How COVID-19 sped the quest for cleaner indoor air p. 612

A sensing neuroprosthesis wins *Science* & PINS Prize p. 634

Cool fabrics for thermal management p. 692

Science

$15
6 AUGUST 2021
sciencemag.org

AAAS



SQUIRREL PARKOUR

Learning to leap and land pp. 620 & 697



NC STATE

An Engineering Powerhouse

Top 10 for research expenditures. Twelfth for graduate programs among public institutions. And sixth among online master's programs. Welcome to one of the leading engineering colleges in the United States.

From biomedical engineering to materials science to quantum computing, NC State is a growing hub of innovation — and one of the only universities to ever host two NSF Engineering Research Centers at once.

Join us at the forefront.

go.ncsu.edu/engineering-powerhouse

Rankings from U.S. News & World Report

CONTENTS

6 AUGUST 2021 • VOLUME 373 • ISSUE 6555



NEWS

IN BRIEF

602 News at a glance

IN DEPTH

605 Taliban's rise puts polio eradication in danger
U.S. departure from Afghanistan raises questions about future of vaccination drives *By L. Roberts*

606 WHO relaunches global drug trial with three new candidates
Solidarity kicks back into action after 6-month pause *By K. Kupferschmidt*

607 Longer days on early Earth set stage for complex life
Extra light spurred oxygen release by mats of photosynthetic bacteria, models and sinkhole data suggest *By E. Pennisi*

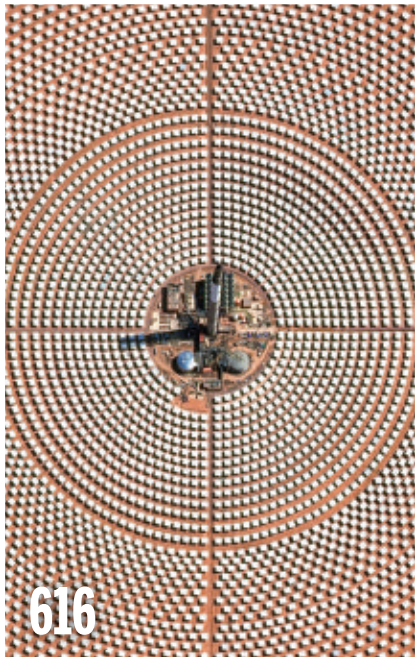
608 Chinese students fight back against visa rejections
A Trump-era proclamation may block thousands from study or research in the United States *By D. Normile*

610 Conflict flares over incidental genetic findings
Bioethicists reject proposal to routinely tell research participants about disease risks *By M. Wadman*

611 Studies probe how payouts affect U.S. vaccination rates
Money can sometimes tip the balance for the hesitant *By A. Oza*

FEATURES

612 The air investigator
For Joseph Allen, the pandemic has made the connection between indoor air quality and human health clearer than ever *By D. Starr*



INSIGHTS

POLICY FORUM

616 An action agenda for Africa's electricity sector
Modernization and expansion require heightened efforts *By D. Puig* et al.

PERSPECTIVES

620 Learning to move in the real world
Leaping squirrels show that locomotion entails perceiving and innovating possibilities for action from moment to moment
By K. E. Adolph and J. W. Young
REPORT p. 697

621 Maturation of HIV-1
Structural transformation of HIV-1 matrix during virus maturation promotes infection
By Y. Hikichi and E. O. Freed
REPORT p. 700

623 Therapy based on functional RNA elements
Fragments of long noncoding RNAs show potential in treating a metabolic disorder in mice
By R. B.-T. Perry and I. Ulitsky
RESEARCH ARTICLE p. 662

624 Treatments for Alzheimer's disease emerge
Anti-amyloid immunotherapies will provide the first disease-modifying therapeutics
By D. J. Selkoe

PHOTOS: (TOP TO BOTTOM) KEN RICHARDSON; DIGITALGLOBE/GETTY IMAGES

626 Phenotyping Alzheimer's disease with blood tests
Affordable and accessible tools could be used for screening and therapy monitoring
By K. Blennow

628 Bioinspired nanofluidic iontronics
Electrolytes in planar nanochannels are predicted to function as nanofluidic memristors *By Y. Hou and X. Hou*
REPORT p. 687

BOOKS *ET AL.*

630 To move beyond the paradigms of Western medicine
A pair of authors aspire to decolonize medicine through liminality, fugitivity, and abolition *By C. L. Ford*

631 Governing the genome
A political scientist proposes a framework for understanding genomic policy debates
By J. Blatt

LETTERS

632 Restore protected status for gray wolves
By P. Kareiva et al.

633 Brazil can protect sharks worldwide
By B. S. Rangel et al.

633 Call for protection of scatter-hoarding rodents
By S. Yi and X. Yi

PRIZE ESSAY

634 Neurorobotics for neurorehabilitation
Peripheral neuromodulation boosts health and cognitive benefits in leg amputees
By S. Raspopovic



RESEARCH

IN BRIEF

637 From *Science* and other journals

REVIEW

640 Nanomaterials
Semiconductor quantum dots: Technological progress and future challenges
F. P. García de Arquer et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.AAZ8541

RESEARCH ARTICLES

Coronavirus

641 Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity *S. M.-C. Gobeil* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABI6226

642 Structural basis for enhanced infectivity and immune evasion of SARS-CoV-2 variants *Y. Cai* et al.

648 SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern *M. McCallum* et al.

655 Plant genomics
De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes *M. B. Hufford* et al.

662 Phenylketonuria
A noncoding RNA modulator potentiates phenylalanine metabolism in mice *Y. Li* et al.
PERSPECTIVE p. 623

673 Quantum sensing
Quantum-enhanced sensing of displacements and electric fields with two-dimensional trapped-ion crystals
K. A. Gilmore et al.

679 Chemical kinetics
Watching a hydroperoxyalkyl radical (•QOOH) dissociate *A. S. Hansen* et al.

REPORTS

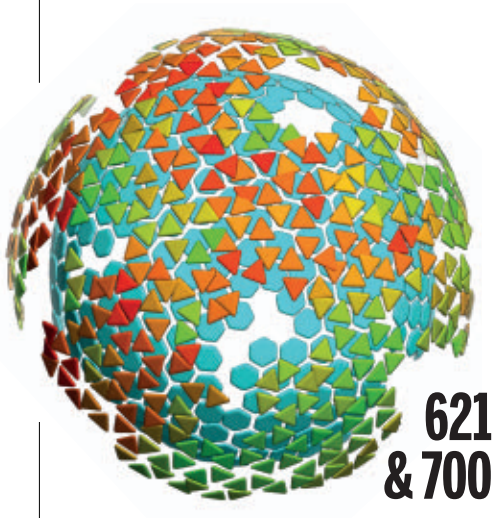
683 Metallurgy
Suppressing atomic diffusion with the Schwarz crystal structure in supersaturated Al–Mg alloys *W. Xu* et al.

