- and heterosynaptic forms of LTP as varieties of gliogenic LTP. (**Ab** and **Bb**) HFS failed to induce LTP at the conditioned site in the presence of A-438079 (10 μM; $n = 8$, $P = 0.006$). A-438079 had no effect on EPSC amplitudes at unconditioned synapses. (**Ac** and **Bc**) In the presence of fluoroacetate, LTP induction by HFS was abolished at conditioned and at unconditioned sites (10 μM; $n = 9$, $P = 0.006$ and $P = 0.034$, respectively). (**Ad** to **Be**) The NMDAR blocker MK-801, which was added to the pipette solution (1 mM; $n = 9$, open bar; $P = 0.044$ and $P = 0.250$, respectively) or DAAO applied to the bath solution (0.2 U·ml^{-1}; $n = 9$, $P = 0.006$ and 0.572, respectively) blocked the induction of LTP on both sites. Insets show individual EPSC traces recorded at indicated time points. Calibration bars indicate 100 pA and 10 ms. Statistical significance was determined by paired *t* test. In case of non-normality, the Wilcoxon signed-rank test was used.

it to the spinal cord dorsum of naïve animals. The maintenance of LTP in the donor animals was not affected by exchanging the superfusate (Fig. 3A). The superfusate induced, however, a robust LTP in the recipient animals (to 173 ± 32% of control at 160 to 180 min; $n = 10$, $P = 0.009$) (Fig. 4A), demonstrating that LTP could be transferred between individuals. The superfusate collected from naïve donor animals had, in contrast, no effect on synaptic transmission in any of the recipient animals (Fig. 4B). When glial cells were blocked in the recipient animals, "transferable LTP" was still induced (to 160 ± 20%; $n = 9$, $P < 0.001$) (Fig. 4C). Blockade of interleukin-1β (IL-1β) signaling in the recipient animals also had no effect on the induction of transferable LTP (to 133 ± 12% at 180 to 240 min; $n = 10$, $P = 0.001$) (Fig. 4D). However, LTP induction was prevented by blocking TNF (Fig. 4E), D-serine signaling (Fig. 4F), or spinal NMDARs (Fig. 4G) in the recipient animals. Application of D-serine to the spinal cord dose-dependently induced a reversible synaptic facilitation (to 152 ± 9% at 220 to 240 min; $n = 10$, $P < 0.001$) (fig. S6), whereas TNF application triggers robust LTP at C-fiber synapses (*21*). These data indicate that transferable LTP required activation of glial cells in the donor but not in the recipient animals and that the combined actions of the gliotransmitters D-serine and TNF were required for its induction.

Collectively, our data indicate that the combined activation of microglia and astrocytes, either by $P2X_7R$ agonists or by HFS of primary afferents, triggered gliogenic LTP at C-fiber synapses with spinal lamina I neurons through the release of D-serine and cytokines such as TNF. Crucially, glial cell–derived signaling molecules accumulated in the extracellular space, including the cerebrospinal fluid, at biologically active but presently unknown concentrations, and induced LTP at C-fiber synapses, constituting the phenomenon of gliogenic LTP.

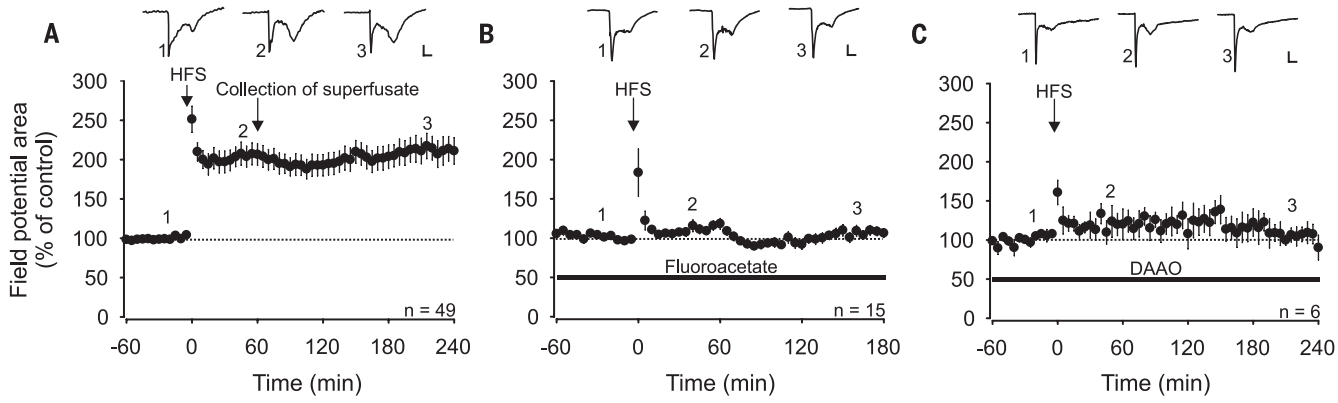


Fig. 3. HFS-induced LTP in vivo depends on spinal glial cells and D-serine signaling. Area of C-fiber–evoked field potentials was normalized to baseline values before conditioning HFS and plotted against time (min). Data are expressed as mean ± 1 SEM. Horizontal bars indicate drug application. (**A**) Mean time course of LTP of C-fiber–evoked field potentials. HFS at time point 0 min (arrow) induced LTP in all animals tested ($n = 49$, $P < 0.001$). One hour after HFS, the superfusate was collected from the lumbar spinal cord dorsum and transferred to animals shown in Fig. 4. (**B**) Spinal superfusion with the glial inhibitor fluoroacetate (10 μM) fully blocked HFS-induced potentiation in all animals tested ($n = 15$, $P = 0.085$). (**C**) HFS-induced LTP was fully prevented by spinal superfusion with DAAO (1 U·ml^{-1}; $n = 6$, $P = 0.365$). Insets show original traces of field potentials recorded at indicated time points. Calibration bars indicate 0.2 mV and 50 ms. RM ANOVA on ranks was performed to determine statistical significance in (A). In all other experiments, data were analyzed by using RM ANOVA.

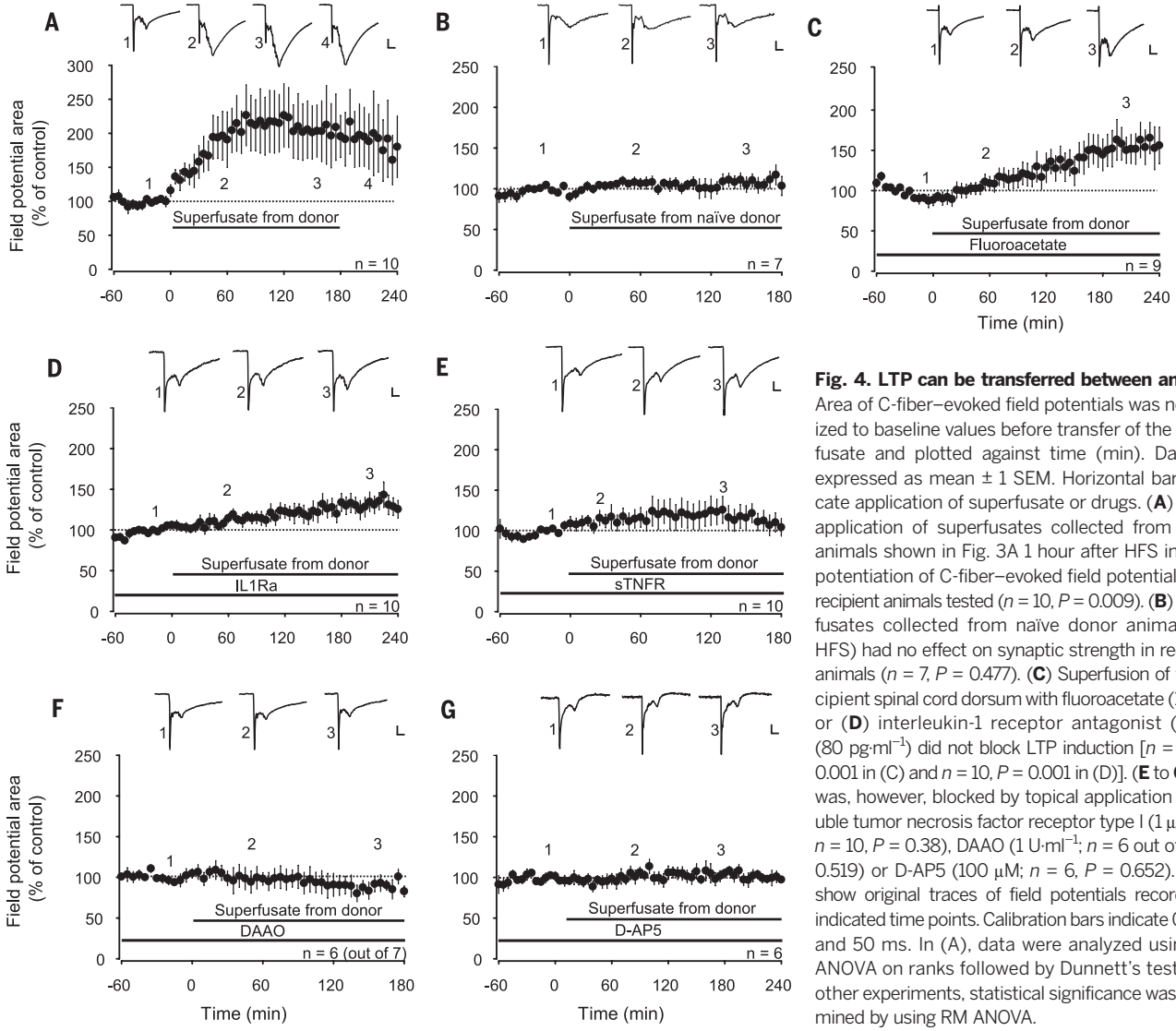


Fig. 4. LTP can be transferred between animals. Area of C-fiber–evoked field potentials was normalized to baseline values before transfer of the superfusate and plotted against time (min). Data are expressed as mean ± 1 SEM. Horizontal bars indicate application of superfusate or drugs. (**A**) Spinal application of superfusates collected from donor animals shown in Fig. 3A 1 hour after HFS induced potentiation of C-fiber–evoked field potentials in all recipient animals tested ($n = 10$, $P = 0.009$). (**B**) Superfusates collected from naïve donor animals (no HFS) had no effect on synaptic strength in recipient animals ($n = 7$, $P = 0.477$). (**C**) Superfusion of the recipient spinal cord dorsum with fluoroacetate (10 μM) or (**D**) interleukin-1 receptor antagonist (IL1Ra) (80 pg·ml^{-1}) did not block LTP induction [$n = 9$, $P < 0.001$ in (C) and $n = 10$, $P = 0.001$ in (D)]. (**E** to **G**) LTP was, however, blocked by topical application of soluble tumor necrosis factor receptor type I (1 μg·ml^{-1}; $n = 10$, $P = 0.38$), DAAO (1 U·ml^{-1}; $n = 6$ out of 7, $P = 0.519$) or D-AP5 (100 μM; $n = 6$, $P = 0.652$). Insets show original traces of field potentials recorded at indicated time points. Calibration bars indicate 0.2 mV and 50 ms. In (A), data were analyzed using RM ANOVA on ranks followed by Dunnett's test. In all other experiments, statistical significance was determined by using RM ANOVA.

Gliogenic LTP is a new form of paracrine synaptic plasticity in the central nervous system and may lead to pain amplification close to and remote from an injury or an inflammation. This is in line with the concept of chronic pain as a gliopathy involving neurogenic neuroinflammation (*7*, *24*). These new insights may pave the way for novel pain therapies (*25*, *26*). $P2X_7Rs$ play a key role in chronic inflammatory and neuropathic pain (*27*) and in other neurodegenerative and neuropsychiatric disorders (*28*). Glial cells display considerable diversity between and within distinct regions of the central nervous system (*29*). If the presently identified gliogenic LTP also existed at some brain areas, it could be of relevance not only for pain but also for other disorders, such as cognitive deficits, fear and stress disorders, and chronic immune-mediated diseases (*24*, *29*, *30*).

REFERENCES AND NOTES

1. T. V. P. Bliss, G. L. Collingridge, *Nature* **361**, 31–39 (1993).
2. H. Ikeda, B. Heinke, R. Ruscheweyh, J. Sandkühler, *Science* **299**, 1237–1240 (2003).
3. H. Ikeda *et al*., *Science* **312**, 1659–1662 (2006).
4. R. Kuner, *Nat. Med.* **16**, 1258–1266 (2010).
5. X.-Y. Li *et al*., *Science* **330**, 1400–1404 (2010).
6. R. Drdla, M. Gassner, E. Gingl, J. Sandkühler, *Science* **325**, 207–210 (2009).
7. R.-R. Ji, T. Berta, M. Nedergaard, *Pain* **154** (Suppl 1), S10–S28 (2013).
8. S. B. McMahon, M. Malcangio, *Neuron* **64**, 46–54 (2009).
9. P. M. Grace, M. R. Hutchinson, S. F. Maier, L. R. Watkins, *Nat. Rev. Immunol.* **14**, 217–231 (2014).
10. Q.-J. Gong *et al*., *Glia* **57**, 583–591 (2009).
11. H. W. Tao, L. I. Zhang, F. Engert, M. Poo, *Neuron* **31**, 569–580 (2001).
12. A. K. Clark *et al*., *J. Neurosci.* **35**, 4552–4570 (2015).
13. J. G. Gu, A. B. MacDermott, *Nature* **389**, 749–753 (1997).
14. Y.-X. Chu, Y. Zhang, Y.-Q. Zhang, Z.-Q. Zhao, *Brain Behav. Immun.* **24**, 1176–1189 (2010).
15. K. Kobayashi, E. Takahashi, Y. Miyagawa, H. Yamanaka, K. Noguchi, *Neurosci. Lett.* **504**, 57–61 (2011).
16. R. Aoyama *et al*., *Pain* **152**, 2085–2097 (2011).
17. W.-J. He *et al*., *Behav. Brain Res.* **226**, 163–170 (2012).
18. C. Ficker *et al*., *Glia* **62**, 1671–1686 (2014).
19. J. Jung, Y. H. Shin, H. Konishi, S. J. Lee, H. Kiyama, *Biochem. Biophys. Res. Commun.* **430**, 488–493 (2013).
20. R. D. Fields, Y. Ni, *Sci. Signal.* **3**, ra73 (2010).
21. D. Gruber-Schoffnegger *et al*., *J. Neurosci.* **33**, 6540–6551 (2013).
22. K. J. Sekiguchi *et al*., *Nat. Commun.* **7**, 11450 (2016).
23. M. Martineau *et al*., *J. Neurosci.* **33**, 3413–3423 (2013).
24. D. N. Xanthos, J. Sandkühler, *Nat. Rev. Neurosci.* **15**, 43–53 (2014).
25. R.-R. Ji, Z. Z. Xu, G. Strichartz, C. N. Serhan, *Trends Neurosci.* **34**, 599–609 (2011).
26. X.-H. Wei *et al*., *J. Neurosci.* **33**, 1540–1551 (2013).
27. I. P. Chessell *et al*., *Pain* **114**, 386–396 (2005).
28. A. M. Basso *et al*., *Behav. Brain Res.* **198**, 83–90 (2009).
29. B. S. Khakh, M. V. Sofroniew, *Nat. Neurosci.* **18**, 942–952 (2015).
30. A. Aguzzi, B. A. Barres, M. L. Bennett, *Science* **339**, 156–161 (2013).

ACKNOWLEDGMENTS

This work was supported by grants P 29206-B27 and W1205 from the Austrian Science Fund (FWF) to J.S. We thank L. Czarnecki for laboratory support and B. Heinke and G. Janeselli for technical support. All of the data are archived on servers of the Center for Brain Research, Medical University of Vienna. The authors declare no conflicts of interest. M.T.K., R.D.-S., and J.S. designed the research. M.T.K., R.D.-S., M.G., S.D.H., and H.L.T. generated and analyzed the data. M.T.K., R.D.-S., and J.S. wrote the paper, with input from the other authors.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1144/suppl/DC1
Materials and Methods
Figs. S1 to S6
Movie S1
References (*31*–*35*)

15 July 2016; accepted 1 November 2016
Published online 10 November 2016
10.1126/science.aah5715

STRUCTURAL BIOLOGY

Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease

Benjamin M. Akiyama,[1] Hannah M. Laurence,[1,2,3]* Aaron R. Massey,[4]* David A. Costantino,[1] Xuping Xie,[5] Yujiao Yang,[5] Pei-Yong Shi,[5] Jay C. Nix,[6] J. David Beckham,[4] Jeffrey S. Kieft[1,7]†

The outbreak of Zika virus (ZIKV) and associated fetal microcephaly mandates efforts to understand the molecular processes of infection. Related flaviviruses produce noncoding subgenomic flaviviral RNAs (sfRNAs) that are linked to pathogenicity in fetal mice. These viruses make sfRNAs by co-opting a cellular exonuclease via structured RNAs called xrRNAs. We found that ZIKV-infected monkey and human epithelial cells, mouse neurons, and mosquito cells produce sfRNAs. The RNA structure that is responsible for ZIKV sfRNA production forms a complex fold that is likely found in many pathogenic flaviviruses. Mutations that disrupt the structure affect exonuclease resistance in vitro and sfRNA formation during infection. The complete ZIKV xrRNA structure clarifies the mechanism of exonuclease resistance and identifies features that may modulate function in diverse flaviviruses.

Globalization, urbanization, and climate change contribute to the spread of pathogenic mosquito-borne viruses, typified by the outbreak of Zika virus (ZIKV) (*1*). ZIKV infection can cause fetal microcephaly and Guillain-Barré syndrome (*2*), motivating efforts to understand the molecular drivers of pathology. ZIKV is a (+)-sense single-stranded RNA mosquito-borne flavivirus (MbFV) related to dengue virus (DENV), yellow fever virus (YFV), and West Nile virus (WNV) (*3*). The structured 3′ untranslated regions (UTRs) of many MbFVs are the source of noncoding subgenomic flaviviral RNAs (sfRNAs) that accumulate during infection when RNA elements resist degradation by the host 5′ → 3′ exonuclease Xrn1 (fig. S1A) (*4*). These sfRNAs are directly linked to cytopathic and pathologic effects (*4*); they dysregulate RNA decay pathways and bind cellular proteins important for antiviral responses (*5*–*14*). Preventing sfRNA production could be a strategy for targeted therapeutics or for generating attenuated virus for vaccines (*15*–*17*).

Because sfRNA formation during ZIKV infection has not been reported, we infected multiple cell lines with ZIKV strain PRVABC59, isolated in 2015 from an infected U.S. mainland–Puerto Rico traveler. Northern blot analysis of total RNA isolated from infected cells showed discrete bands containing parts of the ZIKV 3′UTR, consistent with sfRNAs (Fig. 1A). Mouse primary neuron infection resulted in very little infectious virus and produced three weak sfRNA bands. Infection of C6/36 (*Aedes albopictus* mosquito) cells produced two predominant sfRNAs, whereas Vero (monkey) and A549 (human) epithelial cell infection produced additional bands. Different cell types produced different sfRNA patterns, but the largest sfRNA was present in all. The importance of this cell type–dependent variation in the sfRNA patterns is unknown, although studies with DENV suggest that sfRNA production is modulated to enable host adaptation (*16*, *18*).

The production of ZIKV sfRNAs suggests the existence of Xrn1-resistant structures (xrRNAs) in the viral 3′UTR. Two areas of the UTR match the sequence pattern and potential secondary structure of known MbFV xrRNAs (Fig. 1, B and C, and fig. S1B) (*4*, *18*). The putative xrRNAs are in series near the 5′ end of the UTR—a location and pattern similar to that of other MbFVs (Fig. 1D). Xrn1 halting at putative ZIKV xrRNA1 and xrRNA2 would result in sfRNAs of sizes matching the two produced in all cell types tested (Fig. 1A and fig. S1C). To test whether these putative elements are indeed Xrn1-resistant, we challenged in vitro transcribed full-length ZIKV 3′UTR RNA with recombinant Xrn1 (*19*). Although multiple sfRNAs were observed during ZIKV infection, in vitro the upstream xrRNA1 quantitatively halted the enzyme (Fig. 1E). However, a UTR lacking the upstream xrRNA (ΔxrRNA1) allowed the enzyme to stop at the downstream xrRNA2. The size of the Xrn1-resistant RNAs matched those of the infection-produced sfRNAs (fig. S1D). Thus, ZIKV

[1]Department of Biochemistry and Molecular Genetics, University of Colorado Denver School of Medicine, Aurora, CO 80045, USA. [2]Howard Hughes Medical Institute (HHMI), University of Colorado Denver School of Medicine, Aurora, CO 80045, USA. [3]School of Veterinary Medicine, University of California, Davis, CA 95616, USA. [4]Department of Medicine, Division of Infectious Diseases, University of Colorado Denver School of Medicine, Aurora, CO 80045, USA. [5]Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX 77555, USA. [6]Molecular Biology Consortium, Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [7]RNA BioScience Initiative, University of Colorado Denver School of Medicine, Aurora, CO 80045, USA.
*These authors contributed equally to this work. †Corresponding author. Email: jeffrey.kieft@ucdenver.edu

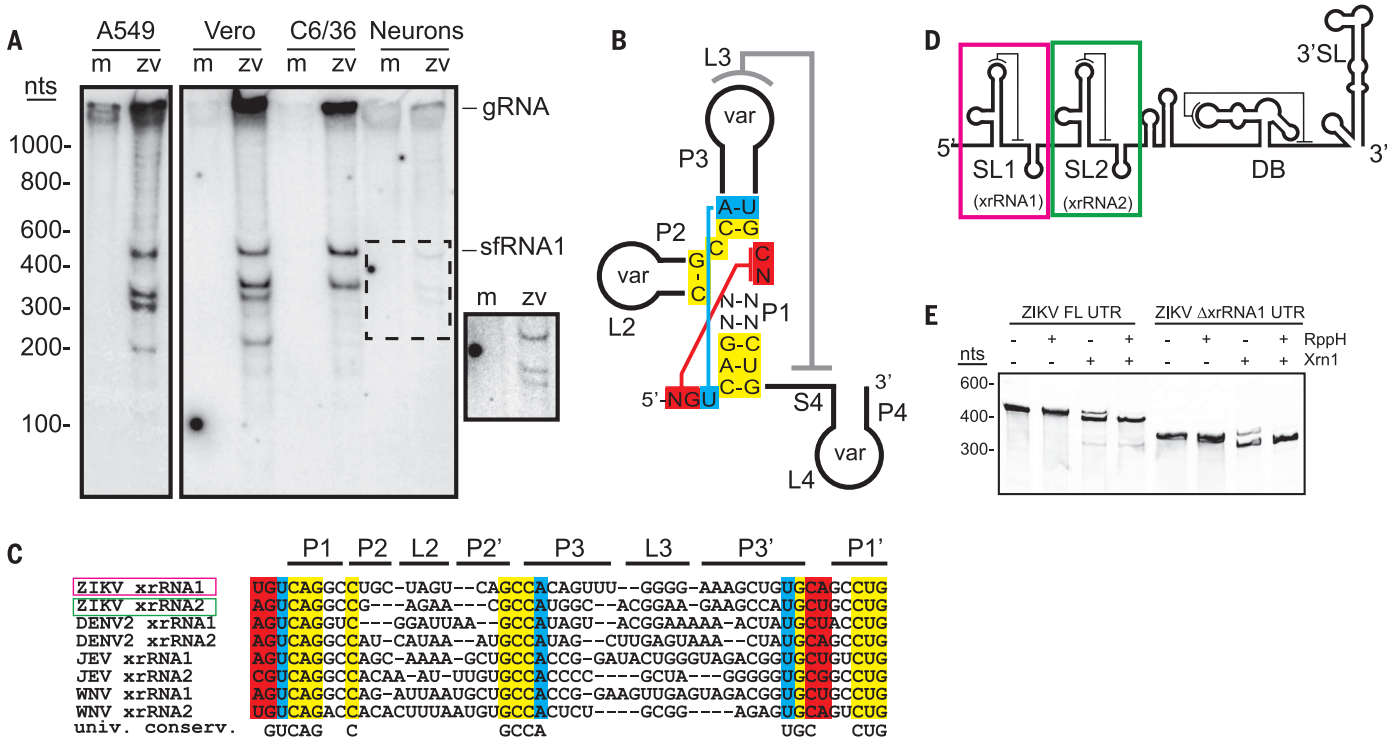


Fig. 1. ZIKV produces sfRNAs by exonuclease resistance. (**A**) Northern blot of purified RNA from mock-infected (m) and ZIKV-infected (zv) human A549, Vero, mosquito C6/36, and cultured primary mouse neuron cells. Probe was complementary to the putative dumbbell sequence in the ZIKV 3′UTR. gRNA, genomic RNA; sfRNA1, band corresponding to largest sfRNA. Inset shows a high-contrast view of the region in the dashed box. Infections and Northern blots were confirmed in three independent experiments. (**B**) Secondary structure diagram with sequence conservation in known stem-loop (SL)–type xrRNAs from MbFVs. Universally conserved nucleotides (yellow) and positions of tertiary interactions (red, blue) are shown. The P2, P3, and P4 stems and associated loops contain variable sequences (var); N represents any nucleotide. (**C**) Sequence alignment of several known or putative MbFV xrRNA sequences with regions colored as in (B). DENV, dengue virus; JEV, Japanese encephalitis virus; WNV, West Nile virus. (**D**) Cartoon of the predicted ZIKV 3′UTR secondary structure, with SL and dumbbell (DB) elements. In the Kunjin strain of WNV (WNV_{Kun}), the SL1 and SL2 RNAs are xrRNAs (*4*, *19*, *21*). (**E**) Ethidium-stained gel of in vitro transcribed ZIKV full-length 3′UTR (FL UTR) or a 3′UTR lacking the first putative xrRNA1 element (ΔxrRNA1) treated with recombinant pyrophosphohydrolase RppH (to generate a 5′ monophosphate) and recombinant exonuclease Xrn1. The smaller band indicates Xrn1 resistance. The presence of Xrn1 activity in the absence of RppH is due to spontaneous loss of the 5′ pyrophosphate moiety that occurs at some level even in the absence of the RppH enzyme, as previously observed (*19*).

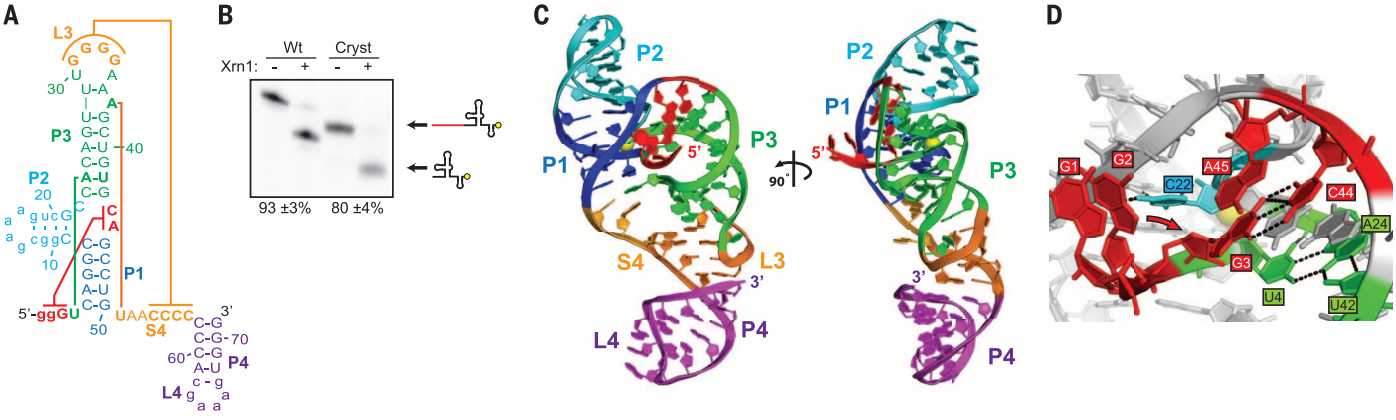


Fig. 2. Structure of the ZIKV xrRNA1. (**A**) Secondary structure diagram of the crystallized RNA. Lowercase letters represent sequences altered to facilitate RNA expression and crystallization. Colored lines indicate interactions discussed in the text. (**B**) Xrn1 resistance assays of wild-type (Wt) ZIKV xrRNA1 and the crystallized RNA (Cryst) using 3′ end-labeled RNA (yellow circle). The percentage of the total RNA that formed an Xrn1-resistant band (listed below the gel) was quantified and is reported as the average ± SD of three independent experiments. (**C**) Ribbon representation of the structure of ZIKV xrRNA1, colored to match (A). Magnesium ions are shown as yellow spheres. (**D**) Detail of interactions at the 5′ end of the RNA. Residue C22 (cyan) contacts the phosphate backbone of neighboring residues, setting up a kink in the RNA critical for folding. Residues U4, A24, and U42 (green) form a base triple interaction orienting the 5′ end. Residue G3 forms a long-range base-pairing interaction with residue C44. Residue G2 was mutated from a U to promote transcription; the wild-type sequence is predicted to form a base pair with residue A45 (predicted position change indicated by arrow) (*21*).

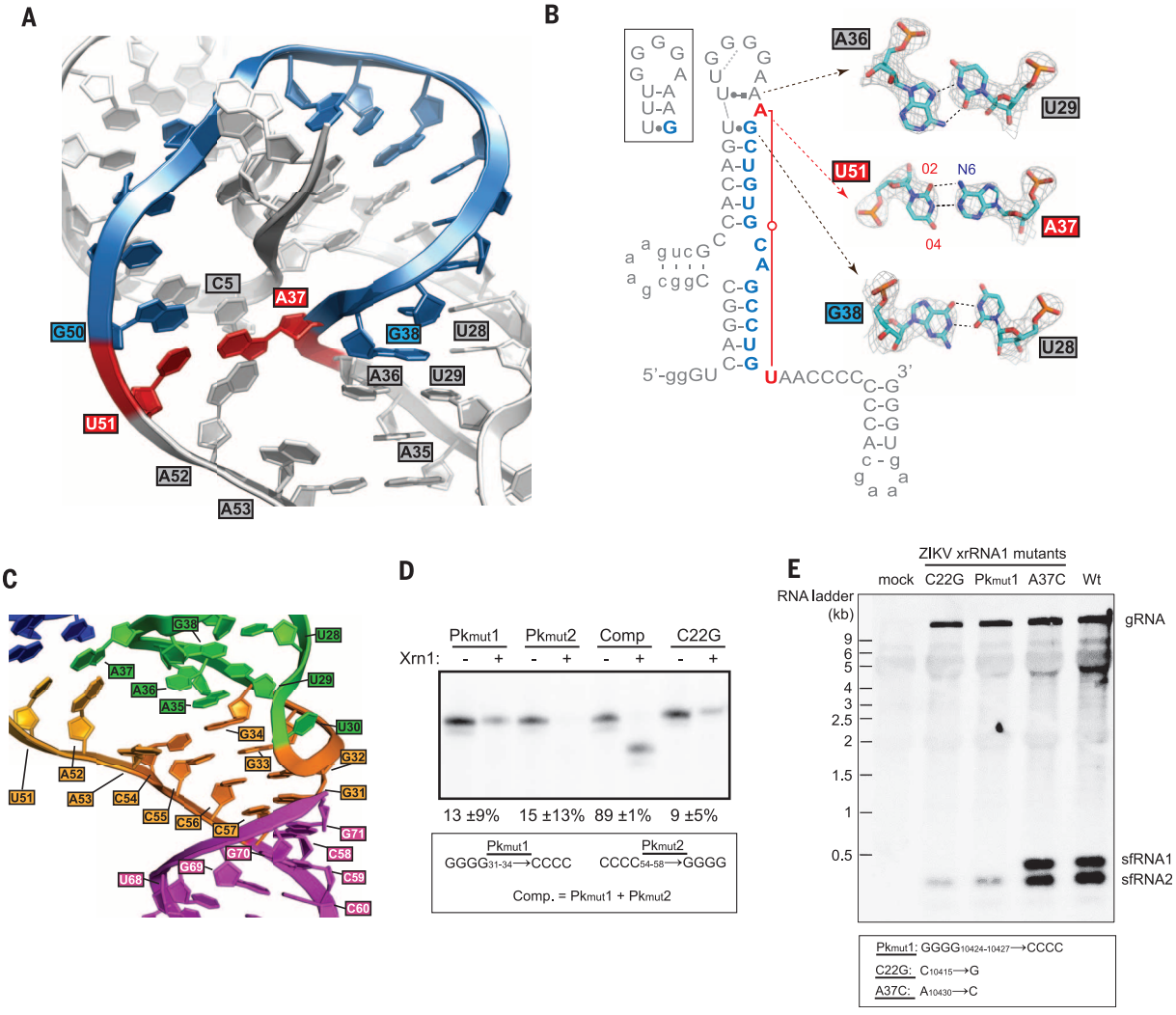


Fig. 3. Details of the ZIKV xrRNA1 structure. (**A**) Detailed view of the A37-U51 base pair (red) and intervening nucleotides (blue), which circle the 5′ end of the RNA. Other nucleotides discussed in the text are labeled. (**B**) The A37-U51 base pair (red) and intervening nucleotides (blue) are highlighted. The box shows the previously predicted secondary structure of the P3-L3 stem-loop. Leontis-Westhof nomenclature is used to indicate noncanonical pairing (*30*). Inset displays details of all three noncanonical base pairs; electron density is displayed at the 2σ contour level. (**C**) The L3-S4 pseudoknot with the P4 stem coaxially stacked. Colors are as in Fig. 2, A and C. (**D**) Xrn1 resistance assays of pseudoknot mutants and a mutant known to disrupt xrRNA folding (C22G) (*19*, *21*). Quantitation of resistance from three experiments is shown, determined as in Fig. 2B. (**E**) Northern blot of viral RNA isolated from viral infection with wild-type virus and virus mutated in the xrRNA1 structure. The mutants are labeled to match the analogous mutants in (D) and fig. S4C; corresponding positions in the viral RNA are provided below.

contains two xrRNAs, and cellular factors could modulate xrRNA function during infection to produce a multiple-sfRNA pattern (*20*).

Insight into the structural basis of sfRNA formation comes from a previously solved structure from Murray Valley encephalitis virus (MVE) (*21*), but our understanding is incomplete. Specifically, in the MVE xrRNA structure, an important pseudoknot (L3-S4; gray line in Fig. 1B) (*4*, *18*, *22*, *23*) was not formed, and thus the structure of a fully folded xrRNA remains unsolved. Also, the MVE structure was of the downstream xrRNA of two in series (xrRNA2), and evidence now suggests differences between xrRNA1s and xrRNA2s (*16*, *18*); xrRNA1s are predicted to have a more stable fold, and WNV infection models suggest that loss of xrRNA1 has a greater effect on viral infection (*4*, *24*). Fully understanding sfRNA formation therefore requires a detailed structural description of a fully folded xrRNA1. ZIKV xrRNA1 is ideal for such studies, as it shows robust Xrn1 resistance in vitro (Fig. 1E), it forms sfRNAs in all tested cell types (Fig. 1A), and its pseudoknot is predicted to be stabilized by four consecutive G-C base pairs.

We solved the structure of ZIKV xrRNA1 by x-ray crystallography (Fig. 2A and table S1), using the MR766-NIID African isolate sequence (PRVABC59 xrRNA1 is 97% identical in this region) with sequence alterations to improve transcription and crystallization (fig. S2). This RNA maintained Xrn1 resistance and is thus correctly folded (Fig. 2B). All 71 nucleotides were visible (fig. S3), and unlike the MVE xrRNA2 structure, the RNA is fully folded; a pseudoknot is formed between L3 and S4, completely encircling and constraining the 5′ end of the RNA (Fig. 2C). In both the MVE and ZIKV xrRNA structures, the fold is organized around the P1-P2-P3 three-way junction; the P1 and P3 helices form a ring through which the 5′ end passes from one side of the structure to the other, positioned by nucleotides at the 5′ end that form base pairs with nucleotides in the junction and a U4•A24-U42 base triple (Fig. 2D; the dot indicates a non–Watson-Crick base interaction). The presence of these interactions in both structures strongly suggests their functional importance; disrupting them may be a way to attenuate diverse MbFVs.

The ZIKV xrRNA1 structure reveals multiple previously unobserved interactions. A37 and U51 form a reverse Watson-Crick (or trans) long-range base pair that closes the ring structure to "lasso" the RNA that passes through (Fig. 3A). Contrary to predictions, A37 is flipped out of the helix; this position is created by an unexpected structure in the P3-L3 stem-loop that contains a G-U

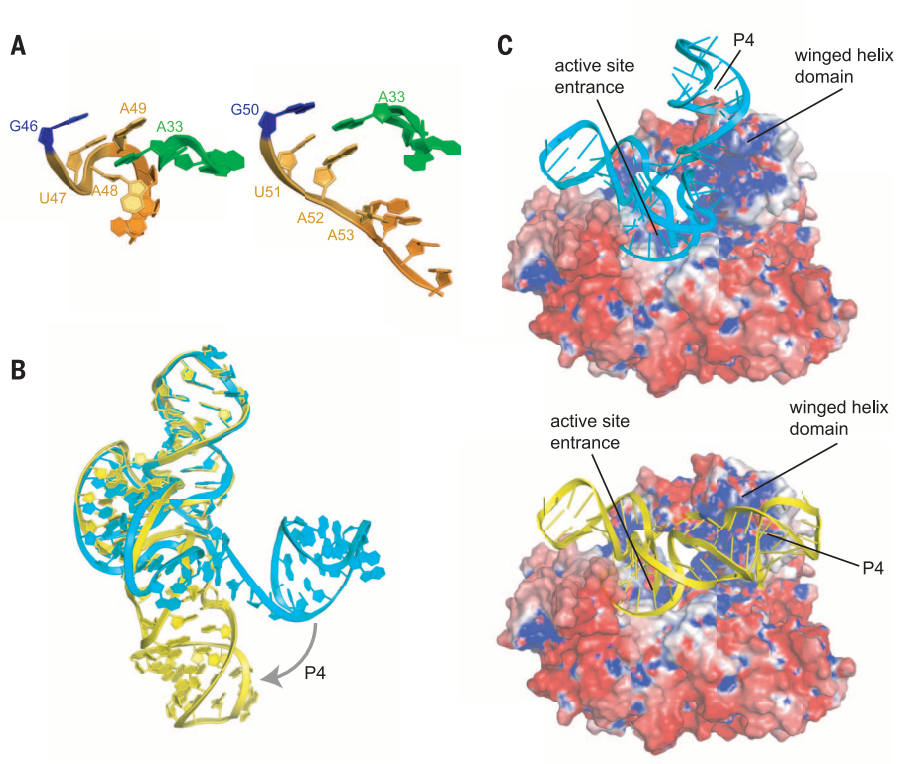


Fig. 4. Model of ZIKV xrRNA–Xrn1 interaction. (**A**) Comparison of the S4 region (orange) and adjacent RNA in the partially folded MVE (left) and fully folded ZIKV (right) xrRNAs. (**B**) Overlay of the MVE (cyan) and ZIKV (yellow) structures, showing the change in the position of the P4-L4 hairpin. (**C**) Models of the MVE (top) and ZIKV (bottom) xrRNAs docked onto the surface of Xrn1, colored according to electrostatic potential (blue, positive; red, negative). Structural features are labeled.

wobble pair between G38 and U28 and a Hoogsteen pair between A36 and U29 (Fig. 3B). The A37-U51 base pair led us to hypothesize that this interaction may precisely define the size of the ring, which is 14 nucleotides in the ZIKV xrRNA1. Indeed, sequence alignment shows that 30 of 33 confirmed or putative xrRNAs have Watson-Crick base-pairing partners 14 nucleotides apart in analogous positions (fig. S4, A and B). Although these are apparent Watson-Crick partners, the structure suggests that they form noncanonical pairs. Indeed, YFV, Sepik virus, and Wesselsbron virus have G bases at both positions, suggestive of a G•G pair and thus implying possible alternative conformations of this long-range base pair.

To explore this long-range interaction, we altered the A37-U51 pair by mutation and tested for Xrn1 resistance in vitro (fig. S4C). Substitution of either nucleotide individually had very little effect, and substitution of both bases to convert the interaction to either a C-G or G•G resulted in only a moderate decrease in resistance. Overall, it is not clear why sequence conservation shows a preference for apparent Watson-Crick pairing partners; perhaps certain pairs improve Xrn1 resistance in the context of specific xrRNAs, or they may contribute to functions other than Xrn1 resistance. Consistent with this, the preference for a purine at the first position of this interaction and a pyrimidine at the second position is reminiscent of a conserved trans R15:Y48 (R, purine; Y, pyrimidine) base pair observed in tRNAs (*25*). G•G pairs have been observed at this position in cysteine tRNAs, and tolerance for switching between G•G and R15:Y48 pairs has been shown to be dependent on the RNA sequence context (*26*).

The A37-U51 base pair defines the size of the ring and continues the progression of stacked bases from the P1 stem to the L3-S4 pseudoknot (Fig. 3, A and C). Specifically, A52 stacks on this pair but is not itself paired. This base is universally a purine (fig. S4, A and B), likely to maximize stacking potential. The next base, A53, is extruded to form a crystal contact (fig. S5), but in solution it likely stacks under A52, supported by the presence of electron density where A53 could stack, consistent with a 1,6-hexanediol molecule from the crystallization solution (fig. S5A). If stacked within the helix, A53 would continue the progression of stacked purines through S4 to the pseudoknot and could form an A•A pair with A35, providing additional stability (Fig. 3C).

As predicted, the L3-S4 pseudoknot contains four consecutive G-C Watson-Crick base pairs (Fig. 3C). The four G's in L3 are made accessible by a U-turn, a motif found in other loops including tRNA anticodon loops. To test the functional importance of the ZIKV xrRNA1 pseudoknot, we generated mutant RNAs (Pkmut1, Pkmut2) and tested them for Xrn1 resistance in vitro (Fig. 3D). Disruption of the pseudoknot severely decreased Xrn1 resistance to the same degree as a previously characterized mutation that disrupts the three-way junction (C22G) (*19*, *21*). Restoring the pseudoknot (Comp) returned nearly wild-type Xrn1 resistance; thus, the pseudoknot is critical for ZIKV xrRNA1 function.

The effects of structure-based mutations to xrRNA1 on Xrn1 resistance in vitro led us to predict that the same mutations would alter sfRNA formation during infection; therefore, we generated mutant ZIKV using an infectious clone based on strain FSS13025 (identical to PRVABC59 in xrRNA1) (*27*). Mutations in xrRNA1 analogous to C22G and Pkmut1 resulted in highly reduced sfRNA formation during infection, matching the in vitro Xrn1 resistance result (Fig. 3E). Both sfRNAs were reduced; this finding underscores the importance of xrRNA1 to sfRNA formation overall. A mutation analogous to A37C had no discernible effect on sfRNA formation, consistent with this mutant's Xrn1 resistance in vitro (fig. S4C).

We also assessed the effect of Xrn1 knockdown on ZIKV sfRNA production during infection of human cells (fig. S6). Xrn1 knockdown resulted in a reduction of one sfRNA species and a change in the overall sfRNA pattern; however, the persistence of some sfRNA bands suggests that Xrn1 could be redundant with other exonucleases in producing ZIKV sfRNAs. Xrn1 resistance in vitro and ZIKV sfRNA production during infection are linked by the strong agreement between the effects of RNA mutations in both assays, which implies that the RNA structure is necessary for sfRNA formation.

One of the most functionally essential interactions is the L3-S4 pseudoknot, whose formation is accompanied by substantial structural differences relative to the MVE xrRNA2 structure (Fig. 4A). The helical element created by the pseudoknot forms a continuous stack with the P4 helix (fig. S7, A and B), placing the P4-L4 stem-loop in a different position relative to the MVE xrRNA2 structure (Fig. 4B) and differing from that predicted by structural modeling (fig. S7, C and D). P4's stacking on the pseudoknot likely stabilizes the overall fold, but there are no obvious sequence-specific roles for this element in forming the structure. However, the P4-L4 stem-loop contains numerous conserved bases found in other MbFVs (fig. S8), and it therefore may be important for some aspect of sfRNA formation or downstream function.

Using the fully folded structure, we constructed a new model of the ZIKV xrRNA1 interacting with Xrn1 from *D. melanogaster* (with bound substrate analog), using previously reported biochemical information (*19*, *28*) and the electrostatic charge distribution and shape of the enzyme's surface (Fig. 4C). The model shows extensive contacts between the enzyme and the RNA. A model of the partially folded MVE xrRNA with Xrn1 suggested that the RNA ring structure prevents Xrn1 from unwinding the structure (*21*), but in that form, the P4 element projected away from the surface of Xrn1 (Fig. 4C). In contrast, the fully folded ZIKV xrRNA1 closely matches the contours

of an electropositive patch, with P4 contacting a conserved charged region on Xrn1's winged helix domain.

This putative P4-Xrn1 interaction could serve to stabilize the xrRNA's pseudoknot interaction and thus enhance resistance to the enzyme, or P4 may form sequence-specific interactions with Xrn1 or with Xrn1-bound proteins. Also, because the winged helix domain is important for processive Xrn1 function (*29*), the bound ZIKV xrRNA may prevent conformational changes in the enzyme that are important for processivity. The new structure and derived hypotheses point the way to future studies that may clarify the formation and function of ZIKV sfRNAs, with implications for the development of interventions or vaccines.

REFERENCES AND NOTES

1. J. S. Mackenzie, D. J. Gubler, L. R. Petersen, *Nat. Med.* **10** (suppl.), S98–S109 (2004).
2. L. R. Petersen, D. J. Jamieson, A. M. Powers, M. A. Honein, *N. Engl. J. Med.* **374**, 1552–1563 (2016).
3. B. D. Lindenbach, H. J. Thiel, C. M. Rice, in *Fields Virology*, D. M. Knipe, P. M. Howley, Eds. (Lippincott Williams & Wilkins, ed. 5, 2007), chap. 33.
4. G. P. Pijlman *et al.*, *Cell Host Microbe* **4**, 579–591 (2008).
5. G. Manokaran *et al.*, *Science* **350**, 217–221 (2015).
6. A. Schuessler *et al.*, *J. Virol.* **86**, 5708–5718 (2012).
7. E. Schnettler *et al.*, *J. Virol.* **86**, 13486–13500 (2012).
8. J. A. Roby, G. P. Pijlman, J. Wilusz, A. A. Khromykh, *Viruses* **6**, 404–427 (2014).
9. K. Bidet, D. Dadlani, M. A. Garcia-Blanco, *PLOS Pathog.* **10**, e1004242 (2014).
10. S. L. Moon *et al.*, *Virology* **485**, 322–329 (2015).
11. S. L. Moon *et al.*, *RNA* **18**, 2029–2040 (2012).
12. A. M. Ward *et al.*, *RNA Biol.* **8**, 1173–1186 (2011).
13. M. M. Emara, H. Liu, W. G. Davis, M. A. Brinton, *J. Virol.* **82**, 10657–10670 (2008).
14. G. P. Göertz *et al.*, *J. Virol.* **90**, 10145–10159 (2016).
15. A. P. Durbin *et al.*, *J. Infect. Dis.* **191**, 710–718 (2005).
16. S. M. Villordo, C. V. Filomatori, I. Sánchez-Vargas, C. D. Blair, A. V. Gamarnik, *PLOS Pathog.* **11**, e1004604 (2015).
17. V. Proutski, E. A. Gould, E. C. Holmes, *Nucleic Acids Res.* **25**, 1194–1202 (1997).
18. J. S. Kieft, J. L. Rabe, E. G. Chapman, *RNA Biol.* **12**, 1169–1177 (2015).
19. E. G. Chapman, S. L. Moon, J. Wilusz, J. S. Kieft, *eLife* **3**, e01892 (2014).
20. S. M. Villordo, J. M. Carballeda, C. V. Filomatori, A. V. Gamarnik, *Trends Microbiol.* **24**, 270–283 (2016).
21. E. G. Chapman *et al.*, *Science* **344**, 307–310 (2014).
22. A. Funk *et al.*, *J. Virol.* **84**, 11407–11417 (2010).
23. P. A. Silva, C. F. Pereira, T. J. Dalebout, W. J. Spaan, P. J. Bredenbeek, *J. Virol.* **84**, 11395–11406 (2010).
24. R. Y. Chang *et al.*, *Vet. Microbiol.* **166**, 11–21 (2013).
25. M. Levitt, *Nature* **224**, 759–763 (1969).
26. L. D. Sherlin *et al.*, *J. Mol. Biol.* **299**, 431–446 (2000).
27. C. Shan *et al.*, *Cell Host Microbe* **19**, 891–900 (2016).
28. M. Jinek, S. M. Coyle, J. A. Doudna, *Mol. Cell* **41**, 600–608 (2011).
29. J. H. Chang, S. Xiang, K. Xiang, J. L. Manley, L. Tong, *Nat. Struct. Mol. Biol.* **18**, 270–276 (2011).
30. N. B. Leontis, E. Westhof, *RNA* **7**, 499–512 (2001).

ACKNOWLEDGMENTS

We thank the members of the Kieft lab for discussions; C. Musselman and M. Stone for critical reading of this manuscript; and R. Soto from the M. Garcia-Blanco and S. Bradrick laboratories for sharing their protocol for sfRNA blots. Supported by NIH fellowship F32GM117730 (B.M.A.); an HHMI Early Career Scientist award and NIH grants R35GM118070 and R01GM081346 (J.S.K.); a HHMI–Burroughs Wellcome Fund medical research fellowship (H.M.L.); University of Colorado School of Medicine and Department of Medicine institutional pilot funds (A.R.M. and J.D.B.); a postdoctoral fellowship from Novartis Institutes for BioMedical Research (X.X.); and a University of Texas Medical Branch startup award, a University of Texas STARs award, and NIH grant R01AI087856 (P.-Y.S.). The UC Denver X-ray Facility is supported by UC Cancer Center support grant P30CA046934 and NIH grant S10OD012033. The Advanced Light Source is supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under contract DE-AC02-05CH11231. The data presented in this manuscript are tabulated in the main paper and in the supplementary materials. Coordinates and structure factors have been deposited with PDB accession code 5TPY.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1148/suppl/DC1
Materials and Methods
Figs. S1 to S8
Table S1
References (*31*–*45*)

20 June 2016; accepted 26 October 2016
Published online 10 November 2016
10.1126/science.aah3963

HEMATOPOIESIS

Depleting dietary valine permits nonmyeloablative mouse hematopoietic stem cell transplantation

Yuki Taya,[1] Yasunori Ota,[2] Adam C. Wilkinson,[1,3] Ayano Kanazawa,[1] Hiroshi Watarai,[4,5] Masataka Kasai,[1] Hiromitsu Nakauchi,[1,3]* Satoshi Yamazaki[1]*

A specialized bone marrow microenvironment (niche) regulates hematopoietic stem cell (HSC) self-renewal and commitment. For successful donor-HSC engraftment, the niche must be emptied via myeloablative irradiation or chemotherapy. However, myeloablation can cause severe complications and even mortality. Here we report that the essential amino acid valine is indispensable for the proliferation and maintenance of HSCs. Both mouse and human HSCs failed to proliferate when cultured in valine-depleted conditions. In mice fed a valine-restricted diet, HSC frequency fell dramatically within 1 week. Furthermore, dietary valine restriction emptied the mouse bone marrow niche and afforded donor-HSC engraftment without chemoirradiative myeloablation. These findings indicate a critical role for valine in HSC maintenance and suggest that dietary valine restriction may reduce iatrogenic complications in HSC transplantation.

Although much is known about the molecules and signaling pathways regulating hematopoietic stem cells (HSCs), our understanding of the HSC bone marrow (BM) niche is less clear. The availability of niche "space" appears to be a limiting factor for engraftment in HSC transplantation (HSCT) (*1*). Ionizing radiation and/or high-dose chemotherapy are most commonly used to empty the BM niche. However, severe side effects limit the success of these treatments. Young patients are particularly at risk of experiencing late effects, including secondary malignancy, endocrinopathy, and reproductive failure (*2*).

[1]Division of Stem Cell Therapy, Center for Stem Cell Biology and Regenerative Medicine, The Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. [2]Department of Pathology, Research Hospital, The Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. [3]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Lorry I. Lokey Stem Cell Research Building, 265 Campus Drive, Stanford, CA 94305-5461, USA. [4]Division of Stem Cell Cellomics, Center for Stem Cell Biology and Regenerative Medicine, The Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. [5]Impulsing Paradigm Change through Disruptive Technologies Program (ImPACT), Japan Science and Technology Agency (JST), Tokyo 102-0076, Japan.
***Corresponding author. Email: y-sato4@ims.u-tokyo.ac.jp (S.Y.); nakauchi@stanford.edu (H.N.)**

Numerous cellular and molecular components of the HSC niche have previously been proposed (*3*). These include several stromal and neuronal cell types, as well as various signaling molecules including the cytokines stem cell factor and thrombopoietin. In 1946, Kornberg *et al.* reported that rats fed a low-protein diet developed severe granulocytopenia or anemia that was corrected by administration of purified amino acids (AAs) (*4*, *5*). On the basis of these findings, we hypothesized that specific AAs may be indispensable for the BM niche and could influence HSC fate.

We initially quantified AA concentrations in BM and peripheral blood (PB) by high-performance liquid chromatography (HPLC). The BM contained >100-fold higher concentrations of all 20 AAs and also displayed a distinct AA profile (or AA balance), as compared with the PB (fig. S1, A and B).

To identify the specific AAs essential for the maintenance of hematopoiesis, $CD34^{-}Kit^{+}Sca1^{+}Lin^{-}$ cells (HSCs) (*6*) were cultured for 1 week in various media lacking single AAs (Fig. 1A and fig. S1C). Proliferation was significantly retarded when cysteine or valine was absent (–Cys, –Val) (Fig. 1B and fig. S2A). We also carried out similar analyses using $CD34^{+}Kit^{+}Sca1^{+}Lin^{-}$ hematopoietic progenitor cells (HPCs). HPCs displayed significant growth retardation in medium lacking cysteine or lysine (fig. S2, B and C). Growth of HPCs

was not appreciably inhibited in medium lacking valine. To determine AA requirements for HSC maintenance, we conducted competitive repopulation assays following in vitro culture (Fig. 1A). No engraftment was detected in mice that received –Cys or –Val cultured HSCs (Fig. 1C).

To assess whether –Cys and –Val media caused reactive oxygen species stress in HSCs, cellular proliferation was determined after in vitro culture in the presence of the antioxidant *N*-acetyl cysteine (NAC). NAC rescued HSC growth in –Cys medium but not in –Val medium (fig. S3). We also failed to identify significant differences in autophagy between HSCs cultured in complete and –Val media (fig. S4). Collectively, these findings support the idea that cysteine and valine are required for HSC survival and HSC self-renewal, respectively.

To investigate how cysteine and valine contribute to hematopoiesis in vivo, mice were fed complete, –Cys, or –Val diets for 4 weeks (fig. S5A). Complete blood counts revealed a significant reduction in white blood cell (WBC) and red blood cell numbers in mice fed a –Val diet (Fig. 2A), whereas platelet counts were unaffected. No differences were observed in mice fed a –Cys diet. To confirm the effect of the AA-deficient diets, we quantified AA levels in PB and BM serum by HPLC. Although cysteine levels were too low to detect, valine concentration decreased by 90% (fig. S5B). Cysteine is a nonessential AA and can be generated from the catabolism of methionine in vivo (*7*). Given this compensation, we focused on valine in our additional studies.

To further understand the effect of dietary valine restriction, we quantified various hematopoietic cell types within the PB and BM. Corresponding with our in vitro data, HSC frequencies (both immunophenotypically defined $CD34^-KSL$ HSCs and $CD150^+CD41^-CD48^-KSL$ HSCs; two

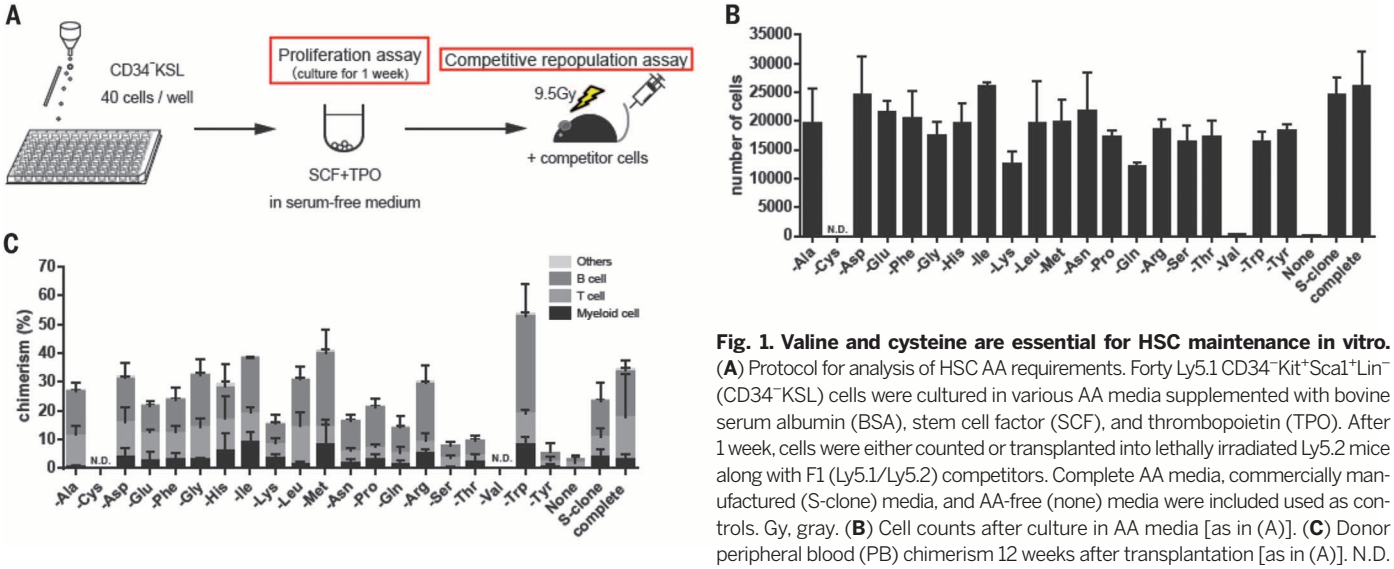


Fig. 1. Valine and cysteine are essential for HSC maintenance in vitro. (**A**) Protocol for analysis of HSC AA requirements. Forty Ly5.1 $CD34^-Kit^+Sca1^+Lin^-$ ($CD34^-KSL$) cells were cultured in various AA media supplemented with bovine serum albumin (BSA), stem cell factor (SCF), and thrombopoietin (TPO). After 1 week, cells were either counted or transplanted into lethally irradiated Ly5.2 mice along with F1 (Ly5.1/Ly5.2) competitors. Complete AA media, commercially manufactured (S-clone) media, and AA-free (none) media were included used as controls. Gy, gray. (**B**) Cell counts after culture in AA media [as in (A)]. (**C**) Donor peripheral blood (PB) chimerism 12 weeks after transplantation [as in (A)]. N.D. denotes nondetected chimerism. For (B) and (C), experiments were performed in triplicate; error bars indicate ± SD.