687 Nanofluidics
Modeling of emergent memory and voltage spiking in ionic transport through angstrom-scale slits *P. Robin* et al.
PERSPECTIVE p. 628

692 Cooling textiles
Hierarchical-morphology metafabric for scalable passive daytime radiative cooling
S. Zeng et al.

697 Biomechanics
Acrobatic squirrels learn to leap and land on tree branches without falling *N. H. Hunt* et al.
PERSPECTIVE p. 620

700 Structural biology
Maturation of the matrix and viral membrane of HIV-1 *K. Qu* et al.
PERSPECTIVE p. 621



DEPARTMENTS

601 Editorial
Blue carbon can't wait *By Fanny Douvere*

710 Working Life
Over the baby bump *By Martha Nelson*

ON THE COVER

A free-ranging fox squirrel leaps from one platform to another. To develop their robust gap-crossing skills, squirrels must learn to rapidly estimate jump distance and landing structure stability. Such combinations of cognitive evaluation and innate physical abilities may play a key role in vertebrate movement. See pages 620 and 697. *Photo: Lawrence Wang, Dominic Pang, Sebastian Lee, Jeremy Snowden, Tina Kuang, and Frank Aliaga Auqui*



Science Careers 705

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2021 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2148; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.sciencemag.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

PHOTO: DENNIS FAST/VWPICS/UNIVERSAL IMAGES GROUP/GETTY IMAGES; GRAPHIC: K. QU *ET AL.*

EDITORIAL

Blue carbon can't wait

When the United Nations released its World Ocean Assessment in 2015, it was clear that the oceans were seriously degraded, with stressors on these environments projected to increase. The 2021 Assessment, released in April, shows that they have further declined, bringing us ever closer to losing the structure, function, and benefits of Earth's marine systems. One way forward might be to focus on "blue carbon" ecosystems and the incentives they offer through carbon credits linked to decreasing carbon emissions.

Blue carbon ecosystems include seagrass meadows, tidal marshes, and mangroves, all of which are among Earth's most efficient absorbers and long-term storers of carbon. This capacity for carbon storage also makes them sources of CO_2 emissions when they are degraded or destroyed.

As the United Nations Environment Programme states in its April 2021 report *Making Peace with Nature*, "Ecosystem restoration can simultaneously mitigate climate change, slow and reverse biodiversity decline and increase the benefits that people get from nature." Restoring blue carbon ecosystems could remove about 0.5% of current global emissions, with co-benefits for local ecosystems and livelihoods. These include improved water quality; increased marine and terrestrial biodiversity; preservation of livelihoods, cultural practices, and values of local and traditional communities; and the protection of shorelines and their resilience in the face of climate change. That's quite a return on investment.

We can start by ensuring that blue carbon ecosystems already identified for protection get the support they need. For example, the 50 United Nations Educational, Scientific and Cultural Organization (UNESCO) marine World Heritage sites make up just 1% of the world's oceans but host 21% of the world's blue carbon ecosystems and 15% of the world's stored blue carbon—equal to 10% of annual global greenhouse gas emissions. Protecting those blue carbon ecosystems will keep billions of additional CO_2 emissions from entering the atmosphere. Their carbon stores are nominally protected from degradation by the legal commitment of over 190 signatory countries to the 1972 UNESCO World Heritage Convention. But local management of these sites is often underfunded and understaffed, and the sites face a variety of threats from pollution, coastal development, and climate change. For the 79% of blue carbon ecosystems outside of UNESCO protection—harboring the remaining 85% of the ocean's stored carbon—marine protected areas (MPAs) can benefit climate mitigation by keeping billions of additional CO_2 emissions from entering the atmosphere. At the same time, MPAs also have been shown to protect biodiversity and boost fishery yields.

Where will the funding come from to support conservation? In part, through the use of blue carbon credits. Blue carbon ecosystems can help nations avoid releasing additional CO_2 and other greenhouse gases. The growing blue carbon credit market allows organizations and countries that conserve and restore blue carbon ecosystems to claim or sell credits in global carbon credit markets. For example, a country that protects seagrass meadows, and hence reduces the risk of future carbon emissions, may receive carbon credits that fund their future protection. Carbon credits can also be claimed for areas damaged by natural events. Wherever blue carbon habitats have suffered losses, blue carbon credits may help support their restoration.

> "Blue carbon ecosystems can help nations avoid releasing ... greenhouse gases."

So far, few countries have incorporated blue carbon strategies into their climate change mitigation policies. The Blue Carbon Initiative, created by UNESCO's Intergovernmental Oceanographic Commission, the International Union for Conservation of Nature, and Conservation International, encourages more countries to develop comprehensive methods for assessing blue carbon stores and emissions.

But momentum is gathering. One of the largest seagrass restorations to date, in the United States, has applied for blue carbon credit certification. On the other side of the globe, Madagascar's Tahiry Honko and Kenya's Mikoko Pamoja mangrove restorations will be the first MPAs funded by blue carbon credits. At Tahiry Honko, revenues generated through carbon credits will cover recurring park costs associated with its mangrove protection and support education and infrastructure projects for the local community.

As the world aims for carbon neutrality in the decades ahead, we can take actions today to help slow climate change. The return on investing in blue carbon ecosystems is clear, meaningful, and immediate. We can't afford to wait.