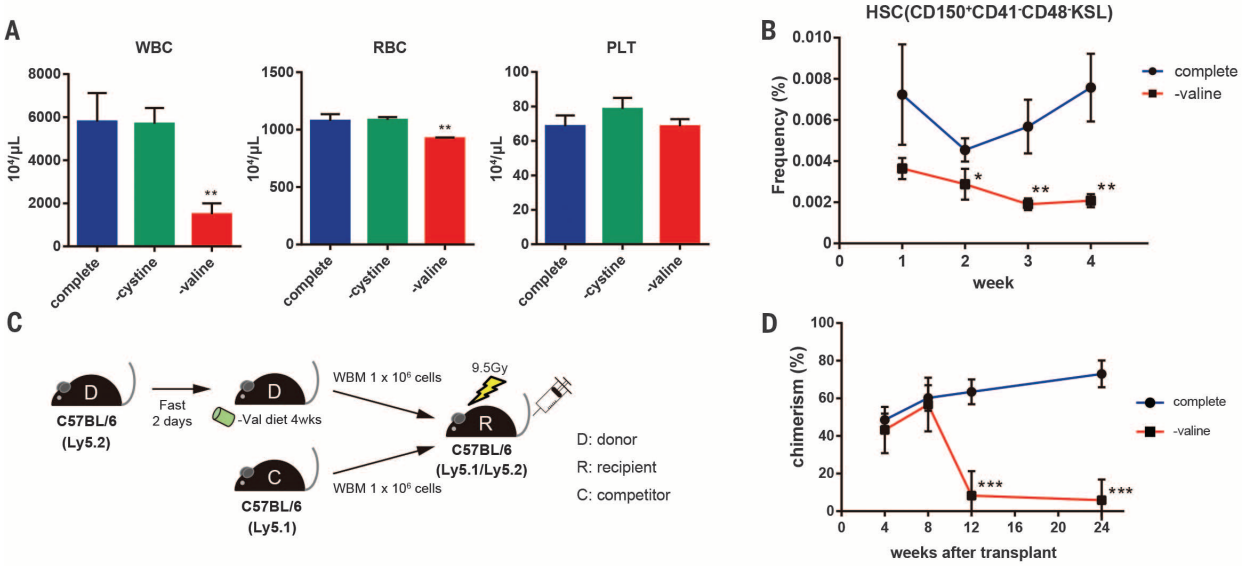
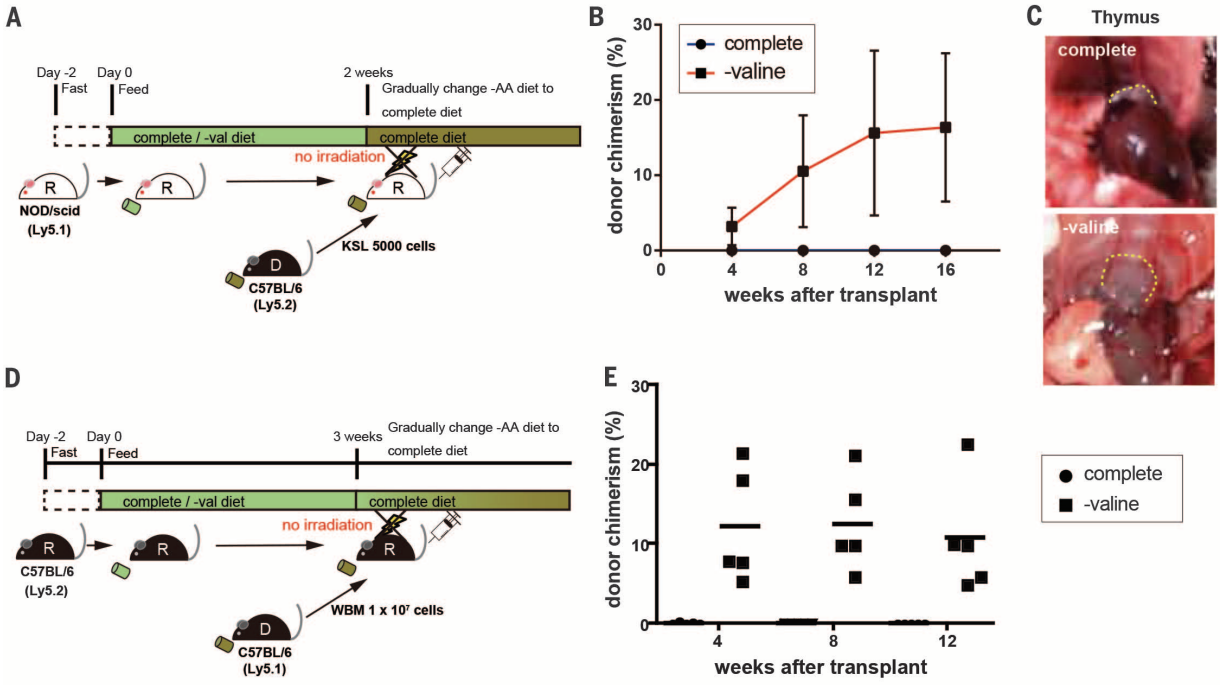


Fig. 2. Dietary valine restriction significantly perturbs hematopoiesis. (**A**) White blood cell (WBC), red blood cell (RBC), and platelet (PLT) counts following a 4-week synthetic diet ($n = 3$ mice). (**B**) Frequency of BM $CD150^+CD41^-CD48^-KSL$ HSCs following a valine-deficient (–Val) diet (complete, $n = 3$; –Val, $n = 4$). (**C**) Competitive repopulation assay protocol. BM cells from Ly5.1 mice fed a complete diet or a –Val diet for 4 weeks were transplanted into lethally irradiated Ly5.2 mice, along with an equal number of F1 competitor BM cells. WBM, whole bone marrow. (**D**) PB donor chimerism after transplantation ($n = 6$), as outlined in (C). For (A), (B), and (D), $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.



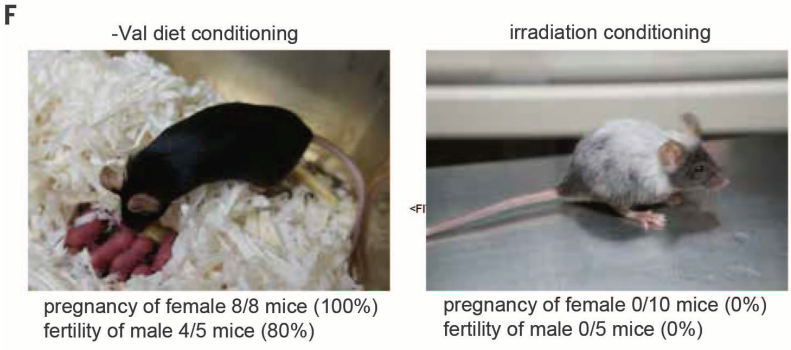


Fig. 3. Valine starvation permits HSC transplantation without irradiation. (**A**) Nonirradiative HSC transplantation into immunocompromised mice. NOD/scid mice fed a complete diet or a –Val diet for 2 weeks were transplanted, with 5×10^3 B6 donor KSL cells isolated, without irradiative conditioning. Complete diets were then returned, and donor PB chimerism was monitored. Engraftment is defined as >1% donor chimerism. (**B**) Donor chimerism after nonirradiative transplantation in NOD/scid mice, as outlined in (A) (complete group, $n = 5$; –Val group, $n = 6$). Error bars indicate ±SD. (**C**) Representative images of the thymus (yellow dotted line) in recipient mice at 12 weeks after nonirradiative transplantation. (**D**) Nonirradiative HSC transplantation into congenic mice. Ly5.2 mice fed a complete diet or a –Val diet for 3 weeks were transplanted with 1×10^7 Ly5.2 whole BM cells, without irradiative conditioning. Complete diets were then gradually returned over 2 weeks, and donor PB chimerism was monitored. Engraftment is defined as >1% donor chimerism. (**E**) Donor chimerism after nonirradiative transplantation, as outlined in (D) (complete group, $n = 5$; –Val group, $n = 5$). (**F**) Images of mice fed a –Val diet (left) or irradiation conditioned (right) 4 months after transplantation.

independent methods to define immunophenotypic HSCs) (*8*) were markedly decreased in mice fed a –Val diet (Fig. 2B and fig. S6, A to C). However, the proportion of HSCs in G_0 did not differ (fig. S6, D and E). Within the PB, numbers of $B220^+$ B cells and $CD3^+$ T cells significantly decreased (fig. S7, A and B). Additionally, cellular frequency decreased for BM B cells, common myeloid progenitor cells, and common lymphoid progenitor cells but not for BM T cells, granulocyte-monocyte progenitor cells, or megakaryocyte-erythroid progenitor cells (*9*) (figs. S7, C to E, and S8). These in vivo data indicate that HSCs display greater sensitivity than HPCs to dietary valine but that other cell populations are also affected.

To characterize the systemic effects of the valine-free diet, we histologically evaluated all organs and tissues. As expected, the BM was hypocellular, and the spleen and thymus were atrophic (fig. S9, A to C). We also observed decreased hair follicle density and a marked increase in brown fat cells in several tissues (fig. S9D). However, no obvious changes were evident in the brain, heart, lung, kidney, stomach, pancreas, or testis. Switching mice back to a complete diet brought a rapid recovery in body weight and in lymphoid populations and tissues (figs. S8 and S10, A and B). Phenotypic HSC frequencies also recovered, as well as BM HPC and $lineage^+$ populations (fig. S10, C to E). Half of the mice fed a –Val diet for 4 weeks died within days of resuming a complete diet. Autopsy identified pulmonary edema, with thickened alveolar walls and fluid-filled alveolar lumina, consistent with refeeding syndrome (*10*) (fig. S10F). Refeeding syndrome is well described in human patients, where refeeding following prolonged fasting or malnourishment can cause fatality.

Next, competitive transplantation assays were performed to determine HSC function following in vivo valine deficiency (Fig. 2C). HSCs from mice fed a –Val diet for 4 weeks were transplanted into lethally irradiated recipients, along with equal numbers of recipient BM cells. Donor chimerism was only slightly decreased for –Val cells 4 to 8 weeks after transplantation but fell sharply at 12 weeks (Fig. 2D and fig. S10G). These data indicate that dietary valine restriction depletes long-term repopulating HSCs rather than shorter-lived HPCs, consistent with our phenotypic and in vitro analyses.

Because mice fed a –Val diet had far fewer BM HSCs, we hypothesized that dietary valine restriction could empty the HSC niche and thereby replace the myeloablative conditioning regimens normally required for successful HSCT. To test this hypothesis, we set up a mouse model of allogenic HSC transplantation in immunodeficient (NOD/Scid) recipients (Fig. 3A). After transplantation of 5000 $Kit^+Sca^+Lin^-$ cells, all mice on the –Val diet survived and displayed good engraftment (Fig. 3B and fig. S11A), with markedly

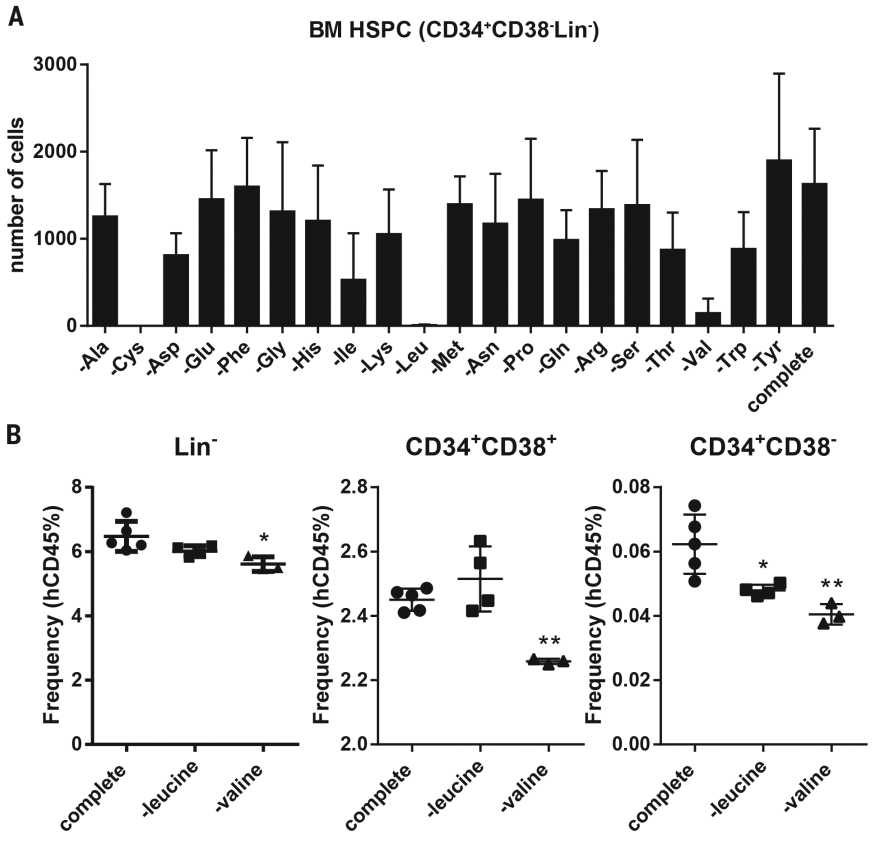


Fig. 4. Human HPSCs are sensitive to AA restriction. (**A**) One hundred human HSPCs were cultured with BSA and SCF/TPO/Flt3L/IL-3/IL-6 in complete medium (containing all AAs) or medium lacking single AAs and were counted after 7 days. Experiments were performed in triplicate. Error bars indicate ±SD. (**B**) Analysis of HSPC fractions (Lin^-, $CD34^+CD38^+$, and $CD34^+CD38^-$) in humanized mice fed a complete (n = 5), –Leu (n = 4), or –Val (n = 3) diet for 2 weeks. *P < 0.05; **P < 0.01.

increased thymus size (Fig. 3C) and gain of discrete splenic white and red pulp (fig. S11B), indicating lymphoid reconstitution.

We also transplanted 1×10^7 donor BM cells into nonirradiated congenic mice fed a –Val diet for 3 weeks. To avoid refeeding syndrome-associated mortality, valine was gradually returned to the diet over 2 weeks after the transplantation (Fig. 3D). Although no engraftment occurred in mice fed a complete diet, 5 of 5 mice displayed long-term donor chimerism following the –Val diet over 3 months after transplantation (Fig. 3E). Two further independent experiments using the same strategy showed long-term engraftment in 5 of 5 and 3 of 5 mice, and no mice died (fig. S11, C and D). HSC engraftment was confirmed by secondary transplantation assays (fig. S11E). Irradiation of recipient mice in HSC transplantation assays causes growth retardation, hair loss, and infertility, and even successfully engrafted mice usually do not survive more than a year after transplantation. By contrast, mice that received a –Val pretreatment diet appeared healthy, were fertile, and lived more than a year (Fig. 3F). Nonirradiative engraftment could also be achieved when mice were switched back to a complete diet immediately rather than gradually, but this led to significant mortality (only 10 of 26 mice survived over the long term) due to refeeding syndrome (fig. S12).

To learn whether similar AA requirements existed for human hematopoietic stem and progenitor cells (HSPCs), we used various AA-deficient media to assess the colony-forming ability of human BM $CD34^+CD38^-$ cells (*11*, *12*) (Fig. 4A and fig. S1D). As with mouse HSCs, no colonies formed under –Cys conditions, and Val deficiency suppressed cellular proliferation. Notably, –Leu conditions also decreased human HSPC proliferation (Fig. 4A).

We also tested –Val and –Leu diets on human hematopoiesis using NOD/Shi-scid/IL-2Rγ^{null} (NOG) mice reconstituted with BM $CD34^+$ cells at nearly 100% human chimerism (*11*, *12*). In mice that received –Val or –Leu diets, absolute numbers of human PB WBCs, BM cells, and $CD34^+CD38^-$ cells decreased (Fig. 4B and fig. S11, F and G). The –Val diet also significantly reduced the human Lin^- and $CD34^+CD38^+$ cell frequency, a finding that was not observed after administration of the –Leu diet (Fig. 4B). These data suggest that valine is also indispensable for human HSPC maintenance and that only valine deprivation exerts a consistent effect on mouse and human HSCs. However, the reason for this valine specificity is not clear. An AA imbalance, rather than changes in the level of a single AA, may exert such an effect in vivo, as previously proposed (*13*, *14*).

Our data indicate that HSC homeostasis relies on valine. Presumably, BM AA levels are optimal for HSC maintenance and are regulated by niche components. We therefore sought to identify AA secreted by representative niche candidates in vitro (*3*, *15*, *16*). The vascular endothelium specifically secreted Thr, Glu, Val, Leu, Ala, and Lys, whereas $PDGFR^+$ cells (platelet-derived growth factor receptor–positive) secreted Thr, Tyr, and Val (fig. S13). These results suggest that secretion of valine, and of other AAs, may be an important function of BM niche cells.

Our study highlights an important role for the AA microenvironment in hematopoiesis. Although the mechanism underlying the HSC-specific valine sensitivity remains elusive, these findings suggest a way to further understand stem cell niche functions. In addition, dietary valine restriction may be applicable for pretreatment in HSCT and the treatment of hematological malignancies.

REFERENCES AND NOTES

1. A. Czechowicz, D. Kraft, I. L. Weissman, D. Bhattacharya, *Science* **318**, 1296–1299 (2007).
2. B. Gyurkocza, B. M. Sandmaier, *Blood* **124**, 344–353 (2014).
3. S. J. Morrison, D. T. Scadden, *Nature* **505**, 327–334 (2014).
4. A. Kornberg, *J. Biol. Chem.* **164**, 203–212 (1946).
5. A. Kornberg, F. S. Daft, W. H. Sebrell, *Science* **103**, 646–648 (1946).
6. H. Ema *et al.*, *Nat. Protoc.* **1**, 2979–2987 (2007).
7. M. H. Stipanuk, *Annu. Rev. Nutr.* **6**, 179–209 (1986).
8. M. J. Kiel *et al.*, *Cell* **121**, 1109–1121 (2005).
9. K. Akashi, D. Traver, T. Miyamoto, I. L. Weissman, *Nature* **404**, 193–197 (2000).
10. H. M. Mehanna, J. Moledina, J. Travis, *BMJ* **336**, 1495–1498 (2008).
11. A. Larochelle *et al.*, *Nat. Med.* **2**, 1329–1337 (1996).
12. M. Bhatia, J. C. Wang, U. Kapp, D. Bonnet, J. E. Dick, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5320–5325 (1997).
13. W. A. Krehl, L. J. Teply, P. S. Sarma, C. A. Elvehjem, *Science* **101**, 489–490 (1945).
14. A. E. Harper, *Ann. N. Y. Acad. Sci.* **69**, 1025–1041 (1958).
15. J. Mitoma, S. Furuya, Y. Hirabayashi, *Neurosci. Res.* **30**, 195–199 (1998).
16. S. Furuya *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11528–11533 (2000).

ACKNOWLEDGMENTS

We thank T. Endo, Y. Yamazaki, Y. Ishii, and K. Chan for technical support; J. E. Dick, R. Yamamoto, and A. Knisely for critical manuscript reading; and S. Satomi, M. Chin, T. Kobayashi, K. Sudo, P. van Galen. N. Mbong, G. Karlsson, A. Reinisch, D. Thomas, K. Takubo, A. Iwama, K. Miharada, M. Otsu, H. Ema, L. Zon, and N. Chandel for advice. This work was supported by grants from JST; the Ministry of Education, Culture, Sport, Science and Technology (Japan); the Japan Society for the Promotion of Science; California Institute of Regenerative Medicine; the Siebel Foundation; and the Ludwig Foundation. H.N., S.Y., and Y.T. are inventors on a patent application (PCT/JP2015/083088) submitted by the University of Tokyo that covers the composition for decreasing HSCs and methods for producing these cells.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1152/suppl/DC1
Materials and Methods
Figs. S1 to S13

8 June 2016; accepted 28 September 2016
Published online 20 October 2016
10.1126/science.aag3145

HEMATOPOIESIS

Self-renewal of a purified *Tie2*$^+$ hematopoietic stem cell population relies on mitochondrial clearance

Kyoko Ito,[1,2,3] Raphaël Turcotte,[4] Jinhua Cui,[1,5]* Samuel E. Zimmerman,[1,6]* Sandra Pinho,[1,5]* Toshihide Mizoguchi,[1,5] Fumio Arai,[7] Judith M. Runnels,[4] Clemens Alt,[4] Julie Teruya-Feldstein,[8] Jessica C. Mar,[1,6,9] Rajat Singh,[2,10,11] Toshio Suda,[7,12] Charles P. Lin,[4] Paul S. Frenette,[1,2,3,5] Keisuke Ito[1,2,3,5,11]†

A single hematopoietic stem cell (HSC) is capable of reconstituting hematopoiesis and maintaining homeostasis by balancing self-renewal and cell differentiation. The mechanisms of HSC division balance, however, are not yet defined. Here we demonstrate, by characterizing at the single-cell level a purified and minimally heterogeneous murine *Tie2*$^+$ HSC population, that these top hierarchical HSCs preferentially undergo symmetric divisions. The induction of mitophagy, a quality control process in mitochondria, plays an essential role in self-renewing expansion of *Tie2*$^+$ HSCs. Activation of the PPAR (peroxisome proliferator–activated receptor)–fatty acid oxidation pathway promotes expansion of *Tie2*$^+$ HSCs through enhanced Parkin recruitment in mitochondria. These metabolic pathways are conserved in human TIE2$^+$ HSCs. Our data thus identify mitophagy as a key mechanism of HSC expansion and suggest potential methods of cell-fate manipulation through metabolic pathways.

Precise mechanisms enable efficient and regulated hematopoietic stem cell (HSC) renewal versus differentiation, and these processes are coordinated. HSCs must withstand various stresses to maintain hematopoiesis for the duration of their life cycles. The repair or clearance of damage is critical to precisely controlling their cell fates and maintaining stemness upon division, especially during HSC expansion through symmetric self-renewing division. HSC exhaustion can result from defective cellular metabolism and/or impaired autophagy (*1*–*6*), a lysosomal degradation pathway that breaks down damaged or unwanted proteins and/or organelles (*7*). The FOXO3A (forkhead box O3a)–driven proautophagy program, for example, protects HSCs from metabolic stress (*5*). In a similar process in mammary epithelial stem-like cells, older mitochondria are pushed into daughter cells fated to differentiate by asymmetric division (AD), thus maintaining high-quality stem cell homeostasis (*8*). Despite extensive study of HSC renewal and differentiation, however, much remains unclear regarding the mechanisms of division balance due to the heterogeneity of HSC-enriched fractions.

In searching for a potential HSC marker, we observed high *Tie2* levels in the CD34$^-$ HSC-enriched compartment (fig. S1A). We therefore established a *Tie2*–green fluorescent protein (GFP) reporter line (*Tie2*-reporter mice or *Tie2* Tg) to validate *Tie2*-expression as an HSC marker (fig. S1, B and C). *Tie2*-GFP was specifically enriched in the HSC fraction but was reduced during differentiation, in parallel with mRNA levels (fig. S1, D to G). *Tie2*-GFP$^+$ and -GFP$^-$ fractions in CD34$^-$CD150$^+$CD48$^{low/-}$CD135$^-$KSL cells were purified from *Tie2* reporter mice (hereafter, *Tie2*-GFP$^+$ or *Tie2*-GFP$^-$ HSCs), and *Tie2* levels were detected in *Tie2*-GFP$^+$ HSCs but not in *Tie2*-GFP$^-$ HSCs (fig. S1H). The cell cycle kinetics (*9*) revealed that cell cycling was slightly slower in *Tie2*-GFP$^+$ versus *Tie2*-GFP$^-$ HSCs (fig. S1I). A higher proportion of phenotypic *Tie2*-GFP$^+$ HSCs than *Tie2*-GFP$^-$ HSCs was closely associated with arteriolar structure (fig. S1J) and expressed endothelial and HSC markers (*10*, *11*) (fig. S1, K to M).

In addition, the *Tie2*-GFP$^+$ fraction displayed a highly significant (5.75-fold) increase in long-term culture–initiating cell (LTC-IC) frequency relative to the *Tie2*-GFP$^-$ fraction (Fig. 1A). Single-cell transplantation showed that a high percentage of single *Tie2*-GFP$^+$ cells (68.0%) exhibited reconstitution capacity without lineage bias and maintained high donor chimerism upon secondary transplantation. Donor contribution was observed in all recipients transplanted with three *Tie2*-GFP$^+$ HSCs, whereas 40% showed reconstitution capacity with three *Tie2*-GFP$^-$ HSCs (Fig. 1B and fig. S2, A to D), demonstrating that *Tie2*-GFP identifies an HSC fraction with minimal heterogeneity. To accurately supply the bone marrow with individual HSCs, we employed multiphoton microscopy guidance (local transplantation) (fig. S2E) to deliver single *Tie2*-GFP$^+$ HSCs near an opening to the marrow of live mice, and we then tracked these cells during subsequent homing to locations some distance from the delivery site (Fig. 1C and movies S1 and S2). Recipients showed reconstitution from locally transplanted single (100%) and multiple *Tie2*-GFP$^+$ HSCs (fig. S2F). A single *Tie2*-GFP$^+$ HSC can exhibit reconstitution capacity in a recipient mouse for more than 6 months, once it is located in the bone marrow.

We also tested, in vivo and in vitro, whether *Tie2* acts as an HSC marker under stress conditions. Of the polyinosinic:polycytidylic acid (pIpC, inducing the interferon response)–treated single *Tie2*-GFP$^+$ HSCs, 65.0% showed repopulation capacity at levels comparable to those seen in control mice (fig. S3, A to D). But whereas the *Tie2*-GFP$^+$/Tie2-Ab$^+$ fraction exhibited high reconstitution capacity, neither the *Tie2*-GFP$^-$/Tie2-Ab$^+$ fraction nor any of the 12 single *Tie2*-GFP$^+$ KSL cells showed high reconstitution. The majority of the re-sorted *Tie2*-GFP$^+$ cells (70.6% after 3-day culture; 85.7% after 7-day culture) exhibited multilineage reconstitution, though *Tie2*-GFP$^-$ cells did not. None of 90 *Tie2*-GFP$^-$ HSCs provided donor-derived *Tie2*-GFP$^+$ HSCs in recipient mice. These data suggest that *Tie2* serves as a useful marker that can prospectively identify HSCs (figs. S2G and S3, E to L).

Single-cell gene expression assays showed that a substantial number of genes—including HSC factors and markers (e.g., *Tie2*, *Hif1a*, *Mpl*, and *Foxo3a*), as well as *Pml* (*promyelocytic leukemia*) and its downstream signaling pathways, which regulate fatty acid oxidation (FAO) via *Cpt1a* (*carnitine palmitoyltransferase 1a*), *Ppard* (*peroxisome proliferator-activated receptor-delta*), and *Acox1* (*acyl–coenzyme A oxidase1*)—were markedly higher in the *Tie2*-GFP$^+$ fraction than the *Tie2*-GFP$^-$ fraction (Fig. 1D). Hierarchical clustering, HSC signature, and signatures for both PPAR signaling and fatty acid metabolism demonstrate that *Tie2*-GFP$^+$ HSCs are molecularly farther away from both KSL cells and *Tie2*-GFP$^-$ HSCs (fig. S4). The functions of the PPAR-FAO pathway in *Tie2*-GFP$^+$ HSCs were then assessed. Impaired reconstitution ability of single *Pml*$^{-/-}$ *Tie2*-GFP$^+$ HSCs could be rescued by GW501516, a PPARδ-specific agonist (fig. S5, A to C). Thus the PPAR-FAO pathway plays critical roles in directly regulating the self-renewal of the top hierarchical HSCs.

Mitochondria are central metabolic organelles responsible for the final steps of FAO, which cleaves two carbons every cycle to form acetyl–coenzyme A. Once the PPAR-FAO pathway is activated, fatty acids have been imported into

[1]Ruth L. and David S. Gottesman Institute for Stem Cell and Regenerative Medicine Research, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [2]Department of Medicine, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [3]Albert Einstein Cancer Center, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [4]Center for Systems Biology, Advanced Microscopy Program, Wellman Center for Photomedicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA. [5]Departments of Cell Biology and Stem Cell Institute, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [6]Department of Systems and Computational Biology, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [7]Department of Cell Differentiation, The Sakaguchi Laboratory of Developmental Biology, School of Medicine, Keio University, Japan. [8]Department of Pathology, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA. [9]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [10]Department of Molecular Pharmacology, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [11]Diabetes Research Center, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [12]Cancer Science Institute of Singapore, National University of Singapore, Singapore. *These authors contributed equally to this work. †Corresponding author. Email: keisuke.ito@einstein.yu.edu

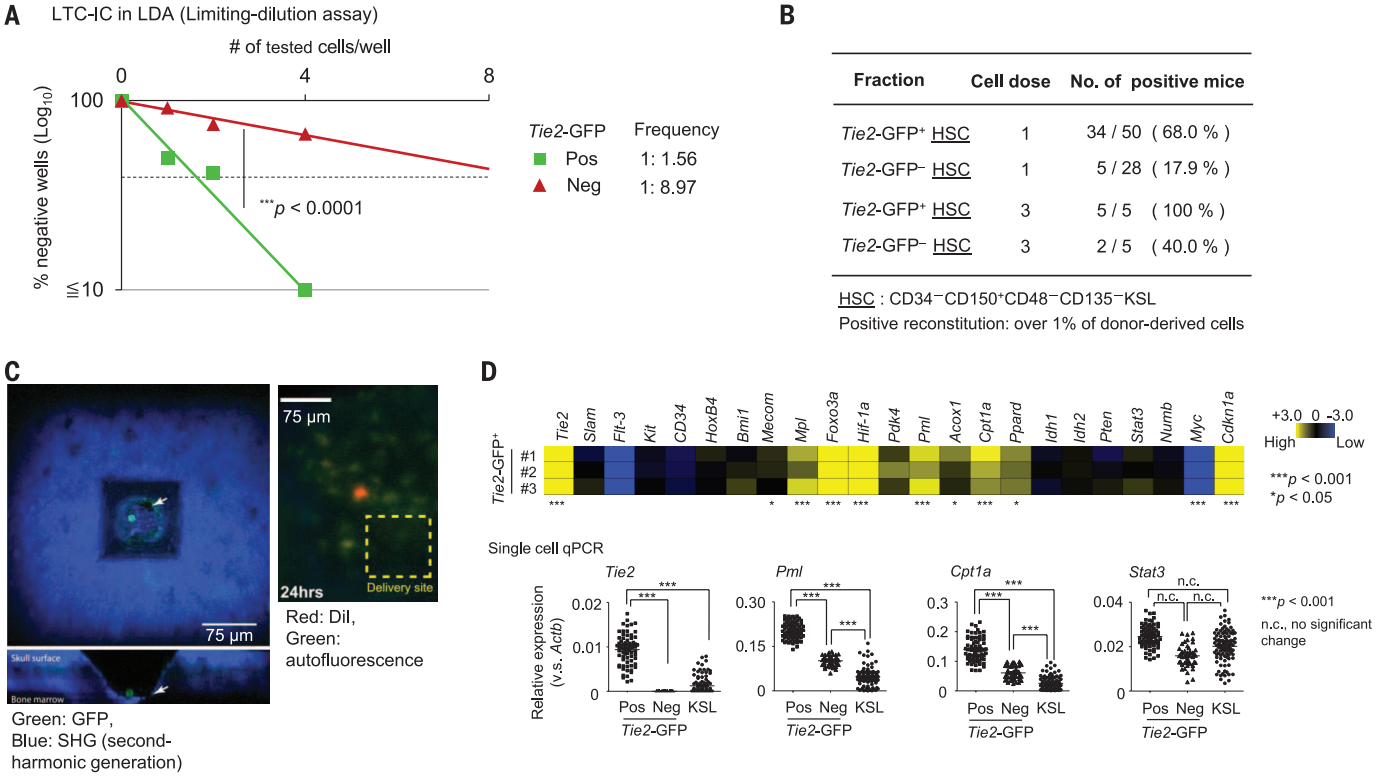


Fig. 1. *Tie2*-GFP marks a distinct HSC subset. (**A**) LTC-IC frequencies of *Tie2*-GFP$^+$ and *Tie2*-GFP$^-$ HSCs were determined in a limiting-dilution assay (LDA). (**B**) Percentages of successful long-term reconstitution. (**C**) In vivo imaging of single *Tie2*-GFP$^+$ HSCs immediately after local delivery in the calvarium. (Left) Top (upper) and side (lower) views of the delivery site. The arrow indicates the opening to the bone marrow. (Right) In vivo imaging of the same cell, 24 hours after local delivery (en face maximum intensity projection). (**D**) Single-cell gene expression analysis using the BioMark System array (Fluidigm) (see materials and methods).