–Fanny Douvere



Fanny Douvere is head of the Marine Programme at UNESCO's World Heritage Centre, Paris, France. f.douvere@unesco.org

PHOTO: FRANCOIS MORI/AP/SHUTTERSTOCK

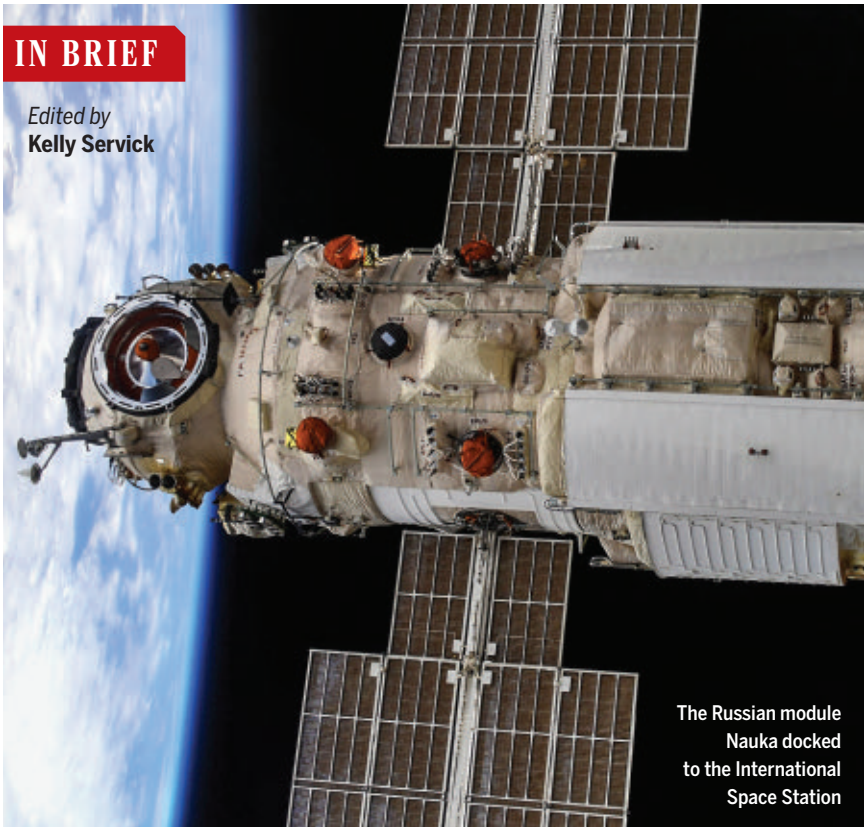
10.1126/science.abl7128

NEWS

> “This pandemic was a big test for AI [artificial intelligence] and medicine. … But I don’t think we passed that test.”
>
> **University of Cambridge machine learning researcher Derek Driggs**, to *MIT Technology Review*, on findings that none of hundreds of AI-based tools to diagnose or triage COVID-19 patients are fit for clinical use.

IN BRIEF

Edited by
Kelly Servick



The Russian module Nauka docked to the International Space Station

SPACE SCIENCE

Dramatic entrance for Russian module

The International Space Station (ISS) had a tense hour last week when thrusters on a newly arrived Russian laboratory module misfired and pushed the station 45° out of its normal orientation. The 20-ton, truck-size Nauka module—named for the Russian word for “science”—was delayed 15 years by funding and technical issues. The troubles continued after its 21 July liftoff on a Proton-M rocket, as controllers struggled to communicate with Nauka after it reached orbit and to fire up its main engine. Three hours after docking with the ISS on 29 July, Nauka’s thrusters switched on unprompted—the result of a “short-term software failure,” according to the Russian space agency Roscosmos—and began to slowly turn the whole station. Russian controllers tried to counteract the turn by firing thrusters on an attached service module and later a cargo freighter. Nauka’s thrusters eventually ran out of fuel and the station was righted. NASA says the ISS crew was never in any danger and could not feel the thruster tug-of-war going on in the station’s Russian section.

CDC reverses mask guidance

COVID-19 | In a dramatic reversal that concedes the power of the highly transmissible Delta variant, the U.S. Centers for Disease Control and Prevention (CDC) on 27 July revised its 13 May guidance on wearing masks, saying fully vaccinated people should again wear masks in public, indoor spaces in areas of substantial or high coronavirus transmission. Three days later, the agency published data from an outbreak in Barnstable county in Massachusetts that it called “pivotal” to its decision. Of 469 people infected there during the first half of July, a time of densely packed indoor and outdoor events, 74% were fully vaccinated. The Delta variant was identified in 89% of 133 sequenced cases. Furthermore, samples from the noses and throats of fully vaccinated people bore as much virus as those from the unvaccinated. That finding “raised concerns that, unlike with other variants, vaccinated people infected with Delta can transmit the virus,” CDC Director Rochelle Walensky said in a 30 July statement. She added that the masking recommendation was updated to ensure vaccinated individuals would “not unknowingly transmit virus to others.” Four of five people who were hospitalized in the Massachusetts outbreak were fully vaccinated. There were no deaths.

Lacks’s family to sue over cells

BIOETHICS | Relatives of Henrietta Lacks, whose cervical cancer cells were harvested without her knowledge at Johns Hopkins Hospital in 1951, plan to sue pharmaceutical companies that have profited from studying those cells. The self-renewing “HeLa” cell line has become a mainstay of basic and applied research in diverse fields including cancer biology and infectious disease. Lacks’s family has hired trial lawyer Christopher Seeger and civil rights attorney Ben Crump, who has also represented families of Black men killed by the police, including George Floyd and Michael Brown, *The Baltimore Sun* reported last week. The lawyers have not disclosed any defendants, but said they plan to file the first lawsuits on the 70th anniversary of Lacks’s death, on 4 October.

PHOTO: NASA/SHANE KIMBROUGH



PHOTO: ANUP SHAH/MINDEN PICTURES

CONSERVATION

Orphaned elephants pay the price of poaching

The killing of adult female elephants reduces the survival chances of offspring even if they are already weaned—and the effect is large enough to slow population growth, researchers report this week. Maternal care is known to be important for long-lived mammals such as elephants, but its impact on population growth hadn't been directly measured for any wild species. By studying 19 years of data on 645 female elephants in a wild population in Samburu county in Kenya, scientists calculated annual survival probabilities for elephants of various ages. An orphaned juvenile—a weaned individual between the ages of 3 and 8 years—had an 86% chance of survival, compared with 96% for a juvenile with a living mother, the researchers report in *Current Biology*. The orphans were less likely to survive than the oldest adult females, which surprised the team, in part because poachers target adults for their large tusks. It's not known why orphaned juveniles are vulnerable, but they tend to face more aggression from their herd.

Coronavirus sequences re-emerge

COVID-19 | The mysterious disappearance of coronavirus sequences from a U.S. database appears to have a mundane explanation: the accidental deletion of a data-sharing statement. In late June, evolutionary biologist Jesse Bloom of the Fred Hutchinson Cancer Research Center suggested in a preprint that Chinese scientists had deliberately hidden viral sequences from early COVID-19 patients in Wuhan, first adding them to a National Institutes of Health (NIH) database but then requesting they be removed. Using data in a paper the group published in the journal *Small*, Bloom was able to recover parts of the missing sequences in Google's Cloud. Last month, a Chinese health minister said editors at the journal had eliminated text that noted where the sequences were deposited, leading the scientists to think the data did not need to be public and to request their removal. Last week, the journal added a correction note to the paper confirming that a "Data Availability" paragraph had been mistakenly deleted during copy editing. The viral sequences have now been added to a public Chinese database, the health minister noted. Bloom says that appears true, but adds the explanation given is "completely inconsistent" with the emailed data removal request sent to NIH last year, which said the sequences had already been deposited elsewhere.