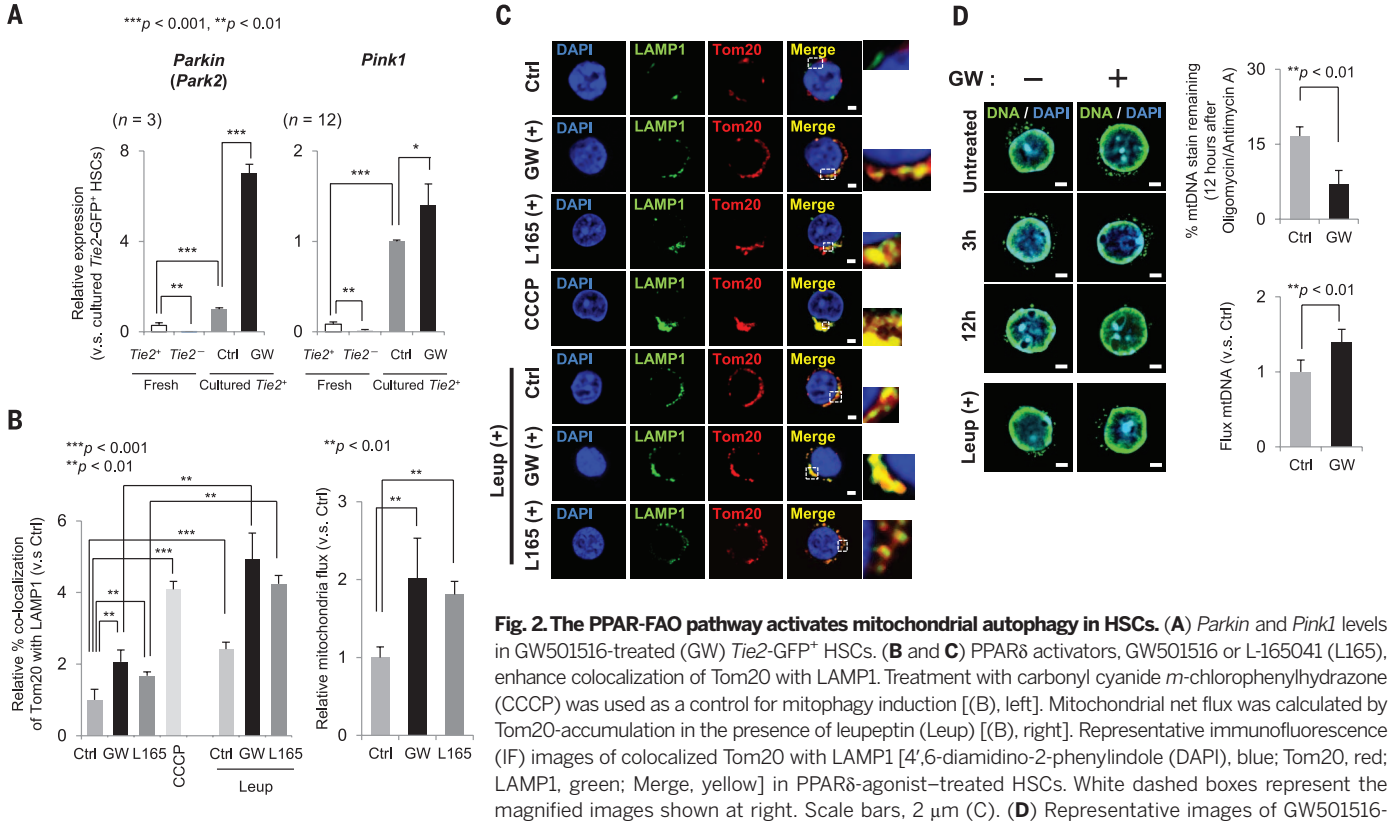


Fig. 2. The PPAR-FAO pathway activates mitochondrial autophagy in HSCs. (**A**) *Parkin* and *Pink1* levels in GW501516-treated (GW) *Tie2*-GFP$^+$ HSCs. (**B** and **C**) PPARδ activators, GW501516 or L-165041 (L165), enhance colocalization of Tom20 with LAMP1. Treatment with carbonyl cyanide *m*-chlorophenylhydrazone (CCCP) was used as a control for mitophagy induction [(B), left]. Mitochondrial net flux was calculated by Tom20-accumulation in the presence of leupeptin (Leup) [(B), right]. Representative immunofluorescence (IF) images of colocalized Tom20 with LAMP1 [4′,6-diamidino-2-phenylindole (DAPI), blue; Tom20, red; LAMP1, green; Merge, yellow] in PPARδ-agonist–treated HSCs. White dashed boxes represent the magnified images shown at right. Scale bars, 2 μm (C). (**D**) Representative images of GW501516-treated HSCs immunostained to label mtDNA (left) and quantified for mitophagy [12 hours oligomycin and antimycin A] (right top) and net mtDNA flux (right bottom). In (A), (B), and (D), error bars represent mean ± SEM (at least 10 cells analyzed per each condition per each experiment).

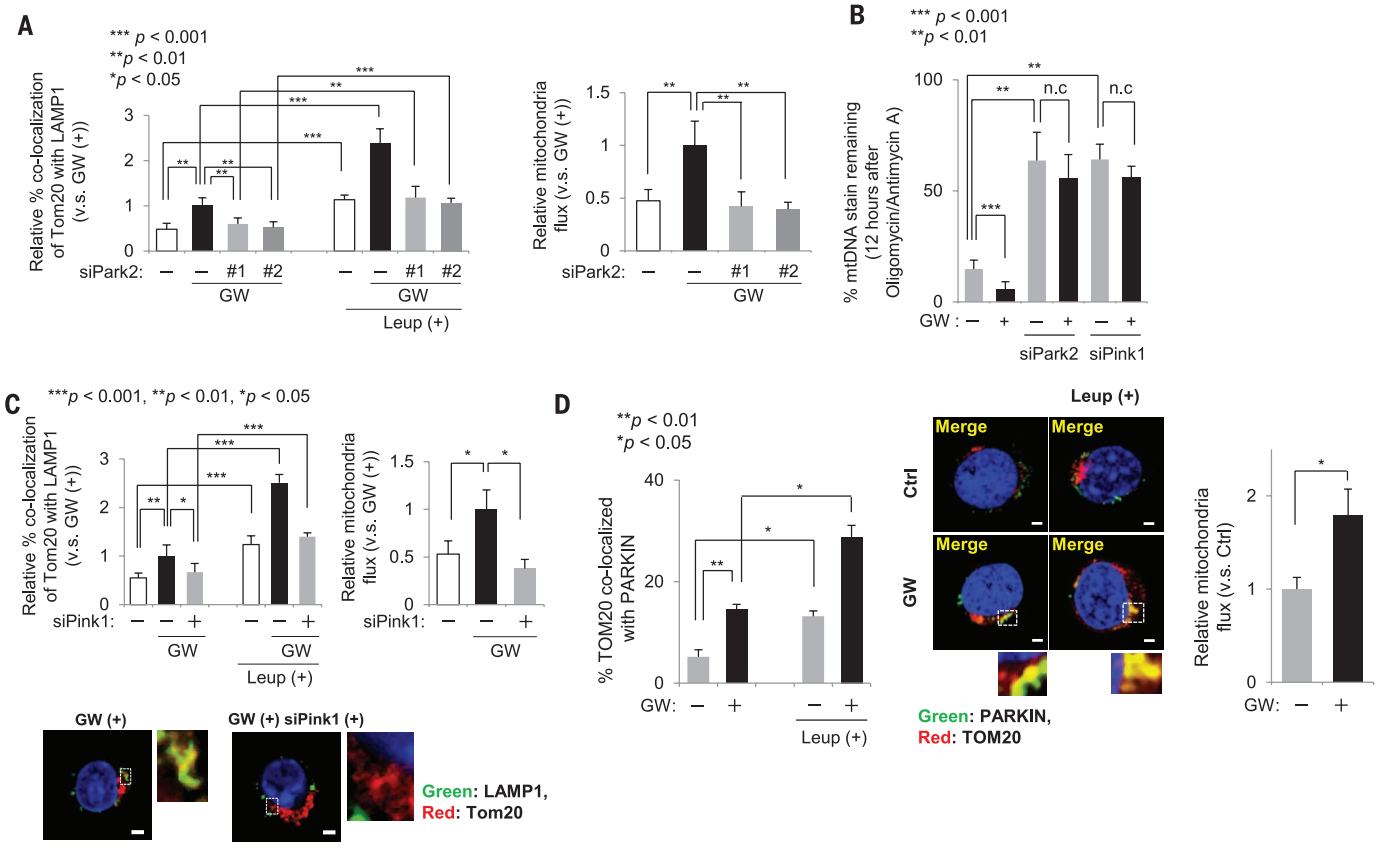


Fig. 3. Parkin is important for PPAR-FAO–induced mitophagy in HSCs. (**A**) Relative Tom20 colocalization (left) and mitochondria net flux (right) in GW501516-treated HSCs in the presence of siPark2. (**B**) Quantification of mitophagy based on the remaining mtDNA staining 12 hours after treatment with oligomycin and antimycin A. (**C**) *Pink1*-silencing reduces PPARδ-enhanced mitophagy. Relative Tom20 colocalization (top left), net mitochondria flux (top right), and representative IF images depicting LAMP1 and Tom20 colocalization in GW501516-treated HSCs in the presence of siPink1 (bottom). (**D**) GW501516 enhances colocalization of TOM20 with PARKIN (left) and increases mitochondrial net flux (right). Scale bars in (C) and (D), 2 μm.

the mitochondrial matrix (*12*). Our gene expression assays revealed higher expression of mitochondrial autophagy (mitophagy)–related genes, including *Parkin* (*Park2*), *Pink1* (*PTEN-induced putative kinase 1*), *Optineurin*, *Tom 7*, *Map1lc3a* (*Lc3*), and *p62/Sqstm1*, in *Tie2*-GFP$^+$ HSCs at steady-state conditions. Most of these genes were further up-regulated by activation of the PPAR-FAO pathway (Fig. 2A and fig. S5, D to F). To assess whether mitophagy-related genes are regulated by the PPAR-FAO pathway at the transcriptional level, *Pink1*-luciferase reporters were used. A PPARδ agonist transactivated *Pink1* reporters, which were reduced by FAO inhibition or mutation of FOXO-response elements. *Foxo3a* was up-regulated in HSCs upon treatment with a PPARδ agonist, and this process was attenuated by the ablation of *Ppard* or *Cpt2*, which are FAO-related genes. These data suggest that the PPAR-FAO pathway transcriptionally regulates PINK1, in part through FOXO3a (fig. S6).

We then sought to determine whether PPARδ-induced mitophagy correlates with a specific metabolic state or profile. Cellular FAO activity was specifically enhanced after PPARδ agonist treatment and was accompanied by Pink1, whereas glucose uptake or levels of pyruvate and tricarboxylic acid cycle metabolites were not altered (fig. S7, A to D). PPARδ agonists also enhanced mitophagy, with the increased colocalization of Tom20 (translocase of outer membrane 20) with LAMP1 (lysosomal-associated membrane protein 1). Flux analyses were then performed to quantify mitochondrial turnover in lysosomes (the organelles where mitophagy occurs) by evaluating the accumulation of Tom20 in the presence of leupeptin, an inhibitor of lysosomal proteolysis. Pharmacological PPAR-FAO activation enhanced the net mitochondrial flux toward lysosomal degradation. These data strongly suggest that the PPAR-FAO pathway activates mitophagy (Fig. 2, B and C).

Other findings supported the enhancement of mitophagy by a PPAR agonist, including increased colocalization of Parkin with pyruvate dehydrogenase (PDH), a matrix mitochondrial protein. Mitophagy enhancement was again confirmed in GW501516-treated HSCs by the increased colocalization of Parkin with both Tom20 and the autophagosome marker LC3 (figs. S7E and S8, A to D). To explore another mitophagy indicator, mitochondrial DNA (mtDNA) nucleoids were quantified after mitochondrial damage with oligomycin and antimycin A. GW501516-treated HSCs were nearly devoid of mtDNA and exhibited increased net mtDNA flux, whereas mtDNA was still partially retained in control HSCs (Fig. 2D). Thus, damaged mitochondria are cleared more quickly due to enhanced mitophagy activation in *Tie2*-GFP$^+$ HSCs treated with a PPAR agonist.

Hematopoiesis-specific *Ppard* conditional knockout mice were then generated. *Ppard*-deleted HSCs exhibited neither enhanced Tom20 and Parkin colocalization nor increased mitochondrial net flux after PPARδ agonist treatment, although these agonists stimulated mitophagy in wild-type HSCs, and a mitochondrial uncoupler, carbonyl cyanide *m*-chlorophenylhydrazone (CCCP), induced mitophagy in both genotyped HSCs (fig. S8E). Conditional heterozygous deletion of *Cpt2* partially but significantly attenuated the enhanced Parkin recruitment in mitochondria induced by a PPAR agonist (fig. S8F). Taken together, these direct genetic approaches reveal that PPARδ agonists stimulate mitophagy in a PPAR-FAO–dependent manner.

To test whether mitophagy is critical for *Tie2*-GFP$^+$ HSC expansion, *Parkin* and *Pink1* were first silenced by RNA interference–mediated knockdown. siPark2 decreased net mitochondrial flux (a hallmark of mitophagy) in GW501516-treated HSCs, as indicated by decreased colocalization of Tom20 and PDH with LAMP1, whereas cell death, proliferation rate, FAO activity, and *Cdk6* levels were not altered (Fig. 3A and figs. S7A and S9). *Parkin*-knockdown in GW501516-treated HSCs led to accumulating signals of Tom20 and PDH,

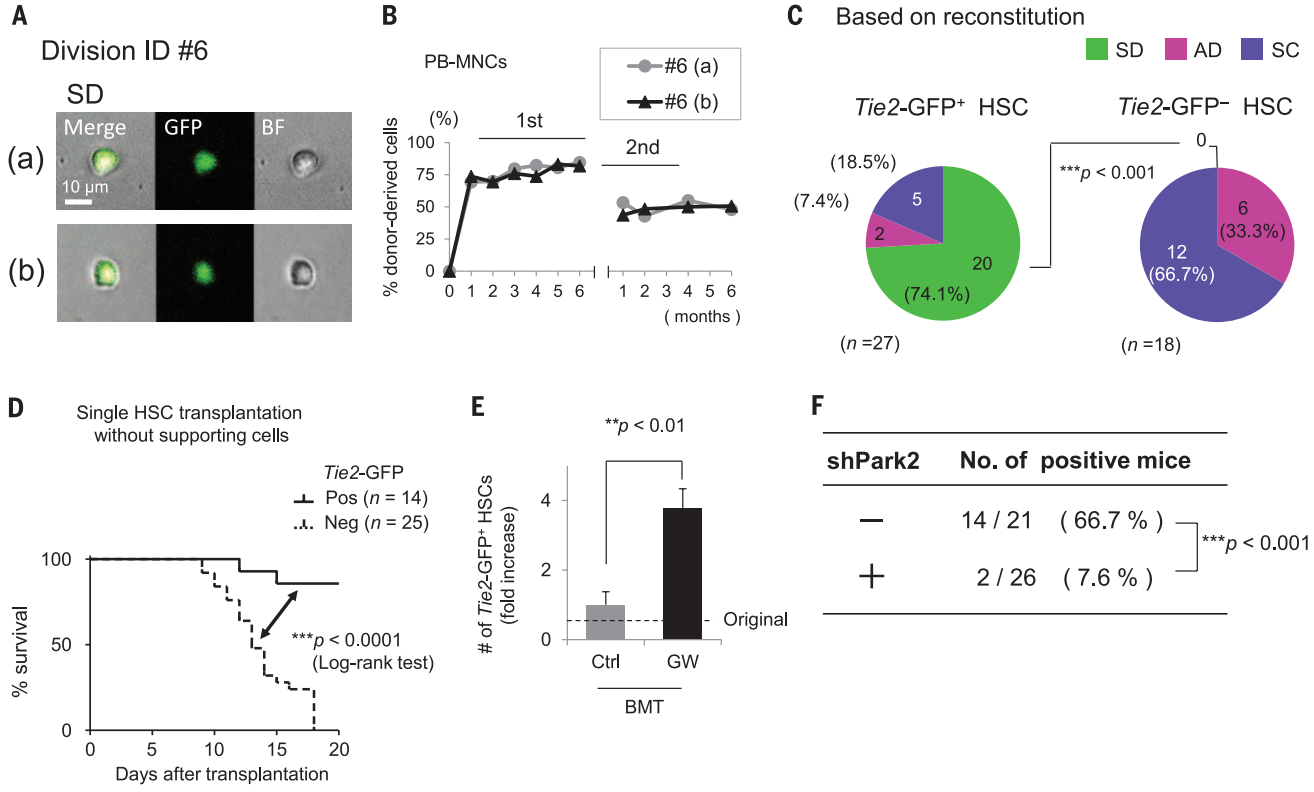


Fig. 4. Self-renewing expansion of *Tie2*$^+$ HSC through mitophagy. (**A**) Representative pictures of *Tie2*-GFP positivity, with bright-field images (BF) collected during symmetric division. (**B**) Representative repopulation kinetics of SD during the first and second bone marrow transplantation (BMT). PB-MNCs, peripheral blood mononuclear cells. (**C**) Division patterns of *Tie2*-GFP$^+$ HSCs (left) and *Tie2*-GFP$^-$ HSCs (right), in terms of reconstitution capacity. SD, symmetric division; AD, asymmetric division; SC, symmetric commitment. (**D**) Single *Tie2*-GFP$^+$ HSC transplantation without supporting cells. Survival of recipient mice was examined by plotting Kaplan-Meier survival curves (also see fig. S13B). (**E**) PPARδ activation induces in vivo expansion of *Tie2*-GFP$^+$ HSCs. The number of transplanted *Tie2*-GFP$^+$ HSCs (donor cells) is denoted by the dashed line marked "Original" (also see fig. S13H). (**F**) Frequency of successful hematopoietic reconstitution of shParkin-silenced *Tie2*-GFP$^+$ HSCs (also see fig. S13J).

as well as reduced colocalization of LAMP1 with LC3 (fig. S10, A to C). mtDNA was retained in *Parkin*-silenced HSCs, and these cells were refractory to PPAR agonist treatment (Fig. 3B and fig. S10D). These data suggest that Parkin has an essential role in PPAR-FAO–induced mitophagy.

Because Pink1 is known to be essential for mitophagy and Pink1 levels were increased by PPAR-FAO activation, we next assessed the effect of *Pink1* knockdown on the process. *Pink1* silencing reduced Tom20 colocalization with both LAMP1 and Parkin and decreased net flux of mitochondria in GW501516-treated HSCs, accompanied by accumulating Tom20 signals. mtDNA was also retained in *Pink1*-silenced HSCs (Fig. 3, B and C, and fig. S10, D to G). Consequently, Parkin and Pink1 play important roles in PPAR-FAO–induced mitophagy in HSCs.

We then explored potential clinical applications of pharmacological FAO modulation in human hematopoiesis. *TIE2* expression is enriched in the HSC fraction, and TIE2-positivity enhances LTC-IC frequency and in vivo repopulation capacity in human bone marrow (fig. S11, A to G). GW501516 enhances LTC-IC frequency with mitophagy activation, whereas low doses of an FAO inhibitor reduce its capacity. Fatty acid metabolism can therefore control self-renewal capacity and the cell fates of human HSCs (Fig. 3D and fig. S11, H to J).

Finally, we explored the effects of FAO-induced mitophagy on *Tie2*-GFP$^+$ HSC expansion. In vivo paired daughter cell (PDC) assays can determine the division patterns of HSCs retrospectively, by assessing the reconstitution capacity of each daughter cell (*3*, *13*). In vivo PDC assays revealed that *Tie2*-GFP$^+$ HSCs preferentially undergo symmetric division (SD) (in 20 out of 27 divisions, or 74.1%, both daughter cells showed a reconstitution capacity) rather than AD or symmetric commitment [(SC), both daughter cells exhibited no reconstitution capacity], whereas *Tie2*-GFP$^-$ HSCs undergo either AD or SC (Fig. 4, A to C, and fig. S12). To further test the restoration capacity of hematopoiesis by single *Tie2*-GFP$^+$ HSCs, we performed transplantation assays under high-stress conditions, as are found in single-cell transplantation without supporting cells. The survival ratio after single *Tie2*-GFP$^+$ HSC transplantation was 85.7%, whereas no recipients of single *Tie2*-GFP$^-$ HSCs survived (Fig. 4D and fig. S13, A and B). Functional HSCs are present in the *Tie2*-GFP$^-$ fraction, but they fail to adequately respond to the demand for hematopoietic recovery under high-stress conditions without the enhanced symmetric expansion provided by *Tie2*-GFP$^+$ HSCs.

As expected, PPAR-FAO activation increased the number of *Tie2*-GFP$^+$ cells in vitro with an increased rate of SD (fig. S13, C to G). *Tie2*-GFP$^+$ HSC numbers were increased in comparison with the original number of transplanted donor cells in recipients and were further increased by PPARδ agonist treatment (about a fourfold increase). Furthermore, PPARδ agonist treatment expanded *Tie2*-GFP$^+$ HSCs under physiological stress, as induced by pIpC (Fig. 4E and figs. S3B and S13H).

However, silencing *Park2* or *Pink1* not only abrogated the expansion of *Tie2*-GFP$^+$ cells in vitro but also inhibited their maintenance (fig. S13I). Most single *Parkin*-silenced *Tie2*-GFP$^+$ cells failed to show reconstitution capacity (92.3%), whereas the majority of control *Tie2*-GFP$^+$ cells (66.7%) exhibited this capacity (Fig. 4F and fig. S13J). *Pink1*$^{-/-}$ HSCs failed to adequately respond to the demand for hematopoietic recovery under high-stress conditions, even though a PPARδ agonist was administered (fig. S13K). In conclusion, *Tie2*-GFP$^+$ HSCs can maintain stem cell potential during cell cycling due to high mitochondrial clearance through mitophagy and up-regulation of *Parkin* and *Pink1* (fig. S14). Our data from this

highly purified HSC fraction identifies mitochondrial clearance by induction of mitophagosome formation as a key mechanism in maintaining stemness.

REFERENCES AND NOTES

1. K. Ito, T. Suda, *Nat. Rev. Mol. Cell Biol.* **15**, 243–256 (2014).
2. N. Shyh-Chang, G. Q. Daley, L. C. Cantley, *Development* **140**, 2535–2547 (2013).
3. K. Ito *et al.*, *Nat. Med.* **18**, 1350–1358 (2012).
4. M. Mohrin *et al.*, *Science* **347**, 1374–1377 (2015).
5. M. R. Warr *et al.*, *Nature* **494**, 323–327 (2013).
6. M. Mortensen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 832–837 (2010).
7. L. Galluzzi, F. Pietrocola, B. Levine, G. Kroemer, *Cell* **159**, 1263–1276 (2014).
8. P. Katajisto *et al.*, *Science* **348**, 340–343 (2015).
9. F. Arai *et al.*, *Cell* **118**, 149–161 (2004).
10. M. Acar *et al.*, *Nature* **526**, 126–130 (2015).
11. Y. Kunisaki *et al.*, *Nature* **502**, 637–643 (2013).
12. W. H. Kunau, V. Dommes, H. Schulz, *Prog. Lipid Res.* **34**, 267–342 (1995).
13. R. Yamamoto *et al.*, *Cell* **154**, 1112–1126 (2013).

ACKNOWLEDGMENTS

We thank all members of the Ito lab and Einstein Stem Cell Institute for their comments on HSC self-renewal and M. Wolfgang, A. Carracedo, H. You, and the Einstein Flow Cytometry and Analytical Imaging core facilities (grant P30CA013330) for help and materials. This work was supported by NIH (grants R01DK98263 and R01DK100689 to Ke.I.) and NYSTEM (New York State Stem Cell Science single-cell-core, grant C029154 to Ke.I.), Harvard Stem Cell Institute (to C.P.L.), NIH and Ellison Medical Foundation (to R.S.), NIH and Leukemia Lymphoma Society (to P.S.F.), and Japan Society for the Promotion of Science (to T.S.). We declare no competing financial interests.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1156/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
References (*14*–*59*)
Movies S1 and S2

26 February 2016; accepted 4 October 2016
Published online 13 October 2016
10.1126/science.aaf5530

T CELL EXHAUSTION

Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade

Kristen E. Pauken,[1] Morgan A. Sammons,[2] Pamela M. Odorizzi,[1] Sasikanth Manne,[1] Jernej Godec,[3,4] Omar Khan,[1] Adam M. Drake,[2] Zeyu Chen,[1] Debattama R. Sen,[3] Makoto Kurachi,[1] R. Anthony Barnitz,[3] Caroline Bartman,[1] Bertram Bengsch,[1] Alexander C. Huang,[5] Jason M. Schenkel,[6] Golnaz Vahedi,[7] W. Nicholas Haining,[3,8,9] Shelley L. Berger,[2] E. John Wherry[1]*

Blocking Programmed Death–1 (PD-1) can reinvigorate exhausted CD8 T cells (T_{EX}) and improve control of chronic infections and cancer. However, whether blocking PD-1 can reprogram T_{EX} into durable memory T cells (T_{MEM}) is unclear. We found that reinvigoration of T_{EX} in mice by PD-L1 blockade caused minimal memory development. After blockade, reinvigorated T_{EX} became reexhausted if antigen concentration remained high and failed to become T_{MEM} upon antigen clearance. T_{EX} acquired an epigenetic profile distinct from that of effector T cells (T_{EFF}) and T_{MEM} cells that was minimally remodeled after PD-L1 blockade. This finding suggests that T_{EX} are a distinct lineage of CD8 T cells. Nevertheless, PD-1 pathway blockade resulted in transcriptional rewiring and reengagement of effector circuitry in the T_{EX} epigenetic landscape. These data indicate that epigenetic fate inflexibility may limit current immunotherapies.

Persisting antigenic stimulation during chronic infections and cancer can result in T cell exhaustion, a state of impaired effector functions, high expression of inhibitory receptors including Programmed Death–1 (PD-1, or CD279), transcriptional reprogramming, and defective immune memory (*1*). Collectively, these properties prevent optimal control of persisting pathogens and tumors. Blocking the PD-1:PD-L1 pathway can reinvigorate exhausted CD8 T cells (T_{EX}), improving effector functions and enhancing viral and tumor control (*1*). Recently developed inhibitors of the PD-1 and cytotoxic T lymphocyte–associated protein 4 (CTLA-4) pathways represent a new paradigm in cancer treatment (*2*–*4*). Although promising, the majority of patients fail to develop durable responses, and most eventually progress (*2*–*4*). Thus, it is unclear whether blocking PD-1 can promote long-lasting improvements and immunological memory development in T_{EX}.

To address this question, we analyzed the cellular, transcriptional, and epigenetic changes associated with PD-1 pathway blockade using the mouse model of chronic lymphocytic choriomeningitis virus (LCMV) infection (fig. S1, A to C) (*5*, *6*). After treatment with antibodies against PD-L1 (anti-PD-L1), 1080 genes were up-regulated and 1686 genes were down-regulated [$P < 0.05$, $\log_2$ fold change (LFC) ≥ 0.2] (Fig. 1A, fig. S1D, and table S1). Previous studies identified transcriptional (*7*) or cellular (*8*, *9*) changes in metabolic pathways after PD-1 pathway blockade. Indeed, several metabolic genes were altered following PD-L1 blockade (table S1). Gene Set Enrichment Analysis (GSEA), however, identified more prominent changes in cell division pathways (Fig. 1B and table S2) (*5*, *10*). In addition, many effector-related genes were biased toward the anti-PD-L1 group (Fig. 1, C and D, and table S3). Other genes of interest included *Cxcl9*, *Il1r2*, and *Il7r* (up-regulated) and *Klra9*, *Tnfrsf9*, and *Cd200r2* (down-regulated) (fig. S1D and table S1). Using leading-edge metagene (LEM) analysis (*11*), we identified two metagenes in anti-PD-L1–treated T_{EX} compared to control T_{EX}; one corresponding to leukocyte activation and one to cell cycle (Fig. 1E; fig. S1, E and F; and table S4). The anti-PD-L1–treated T_{EX} metagenes displayed some overlap with effector T cells (T_{EFF}), largely driven by cell cycle pathways, but minimal overlap with T_{MEM} (Fig. 1E and table S4), suggesting limited acquisition of memory potential upon T_{EX} reinvigoration.

PD-1 pathway blockade can reactivate functions in T_{EX}, but whether reinvigoration is sustained is unclear. There was a robust reinvigoration of T_{EX}, as expected (Fig. 1, F and G, and figs. S1, A and B, and S2) (*5*), and expansion peaked ~3 weeks after initiation of blockade. By 8 to 11 weeks after treatment, however, this reinvigoration was lost, and the quantity, proliferation, effector function, and inhibitory receptor expression of LCMV-specific CD8 T cells in the anti-PD-L1–treated mice were comparable to those in control-treated mice (Fig. 1, F to H, and figs. S2 to S4). Moreover, although anti-PD-L1 treatment reduced viral load immediately after treatment, 4 months later, viral load was similar to that in control-treated mice (Fig. 1I). Lastly, 18 to 29 weeks after cessation of blockade, the transcriptional profiles of control- and anti-PD-L1–treated groups were similar (Fig. 1J, figs. S5 and S6, and tables S5 and S6). Collectively, these data indicate that when antigen concentration remains high, T_{EX} that were reinvigorated by PD-1 pathway blockade become "reexhausted."

One possible reason the effects of PD-L1 blockade were not sustained is that the infection persisted. We hypothesized that if the infection were cleared, anti-PD-L1 might induce differentiation into T_{MEM}.

[1]Department of Microbiology and Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [2]Departments of Cell and Developmental Biology, Genetics, and Biology, Penn Epigenetics Program, University of Pennsylvania, Philadelphia, PA, USA. [3]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [4]Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA, USA. [5]Department of Medicine and Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [6]Department of Microbiology and Immunology, University of Minnesota, Minneapolis, MN, USA. [7]Department of Genetics and Institute for Immunology, University of Pennsylvania, Philadelphia, PA, USA. [8]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [9]Division of Hematology/Oncology, Children's Hospital, Harvard Medical School, Boston, MA, USA.
***Corresponding author. Email: wherry@mail.med.upenn.edu**

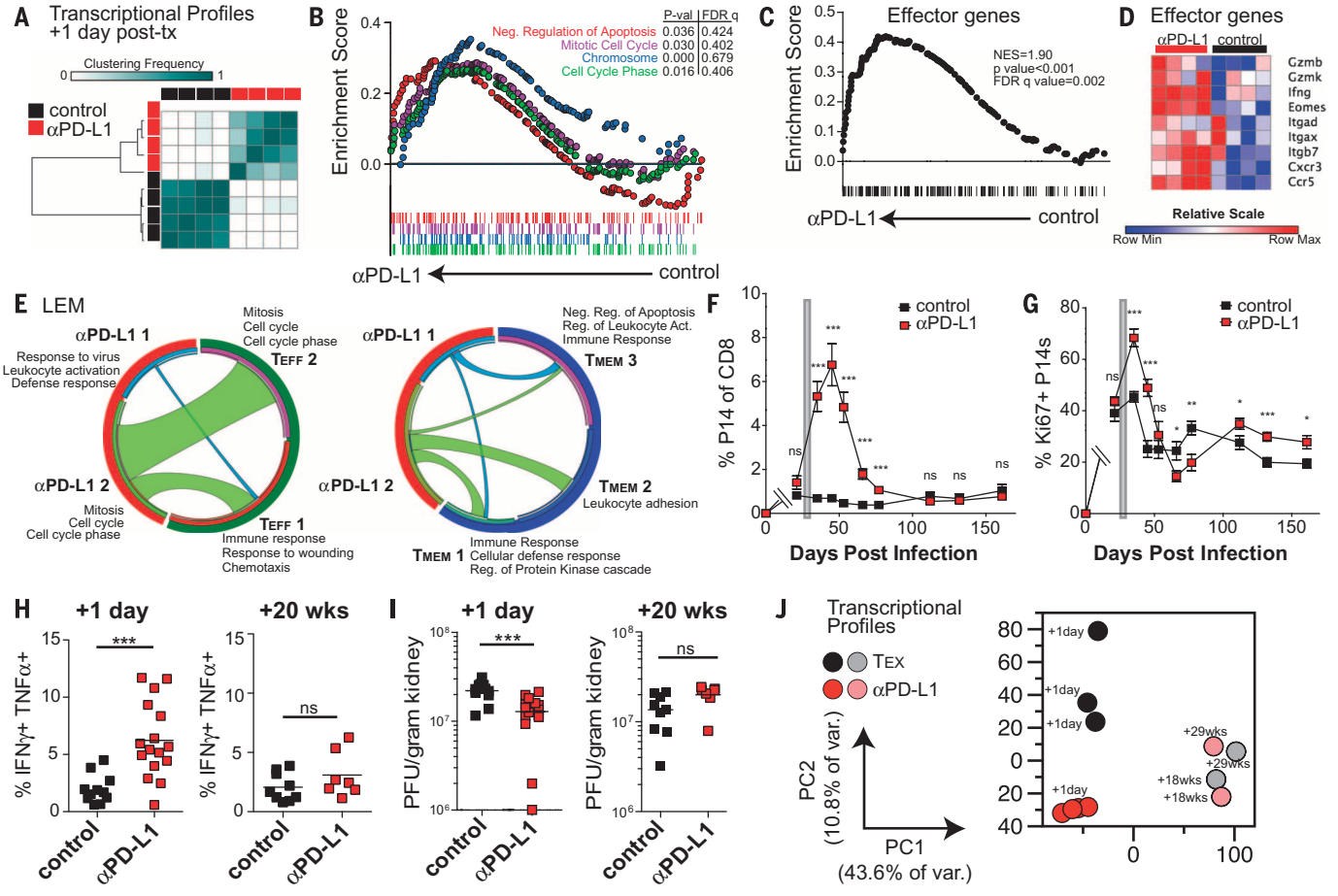


Fig. 1. Anti-PD-L1 induces an effector-like transcriptional program in T_{EX} cells that is not sustained after cessation of treatment. (**A**) Consensus hierarchical clustering by 1-Pearson correlation from the microarray on control- or anti-PD-L1–treated T_{EX}. (**B**) GSEA of representative Gene Ontology (GO) terms. (**C**) GSEA of effector genes. (**D**) Row-normalized heat map of effector-associated genes. (**E**) Circos plots showing overlap in metagenes identified in anti-PD-L1–treated T_{EX} compared to metagenes in T_{EFF} (left) and T_{MEM} cells (right). Ribbons connecting areas of the Circos plots indicate shared genes between groups. The microarray includes four independent experiments with 10 to 12 mice per group per experiment. (**F**) Frequency of LCMV GP33-specific T cell receptor transgenic "P14" cells among CD8 T cells and (**G**) Ki67$^+$ P14 cells in the peripheral blood. Gray bar indicates antibody treatment period. Error bars represent SEM. (**H**) Quantification of IFNγ$^+$ TNFα$^+$ P14 cells. (**I**) Viral load in the kidney. Data in (F) and (G) are one representative experiment. In (H) and (I), the $^+$1 day time point is combined from two representative experiments, and the $^+$20 week time point is from one representative experiment. Data in (F) to (I) are representative of at least two independent experiments with at least four mice per group per experiment. (**J**) Principle-component analysis of RNA sequencing (RNA-seq), percentage of variance (% of var.) indicated. The RNA-seq was performed on two to four independent experiments with 5 to 13 mice per group as indicated in the supplementary methods. Each dot represents an independent replicate. Asterisks indicate statistical significance determined by unpaired *t* tests between groups (*$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$). ns, not significant.

To test this idea, we transferred equal numbers of control T_{EX}, anti-PD-L1–treated T_{EX}, or T_{MEM} into antigen-free mice and monitored persistence (fig. S7A). Consistent with previous studies (*12*, *13*), T_{EX} survived poorly in antigen-free recipients compared to functional T_{MEM} (Fig. 2, A and B). There was a trend toward anti-PD-L1–treated T_{EX} persisting somewhat longer, though survival was poor compared to T_{MEM} (Fig. 2, A and B). We next investigated potential mechanisms for this trend. After PD-1 pathway blockade, the number of interleukin-7 (IL-7) receptor transcripts (*Il7r*; CD127) increased significantly (fig. S1D and table S1). There was also a modest increase in CD127 protein abundance on a subset of T_{EX} after anti-PD-L1 treatment (Fig. 2, C to E). Upon stimulation with IL-7, anti-PD-L1–treated T_{EX} also showed more phospho-STAT5 (signal transducer and activator of transcription 5) compared to control-treated T_{EX} (Fig. 2F and fig. S7B). By contrast, expression of the IL-15 receptor subunit CD122 and responsiveness to IL-15 in vitro were not substantially altered (Fig. 2, C and F, and fig. S7B). These data suggest that anti-PD-L1 treatment may augment activity of the memory-biased IL-7R pathway.

Treatment with IL-7 starting in the effector phase can prevent development of exhaustion (*14*, *15*). However, later in chronic infection, T_{EX} respond poorly to IL-7 (*12*, *13*). Because anti-PD-L1 improved IL-7R signaling, we tested whether combined treatment had additional benefit (fig. S7C). Indeed, although other aspects of the response were less affected, treatment with IL-7 and anti-PD-L1, but not IL-7 alone, resulted in more antigen-specific CD8 T cells and improved coproduction of interferon-γ (IFN-γ) and tumor necrosis factor–α (TNF-α) (Fig. 2, G to H, and fig. S7, D to H). Thus, it may be possible to exploit pathways up-regulated by PD-L1 blockade, including IL-7R, to improve checkpoint blockade.

We next examined whether PD-1 pathway blockade could restore robust recall potential upon reinfection, a defining property of T_{MEM}. Equal numbers of D^{b}GP33$^+$ CD8 T_{EX}, anti-PD-L1–treated T_{EX}, or T_{MEM} were transferred into antigen-free mice, rested, and then rechallenged with *Listeria monocytogenes* expressing GP33-41. T_{MEM} robustly expanded and efficiently produced IFN-γ (Fig. 3, A to D). By contrast, both control- and anti-PD-L1–treated T_{EX} mounted poor responses to *Listeria*-GP33 challenge, and reinvigorated T_{EX} were as defective as control T_{EX} in these key properties (Fig. 3, A to D).

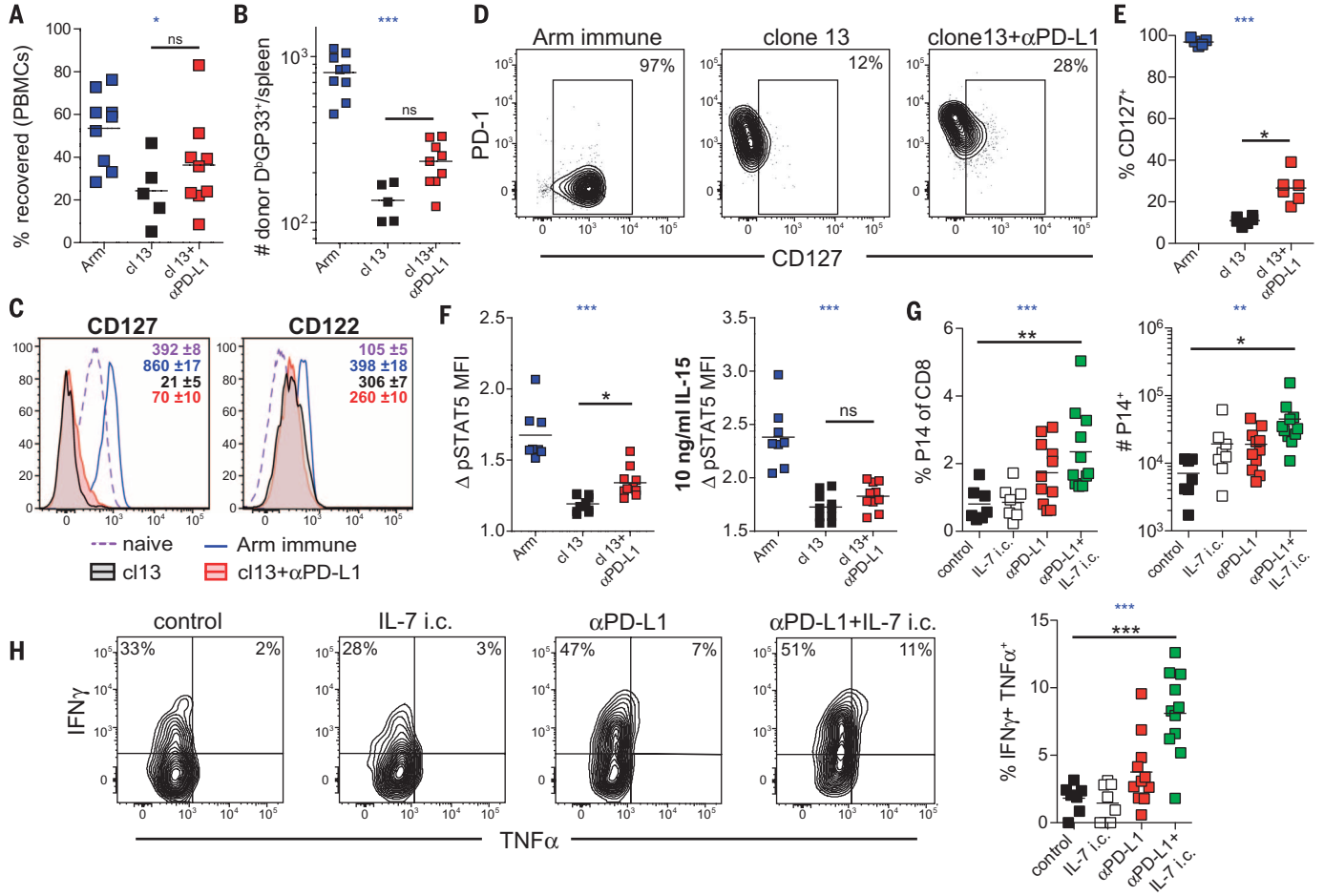


Fig. 2. PD-1 pathway blockade moderately improves antigen-independent persistence and IL-7 signaling in T_{EX}. (**A**) Number of D^bGP33^+ donor CD8 T cells per million peripheral blood mononuclear cells (PBMCs) at day 27 (compared to day 1) after transfer and (**B**) number recovered from the spleen. (**C**) Histograms of CD127 and CD122 expression on T_{EX} P14 cells (day 35 after clone 13) compared to T_{MEM} P14 cells or bulk $CD44^{lo}$ $CD62L^+$ T_N cells [day 167 after LCMV strain Armstrong (Arm)]. Values indicate average geometric mean fluorescence intensity (MFI) and SEM. (**D**) Contour plots of PD-1 versus CD127 from mice in (C). (**E**) Quantification of (D). Data in (A) to (E) are representative of at least two independent experiments with at least four mice per group. (**F**) Quantification of phospho-STAT5 induction by P14 cells at day 39 after infection after ex vivo stimulation with IL-7 or IL-15 for 30 min. Values indicate fold change over unstimulated controls. (**G**) Frequency (of $CD8^+$, left) and number (right) of P14 cells in the spleen after 2 weeks of treatment. (**H**) Plots (left) and quantification (right) of $IFN\gamma^+$ $TNF\alpha^+$ P14 cells from (G) after ex vivo peptide stimulation. Data in (F) to (H) are combined from two independent experiments with at least four mice per group. Asterisks indicate statistical significance (*$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$) determined as described in the supplementary methods. Blue asterisks indicate analysis of variance (ANOVA) P values; black asterisks indicate post-test P values.

After antigen withdrawal, T_{EX} and anti-PD-L1–treated T_{EX} failed to down-regulate PD-1 (Fig. 3E), consistent with *Pdcd1* locus DNA methylation and long-term expression of PD-1 (*16*–*18*). T_{EX} also have lower global diacetylated histone H3 (*19*), but how this relates to differentiation is unclear. We hypothesized that the genome-wide epigenetic landscape of T_{EX} may contribute to the lack of durable improvements after PD-1 pathway blockade. Thus, we performed global chromatin landscape mapping using assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) (*20*) (fig. S8). The majority of open chromatin regions (OCRs) identified were in intergenic regions (33.3 to 43.3%) or introns (43.4 to 48.5%) (fig. S9A), as expected (*21*). T_{EFF}, T_{MEM}, and T_{EX} showed substantial chromatin remodeling compared to T_N (Fig. 3F and fig. S9, B and C), and genes with transcriptional start sites (TSS) within 20 kb of OCRs tended to be more highly expressed (fig. S10). OCRs at specific genes illustrated distinct patterns for T_{EFF}, T_{MEM}, and T_{EX}. For example, T_{EX} lacked several OCRs present in the *Ifng* locus in T_{EFF} and T_{MEM} (Fig. 3G, blue boxes). Similarly, for *Pdcd1*, T_{EX}-specific OCRs were identified in the "B" and "C" regions (Fig. 3G, black box) (*9*, *22*, *23*) and a previously unidentified OCR ~23 kb from the TSS (Fig. 3G, red box). Global hierarchical clustering and cocluster analysis showed that T_{EFF} and T_{MEM} were more similar to each other than to T_{EX} and that T_{EX} had a distinct global epigenetic landscape (Fig. 3, H to J, and figs. S11 to S13). These data suggest that T_{EX} may represent a distinct lineage of CD8 T cells.

Two subsets of T_{EX} have been defined based on expression of Eomes, T-bet, and PD-1 (*24*, *25*), and additional heterogeneity has recently been described (*26*–*28*). The $T\text{-bet}^{hi}Eomes^{lo}$ $PD\text{-}1^{int}$ subset can be reinvigorated by PD-1 pathway blockade, whereas the $Eomes^{hi}PD\text{-}1^{hi}$ subset is more terminal and responds poorly to blocking PD-1 (*24*, *25*). In this study, T_{EX} were ~80% $Eomes^{hi}$ and ~20% $T\text{-bet}^{hi}$, and this distribution changed minimally upon anti-PD-L1 treatment (fig. S14, A to C). The transcriptional and epigenetic profiles of T_{EX} and anti-PD-L1–treated T_{EX} were significantly enriched for genes from the $Eomes^{hi}$ subset (fig. S14, D and E) (*29*). However, there was also a trend toward enrichment of genes from the $PD\text{-}1^{intr}Tbet^{hi}$ T_{EX} subset in the anti-PD-L1–treated group (fig. S14, F and G), perhaps reflecting recent conversion of $Tbet^{hi}$ cells into $Eomes^{hi}$ cells or additional heterogeneity.

We next examined the ability of PD-1 pathway blockade to reprogram the epigenetic landscape of T_{EX}. Hierarchical clustering, co-clustering, and principle-component analysis showed considerable

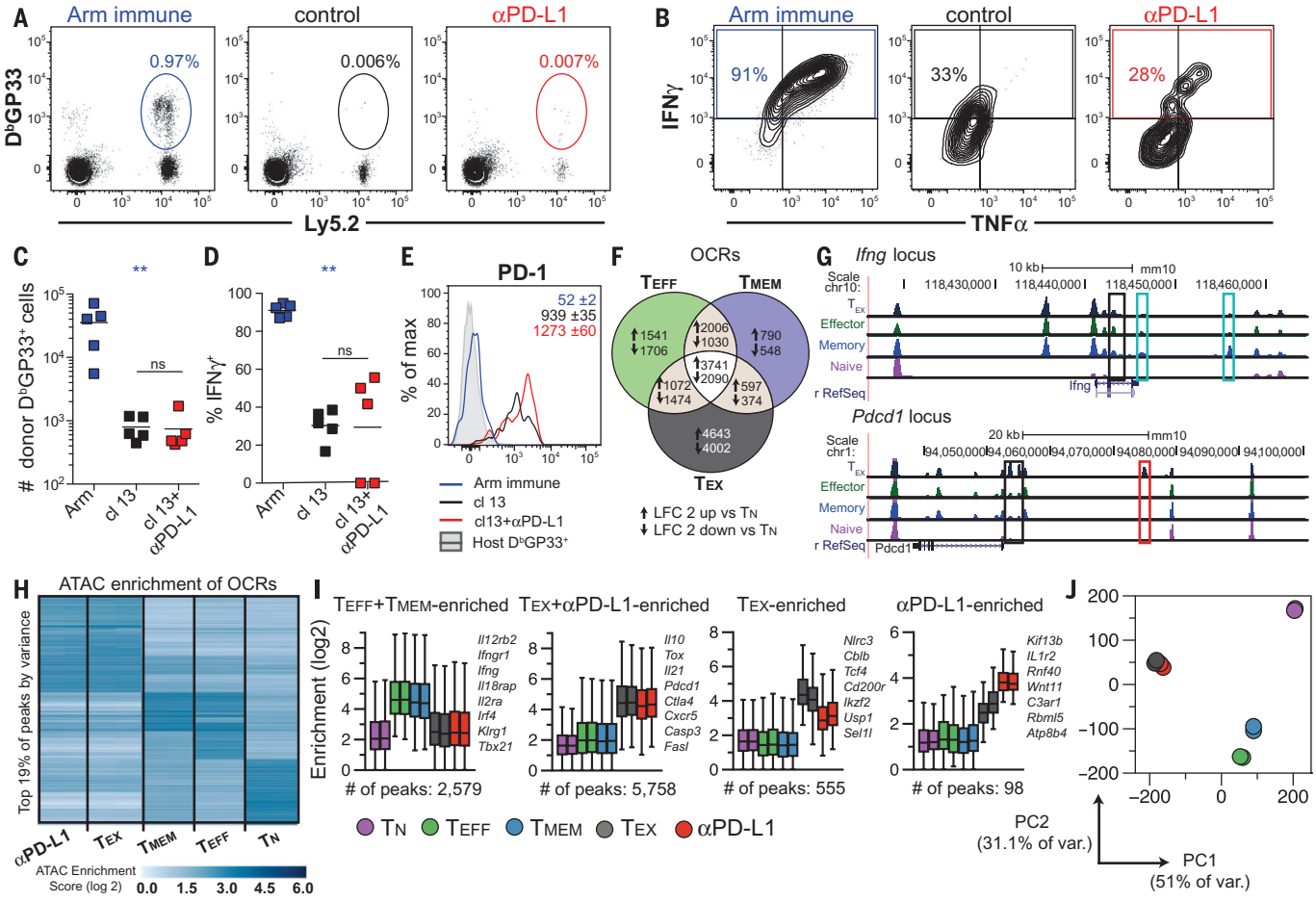


Fig. 3. PD-1 pathway blockade fails to restore memory-like recall capacity or reprogram the epigenetic landscape of T_{EX} into T_{EFF} or T_{MEM} cells. (**A** to **D**) The experimental design outlined in fig. S7A was used except that recipient mice were rechallenged with *Listeria*-GP33 3.5 weeks after transfer. (A) Flow cytometry plots of responding T_{MEM}, T_{EX}, or anti-PD-L1–treated T_{EX} at 6 days after rechallenge with *Listeria*-GP33. (B) Concatenated flow cytometry plots gated on P14 cells from mice in (A) after ex vivo peptide stimulation. (C) Quantification of donor (Ly5.2$^+$) D^bGP33^+ CD8 T cells in the spleens shown in (A). (D) Quantification of IFNγ$^+$ P14 cells shown in (B). (**E**) Histograms of PD-1 on donor D^bGP33^+ cells from mice shown in fig. 2B. Values indicate average geometric MFI and SEM. Data are representative of two independent experiments with at least four mice per group. Asterisks indicate statistical significance (***P < 0.01, and ***P < 0.001) determined as described in the supplementary methods. Blue asterisks indicate ANOVA P values; black asterisks indicate post-test P values. (**F**) Venn diagrams of ATAC-seq OCRs compared to T_N cells (LFC ≥ 2). Data from the two replicates are combined. (**G**) Representative ATAC-seq tracks from one independent replicate per group shown at the *Ifng* and *Pcdc1* loci. (**H**) Cocluster analysis of variance showing enrichment of OCRs in ATAC-seq data set. Solid lines separate cell types, and replicates are shown side by side. (**I**) Box and whisker plots showing ATAC-seq enrichment from (H). Whiskers represent the interquartile range. (**J**) Principle-component analysis of all OCRs. For (I) and (J), each replicate is shown. ATAC seq data are from two independent experiments with 2 to 15 mice per group as described in the supplementary methods.

similarity between control and anti-PD-L1–treated T_{EX} (Fig. 3, H to J, and fig. S11). OCRs preferentially found in both T_{EX} and anti-PD-L1–treated T_{EX} were located near *Pdcd1*, *Il10*, *Ctla4*, *Cxcr5*, and elsewhere, suggesting state-specific regulation that was not substantially altered after PD-L1 blockade (fig. S13). Although globally, the epigenetic changes were modest, cocluster analysis identified a small subset of OCRs specifically enriched in T_{EX} (555 peaks) or anti-PD-L1–treated T_{EX} (98 peaks) (Fig. 3, H to I; fig. S13; and table S7). Some of these genes showed the same trend epigenetically and transcriptionally (e.g., *CD200r*; fig. S10E), and specific biological pathways were enriched in sets of genes near OCRs that were unique to each cell type. (fig. S15).

T_{EX} displayed ~6000 unique OCR changes compared to T_{EFF} and T_{MEM} (Fig. 3, F to I). Thus, the ~650 OCR changes induced by PD-L1 blockade were modest by comparison. To determine whether these changes affected specific transcriptional circuits, we identified transcription factor (TF) motifs enriched in peaks gained [e.g., nuclear factor κB (NF-κB), Jun:AP-1 (activator protein 1), and CCCTC-binding factor (CTCF)] or lost [e.g., nuclear factor of activated T cells, cytoplasmic 1 (NFATc1), NFAT:AP1, Nur77, Eomes, and Egr2] (Fig. 4A). We hypothesized that reinvigoration resulted from rewired transcriptional control within the existing T_{EX} epigenetic landscape. To test this notion, we performed Wellington bootstrap analysis to predict TF binding activity (Fig. 4B and table S10). T_{EX} and anti-PD-L1–treated T_{EX} were more similar to each other than to T_N, T_{EFF}, or T_{MEM}. However, TF motifs biased toward T_{EX} or anti-PD-L1–treated T_{EX} were identified (Fig. 4B and table S10). TF footprinting was then performed to identify TFs with evidence of likely binding (Fig. 4C and figs. S16 and S17). An integrated network was then constructed for transcriptional circuitry based on predicted TF activity (Fig. 4D and table S11). This network identified augmented activity of NF-κB, interferon regulatory factors (IRFs), and bZip factors (AP-1 family) and decreased activity of NFAT, Egr2, and Nur77 upon PD-L1 blockade. Major features of this transcriptional network were recapitulated with a second network approach in which additional TF families were identified (e.g., Runx, Nr2f6, Prdm1, Rarb, Pparg, Rxra, and homeobox TFs; fig. S18 and table S12). To further investigate how these changes might affect a specific TF, we examined NFAT. NFAT working with AP-1 transactivates many effector-phase

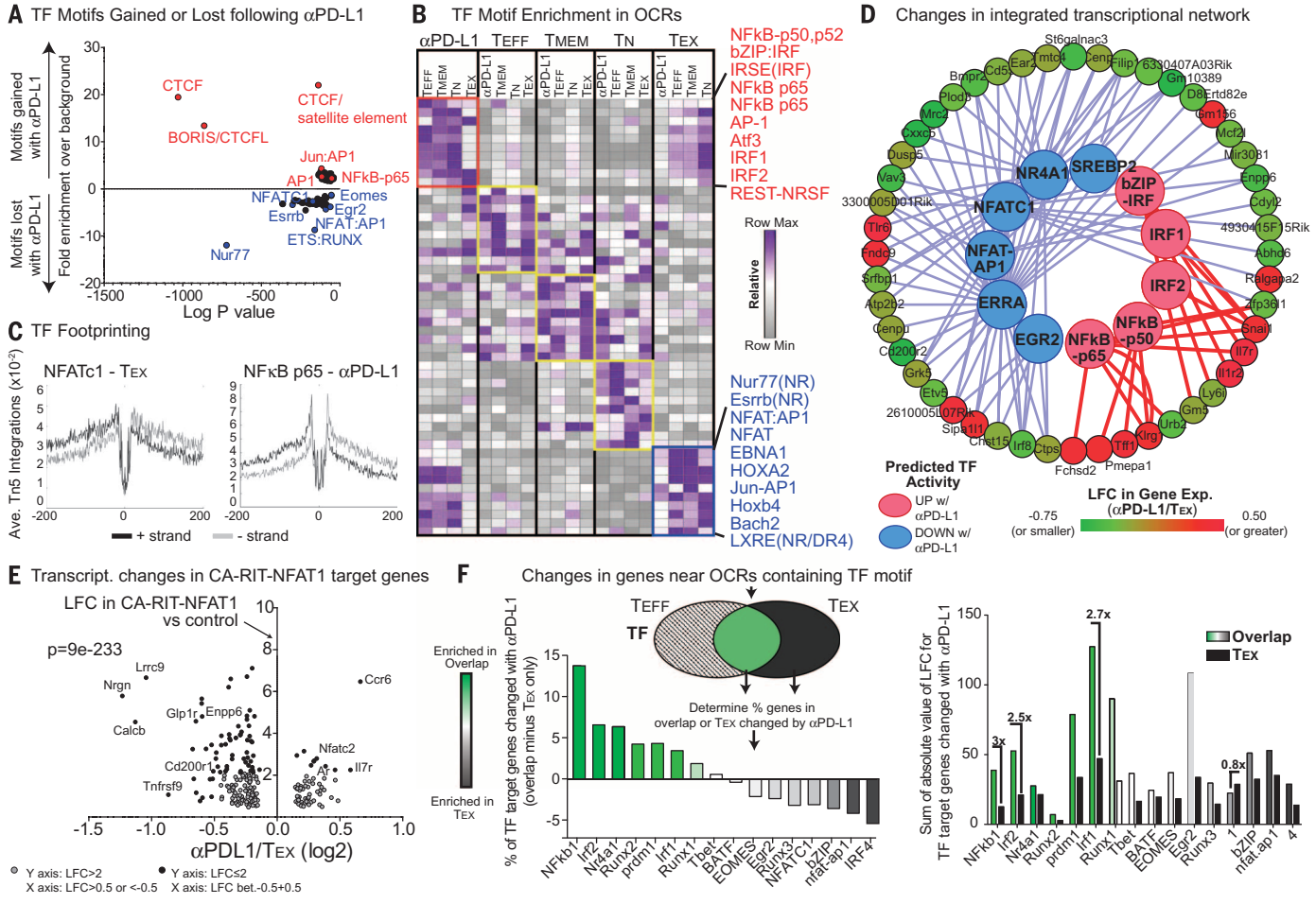


Fig. 4. Differential transcription factor binding after PD-1 pathway blockade contributes to an altered transcriptional network during T_{EX} reinvigoration. (**A**) Enrichment of TF binding motifs in OCRs lost or gained after anti-PD-L1 treatment. (**B**) Wellington bootstrap analysis of TF binding in pairwise comparisons for each cell type; the top 10 TFs (in boxes) enriched in all OCRs are shown. Full list in table S10. (**C**) TF footprint for NFATc1 in T_{EX} and NFκB-p65 in anti-PD-L1–treated T_{EX}. (**D**) Integrated network analysis of the transcriptional and epigenetic changes after anti-PD-L1 treatment. Lines connect TFs predicted to have altered activity to corresponding genes regulated. Details in table S11. (**E**) LFC of genes significantly changed by anti-PD-L1 treatment compared to genes significantly induced by the "partnerless" NFAT construct CA-RIT-NFAT1 (*30*). (**F**) Venn diagram showing genes near OCRs containing given TF motifs in T_{EFF}, T_{EX}, or both (overlap) (top left). Percentage difference in TF target genes changed ($P < 0.05$, LFC ≥ 0.3) with anti-PD-L1 in overlap compared to T_{EX} only (bottom left). Sum of the absolute value of the LFC in expression in TF target genes after anti-PD-L1 treatment (right). ATAC-seq data shown are combined replicates for each condition.

genes. By contrast, "partnerless" NFAT that fails to bind AP-1 induces a subset of T_{EX} genes (*30*). Here, upon anti-PD-L1 treatment, there was significantly reduced expression of targets of partnerless NFAT in reinvigorated T_{EX} (Fig. 4E), suggesting a rewiring of this transcriptional circuit after blockade.