China ties case faces retrial

RESEARCH COLLABORATION | The U.S. government wants to retry a former faculty member at the University of Tennessee, Knoxville, it had previously charged with concealing his ties to Chinese entities. On 16 June, a federal judge declared a mistrial in the case against Anming Hu, the first to go before a jury under the government's

BY THE NUMBERS

$33.5 billion

Anticipated 2021 revenue from Pfizer's COVID-19 vaccine, according to an updated projection the company released last week.

40%

Proportion of wild white-tailed deer tested in four U.S. states between January and March that had SARS-CoV-2 antibodies, suggesting they had been infected with the virus. (bioRxiv)

THREE QS

A test for anonymous hiring

In a bid to reduce bias during faculty hiring, Yale University's molecular biophysics and biochemistry department conducted an experiment last year: Scientists applying for a tenure-track position were asked to submit anonymized materials—omitting their names and the names of institutions they'd trained at and journals they'd published in. *Science* spoke with the department chair, Enrique De La Cruz, to find out how the search—which was ultimately successful—went. (A longer version of this interview is at https://scim.ag/CareersFacultySearch.)

Q: Was there any pushback to the idea?
A: No. There was skepticism. I think there was concern that we would de-emphasize the significance of scholarship for other things. But the fact that it was an experiment helped. We're scientists, so if you phrase it that way everyone's on board. "OK, it's an experiment? Sure, sounds good. That's what we do."

Q: How did it work?
A: We asked applicants to make their applications anonymous. It required them to think deeply about what they've done and articulate their contributions without relying on the shorthand—for example, I worked for this Nobel laureate at this prestigious institution, and I published in these fantastic journals. The hiring committee made a first cut based only on anonymized statements about the applicants' past and future research, and then another based on teaching and diversity, equity, inclusion statements, which were also anonymized. After all that was done, they looked at CVs and letters of recommendation, which were not anonymous. … It was a lot of work, but I do think it'll be easier next time.

Q: After the initial selection, you ended up with a larger percentage of women and members of underrepresented racial and ethnic groups compared with the applicant pool. Why do you think that is?
A: I'm not going to draw any firm conclusions based on an experiment I've done once. But I can imagine that it is possible that people from underrepresented groups were energized and motivated by being evaluated without identifying information. Maybe they took it a little more seriously because they saw it as an opportunity as opposed to an obstacle.

3-year-old China Initiative, which has led to the prosecution of several scientists of Chinese heritage. But on 30 July the Department of Justice filed "a notice of intent" to retry Hu, who lost his job after his arrest in February 2020 and has remained under house arrest. Civil rights groups that have accused the government of racial profiling condemned the move. On 29 July, nearly 100 Democratic members of Congress complained to U.S. Attorney General Merrick Garland that Hu has been one of "many people of Asian descent … falsely accused of espionage" under the initiative.

Gene drive passes cage test

BIOTECHNOLOGY | A genetic engineering technique called gene drive successfully collapsed captive populations of the malaria-spreading mosquito *Anopheles gambiae* in the largest test yet of the strategy. Researchers inserted a mutation that renders females unable to reproduce, along with another genetic element that spreads that mutation quickly through a population. A 2018 study showed the approach could suppress populations of mosquitoes housed in small cages (about .016 cubic meters). The new study tested the strategy in larger cages (nearly 5 cubic meters), where multiple generations of mosquitoes could mate, forage, and lay eggs more like they would in the wild. Introducing the engineered mosquitoes crashed the caged populations within 1 year, the researchers reported last week in *Nature Communications*, and the insects did not develop resistance to the sterilizing effects of the mutation. Experimental releases of the insects in the wild face regulatory hurdles and are likely still years away.

Red List adds recovery metric

CONSERVATION | The International Union for Conservation of Nature (IUCN) last week announced that its long-running "Red List of Threatened Species" will be joined by a new measure, "Green Status of Species," that looks beyond the Red List's ratings of extinction risk to measure recovery and the impact of conservation efforts. Species currently on the Red List will now also get a green score that puts them into one of several categories ranging from "extinct in the wild" to "fully recovered." Researchers will also estimate the impact of past, current, and future conservation efforts on a species. They hope this metric, which is more success-oriented and sensitive to change than Red List status, will provide a road map for future protective measures. The first batch of roughly 50 assessments is slated to go live on the IUCN website this fall.

SCIENCEMAG.ORG/NEWS
Read more news from *Science* online.



IN FOCUS The Shanghai Astronomy Museum opened last month with 39,000 square meters of floor space, making it the world's largest museum devoted to astronomy and one of the biggest of any kind. The building's curving design is inspired by the trajectories calculated for a trio of hypothetical celestial bodies interacting gravitationally.

PHOTO: ENNEAD ARCHITECTS



Members of the Taliban in eastern Afghanistan in March 2020. Some are hopeful that the extremist group will allow polio campaigns to continue if it consolidates power.

PHOTO: JIM HUYLEBROEK/*THE NEW YORK TIMES*

GLOBAL HEALTH

Taliban's rise puts polio eradication in danger

U.S. departure from Afghanistan raises questions about future of vaccination drives

By **Leslie Roberts**

The U.S. troop withdrawal from Afghanistan—along with the surge of the Taliban there—is imperiling the 3-decade global quest to eradicate polio. The Taliban has blocked house-to-house polio vaccination in areas under its reign for the past 3 years, putting up to 3 million children out of reach of the campaign and leaving Afghanistan one of only two countries, along with Pakistan, where the wild polio virus survives. Since the United States began to pull out troops, the Taliban has made rapid military gains and now controls much of the country. Some fear it may wrest complete control from the Afghan government—which supports the eradication campaign—after U.S. forces are gone in September.

That's a frightening prospect to many polio watchers. But some inside and outside the Global Polio Eradication Initiative (GPEI) say a clear resolution to the conflict, regardless of who prevails, may actually aid eradication efforts. They hope the campaign will be able to work with the Taliban to keep vaccinations going. Until the conflict subsides, though, chaos is likely to interfere with vaccination drives.