Together, these data suggested that, although PD-1 pathway blockade did not fully reprogram T_{EX} into T_{MEM} or T_{EFF}, these cells may (re)acquire some features of T_{EFF} biology. One hypothesis is that upon PD-L1 blockade, the rewired transcriptional network allows T_{EX} to preferentially reengage features of their epigenomic program that overlap with T_{EFF}. To test this idea, we separated TF target genes into those containing OCRs that were (i) specific to T_{EFF}, (ii) specific to T_{EX}, or (iii) shared between T_{EFF} and T_{EX} (Fig. 4F). We then examined the change in genes expressed in each category after PD-L1 blockade. For several TFs, including T-bet and Eomes, there was no redistribution of the pattern of target gene expression (Fig. 4F). However, for many TFs identified above that have a key role in effector biology, such as NF-κB, IRF1, IRF2, Nur77, and Blimp-1 (encoded by *Prdm1*), there was an increase in the number of target genes expressed in the T_{EFF} and T_{EX} overlap group compared to the T_{EX}-only group upon PD-L1 blockade (Fig. 4F). Moreover, genes in the shared T_{EFF} and T_{EX} epigenetic module displayed a substantially greater magnitude of change in expression than genes in the T_{EX}-only group (Fig. 4F). These data indicate that PD-1 pathway blockade induces rewired transcriptional activity, allowing T_{EX} to more effectively reengage modules of effector genes contained within the epigenetic landscape of T_{EX}. Specific TF circuits that are altered, such as NF-κB, may have implications for cotargeting PD-1 and TNFR family pathways (*1*, *3*, *31*) and may be relevant for the design of future therapeutics.

Our data suggest that in settings of severe T cell exhaustion, reacquiring durable immune memory may be challenging, especially if tumor or viral antigen persists. However, our data also indicate that PD-1 pathway blockade may reveal opportunities to further augment T cell quality or effector activity (e.g., NF-κB, IL-7R). Additional strategies, such as priming new T cell responses (*3*), selectively expanding less exhausted subsets (*25*), or targeting multiple immunoregulatory or homeostatic pathways (e.g., IL-7, IL-2) simultaneously (*1*, *3*), may also augment acquisition of durable immunity. Finally, these studies provide the impetus for extending epigenetic landscape mapping to human T_{EX}, future evaluation of checkpoint blockade combined with epigenetic modifiers, or epigenomic engineering for T cells. Thus, integrated cellular, transcriptional, and epigenetic profiling of T_{EX} not only reveals mechanistic insights into PD-1 pathway blockade–mediated reinvigoration,

but also points to key opportunities for improving the long-term durability of these effects.

REFERENCES AND NOTES

1. E. J. Wherry, M. Kurachi, *Nat. Rev. Immunol.* **15**, 486–499 (2015).
2. D. B. Page, M. A. Postow, M. K. Callahan, J. P. Allison, J. D. Wolchok, *Annu. Rev. Med.* **65**, 185–202 (2014).
3. P. Sharma, J. P. Allison, *Science* **348**, 56–61 (2015).
4. D. S. Shin, A. Ribas, *Curr. Opin. Immunol.* **33**, 23–35 (2015).
5. D. L. Barber *et al.*, *Nature* **439**, 682–687 (2006).
6. See supplementary information on *Science* Online.
7. M. M. Gubin *et al.*, *Nature* **515**, 577–581 (2014).
8. B. Bengsch *et al.*, *Immunity* **45**, 358–373 (2016).
9. M. M. Staron *et al.*, *Immunity* **41**, 802–814 (2014).
10. N. Patsoukis *et al.*, *Sci. Signal.* **5**, ra46 (2012).
11. J. Godec *et al.*, *Immunity* **44**, 194–206 (2016).
12. H. Shin, S. D. Blackburn, J. N. Blattman, E. J. Wherry, *J. Exp. Med.* **204**, 941–949 (2007).
13. E. J. Wherry, D. L. Barber, S. M. Kaech, J. N. Blattman, R. Ahmed, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16004–16009 (2004).
14. M. Pellegrini *et al.*, *Cell* **144**, 601–613 (2011).
15. S. G. Nanjappa, E. H. Kim, M. Suresh, *Blood* **117**, 5123–5132 (2011).
16. B. Youngblood *et al.*, *Immunity* **35**, 400–412 (2011).
17. D. T. Utzschneider *et al.*, *Nat. Immunol.* **14**, 603–610 (2013).
18. J. M. Angelosanto, S. D. Blackburn, A. Crawford, E. J. Wherry, *J. Virol.* **86**, 8161–8170 (2012).
19. F. Zhang *et al.*, *Mol. Ther.* **22**, 1698–1706 (2014).
20. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, *Nat. Methods* **10**, 1213–1218 (2013).
21. D. R. Winter, I. Amit, *Immunol. Rev.* **261**, 9–22 (2014).
22. K. J. Oestreich, H. Yoon, R. Ahmed, J. M. Boss, *J. Immunol.* **181**, 4832–4839 (2008).
23. C. Kao *et al.*, *Nat. Immunol.* **12**, 663–671 (2011).
24. M. A. Paley *et al.*, *Science* **338**, 1220–1225 (2012).
25. S. D. Blackburn, H. Shin, G. J. Freeman, E. J. Wherry, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15016–15021 (2008).
26. R. He *et al.*, *Nature* **537**, 412–428 (2016).
27. S. J. Im *et al.*, *Nature* **537**, 417–421 (2016).
28. D. T. Utzschneider *et al.*, *Immunity* **45**, 415–427 (2016).
29. T. A. Doering *et al.*, *Immunity* **37**, 1130–1144 (2012).
30. G. J. Martinez *et al.*, *Immunity* **42**, 265–278 (2015).
31. L. K. Ward-Kavanagh, W. W. Lin, J. R. Šedý, C. F. Ware, *Immunity* **44**, 1005–1019 (2016).

ACKNOWLEDGMENTS

We thank the Wherry lab for discussions and critically reading the manuscript. We thank C. Surh for providing the antibody against IL-7 (anti-IL-7). The anti-IL-7 antibody is available from La Jolla Institute of Allergy and Immunology, and the IL-7 used is available from the National Cancer Institute, both under material transfer agreements with the University of Pennsylvania. The data presented in this manuscript are tabulated in the main paper and in the supplementary materials. Sequencing data are available at Gene Expression Omnibus [accession numbers GSE86796 (microarray), GSE86881 (RNA seq), and GSE86797 (ATAC seq)]. This study was supported by a Robertson Foundation–Cancer Research Institute Irvington Fellowship (K.E.P.), an American Cancer Society Postdoctoral Fellowship (M.A.S.), National Institutes of Health grant F30DK100159 (J.M.S.), German Research Foundation Fellowship BE5496/1-1 (B.B.), and National Institutes of Health grant T32 2T32CA009615-26 (A.C.H.). This work was funded by the National Institutes of Health (grant CA78831 to S.L.B. and grants AI105343, AI112521, AI082630, AI115712, AI117950, and AI108545 to E.J.W.). This research was also supported by the Parker Institute for Cancer Immunotherapy. E.J.W. has a patent licensing agreement on the PD-1 pathway. The authors declare no additional conflicts of interest.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1160/suppl/DC1
Materials and Methods
Figs. S1 to S18
Tables S1 to S12
References (*32*–*38*)

18 January 2016; accepted 19 September 2016
Published online 27 October 2016
10.1126/science.aaf2807

T CELL EXHAUSTION

The epigenetic landscape of T cell exhaustion

Debattama R. Sen,[1,2]* James Kaminski,[3]* R. Anthony Barnitz,[1] Makoto Kurachi,[4,5] Ulrike Gerdemann,[1] Kathleen B. Yates,[1] Hsiao-Wei Tsao,[1] Jernej Godec,[1,2] Martin W. LaFleur,[1,2] Flavian D. Brown,[1,2] Pierre Tonnerre,[6] Raymond T. Chung,[6] Damien C. Tully,[7] Todd M. Allen,[7] Nicole Frahm,[8] Georg M. Lauer,[6] E. John Wherry,[4,5] Nir Yosef,[3,7,9]†‡ W. Nicholas Haining[1,10,11]†‡

Exhausted T cells in cancer and chronic viral infection express distinctive patterns of genes, including sustained expression of programmed cell death protein 1 (PD-1). However, the regulation of gene expression in exhausted T cells is poorly understood. Here, we define the accessible chromatin landscape in exhausted CD8$^+$ T cells and show that it is distinct from functional memory CD8$^+$ T cells. Exhausted CD8$^+$ T cells in humans and a mouse model of chronic viral infection acquire a state-specific epigenetic landscape organized into functional modules of enhancers. Genome editing shows that PD-1 expression is regulated in part by an exhaustion-specific enhancer that contains essential RAR, T-bet, and Sox3 motifs. Functional enhancer maps may offer targets for genome editing that alter gene expression preferentially in exhausted CD8$^+$ T cells.

T cell exhaustion—an acquired state of T cell dysfunction—is a hallmark of cancer and chronic viral infection (*1*, *2*), and clinical trials of checkpoint blockade, which aim to reverse T cell exhaustion in cancer, have proven strikingly effective (*3*, *4*). Chimeric antigen receptor (CAR)–T cell therapy has also proven highly effective for hematologic malignancies (*5*), but the development of exhaustion in T cells engineered to treat solid tumors remains a substantial barrier to its broader use (*6*). The identification of mechanisms that regulate exhausted T cells is therefore a major goal in cancer immunotherapy.

To identify regulatory regions in the genome of exhausted CD8$^+$ T cells, we used an assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) (*7*) to demarcate areas of accessible chromatin in mouse antigen-specific CD8$^+$ T cells differentiating in response to lymphocytic choriomeningitis virus (LCMV) infection (fig. S1A and table S1). Acute LCMV infection elicits highly functional effector CD8$^+$ T cells, whereas chronic LCMV infection gives rise to exhausted CD8$^+$ T cells (*1*–*3*, *8*, *9*). Analysis of high-quality ATAC-seq profiles (fig. S1, B to H) from naïve CD8$^+$ T cells and those at day 8 and day 27 postinfection (p.i.) (d8 and d27, respectively) revealed that naïve CD8$^+$ T cells underwent large-scale remodeling (Fig. 1A and fig. S2A) during differentiation [as detected by DESeq2, with a false discovery rate (FDR) < 0.05]. The majority (71%) (fig. S2A) of chromatin-accessible regions (ChARs) either emerged (e.g., those at the *Ifng* locus) or disappeared (e.g., *Ccr7*) (Fig. 1A) as naïve CD8$^+$ T cells underwent differentiation. The gain and loss of ChARs were not balanced; a much larger fraction of regions emerged at d8 p.i. and persisted or emerged only at d27 than were either transiently detected at d8 p.i. or lost from naïve cells (Fig. 1B). Thus, differentiation from a naïve CD8$^+$ T cell state is associated with a net increase, rather than decrease, in chromatin accessibility (fig. S2B).

Comparison of ChARs from exhausted CD8$^+$ T cells with those found in functional effector or memory CD8$^+$ T cells revealed marked differences in the pattern of regulatory regions. Differential regulatory regions between acute and chronic infection (Fig. 1C and fig. S2C) showed features of enhancers: They tended to be depleted of transcription start sites (TSSs) and enriched for intergenic and intronic areas (Fig. 1D), and found distal to gene promoters (fig. S2D). The magnitude of difference in the profile of regulatory regions between exhausted and functional CD8$^+$ T cells was greater than that seen in gene expression. We found that 44.48% of all ChARs

[1]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA. [2]Division of Medical Sciences, Harvard Medical School, Boston, MA 02115, USA. [3]Center for Computational Biology, University of California, Berkeley, Berkeley, CA 94720, USA. [4]Institute of Immunology, University of Pennsylvania, Philadelphia, PA 19104, USA. [5]Department of Microbiology, University of Pennsylvania, Philadelphia, PA 19104, USA. [6]Gastrointestinal Unit and Liver Center, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02115, USA. [7]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard University, Boston, MA 02139, USA. [8]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [9]Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, CA 94720, USA. [10]Division of Pediatric Hematology and Oncology, Children's Hospital, Boston, MA 02115, USA. [11]Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02142, USA.
*These authors contributed equally to this work. †These authors contributed equally to this work. **‡Corresponding author. Email: niryosef@berkeley.edu (N.Y.); nicholas_haining@dfci.harvard.edu (W.N.H.)**

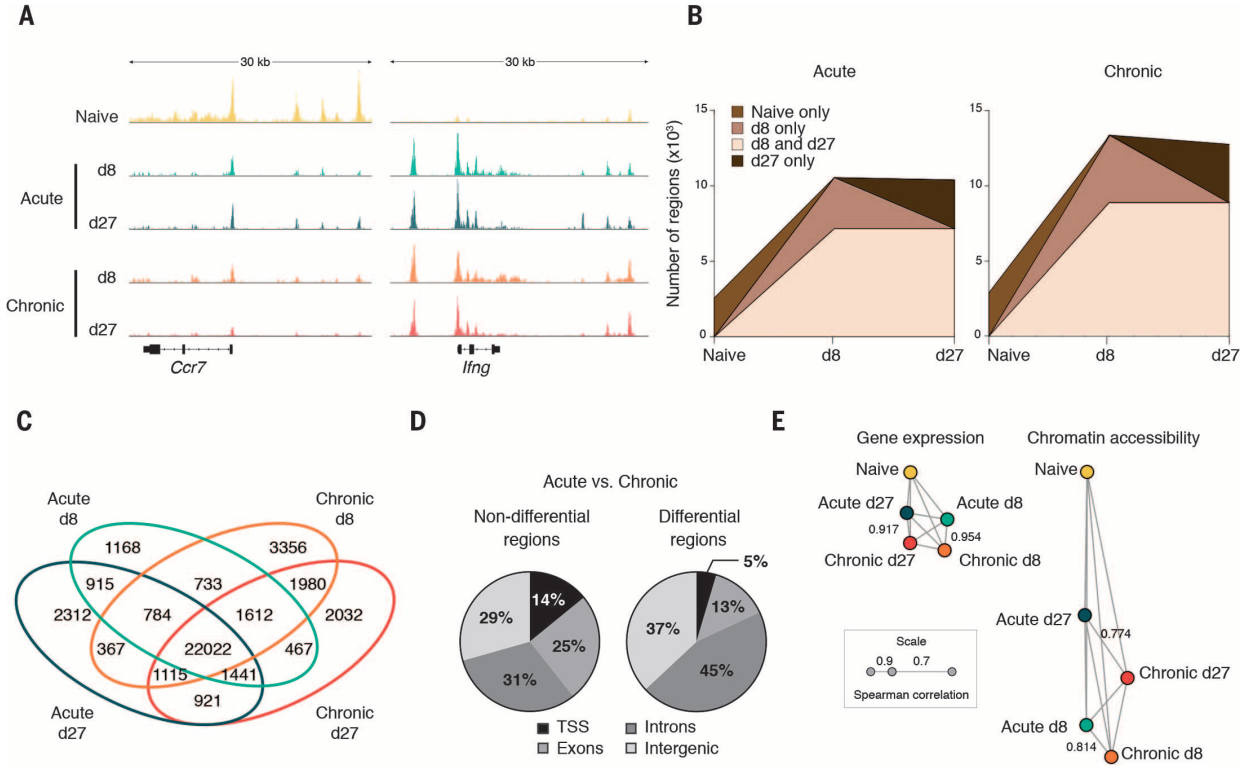


Fig. 1. CD8⁺ T cell exhaustion is associated with extensive changes in accessible chromatin. (**A**) Representative ATAC-seq tracks at the *Ccr7* and *Ifng* gene loci. (**B**) Developmental trajectory of new regions at each time point. (**C**) Overlap in ChARs between cell states. (**D**) Distribution of nondifferential (left) and differential (right) regions between acute and chronic CD8⁺ T cell states. TSS, transcription start site. (**E**) Correlation network of similarity between states measured by gene expression (left) and chromatin accessibility (right). Edge length corresponds to similarity (Spearman correlation).

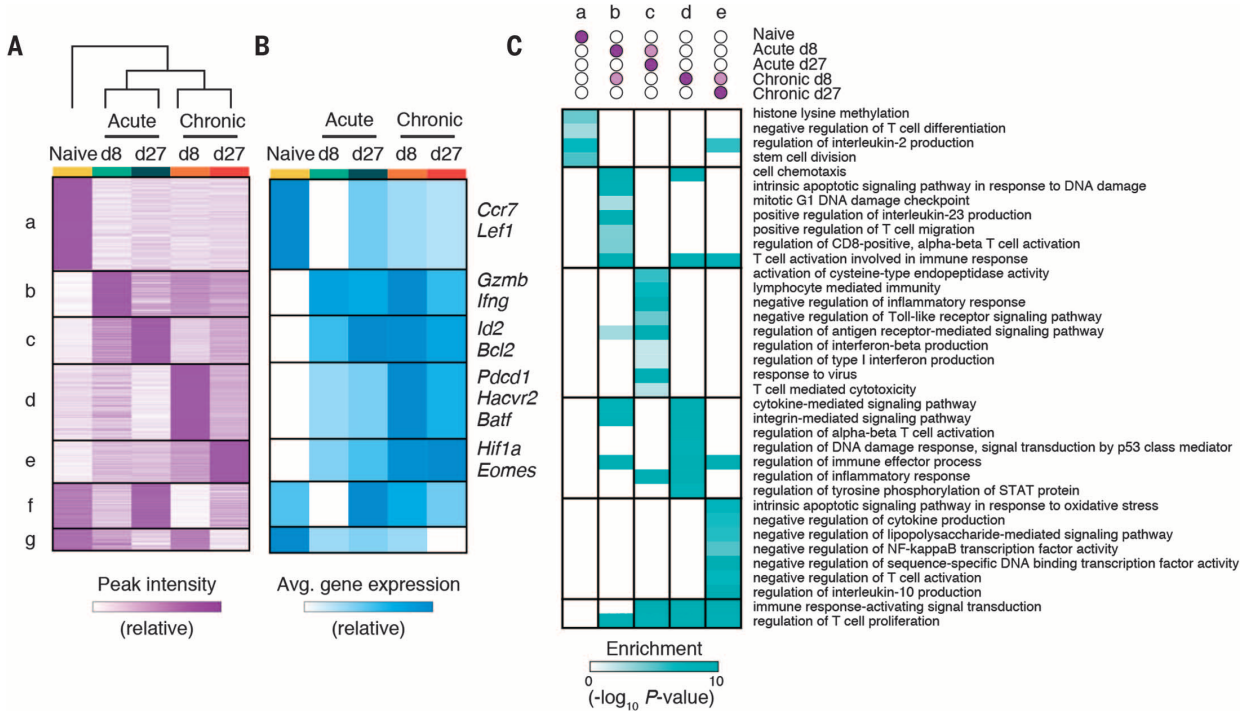


Fig. 2. State-specific enhancers in CD8⁺ T cells form modules that map to functionally distinct classes of genes. (**A**) Heat map of peak intensity for all differentially accessible regions (rows) clustered by similarity across cell states (columns). Shown are normalized numbers of cut sites (supplementary methods), scaled linearly from row minimum (white) to maximum (purple). (**B**) Heat map showing row-normalized average mRNA expression of neighboring genes within each module in (A) in each cell state. Informative genes from each module are shown on right. (**C**) Heat map showing enrichment of Gene Ontology (GO) terms (rows) in each module (columns). *P*-values (hypergeometric test) presented as $-\log_{10}$.

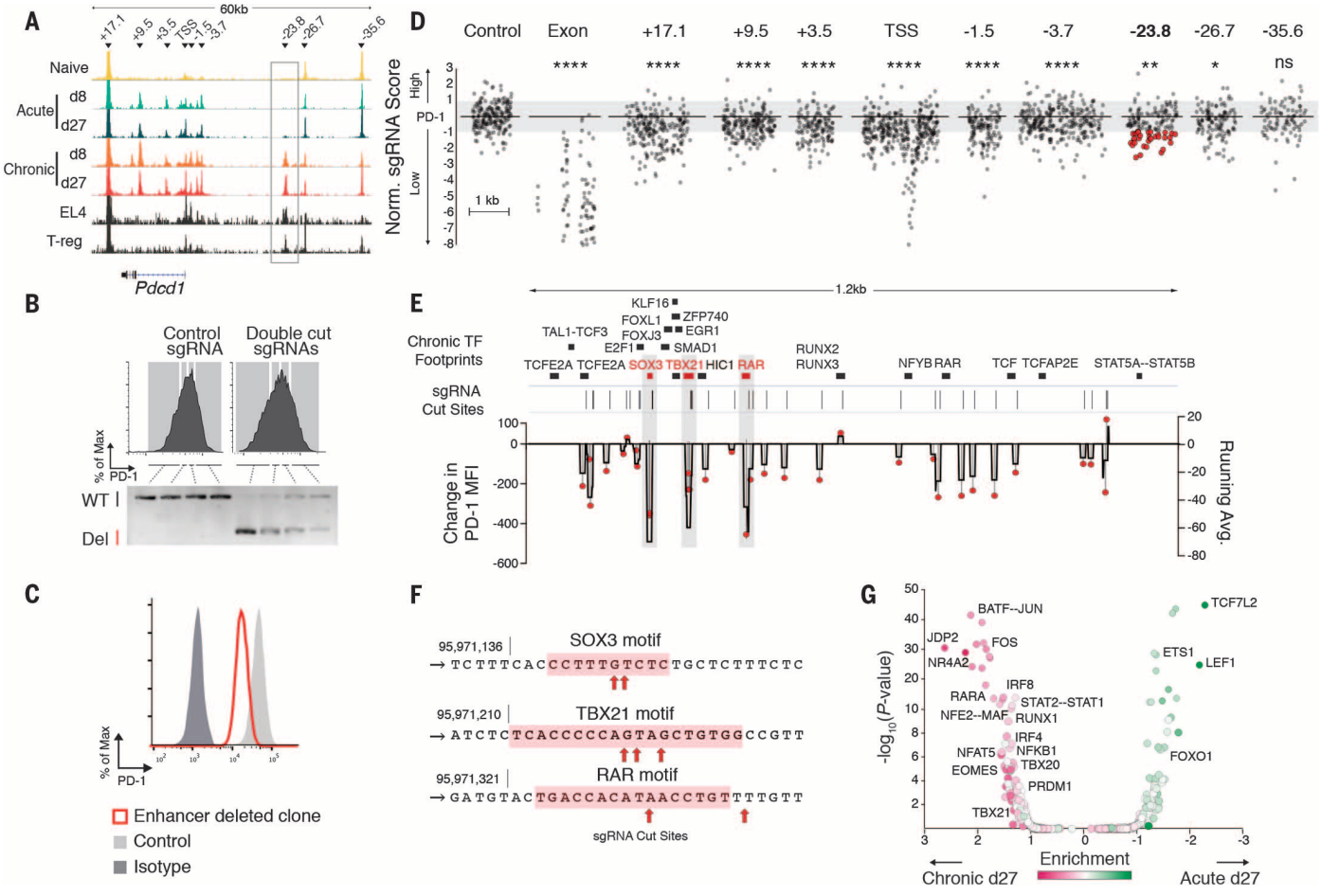


Fig. 3. High-resolution functional mapping of an exhaustion-specific enhancer identifies minimal sequences that regulate PD-1. (**A**) ATAC-seq tracks from CD8$^+$ T cells, EL4 cell line, and regulatory CD4$^+$ T cells (*12*). Arrowheads indicate individual ChARs. (**B**) Cell sorting gates (top) and corresponding genomic polymerase chain reaction amplification for the PD-1 enhancer region (bottom) showing proportion of wild-type (WT) or deleted (Del) alleles in EL4 cells transfected with control (left) or double-cut sgRNAs (right). Representative data shown from two replicates. (**C**) PD-1 expression of EL4 WT (light gray) or representative enhancer-deleted (red) single-cell clone out of 46 clones. (**D**) Normalized enrichment of sgRNAs (gray symbols) within PD-1–high and PD-1–low populations at locations shown (supplementary methods). Control nontargeting sgRNAs are pseudo mapped with 5-bp spacing. Red symbols correspond to the 21 sgRNAs with the largest effect within the −23.8 kb enhancer, for which isogenic cell lines were later produced. (**E**) Overlap of TF footprints and sgRNA activity within the −23.8 kb enhancer. TF footprints with binding probability >0.9 in chronic d27 are shown on top. Lines represent cut sites of top-scoring sgRNAs. Change in PD-1 mean fluorescence intensity (MFI) relative to control guide transfected populations for each sgRNA (red symbol, left axis); 10-bp running average of PD-1 MFI changes caused by sgRNA activity shown in black (right axis). (**F**) sgRNA cut sites within the SOX3, TBX21, and RAR motifs. (**G**) Log fold enrichment of predicted TF footprints in acute d27 versus chronic d27 CD8$^+$ T cells (*x* axis) (see supplementary methods) are plotted against the corresponding *P*-value (hypergeometric).

were differentially present between functional and exhausted cells at each time point, compared with only 9.75% of differentially expressed genes (both values estimated at FDR < 0.05). Consistent with this, the rank correlation between each T cell state by gene expression was much higher than at the level of regulatory regions (Fig. 1E). Thus, state change during CD8$^+$ T cell differentiation is accompanied by a larger reorganization of accessible chromatin than is apparent by examination of gene expression.

Unsupervised clustering identified "modules" of differential ChARs with similar patterns of activity across T cell states (Fig. 2A and fig. S3A). We found a highly significant positive correlation between the average peak intensity of ChARs within each module and the average gene expression of the adjacent genes (*F* test, $P < 0.001$) (Fig. 2B and fig. S3B). This suggests that, on average, the ChARs contained in each module tended to be associated with the activation, rather than repression, of corresponding genes.

Genes adjacent to ChARs in each state-specific module included many with known functions in the corresponding T cell state. For example, module d, active in mouse T cells experiencing chronic LCMV infection on d8 and d27 p.i., contained ChARs adjacent to the inhibitory receptors *Pdcd1* and *Havcr2* (which encodes Tim3) and the transcription factor *Batf*, all genes that are upregulated in exhausted CD8$^+$ T cells (Fig. 2B) (*1*, *8*). Moreover, the functional classes of genes in each module were distinct on the basis of pathway enrichment (Fig. 2C and table S2). Thus, ChARs that distinguish naïve, effector, memory, and exhausted CD8$^+$ T cells are organized into state-specific modules that positively regulate functionally distinct programs of genes.

We next sought to test whether regulatory regions specific to exhausted cells could regulate genes differentially expressed in exhausted CD8$^+$ T cells. Persistent expression of PD-1 is a cardinal feature of exhausted CD8$^+$ T cells, but PD-1 is also transiently expressed by effector CD8$^+$ T cells during acute LCMV infection (*3*, *8*). We identified nine ChARs within 45 kb of the *Pdcd1* gene locus (Fig. 3A) and found several that correspond to

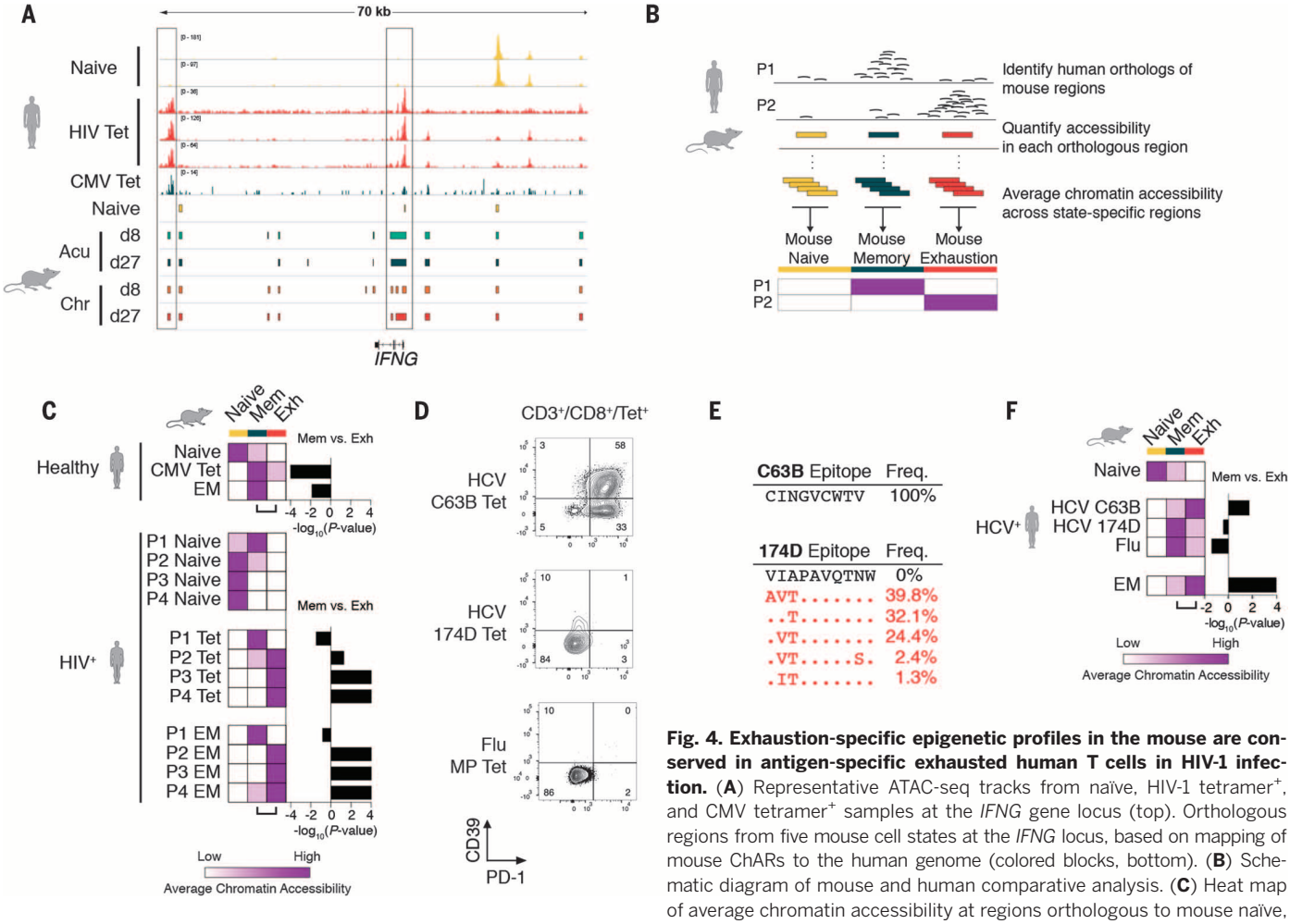


Fig. 4. Exhaustion-specific epigenetic profiles in the mouse are conserved in antigen-specific exhausted human T cells in HIV-1 infection. (**A**) Representative ATAC-seq tracks from naïve, HIV-1 tetramer$^+$, and CMV tetramer$^+$ samples at the *IFNG* gene locus (top). Orthologous regions from five mouse cell states at the *IFNG* locus, based on mapping of mouse ChARs to the human genome (colored blocks, bottom). (**B**) Schematic diagram of mouse and human comparative analysis. (**C**) Heat map of average chromatin accessibility at regions orthologous to mouse naïve, memory, and exhaustion enhancers in human samples indicated. Color scale as in Fig. 2A. (**D**) PD-1 and CD39 expression measured by flow cytometry in HCV C63B tetramer$^+$, HCV 174D tetramer$^+$, and influenza (flu) matrix peptide (MP) tetramer$^+$ populations from a single HCV-infected donor. (**E**) Viral sequences encoding C63B and 174D epitopes. (**F**) Heat map of average chromatin accessibility at regions orthologous to mouse naïve, memory, and exhaustion enhancers in human samples indicated from a single HCV-infected donor.

previously described regions with enhancer activity (–1.5 kb and –3.7 kb) (Fig. 3A) (*10*); these were present in both acute and chronic infection. We also identified an additional region (–23.8 kb) that only showed appreciable chromatin accessibility in exhausted CD8$^+$ T cells at d8 and d27 p.i. from chronic infection (Fig. 3A).