The U.S. withdrawal comes at a time when the program is making strides against the wild virus after several years of setbacks. Cases in Afghanistan almost tripled to 56 between 2018 and 2020, and the country also saw explosive outbreaks of polio virus derived from the live vaccine, which paralyzed more than 300 children in 2020. Roughly 85% of all cases occur in areas inaccessible because of Taliban control, says Aidan O'Leary, who in January took over as head of GPEI, headquartered at the World Health Organization in Geneva. The COVID-19 pandemic initially made things worse (*Science*, 24 July 2020, p. 360).

But so far this year, there has only been one case caused by the wild virus. That may be partly due to reduced population movement during the pandemic and more hand washing, says GPEI's Hamid Jafari, who directs eradication efforts in the region. Even so, "The trend is very real."

Afghanistan's fate is closely tied to that of Pakistan, with which it shares a porous, 2670-kilometer border. That country has also reported just one wild virus case in this year, after a similar upsurge from 12 in 2018 to 84 last year. (The spike there was largely because of vaccine refusals driven by rumors and a virulent disinformation campaign, along with a sometimes-inefficient eradication program.)

Some optimism about Afghanistan stems from the belief that the Taliban is not opposed to polio vaccination per se—in fact, it has collaborated with the program in the past. "It was purely for security reasons in specific areas where it imposed the ban," Jafari says. The insurgents accused polio vaccinators of passing along information to help the United States target airstrikes that killed Taliban leadership. "Whether right or wrong, if that is the perception, that is reality," O'Leary says. "You have to admit, the airstrikes have been brutal" for the Taliban and civilians, adds Stephen Morrison, senior vice president and director of the global health policy center at the Center for Strategic and International Studies in Washington, D.C.

The hope is that if the Taliban continues to consolidate power, its suspicions may ease and GPEI may be able to operate unimpeded. Following years of negotiations, the Taliban recently gave GPEI the green light to conduct vaccination in mosques in the provinces where it has imposed the house-to-house ban. The program is awaiting "a more formal statement" to proceed this month, O'Leary says—if the security situation allows. The Taliban will select people it trusts to be trained as vaccinators by GPEI, Jafari says. Mosque campaigns are usually not as effective as going house to house, O'Leary says, reaching perhaps 40% to 50% of the target population, "but we can hope to build on that."

Given this opening, Jafari thinks it unlikely the Taliban will issue new bans on polio vaccination. "We are on a very different trajectory in negotiations with them," he says. Others decline to speculate. "The Taliban's endgame remains to be determined. There are more dire and more benevolent views," O'Leary says. And Morrison cautions that "U.S. aerial campaigns won't necessarily end with the withdrawal."

Although fighting between the Taliban and the government is a major obstacle, "the program is not at a standstill," Jafari says. But it has had to cease activities where there is active fighting, says Godwin Mindra, UNICEF's polio team lead in Kabul. And districts that are accessible today might not be tomorrow, Mindra adds. A nationwide polio vaccination campaign is scheduled for September—if it can be conducted safely—with smaller campaigns scheduled for November and December. "We will look very carefully at how the situation is evolving," O'Leary says.

The worst-case scenario is a descent into full civil war, with escalating violence, large numbers of refugees, and a broader public health crisis. Even then, GPEI leaders point out that the program has lots of experience operating during conflict, in Syria and many other countries.

GPEI's new strategic plan for 2022–26 aims to interrupt all polio transmission in Afghanistan and Pakistan by the end of 2023. "If we can continue to vaccinate through this year, we can make good progress" toward that goal, Jafari says. But success also depends on stopping the virus in Pakistan, as the virus has often jumped back into Afghanistan just as the country was making gains. Although the polio program is "very resilient, very innovative," Morrison thinks the 2023 time frame may be "a bit delusional." For now, O'Leary says, "We are hostage to events on the security side." ■

Leslie Roberts is a science writer in Washington, D.C.



A suspected COVID-19 patient receives care in Turku, Finland, the first country to join Solidarity's new phase.

PHOTO: RONI LEHTI/LEHTIKUVA/AFP/GETTY IMAGES

COVID-19

WHO relaunches global drug trial with three new candidates

Solidarity kicks back into action after 6-month pause

By **Kai Kupferschmidt**

After months in the doldrums, one of the world's largest trials of COVID-19 treatments is finally restarting. Solidarity, a global study led by the World Health Organization (WHO), will test three new drugs in hospitalized COVID-19 patients: the cancer drug imatinib, an antibody named infliximab that is used to treat autoimmune diseases, and artesunate, an antimalarial.

The medicines have been shipped to Finland, the first country to have all approvals in place, says John-Arne Røttingen of the Norwegian Institute of Public Health, who chairs the study's executive group. "I expect that the first patients will probably be recruited there any day," he says. Other countries could soon join SolidarityPlus, as the new phase has been dubbed; more than 40 are in the process of getting ethical and regulatory approvals.

When the original Solidarity trial started in March 2020 it was a first: an effort to test drugs in dozens of countries simultaneously in the middle of a pandemic. By late in the year it had delivered verdicts on four treatments—none showed a benefit—but then became mired in negotiations with pharmaceutical companies and regulatory delays. "It's great that Solidarity is proceeding with randomized clinical trials again, as they have already made an important contribution to our therapeutic approach during the pandemic," says Eric Topol, director of the Scripps Research Translational Institute. "We can't be at all complacent about needing better therapies for patients with severe COVID."

Although COVID-19 vaccine development has been a huge success story, only two drugs have proved to reduce COVID-19 mortality in hospitalized patients. In June 2020, the United Kingdom's Recovery trial found that dexamethasone, a cheap steroid, reduced deaths in that group by up to one-third. In February, Recovery investigators announced that tocilizumab, a monoclonal antibody that blocks the receptor for interleukin-6, reduced mortality a bit further. Both drugs work by dampening the overshooting immune response in severely sick patients.

The new drugs also target the immune system rather than the virus itself. In the severely ill patients included in Solidarity, it's probably too late for an antiviral drug to work, Røttingen explains. (Monoclonal antibodies to SARS-CoV-2, for example, are most effective when given before serious disease develops.) But sicker patients could benefit from additional drugs that target the immune system, says Anthony Gordon, a critical care specialist at Imperial College London. Although dexamethasone broadly dampens the immune response and tocilizumab powerfully shuts off one particular pathway, "There are still other pathways that we can block and maybe make a difference," Gordon says.

Imatinib, an oral drug used to treat some leukemias and other types of cancer, can also protect the epithelium lining the alveoli, where oxygen crosses from the lungs into the blood. A placebo-controlled trial in 400 hospitalized COVID-19 patients in the Netherlands, published in June, showed patients on the drug spent less time on ventilators and were less likely to die. Although not statistically significant, the data were encouraging enough to spur larger studies, says Gordon, who is part of another international trial called REMAP-CAP that is also planning to test the drug.

Infliximab is an antibody given as a single infusion that blocks tumor necrosis factor alpha, a pivotal signaling molecule in the immune system, and is used to treat autoimmune diseases such as rheumatoid arthritis and inflammatory bowel disease. Some observational data from large patient populations suggest the drug can also protect against COVID-19, Røttingen says.

Artesunate, an injected derivative of artemisinin and a powerful killer of malaria parasites, has also shown some antiviral activity in laboratory studies of SARS-CoV-2. But Solidarity is testing it because of another effect: The drug appears to reduce inflammation and counteract signals that attract immune cells into tissues. That could stop the immune reactions that damage the lungs in severe COVID-19.

Solidarity's revival was a long time coming. In October 2020, it published results from more than 11,000 patients in 400 hospitals that deflated hopes—and punctured hype—by showing no benefit for four treatments: the HIV combination therapy lopinavir/ritonavir, the malaria drug hydroxychloroquine, interferon-beta, and Gilead Sciences's antiviral drug remdesivir. The remdesivir arm was continued for a while to gather more data—full results are expected in the coming weeks—but by late January all arms had been stopped.

An independent expert committee picked the three new drugs soon after. The delay is due partly to negotiations with the manufacturers to ensure that the drugs would be available at affordable prices worldwide if they turned out to work, Røttingen says, and partly due to the time needed for regulatory and ethical approvals in participating countries.

"We have definitely seen that there was a strong willingness to sort of work outside the normal system and really speed up processes in the beginning of the epidemic, and that seems to be less the case now," Røttingen says. That's understandable, he adds, "But it also demonstrates that these processes are not fit for emergencies. We need fast-track systems for the future, in all countries." ■

CREDITS: (GRAPHIC) K. FRANKLIN/*SCIENCE*; (DATA) J. M. KLATT *ET AL.*, *NAT. GEOSCI.* (2021). DOI: 10.1038/S41561-021-00784-3

GEOSCIENCE

Longer days on early Earth set stage for complex life

Extra light spurred oxygen release by mats of photosynthetic bacteria, models and sinkhole data suggest

By **Elizabeth Pennisi**

Earth wasn't born with the oxygen-rich atmosphere that fuels life today. Living things supplied the gas, but scientists have struggled to find a satisfying explanation of what triggered the buildup and why it didn't start until well after the first photosynthetic life. Now, based on modeling of Earth's early rotation and evidence from microbial mats coating the bottom of a shallow, sunlit sinkhole in Lake Huron, researchers have identified a surprising potential trigger: the increasing length of a day as ancient Earth's spin slowed. Longer days could have coaxed more photosynthesis from similar mats, allowing oxygen to build up in ancient seas and diffuse up into the atmosphere.

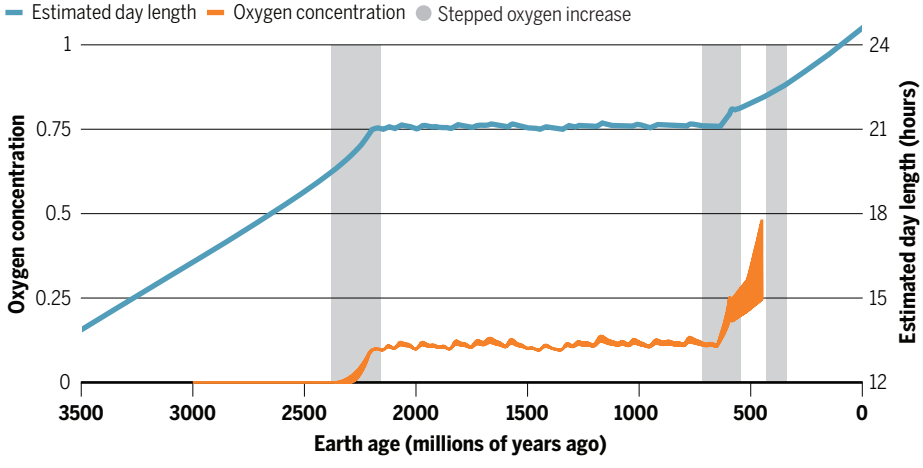
That proposal, described this week in *Nature Geoscience*, has intrigued some other scientists. "The rise of oxygen [on Earth] is easily the most substantial environmental change in the history of our planet," says Woodward Fischer, a geobiologist at the California Institute of Technology who was not involved with the work. This study offers "a totally new flavor of an idea. It's making a connection that people haven't made before."

By 3.5 billion years ago the planet's vast shallow seas teemed with cyanobacteria, which can form mats on sediments and rock surfaces and today sometimes cause "algal" blooms deadly to fish and other aquatic animals. These microbes had evolved the molecular machinery for photosynthesis, enabling them to convert carbon dioxide and water into sugars and oxygen. They presumably provided Earth's initial supply of oxygen, creating an environment that favored the evolution of aerobic life in all its forms. But that picture left a puzzle: Why did another billion years pass before the first good geological evidence for a buildup of oxygen appears?

That puzzle led Judith Klatt, a biogeochemist now at the Max Planck Institute for Marine Microbiology, to a seemingly unrelated phenomenon: the drag that the Moon's gravity exerts on the spinning Earth by tugging at its surface and raising tides. That effect has been slowing Earth's rotation and lengthening days since the beginning. Many agree that 4.5 billion years ago, 1 day was only about 6 hours long. By about

Light and air on early Earth

Oxygen in Earth's atmosphere began to build up only well after the first oxygen-producing microbes evolved 3.5 billion years ago, then continued to rise in steps, separated by a long plateau. Estimated day length rose in the same stepped pattern, driven by gravitational interactions between Earth and the Moon. A new study suggests the changing day length drove the oxygen bumps as longer periods of daylight allowed microbes to produce ever-larger surpluses of the gas.





Divers collected samples of these colorful microbial mats from Lake Huron to understand how similar mats behaved on early Earth.

2.4 billion years ago, computer models suggest, the pull of the Moon had slowed that spin to about a 21-hour day. The models predict Earth's rotational speed then stayed constant for about 1 billion years, as other forces countered the Moon's pull on Earth. Those forces fell out of balance about 700 million years ago, models suggest, and the planet's spin resumed slowing until it reached its current speed, creating a 24-hour day.