We hypothesized that this ChAR might function as an enhancer of PD-1 that is required for persistent, high levels of expression in exhausted CD8$^+$ T cells. Analysis of chromatin accessibility at this region in previously published deoxyribonuclease I–hypersensitive site–mapping (*11*) or ATAC-seq data (*12*) showed that it was not active in other types of hematopoietic cells, except the murine T cell line EL4 and regulatory CD4$^+$ T cells, both of which can constitutively express high levels of PD-1 (*10*, *13*) (Fig. 3A and fig. S4A). We cloned a 781–base pair (bp) fragment corresponding to this region into a reporter construct and found that it induced a 10- to 12-fold increase in reporter gene expression, confirming that it could function as an enhancer (fig. S4B).

We then tested whether the –23.8 kb enhancer was necessary for high-level PD-1 expression. We used the CRISPR-Cas9 nuclease to delete a 1.2-kb fragment at that position in EL4 cells, which have both sustained high-level PD-1 expression and open chromatin at that enhancer site (*14*, *15*) (fig. S4, C to G). In Cas9-expressing EL4 cells transduced with a pair of single-guide RNAs (sgRNAs) flanking the enhancer, cells with the lowest PD-1 expression had the highest amount of the enhancer deletion (Fig. 3B). We confirmed this finding in single-cell clones and found that the expression of PD-1 in clones with a biallelic deletion of the target ChAR was significantly lower ($P > 0.0002$, Mann-Whitney U test) than expression in nondeleted clones (fig. S4, H to J). Deletion of this region resulted in decreased but not abrogated PD-1, suggesting that additional regulatory regions in EL4 cells are also involved in regulating PD-1 expression (Fig. 3C). Among all genes within 1.5 Mb of the *Pdcd1* locus, only PD-1 mRNA expression was significantly decreased by deletion of the –23.8 kb ChAR (fig. S3K). This suggests that the –23.8 kb ChAR present in exhausted, but not functional, CD8$^+$ T cells serves as an enhancer that is required to maintain high levels of PD-1 expression.

We next sought to identify the functional contribution of specific sequences within enhancer regions to the regulation of PD-1 expression. We used Cas9-mediated in situ saturation mutagenesis and designed all possible sgRNAs within the –23.8 kb enhancer and eight other regulatory sequences near the *Pdcd1* locus (*15*, *16*) (Fig. 3A). We transduced Cas9-expressing EL4 cells with a pool of 1754 enhancer-targeting sgRNAs, 117 sgRNAs targeting the *Pdcd1* exons as positive controls, and 200 nontargeting sgRNAs as negative controls (fig. S5, A and B). We sorted transduced EL4s into populations on the basis of high or low PD-1 expression and quantified the abundance of individual sgRNAs (fig. S5C).

In comparison with nontargeting sgRNAs, which were equivalently distributed between PD-1–high and PD-1–low fractions, sgRNAs targeting *Pdcd1* exons were highly enriched in the

PD-1–low fraction as expected (Fig. 3D and fig. S5, D and E). sgRNAs targeting eight of the nine regulatory regions were also significantly enriched in the PD-1–low fraction to varying degrees ($P < 0.00001$ to $P < 0.01$, see supplementary methods), suggesting that critical sequences affecting PD-1 expression are densely represented within each of the eight regulatory regions. However, sgRNAs in the –35.6 kb ChAR had no significant effect on PD-1 expression, consistent with prior observations that this region falls outside the CCCTC-binding factor (CTCF)–mediated boundaries of the *Pdcd1* locus (*10*).

We focused on sgRNAs inducing cleavage in the –23.8 kb enhancer (fig. S5F) and found a strong correlation between the predicted activity in a pooled setting (PD-1 high:low ratio >1 SD below mean) and their effect on PD-1 mean fluorescence intensity in individual cell lines ($P = 0.0041$) (fig. S5, G and H). Inspection of the predicted cleavage-site locations revealed three critical regions of the enhancer in which cleavage markedly affected PD-1 expression (Fig. 3E, gray shading).

We next asked whether these critical regions in the –23.8 kb enhancer were associated with distinct patterns of transcription factor (TF) binding in exhausted $CD8^+$ T cells in vivo. We identified TF footprints (*17*) using ATAC-seq cut sites from $CD8^+$ T cells experiencing chronic infection, which allowed us to infer TF binding within the –23.8 kb enhancer (Fig. 3E; fig. S6, A to D; fig. S7A; and tables S3 to S6). We found that cleavage sites of sgRNAs that reduced PD-1 expression in EL4 cells were significantly enriched in TF footprints found in exhausted $CD8^+$ cells in vivo ($P = 8.63 \times 10^{4}$, hypergeometric test). The three TF footprints with greatest sensitivity to disruption corresponded to motifs for Sox3, T-bet (encoded by *Tbx21*), and retinoic acid receptor (RAR) in exhausted $CD8^+$ T cells in vivo (Fig. 3F and fig. S7B). Indeed, comparison of genome-wide TF footprinting between chronic and acute infection at d27 to identify TF motifs that showed significantly differential inferred binding (Fig. 3G, fig. S7C, and tables S3 and S5) confirmed that *Rara* binding was significantly enriched in exhausted $CD8^+$ T cells (FDR $= 3.14 \times 10^{-13}$) compared with their functional counterparts.

To test whether T cell exhaustion is also associated with a distinct epigenetic state in human exhausted $CD8^+$ T cells, we analyzed global patterns of chromatin accessibility in $tetramer^+$ $CD8^+$ T cells from four subjects with chronic progressive HIV-1 who were not on therapy (Fig. 4, A and B; fig. S8, A and B; and table S7). We successfully mapped 80 to 85% of ChARs identified in the mouse model to their human orthologous regions (Fig. 4A, colored blocks, and fig. S8C) (*18*, *19*) and found them to be enriched for disease-associated single-nucleotide polymorphisms (SNPs) (probabilistic identification of causal SNPs, $P < 2.77 \times 10^{-8}$; hypergeometric test) (fig. S8D) (*20*) and, in particular, immune-related National Human Genome Research Institute genome-wide association study SNPs ($P < 3.70 \times 10^{-3}$) (fig. S8, E to G). This enrichment strongly suggested that mapped regions corresponded to functional regulatory regions within the immune system. Regions at the *Pdcd1* locus were not among those mapped from the mouse model, as previously observed (*10*), which limited our ability to detect an ortholog to the –23.8 kb enhancer observed in the mouse model.

Human naïve $CD8^+$ T cells from the majority of donors showed greater chromatin accessibility in naïve-specific regions defined in the mouse than in memory- or exhaustion-specific regions. In the healthy donor, CMV-specific $tetramer^+$ $CD8^+$ T cells, and effector memory cells were enriched for memory-specific regions (Mann-Whitney *U* test, $P = 0.01$ to $P < 0.0001$) (Fig. 4C and fig. S8H). In contrast, HIV-specific $tetramer^+$ cells from three out of the four subjects showed significantly greater chromatin accessibility in exhaustion-specific regions (Mann-Whitney *U* test, $P = 0.05$ to $P < 0.001$) than in memory-specific regions.

Finally, we confirmed these findings in a subject with chronic hepatitis C virus (HCV) infection in whom $CD8^+$ T cell responses to two epitopes of HCV could be detected (Fig. 4D). Sequencing of the HCV genome in this subject revealed that, unlike the C63B epitope, the 174D epitope had undergone extensive viral escape, and no wild-type viral sequence could be detected (Fig. 4E). We found that the C63B $tetramer^+$ cells had a phenotype consistent with exhaustion and showed significantly greater chromatin accessibility at exhaustion-specific regions (Mann-Whitney *U* test, $P = 0.01$) than memory regions (Fig. 4F). In contrast, 174D $tetramer^+$ cells, which were specific for the escape mutant epitope, lacked exhaustion-specific surface markers and showed greater chromatin accessibility in memory-specific regions, as did influenza-specific $CD8^+$ T cells (Mann-Whitney *U* test, $P = 0.04$) (Fig. 4F). Thus, the state-specific pattern of chromatin accessibility found in mouse exhausted $CD8^+$ T cells is conserved in human exhausted $CD8^+$ T cells.

We find that $CD8^+$ T cell exhaustion occurs with a broad remodeling of the enhancer landscape and TF binding. This suggests that exhausted $CD8^+$ T cells occupy a differentiation state distinct from functional memory $CD8^+$ T cells. Identifying the plasticity of this state and whether or how it could be reverted becomes a critical question for immunotherapy applications. Our data also suggest that mapping state-specific enhancers in exhausted T cells could enable more precise genome editing for adoptive T cell therapy. Genome editing of CAR-T cells to make them resistant to exhaustion is an appealing concept and has led to recent studies investigating the deletion of the PD-1 gene locus (*21*, *22*). Editing exhaustion-specific enhancers (*15*) may provide a more "tunable" and state-specific approach to modulate T cell function than deleting coding regions of genes. Functional maps of enhancers specific to exhausted $CD8^+$ T cells may therefore provide a crucial step toward the rational engineering of T cells for therapeutic use.

REFERENCES AND NOTES

1. E. J. Wherry *et al.*, *Immunity* **27**, 670–684 (2007).
2. A. J. Zajac *et al.*, *J. Exp. Med.* **188**, 2205–2213 (1998).
3. D. L. Barber *et al.*, *Nature* **439**, 682–687 (2006).
4. S. L. Topalian *et al.*, *N. Engl. J. Med.* **366**, 2443–2454 (2012).
5. D. L. Porter, B. L. Levine, M. Kalos, A. Bagg, C. H. June, *N. Engl. J. Med.* **365**, 725–733 (2011).
6. A. H. Long *et al.*, *Nat. Med.* **21**, 581–590 (2015).
7. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, *Nat. Methods* **10**, 1213–1218 (2013).
8. T. A. Doering *et al.*, *Immunity* **37**, 1130–1144 (2012).
9. M. A. Paley *et al.*, *Science* **338**, 1220–1225 (2012).
10. J. W. Austin, P. Lu, P. Majumder, R. Ahmed, J. M. Boss, *J. Immunol.* **192**, 4876–4886 (2014).
11. J. Vierstra *et al.*, *Science* **346**, 1007–1012 (2014).
12. D. Lara-Astiaso *et al.*, *Science* **345**, 943–949 (2014).
13. G. Raimondi, W. J. Shufesky, D. Tokita, A. E. Morelli, A. W. Thomson, *J. Immunol.* **176**, 2808–2816 (2006).
14. O. Shalem *et al.*, *Science* **343**, 84–87 (2014).
15. M. C. Canver *et al.*, *Nature* **527**, 192–197 (2015).
16. J. Vierstra *et al.*, *Nat. Methods* **12**, 927–930 (2015).
17. R. Pique-Regi *et al.*, *Genome Res.* **21**, 447–455 (2011).
18. E. Gjoneska *et al.*, *Nature* **518**, 365–369 (2015).
19. D. Villar *et al.*, *Cell* **160**, 554–566 (2015).
20. K. K. Farh *et al.*, *Nature* **518**, 337–343 (2015).
21. A. Hendel *et al.*, *Nat. Biotechnol.* **33**, 985–989 (2015).
22. K. Schumann *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10437–10442 (2015).

ACKNOWLEDGMENTS

The authors thank members of the Yosef and Haining laboratories for their input, the research subjects for their participation, and J. Doench and the entire Genetic Perturbation Platform at the Broad Institute for their advice on Cas9-mediated screening technology. The authors are grateful for input from the Cancer Center for Genome Discovery. The data reported in this manuscript are tabulated in the main paper and in the supplementary materials. Genome-wide data generated in this study can be accessed via GEO accession no. GSE87646. This research was supported by AI115712, AI091493, and AI082630 to W.N.H. from the NIH; by the BRAIN Initiative grants MH105979 and HG007910 from the NIH to N.Y.; and by 1R21AI078809-01 and UM1 AI068618 from the NIH to N.F. The authors declare no potential conflicts of interest. D.R.S., J.K., N.Y., E.J.W., and W.N.H. are inventors on a patent application (U.S. Patent Application no. 62/310,903) held and submitted by Dana-Farber Cancer Institute.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1165/suppl/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S7
References (*23*–*43*)

18 December 2015; accepted 7 October 2016
Published online 27 October 2016
10.1126/science.aae0491

VACCINATION

Generation of influenza A viruses as live but replication-incompetent virus vaccines

Longlong Si,* Huan Xu,* Xueying Zhou, Ziwei Zhang, Zhenyu Tian, Yan Wang, Yiming Wu, Bo Zhang, Zhenlan Niu, Chuanling Zhang, Ge Fu, Sulong Xiao, Qing Xia, Lihe Zhang, Demin Zhou†

The conversion of life-threatening viruses into live but avirulent vaccines represents a revolution in vaccinology. In a proof-of-principle study, we expanded the genetic code of the genome of influenza A virus via a transgenic cell line containing orthogonal translation machinery. This generated premature termination codon (PTC)–harboring viruses that exerted full infectivity but were replication-incompetent in conventional cells. Genome-wide optimization of the sites for incorporation of multiple PTCs resulted in highly reproductive and genetically stable progeny viruses in transgenic cells. In mouse, ferret, and guinea pig models, vaccination with PTC viruses elicited robust humoral, mucosal, and T cell–mediated immunity against antigenically distinct influenza viruses and even neutralized existing infecting strains. The methods presented here may become a general approach for generating live virus vaccines that can be adapted to almost any virus.

The greatest challenge for converting viruses—such as those responsible for influenza, Ebola, and AIDS pandemics—into live whole-virion vaccines is to render them as avirulent as possible while maintaining their full infectivity to elicit sufficient immunity (*1*–*4*). However, inactivated and attenuated virus vaccines against influenza often exhibit a decrease or even loss of productivity and efficacy (*5*). In addition, immune escape due to antigenic drift and shift (*5*, *6*) introduces a further challenge for the efficacy of conventional influenza vaccines. The modern technology of genetic code expansion (*7*, *8*) and its application (*9*–*13*) to the viral genome may provide the potential to generate live but replication-incompetent virus vaccines eliciting strong and broad immunity (Fig. 1A). A key challenge in designing such a vaccine is to maintain the reproductive potential of the progeny viruses during vaccine production, which requires a special packaging cell line.

We first tested the compatibility of the orthogonal translation system—the *Methanosarcina barkeri* MS pyrrolysyl tRNA synthetase/$tRNA_{CUA}$ pair (MbpylRS/$tRNA_{CUA}$) and the orthogonal unnatural amino acid (UAA) N^{ε}-2-azidoethyloxycarbonyl-L-lysine—with the viral packaging cells. Human embryonic kidney (HEK) 293T cells were transduced with lentivector pSD31s (*14*) for integration of the genes encoding MbpylRS and an amber codon–containing green fluorescent protein (GFP^{39TAG}) into the host genome (fig. S1). The resultant transduced cells were stably transfected with the bjmu-12t-zeo vector (fig. S1), which harbored 12 tandem tRNA-expression cassettes (*15*, *16*). The final transgenic cells, HEK293T-tRNA/pylRS/GFP^{39TAG}, were selected according to the UAA-dependent GFP phenotype and verified by the constitutive expression of the orthogonal tRNA/pylRS pair as tested in their 200th generation (Fig. 1B). The capability of the transgenic cells to propagate viruses was tested in parallel with parental cells by transfection with the packaging plasmids of influenza A/WSN/33 (H1N1; WSN) virus (Fig. 1A) (*17*). The plaque formation assay (*18*, *19*) indicated almost identical viral titers and thus identical packaging efficiency from each transfection (Fig. 1C).

We then randomly selected one codon in the gene encoding the viral NP protein, Asp^{101}, for amber codon replacement via site-directed mutagenesis. The mutant viral genome was reciprocally packaged by the transgenic and parental cells to generate the PTC viruses (Fig. 1A). The production and infectivity of the putative PTC viruses were verified by the cytopathic effect (CPE) assay (*18*, *19*). The CPE phenotype was observed in the transgenic cells only in the presence of 1 mM UAA (Fig. 1D). Clearly, the generation and stepwise reproduction of PTC viruses, as shown in the time-course of CPE accumulation and viral growth curve, occurred only in the transgenic cells and not in the conventional cells (Fig. 1E and fig. S2). This result suggests that PTC viruses had a higher level of safety (i.e., the viruses were replication-incompetent in conventional cells) than did the clinically used cold-adapted live attenuated influenza vaccine (CAIV) (fig. S2) and the attenuated viruses by codon deoptimization, which were still able to replicate in conventional cells and kill the mice at a high dose (*20*–*22*).

One concern over such a design is the potential reversion of the amber codon to a sense codon during PTC virus replication and propagation. To generate genetically stable PTC viruses, we individually replaced 21 extra codons across the NP gene, most of which were located at conserved sites (*23*), with an amber codon (Fig. 2A and figs. S3 to S5). Seven of them—NP-Asp^{101}, NP-Gly^{102}, NP-Gly^{126}, NP-Asp^{128}, NP-Arg^{150}, NP-Met^{163}, and NP-Gly^{169}—caused UAA-dependent CPE, and such dependence was stable with a low escape frequency (7×10^{-10} to 5.9×10^{-7}) even over 20 passages (Fig. 2A, fig. S4, and table S1). In addition, the amber codon substitutions at Asp^{101}, Gly^{102}, and Met^{163} had much less effect on virus packaging efficiency and replication kinetics relative to the wild-type virus (Fig. 2A and fig. S6).

Using the same strategy, we systematically explored the following codons for the generation of replication-incompetent viruses: 22 codons in PB1, HA, NA, and NS; 8 codons in PA, PB2, and M2; and 14 codons in M1 (fig. S3). Different degrees of packaging and propagation efficiency in transgenic cells were observed (figs. S4 to S6 and table S1): 11 UAG codon mutations in PB1, 10 in NA, 5 in HA and PB2, 18 in NS, 6 in PA and M2, and 0 in M1 caused clearly UAA-dependent CPE (figs. S4 and S5). As expected, most of them replicated only in the transgenic cells in the presence of UAA with various replication kinetics and low escape frequencies ($<10^{-6}$) (fig. S6 and table S1). However, six of them (HA-Lys^{57}, NS-Phe^{103}, PB2-Gln^{13}, PB2-Thr^{35}, M2-Lys^{49}, and M2-Lys^{60}) showed relatively high escape frequencies (3.5×10^{-4} to 8.9×10^{-3}) and finally lost UAA dependency, which could be ascribed to mutation of PTC codons as verified by sequencing (fig. S5 and tables S1 and S2). Given the odds of reversion during viral propagation, multiple UAG codons, from two to eight, were introduced into viral genome with each at one RNA segment. The packaging and propagation efficiency of the resultant PTC viruses generally decreased after the introduction of more PTCs into the viral genome (Fig. 2B and table S1). Their escape frequencies also decreased: 1.0×10^{-8} for mono-PTC virus (PTC-1), 1.2×10^{-10} for dual-PTC virus (PTC-2), and undetectable ($<10^{-11}$) for strains containing three or more stop codons (Fig. 2C and table S1).

We next explored the in vivo safety of PTC viruses by intranasally infecting BALB/c mice, ferrets, and guinea pigs. PTC-4A was chosen as a representative (Fig. 2, D and E) because it maintained the intact surface antigens by harboring four stop codons in PA, PB2, PB1, and NP, rather than touching envelope HA and NA genes (figs. S7 and S8 and table S1). In the mouse model study, the median lethal dose (LD_{50}) of wild-type WSN virus was 8×10^3 plaque-forming units (PFU), and $10 \times LD_{50}$ of wild-type viruses killed all mice, with a significant loss of body weight preceding death (fig. S9A). In contrast, no mice were killed by PTC-4A even at a dosage of 10^9 PFU, a factor of 10^5 higher than the LD_{50} of the wild-type viruses; no body weight loss or other health issues were observed (Fig. 3A). Detection of viruses in the turbinate, trachea, or lung of mice 3 days after inoculation of 10^5 PFU of viruses indicated that the mean viral titers were $10^{1.6}$, $10^{1.5}$, and $10^{1.3}$ PFU/g (turbinate, trachea, and lung, respectively) for the

State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, 38 Xueyuan Road, Beijing 100191, China.
*These authors contributed equally to this work. †Corresponding author. Email: deminzhou@bjmu.edu.cn

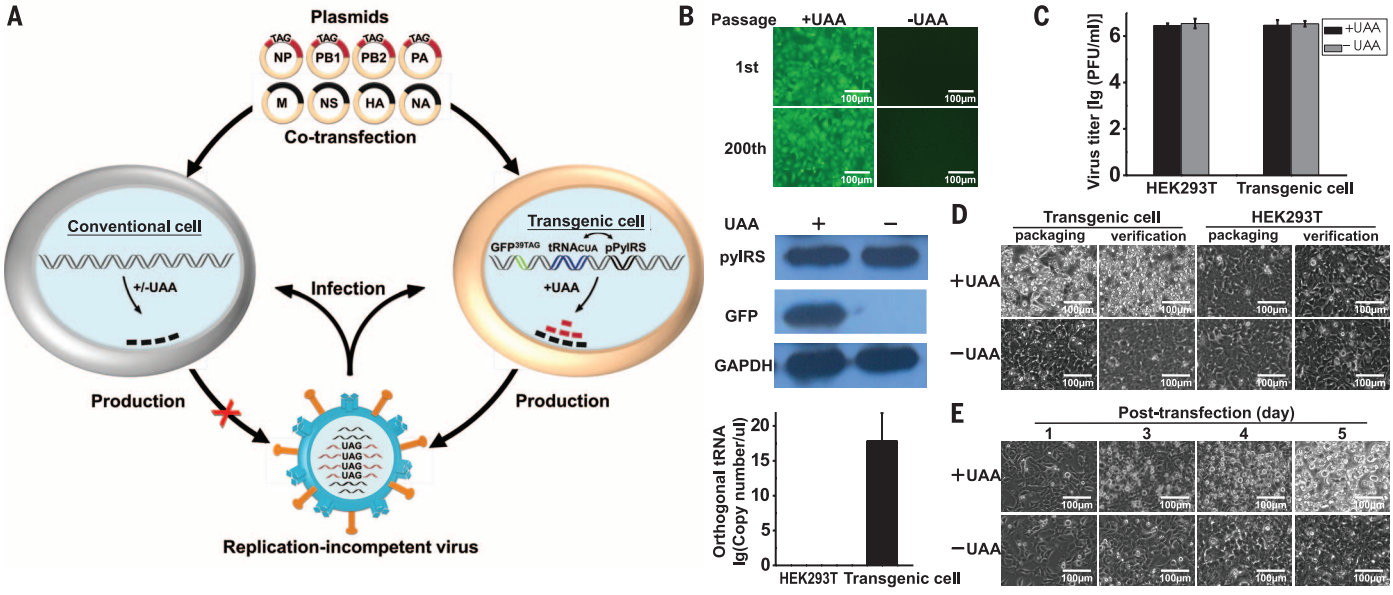


Fig. 1. Establishment of a virion packaging system that is compatible with the orthogonal translation machinery. (**A**) Schematic representation of the generation of premature termination codon (PTC) influenza viruses that are characterized by replication incompetence in conventional cells but high reproduction in transgenic cells that contain integrated cassettes for the expression of orthogonal tRNA ($tRNA_{CUA}$), tRNA synthase (pylRS), and a gene encoding an amber codon–containing GFP (GFP^{39TAG}). NP, nucleoprotein; PB1 and PB2, polymerase basic proteins 1 and 2; PA, polymerase acidic protein; M, matrix protein; NS, nonstructural protein; HA, hemagglutinin; NA, neuraminidase. (**B**) Characterization of the transgenic cells and their genetic stability by constitutive expression of GFP in the presence of UAA (top) and expression of pylRS (middle) and tRNA (bottom) ($N = 3$). (**C**) Functional evaluation of the effect of the orthogonal pylRS/tRNA pair on the propagation of the wild-type WSN viruses by comparing the parental HEK293T and transgenic tRNA/pylRS/GFP^{39TAG} cells in the presence or absence of UAA ($N = 3$). (**D**) Characterization of the infectivity of the progeny viruses by the CPE assay and the replication incompetence in the presence of UAA by comparing the transgenic and parental cells ($N = 3$). (**E**) Reproduction and stepwise accumulation of the PTC viruses in transgenic cells; 10 ± 3%, 50 ± 5%, and 99 ± 1% CPE were observed at days 3, 4, and 5 after inoculation, respectively, conditional on the presence of UAA ($N = 3$). Error bars denote SEM.

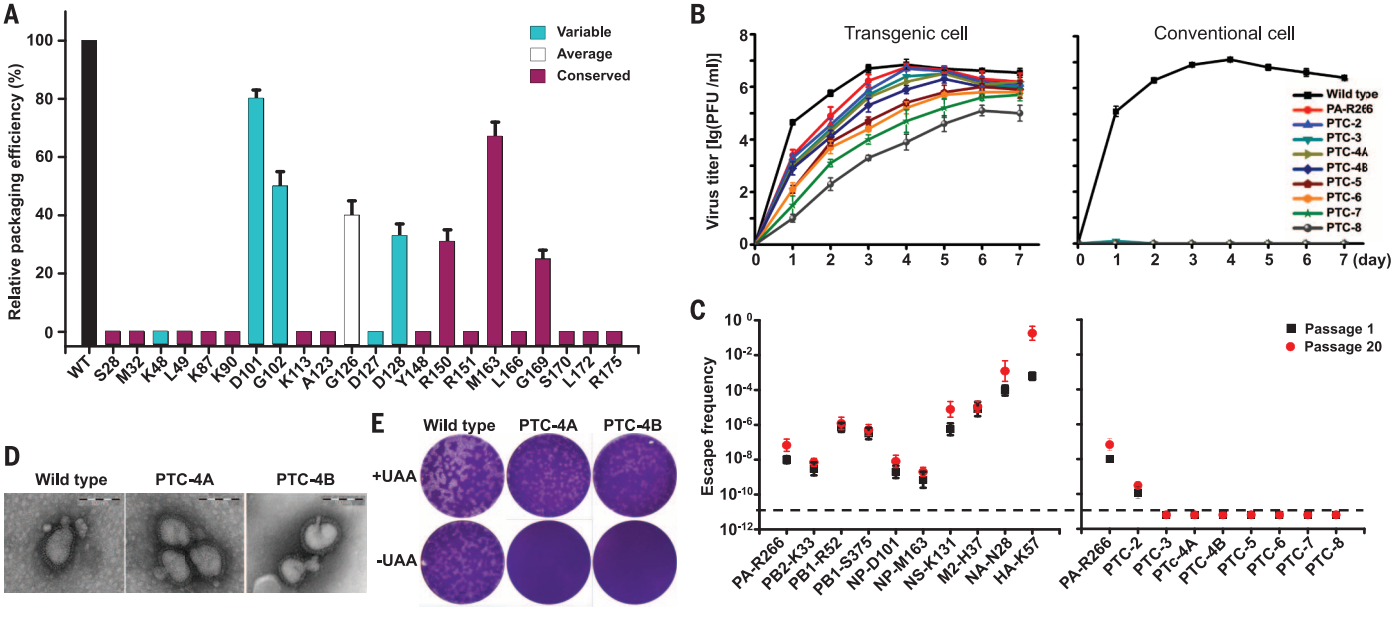


Fig. 2. Genome-wide investigation and characterization of the PTC influenza viruses. (**A**) Systematic exploration of the effect of the introduction of the amber codon at different test sites located in variable, average, or conserved domains, based on the ConSurf analysis (*23*), on PTC virus production. Relative efficiency represents a normalization of the days required to attain ~100% CPE at each test site compared to the wild-type WSN virus. (**B**) Multicycle growth kinetic curves of different PTC viruses in transgenic and conventional cells in the presence of UAA. (**C**) Escape frequencies of different PTC viruses at the first and 20th passage. (**D**) Comparisons of the morphologies of the PTC viruses and the wild-type WSN by transmission electron microscopy Scale bars, 100 nm. (**E**) Direct observations of the UAA-dependent plaque phenotypes of PTC viruses; $N = 3$. Abbreviations for amino acid residues: A, Ala; D, Asp; G, Gly; H, His; K, Lys; L, Leu; M, Met; N, Asn; R, Arg; S, Ser; Y, Tyr. Error bars denote SEM.