In 2016, after a chance suggestion, Klatt realized those slowdowns in Earth's rate of spin mirrored big leaps in atmospheric oxygen. For example, oxygen first jumped during what's called the Great Oxygenation Event, some 2.4 billion years ago, and then again during the Neoproterozoic era, more than 1 billion years later. The Paleozoic, about 400 million years ago, brought a final major increase in atmospheric oxygen.

As a postdoc at the University of Michigan, Ann Arbor, Klatt had studied the Middle Island Sinkhole in Lake Huron, where oxygen-depleted water and sulfur gas bubble up from the lake floor, creating anoxic conditions that roughly approximate conditions of early Earth. The shallow sinkhole also hosts microbial mats, rich in cyanobacteria, that get enough sunlight for photosynthesis.

Scuba divers collected samples of the microbial mats and in the lab, Klatt tracked the amount of oxygen they released under various day lengths simulated with halogen lamps. The longer the exposure to light, the more of the gas the mats released.

Excited, Klatt and Arjun Chennu, a modeler from the Leibniz Centre for Tropical Marine Research, set up a numerical model to calculate how much oxygen ancient cyanobacteria could have produced on a global scale. When the microbial mat results and other data were plugged into this computer program, it revealed a key interaction between light exposure and the microbial mats.

Typically, microbial mats "breathe" in almost as much oxygen at night as they produce during the day. But as Earth's spin slowed, the additional continuous hours of daylight allowed the simulated mats to build up a surplus, releasing oxygen into the water. As a result, atmospheric oxygen tracked estimated day length over the eons: Both rose in a stepped fashion with a long plateau (see graphic, p. 607).

This "elegant" idea helps explain why oxygen didn't build up in the atmosphere as soon as cyanobacteria appeared on the scene 3.5 billion years ago, says Timothy Lyons, a biogeochemist at the University of California, Riverside. Because day length was still so short back then, oxygen in the mats never had a chance to build up enough to diffuse out. "Long daytimes simply allow more oxygen to escape to the overlying waters and eventually the atmosphere," Lyons says.

Still, Lyons and others say, many factors likely contributed to the rise in oxygen. For example, Fischer suspects free-floating cyanobacteria, not just those in rock-affixed mats, were big players. Benjamin Mills, an Earth system modeler at the University of Leeds, thinks the release of oxygen-binding minerals by ancient volcanoes likely countered the early buildup of the gas at times and should be factored into oxygen calculations.

Nonetheless, changing day length "is something that should be considered in more detail," Mills says. "I'll try to add it to our Earth system models." ■

IMMIGRATION POLICY

Chinese students fight back against visa rejections

A Trump-era proclamation may block thousands from study or research in the United States

By **Dennis Normile**

When Chen Siyu met a consular official at the U.S. embassy in Beijing in March to review her qualifications for a student visa, "Everything was going well," she says—or so it seemed. Chen, who has a master's in public health from the University of Hong Kong, had won a fully funded slot in an epidemiology Ph.D. program at the University of Florida. When the consular officer asked about her current employment, Chen explained that she had worked as an epidemiology research assistant at a major hospital for 5 years. She mentioned that the hospital is affiliated with a military medical university.

The consular officer thanked Chen for the information and moments later handed her a rejection form letter with "Other: 212(f)" ticked off from among a selection of reasons. The interview was over, as were her dreams of earning a Ph.D. in the United States.

Chen is one of a growing group of Chinese students barred from the United States based on 212(f), a clause in the decades-old Immigration and Nationality Act (INA) that allows the U.S. president to identify aliens whose entry would be "detrimental to the interests of the United States." In May 2020, then-President Donald Trump signed a proclamation that invoked the clause to bar Chinese graduate students and postgraduate researchers with ties to an entity in China "that implements or supports China's 'military-civil fusion strategy.'" The proclamation exempts those working in fields that don't contribute to that strategy—but apparently epidemiology is not among them.

Now, Chen is one of 2500 activists—Chinese students with visa problems and

PHOTO: PHIL HARTMEYER/NOAA THUNDER BAY NATIONAL MARINE SANCTUARY

their supporters—who are fighting back against what they see as an arbitrary and discriminatory policy. Armed with a website and a Twitter account, the students have written to more than 50 top U.S. research universities to focus attention on their plight. They are getting a sympathetic hearing in the U.S. academic world: A 10 June letter from the American Council on Education to the Department of State warned of "delays in students' academic careers and critical projects."

The group is also discussing legal action with a U.S. immigration lawyer and recently launched a fundraising campaign to try to cover the costs. "We think this is a policy of discrimination based on nationality," says Hu Desheng, a doctoral candidate in computer science at Northeastern University who got stuck in China because of pandemic-related travel restrictions in early 2020, and whose visa application is now backlogged.

Trump's proclamation initially had little impact because the pandemic disrupted academic travel globally. But after more than a year, the U.S. embassy and consulates in China resumed processing routine visa applications on 4 May. Between then and mid-June, more than 500 visa applications have been rejected, according to the students' tally. More than 1000 Chinese scholars already in the United States reportedly had their visas revoked by September 2020. Many others hesitate to leave the United States, fearing they won't get back in.

How many students will be affected annually is unclear, in part because the U.S. government has not said which Chinese entities are deemed to be supporting the military-civil fusion strategy and which fields of study are considered sensitive or exempt.

A study of the measure's potential impact published in February by Georgetown University's Center for Security and Emerging Technology (CSET) assumed the designated entities include 11 universities subject to stringent export control restrictions by the U.S. Department of Commerce, including the so-called Seven Sons of National Defence—schools with historical ties to China's defense establishment. The study also assumed the sensitive fields mentioned in the proclamation will cover all areas of science, technology, engineering, and math (STEM). If so, it could block 3000 to 5000 of the roughly 19,000 Chinese students who start graduate programs each year, CSET estimated. The report did not cover postdoctoral and visiting researchers, graduates of other universities, or those in non-STEM fields. (The proclamation exempts undergraduate students from scrutiny.)

A spokesperson for the State Department declined to name which institutions are blacklisted, but said the sensitive technologies include quantum computing, big data, semiconductors, biotechnology, 5G, advanced nuclear technology, aerospace technology, and artificial intelligence. "By design, the policy is narrowly targeted," the spokesperson says.

But the Chinese students say rejections are broad. Even those intending to study finance, obstetrics and gynecology, water conservation, medicine, agronomy, and other seemingly nonmilitary topics have had visas rejected under clause 212(f), they say. Li Xiang, for example, earned a master's in linguistics from the Harbin Institute of Technology, one of the schools with historical defense ties, then studied at an art school to prepare for a master's program in game development at the Academy of Art University in San Francisco. "To be an artist in the game and film industry is my dream," she says. Her application was rejected and she was told she is not even eligible for a visa to visit her husband, who is working in the United States.



The visa of another student, Xue Shilue, was revoked in the summer of 2020 after she had completed the first year of a master's program in "user experience design" at the University of Texas, Austin. She happened to be in China at the time and can't go back to Austin to complete her degree or even collect her personal belongings.

The proclamation also appears to target students supported by the China Scholarship Council (CSC), which falls under China's Ministry of Education but has been under scrutiny for supposed links to the defense establishment, according to a separate CSET study. Blacklisting CSC could have dramatic implications. CSET estimates that during the 2017–18 academic year, the council supported 26,000 Chinese scholars in all disciplines in the United States. Huang Yunan, who last year started a Ph.D. program in food science at Cornell University remotely because of the pandemic, was denied a visa after telling a consular officer about her CSC support during a May interview. More than 100 of some 500 CSC-supported members of a chat group she belongs to have recently had visa applications rejected, she says.

The students object to the absence of any individual assessment. "There is a presumption of guilt on the part of every Chinese student who has studied at a targeted university," Hu says. As to the Seven Sons, "We go to those schools because they are top-ranked universities," Hu says, not because of their military ties.

Wendy Wolford, vice provost for International Affairs at Cornell University, asked U.S. Secretary of State Antony Blinken in a 26 May letter to rectify the "capricious, unclear, and excessive" interpretations of the proclamation that are "creating tremendous uncertainty and confusion for international students and their U.S. universities." (Wolford did not respond to an email asking whether she had heard back from Blinken.)

A lawsuit, however, is a long shot, says Charles Kuck, a U.S. immigration lawyer who has advised the students. "The Supreme Court has given a literal carte blanche to the president to use INA 212(f), along with a 'reasonable' explanation, for whatever entry ban the president wants to put into place," Kuck says.

The problems are driving some students to pursue advanced degrees elsewhere; Chen, for one, will now get her Ph.D. at the University of Hong Kong. Moves like hers should be a bigger worry than the possibility that graduate students are stealing U.S. technology, says Denis Simon, an expert in innovation at Duke University who studies China's research efforts. "The notion of there being a conspiratorial effort [to acquire advanced technology] is just far beyond the reality." In contrast, he says, slowing the flow of Chinese students will harm the United States, where they help sustain many research programs. "It's a pipeline that has been built over 40 years, and by deconstructing it, we will do some very serious damage to our ability to have the kind of talent needed to drive our innovation system forward." ■

PHOTO: JOSEPHINE PETTIGREW/NORTHEASTERN UNIVERSITY

RESEARCH ETHICS

Conflict flares over incidental genetic findings

Bioethicists reject proposal to routinely tell research participants about disease risks

By **Meredith Wadman**

Should people who volunteer for genomic studies be told about unrelated disease mutations that turn up in their sequence data? The decadeslong debate about such "incidental findings," which can include genes that boost risk for cancer or heart disease, flared up again last week after bioethicists at the National Institutes of Health (NIH) published a study showing many participants who at first refuse those findings can change their minds. Controversially, it went on to suggest all research participants routinely be told about their genetic risks for conditions that can be prevented or treated—a change from current practice.

The controversy pits researchers, many of them physicians who see incidental findings as an opportunity to boost the health of the millions who have had their genomes analyzed, against others, mainly bioethicists, who stress the need to respect study participants' right to reject such information. Deepening the divide, the study showed Black participants were more likely to refuse incidental results. "That strengthens the argument for saying we've really got to get true consent, opt-in consent from everyone," says Susan Wolf, a lawyer who teaches health law and bioethics at the University of Minnesota Law School.

In the study, investigators recontacted research participants in a large NIH study 1 to 3 years after they enrolled. Initially, 1.9% of participants declined to receive incidental findings. The team reports in *Genetics in Medicine* that of the 83 initial refusers, 41 changed their minds after being presented with new information, including an assurance that researchers would only return results on genes that raised the risk for serious conditions that were preventable or treatable. "I had a hypothesis that we would have a surprising number of people who would be willing to change their mind, but I had no idea how strong that would be," says senior author Ben Berkman, a lawyer in NIH's Department of Bioethics.

In current research studies that offer to return incidental findings, participants typically need to opt in, affirming their desire to receive such results. Many bioethicists say the ability to actively choose whether to be told protects patients' rights to decide what information is generated about them and to guard their privacy.

Berkman and his co-authors conclude that, given the number of minds that were changed during their study, an opt-out system would be better: Researchers should notify participants that they will receive incidental results, and withhold the findings only if they actively refuse, the authors recommend. Berkman thinks the change could be lifesaving, spurring people to get medical care or early, preventive screenings for diseases like colon cancer. NIH itself appears open to that approach; the press release for the study announced, "New study brings into question current policies on receiving secondary genomic findings."



NIH's All of Us study requires active consent to return disease risk information.

PHOTO: DAKE KANG/AP IMAGES

In the study, however, Black participants were significantly more likely than others to refuse incidental findings. They initially rejected receipt of secondary findings at twice the rate of white participants and were less likely to change their minds after being reapproached.

"This research makes a recommendation without any regard to the Black participants' answers," says Keisha Ray, a bioethicist at the McGovern Center for Humanities & Ethics at the University of Texas Health Science Center, Houston. "Why isn't this their recommendation: 'Based on the answers and hesitation expressed by the Black participants and their fear that their genetic information will be misused by the NIH, we recommend that secondary genetic information should not be given to participants unless they directly consent to knowing?'"

Berkman responds that the small number of Black participants—57—makes him "wary of making a concrete policy suggestion" without more data.

Bioethicist Faith Fletcher of the Baylor College of Medicine says the new study points to the urgent need for more research. She calls for "work to find out why participants refused and how we use information from the study to figure out ethically informed ways to handle secondary findings."

To Robert Green, a medical geneticist at Harvard University who was not involved with the new study, it is "stunning" that only a small number of research studies return incidental results in the first place. "Only a tiny, tiny fraction of [genetic research subjects] have ever been offered the opportunity to have any of those results returned to them for their potential medical benefit." But he thinks the zeitgeist is changing. For instance, NIH's huge All of Us study is returning incidental findings to those who opt in.

Green and others stirred controversy in 2013 when they published for the American College of Medical Genetics and Genomics a list of more than