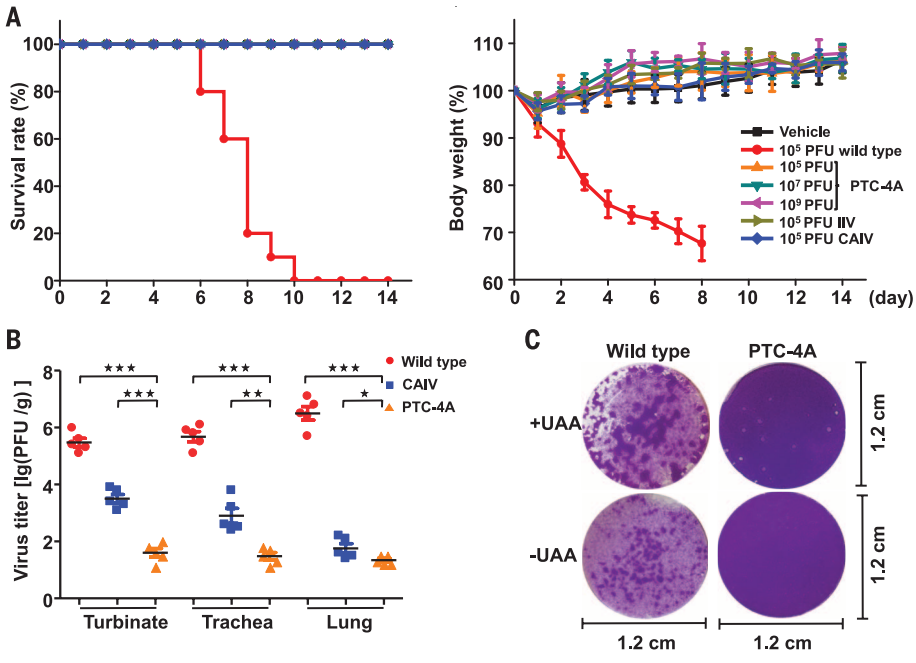


Fig. 3. Characterization of the in vivo safety of the PTC viruses. (**A**) Effect of intranasal virus infection with the indicated viruses or vaccines on the survival rates and body weights of BALB/c mice (n = 10). Error bars denote SEM. (**B**) Detection of the virus titers in mouse tissues (n = 5) on day 3 after infection with 10^5 PFU of the wild-type, CAIV, or PTC-4A viruses. Data are plotted for individual mice (n = 5) and overlaid with means ± SD. *P < 0.05, **P < 0.01, ***P < 0.001 [one-way analysis of variance (ANOVA) with Newman-Keuls multiple-comparisons test]; n.s., not significant. (**C**) Plaque phenotypes of the rescued wild-type and PTC-4A viruses in mice.

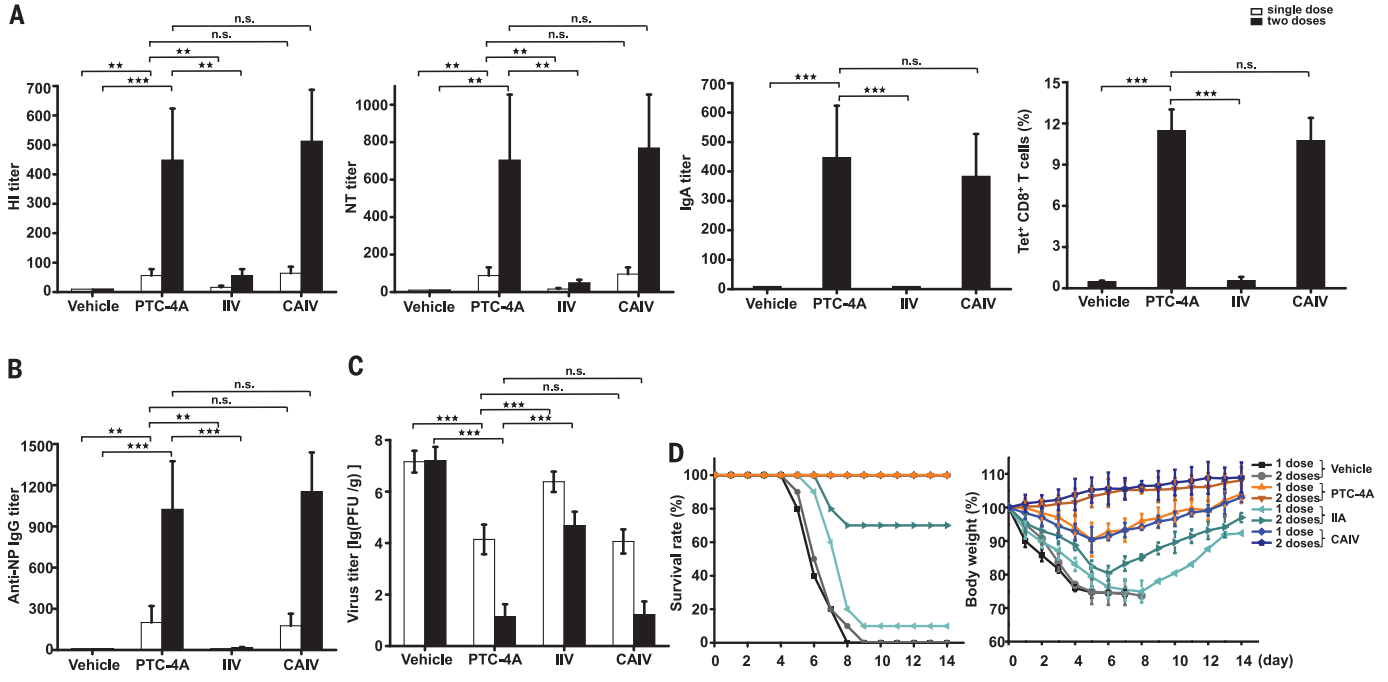


Fig. 4. Characterization of the immunogenicity and protective efficacy of the PTC-4A virus in BALB/c mice that were intranasally inoculated with one or two doses. (**A**) Antibody responses (including serum IgG and mucosal IgA) and virus-specific $CD8^+$ T cell responses, induced by PTC-4A (n = 5). Error bars denote SD. (**B**) Mouse serum IgG antibody responses to the influenza internal protein NP (n = 5). Error bars denote SD. (**C**) Plaque titration of the WSN viruses in the lungs (n = 5) to compare the protective efficacy of the PTC-4A virus versus IIV or CAIV at one or two doses. Error bars denote SD. (**D**) Characterization of the protective efficacy of the PTC-4A viruses in terms of survival rates and body weights of the vaccinated mice (n = 10). Error bars denote SEM. *P < 0.05, **P < 0.01, ***P < 0.001 (one-way ANOVA with Newman-Keuls multiple-comparisons test).

PTC group, $10^{3.5}$, $10^{2.9}$, and $10^{1.8}$ PFU/g for the CAIV group, and $10^{5.5}$, $10^{5.7}$, and $10^{6.5}$ PFU/g for the wild-type group (Fig. 3B). Unlike the rescued wild-type viruses, no plaque was observed for the rescued PTC viruses from mice in the absence of UAA (Fig. 3C), which suggests that PTC reversion did not occur in vivo. In addition, respiratory droplet transmission experiments indicated that PTC viruses were not detected in noninoculated guinea pigs when caged together with inoculated guinea pigs, in contrast to the findings for wild-type viruses and CAIV (fig. S9B). All data indicated that influenza viruses have been rendered avirulent via the introduction of amber codons, as confirmed by ferret and guinea pig model studies (fig. S9, C and D).

Next, we tested the immunogenicity of PTC-4A by comparison with the commercially available

inactivated influenza vaccine (IIV) and CAIV. Three weeks after immunization of 10^6 PFU for mice and 10^7 PFU for ferrets and guinea pigs, PTC-4A and CAIV, both as whole virions, induced robust antibodies in sera according to hemagglutination inhibition (HAI) and neutralization (NT) antibody assay, but IIV did not do so, even after the second vaccination (Fig. 4A and figs. S10 to S12). Impressively, the second vaccination by PTC-4A caused the HI and NT antibody titers to increase further by a factor of approximately 6 to 8, comparable to the effect of CAIV. Such a PTC virus–mediated immune response was also observed against internal NP antigen; this effect was not observed for IIV even after a second vaccination (Fig. 4B and fig. S11A). Furthermore, a high level of secretory immunoglobulin A (IgA), representing the mucosal immune response, was elicited in the lungs by both PTC-4A and CAIV but not by IIV vaccine (Fig. 4A and fig. S11A). Moreover, approximately 10 times as many virus-specific cytotoxic T lymphocytes (CTLs) were elicited by PTC-4A and CAIV than by IIV, which only elicited a basic level of CTLs in the lungs (Fig. 4A). Our results indicate that the PTC virus vaccines elicit robust humoral, mucosal, and cell-mediated immunity, which may be due to their full infectious form carrying the native conformation of all surface antigens and internal viral components.

The protective efficacy of PTC-4A virus was tested and compared with that of CAIV and IIV. Three weeks after vaccination, 15 mice in each group were challenged intranasally with $50 \times LD_{50}$ of wild-type viruses. Three days after challenge, five mice from each group were killed for plaque titration of wild-type viruses in the lungs (Fig. 4C). We found a significant decrease of viral titers in both PTC-4A and CAIV groups relative to the IIV group, in particular for dual vaccinations, which led to a further decrease of viral titers by a factor of 10^3 for both PTC-4A and CAIV. In terms of survival rate and body weight for the remaining mice, all mice in the vehicle group succumbed by 9 days after challenge, whereas all mice survived after immunization even with just one dose of PTC-4A or CAIV (Fig. 4D). The mice inoculated with one dose of PTC-4A regained their body weight 5 days after challenge, whereas no weight loss was observed in mice inoculated with two doses of PTC-4A or CAIV, which we also confirmed in ferret and guinea pig model studies (figs. S11 and S12).

We also tested whether the PTC viruses exhibit heterologous protection against antigenically distant influenza strains. Fifteen mice, vaccinated with a single dose or two doses of PTC-4A, were challenged with 10^6 PFU of influenza A/reassortant/NYMC X-179A (pH1N1) or A/Aichi/2/68 (H3N2) viruses. Significant heterologous protection was observed, according to plaque titration and observations of the survival rates and body weight changes (fig. S13). Such heterologous protection was ascribed to the conserved regions of viral proteins, including surface and internal antigens among distant influenza viruses, and to virus-specific CTLs that collectively contribute to the broad protective effect (*5*, *24*–*26*).

We then tested whether the existing viral virulence of PTC viruses could be enhanced by co-propagating wild-type and PTC viruses in conventional cells. Rather than more plaques, we observed remarkably fewer plaques or even undetectable plague formation, depending on the ratio of wild-type versus PTC viruses and the number of PTC codons in the PTC viruses (fig. S14, A and B). This contrasted with CAIV, which exerted no inhibitory effect on the propagation of wild-type viruses. Furthermore, the capability of attenuating wild-type viruses by PTC viruses was verified by in vivo experiments (fig. S14C). This was due to the genetic reassortment between the existing and PTC viruses, as verified by sequencing (fig. S14D). Progeny reassortants harbored at least one replication-incompetent gene segment, providing an effective approach for "neutralization" of infectious viruses.

Our results show that we succeeded in generating live but replication-incompetent virus vaccines by applying genetic code expansion to the influenza virus genome. Such live vaccines elicited robust immunity against both parental and antigenically distinct strains. Generation of such PTC virus vaccines can be potentially adapted to almost any virus (*27*) so long as their genome could be manipulated and packaged in a cell line. Furthermore, the multiple PTC-harboring viruses are not only prophylactic but also therapeutic vaccines in the neutralization of the replicating viruses.

REFERENCES AND NOTES

1. F. Krammer, P. Palese, *Nat. Rev. Drug Discov.* **14**, 167–182 (2015).
2. A. Marzi *et al.*, *Science* **348**, 439–442 (2015).
3. J. B. Ulmer, U. Valley, R. Rappuoli, *Nat. Biotechnol.* **24**, 1377–1383 (2006).
4. S. A. Plotkin, *Nat. Med.* **11**, S5–S11 (2005).
5. Y. H. Jang, B. L. Seong, *Clin. Exp. Vaccine Res.* **1**, 35–49 (2012).
6. S. Tong *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4269–4274 (2012).
7. L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **292**, 498–500 (2001).
8. J. W. Chin *et al.*, *Science* **301**, 964–967 (2003).
9. A. J. Rovner *et al.*, *Nature* **518**, 89–93 (2015).
10. S. B. Sun, P. G. Schultz, C. H. Kim, *ChemBioChem* **15**, 1721–1729 (2014).
11. Y. Zheng *et al.*, *Nucleic Acids Res.* **43**, e73 (2015).
12. N. Wang *et al.*, *Angew. Chem. Int. Ed.* **53**, 4867–4871 (2014).
13. C. Zhang *et al.*, *Biomaterials* **80**, 134–145 (2016).
14. X. Jin *et al.*, *Nucleic Acids Res.* **40**, 1797–1806 (2012).
15. J. Hsieh, A. Fire, *Annu. Rev. Genet.* **34**, 187–204 (2000).
16. W. Wang *et al.*, *Nat. Neurosci.* **10**, 1063–1072 (2007).
17. G. Neumann *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9345–9350 (1999).
18. M. Yu *et al.*, *J. Med. Chem.* **57**, 10058–10071 (2014).
19. S. Xiao *et al.*, *Biomaterials* **78**, 74–85 (2016).
20. E. J. Jung, K. H. Lee, B. L. Seong, *Exp. Mol. Med.* **42**, 116–121 (2010).
21. J. R. Coleman *et al.*, *Science* **320**, 1784–1787 (2008).
22. A. Nogales *et al.*, *J. Virol.* **88**, 10525–10540 (2014).
23. H. Ashkenazy, E. Erez, E. Martz, T. Pupko, N. Ben-Tal, *Nucleic Acids Res.* **38**, W529–W533 (2010).
24. A. Impagliazzo *et al.*, *Science* **349**, 1301–1306 (2015).
25. D. R. Burton, P. Poignard, R. L. Stanfield, I. A. Wilson, *Science* **337**, 183–186 (2012).
26. Z. Staneková, E. Varečková, *Virol. J.* **7**, 351 (2010).
27. J. Cohen, *Science* **351**, 16–19 (2016).

ACKNOWLEDGMENTS

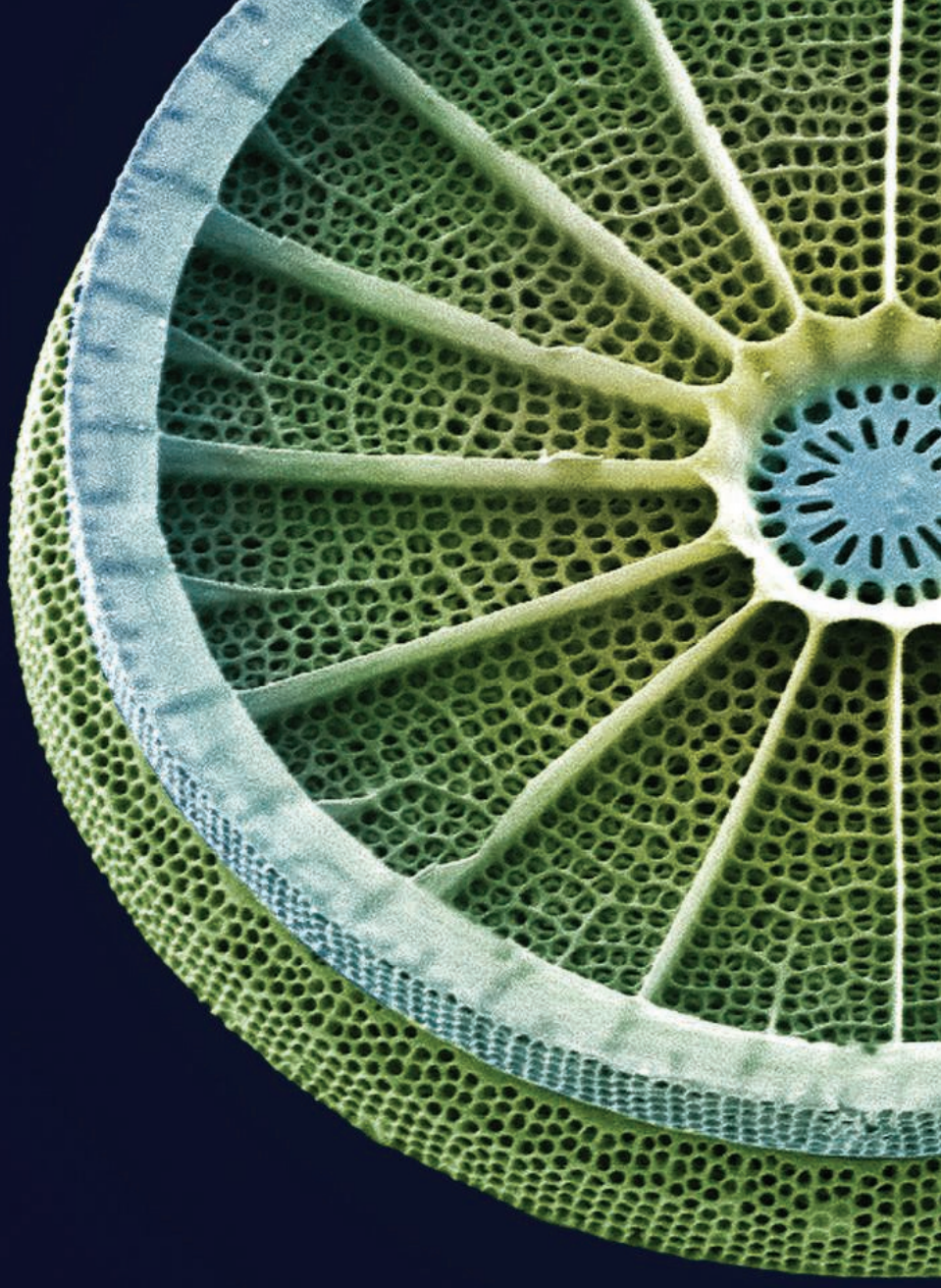
We thank Sinovac Biotech Ltd. (Beijing) for technical assistance with the animal work. Supported by National Natural Science Foundation of China grants 81530090 and 21572015 and by National Key Research and Development Program of China grant 2016YFA0501500. All authors declare no competing financial interests. D.Z. and L.S. are inventors on a patent application (PCT/CN2016/092778) held by Peking University that covers the preparation of PTC viruses through genetic codon expansion. Sharing of materials will be subject to standard material transfer agreements. The nucleotide sequences used in the study have been deposited in GenBank under accession numbers CY034139.1, CY034138.1, X17336.1, HE802059.1, CY034135.1, CY034134.1, D10598.1, and M12597.1; additional data are presented in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6316/1170/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 and S2
References (*28*–*41*)

18 July 2016; accepted 8 November 2016
10.1126/science.aah5869

NEW ENGLAND BioLabs® Inc.



Target with precision.

Introducing the NEBNext Direct™ Cancer HotSpot Panel

Using a unique approach, the NEBNext Direct Cancer HotSpot Panel enriches for 190 common cancer targets from 50 genes prior to next generation sequencing. Combining a novel method for hybridization-based target enrichment with library preparation, the NEBNext Direct technology reduces processing time and minimizes sample loss. Ideal for automation, NEBNext Direct enables highly-specific deep sequencing of genomic regions of interest for the discovery and identification of low frequency variants from challenging sample types.

Visit **NEBNextDirect.com** to learn more and to inquire about sampling this product.

TARGETS INCLUDE REGIONS FROM THE FOLLOWING CANCER-RELATED GENES, INCLUDING >18,000 COSMIC FEATURES:

ABL1	EGFR	GNAQ	KRAS	PTPN11
AKT1	ERBB2	GNAS	MET	RB1
ALK	ERBB4	HNF1A	MLH1	RET
APC	EZH2	HRAS	MPL	SMAD4
ATM	FBXW7	IDH1	NOTCH1	SMARCB1
BRAF	FGFR1	IDH2	NPM1	SMO
CDH1	FGFR2	JAK2	NRAS	SRC
CDKN2A	FGFR3	JAK3	PDGFRA	STK11
CSF1R	FLT3	KDR	PIK3CA	TP53
CTNNB1	GNA11	KIT	PTEN	VHL

For research use only; not intended for diagnostic use.

Explore a World of Cellular Information

The human cell like never before.
A protein atlas of high-resolution microscopy images
for you to explore for free.

THE CELL ATLAS - created by researchers for researchers

THE HUMAN PROTEIN ATLAS

www.proteinatlas.org

High-Throughput Evaporator

The Series 3 HT evaporator range represents the ultimate in solvent removal technology. Incorporating a high-performance vacuum pump, the latest touchscreen technology, and a sleek ergonomic design, the Series 3 HT evaporator optimizes evaporation processes and makes them effortless. Available in 6- and 12-sample holder configurations, these evaporators are the system of choice for high-throughput (HT) sample drying, offering high capacity, unique technology that eliminates cross contamination due to sample bumping, and automated temperature and pressure monitoring to prevent sample degradation. Intuitive touchscreen controls enhance monitoring and review of the whole evaporation process. For popular solvent removal protocols, preset "press and go" methods make operation easy and productive even for occasional users. Simplified manual and automatic programming means even the most complex multistage evaporation methods can be set up and run quickly and easily.

Genevac
For info: +44-(0)-1473-240000
www.genevac.com

Biomanufacturing Platform

ProteoNic's CHOGUN expression vector is included alongside Horizon's HD-BIOP3 GS Null CHO cell line, providing a complete biomanufacturing platform. Biomanufacturing in Chinese hamster ovary (CHO) cells is performed based on high-expressing clones, the identification of which can be streamlined through metabolic selection. The only source of glutamine production in the cell is through the enzyme glutamine synthetase (GS). Methionine sulfoximine (MSX) inhibits GS activity in cells so that only those clones that are most efficient at producing recombinant protein are able to proliferate. However, MSX is toxic. An alternative approach is to use a cell line that lacks GS (GS Null). When the gene for a biotherapeutic product is provided on a vector alongside GS, such as in the CHOGUN vector, production of GS is linked to the production of the protein of interest, so cells that survive selection are more likely to be expressing the product at high levels.

Horizon Discovery
For info: 844-655-7800
www.horizondiscovery.com

β-Glucuronidase Removal Kit

β-Gone sorbent delivers fast, one-step β-glucuronidase removal from hydrolyzed urine. Analysts in forensic toxicology and clinical testing labs have wrestled for years to find a solution that is rapid yet sensitive when working with hydrolyzed urine samples. Within the body, a glucuronidation reaction during metabolism promotes the elimination of drugs through urine. To accurately analyze drug metabolites, the glucuronide must be cleaved, typically with hydrolysis using an enzyme. Hydrolysis with β-glucuronidase produces a urine sample with a significant amount of solubilized enzyme that must be removed before HPLC analysis. If the enzymes are not removed from the sample, they can build up on the HPLC column, reducing its lifetime. β-Gone also offers a better alternative to the "dilute-and-shoot" method in which sensitivity and accuracy are compromised when the sample is diluted. It also improves assay performance and column lifetime, and reduces mass spectrometer maintenance.

Phenomenex
For info: 310-212-0555
www.phenomenex.com

Zika Virus Recombinant Antigens and Antibodies

Recombinant antigens and antibodies for Zika virus (ZIKV) are available for the development of rapid assays, and will aid researchers to overcome the problem of cross-reactivity with related viruses. ZIKV is causing worldwide concern because of its connection to a neurological birth disorder and its rapid spread. The virus is closely related to other mosquito-borne flaviviruses. PCR assays and antibody (IgM/IgG) detection assays have difficulty in differentiating between mosquito-borne flaviviruses, and only detect infections up to five days after the onset of symptoms. Recombinant antigens of envelope and NS1 proteins can overcome the problem of flavivirus cross-reactivity with the closely related dengue, yellow fever, Japanese encephalitis, and West Nile viruses in diagnostics. The recombinant proteins are produced in Sf9 insect cells and have been tested to work in lateral-flow formats. These antigens are available in 1-mg size for evaluation, and in gram-bulk quantiies for the development of diagnostic assays.

AMS Biotechnology
For info: +44-(0)-1235-828200
www.amsbio.com

Triple Quadrupole GC/MS

Laboratories that test the safety of our food and water are turning to triple quadrupole GC/MS systems because their selectivity enables them to better handle complex matrices than standard single quadrupole systems. Agilent's 7000D and 7010B include enhancements that not only improve analytical performance but make them easier to use than any previous triple quad GC/MS system. Those who want to run samples on their triple quad can open existing single quad GC/MS methods with a single mouse click using Agilent MassHunter software. This capability also streamlines the transition from single-ion monitoring on the single quad to tandem mass spectrometry (MS/MS) on the triple quad. The 7010B includes Agilent's exclusive high-efficiency ion source, which enhances the instrument's sensitivity by a factor of 10. Both models are also available with Agilent's patented JetClean self-cleaning ion source, which readily removes matrix deposits that can build up over time and compromise instrument performance.

Agilent Technologies
For info: 877-424-4536
www.agilent.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

Soochow University Seeks Global Talents

SOOCHOW UNIVERSITY

About Soochow University

Established in 1900, Soochow University (SU) is one of the oldest modern universities in China. It is located in the historical and cultural city of Suzhou, which is reputed as the "Venice of the Orient" and "Paradise on Earth". After 116 years of development, Soochow University has become a comprehensive research-oriented university with distinctive features and growing impact at home and abroad. As a member university of the national "211 Project" and "2011 Plan", it has embarked on a new journey to become a well-known university in the world and a first-class university in China.

SU has always been concentrating on cultivating talents with independent personality, free spirit, remarkable ability, and social responsibility. At present, it has over 50,000 students who study in various majors for their PhD, master and bachelor degrees. There are altogether 24 Level-I doctoral degree programs and 29 post-doctoral research stations in SU.

SU is deeply committed to academic and scientific research. With its creative, interdisciplinary and international approaches to research, SU has made great contribution to fundamental discoveries and technological breakthroughs. For the past three consecutive years, its total number of projects founded by the National Science Foundation of China (NSFC) has ranked top 20 among 2595 domestic universities, and 7 disciplines are listed in Thomson Reuter's Essential Science Indicators (ESI) as world's top 1% disciplines.

Contact Us
Human Resource Department, SU
E-mail: sdrcjl@suda.edu.cn
Tel: +86-512-67503248
+86-512-67503045
Website: http://rsc.suda.edu.cn/

About the "Distinguished Professor" Program

SU's "Distinguished Professor" Program aims at recruiting high-end experts, and young and middle-aged scholars on a global scale. Professors of the program will be offered a competitive and personalized salary and compensation, as well as favorable start-up research funding, workplace and research assistants. The university also assists the professors to apply for all forms of talent programs, and the grant will be allocated to individual applicants.

Fields of Recruitment

Mathematics, Physics, Chemistry, Information Science, Engineering, Material Science, Life Science, Medicine, Philosophy, Economics, Law, Pedagogy, Literature, History, Management, Art

Qualifications & Requirements

Experts and scholars from overseas are welcomed to join us and work full-time at SU.
For young scholars, applicants are expected to be under 35 years old, holding a doctoral degree of a world-renowned university overseas, or a doctoral degree of a domestic university with over three years of research experience overseas.

Application documents

Qualified applicants can submit the application materials to sdrcjl@suda.edu.cn.
Application materials include: a CV, at least 3 recommendation letters, and at least 5 representative papers, etc.

By **Matthew Tuthill**

Making a difference, differently

Ten years ago, after 2 years as a postdoc, I found myself wondering whether I should take a different road. Up to that point, I had stuck to a pretty traditional path investigating cancer genetics, but the research grind had taken its toll and my interest was waning. At the same time, federal funding had flattened, and the scarcity of secure, long-term employment in academic research added to my dissatisfaction. So I decided to take the leap and leave the typical academic path to find a better match. Then came the hard part: identifying a new career that would nurture my passion for science and allow me to make an impact with my work. I found it in a place many would not expect: a 2-year community college.

As I was considering my options, I found inspiration in my first graduate school research mentor, Randal Wada, whose work reminded me that scientists' efforts away from the bench can be incredibly powerful. In addition to developing neuroblastoma therapies, Randy had created the Hawaii Cord Blood Bank, which helps improve stem cell transplant matches for bone marrow recipients. His work made me realize that I could serve science and society in lots of ways if I simply opened my mind.

But I still didn't know exactly what I should do. Then a second bit of inspiration came my way. A colleague mentioned that a professor at a nearby 2-year college was training students to produce monoclonal antibodies for labs on campus and for paying clients. I was impressed that the professor, John Berestecky, had taken on this type of ambitious project with relatively inexperienced students. Curious to find out more, I set up a meeting with John and was struck by his sincerity and the way he prioritized student training above grants, publications, and personal ambition. I could also see his passion for teaching, which reminded me of my own back when I started college, planning to become a high school biology teacher. Since graduate school I had not thought about teaching, let alone working at a 2-year college—many of my research mentors and peers considered it career suicide—but meeting John made me see things differently.

Here, at last, was a way to fuse my interest in science with my passion for teaching. I found a faculty position and joined John at the same quiet junior college in the middle of the Pacific. Now, I effectively hold two positions: classroom instructor and research co-adviser of 15 inexperienced but eager undergraduates. Both roles give me a chance to help students transform themselves, which is enormously rewarding. They learn lab skills and concepts impressively fast, and it's gratifying to see them gain confidence in their abilities. The deeper emotional connections I have forged with students over the years have been unexpected gifts that continually inspire me.

"It's gratifying to see [my students] gain confidence in their abilities."

My work has its challenges. It's disheartening when others see both my students and me as less worthy because we are not at universities. We sometimes struggle to get access to adequate research space, federal funding, scientific conferences, and other resources and opportunities. Because my students are novices, investigating complex biomedical questions requires creativity and restraint. Long-range projects take even longer than they would in other settings because students leave for universities just as their skill sets have finally matured. My pay is below the standard at 4-year research institutions, even though my teaching workload is greater. But my occasional frustration is buffered by the thought of the students, many of whom are considered underrepresented minorities, that I have helped train.

Looking back at these 10 years, I realize how much my work on this campus has helped me grow, both as an academic and a mentor, just as it has for many of my students. For my sake as much as for theirs, I'm grateful that I stepped away from a prescribed career path and found a way to serve both the student and research communities in my own way, modest though it may be. ■

Matthew Tuthill is an associate professor at Kapi'olani Community College in Honolulu. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER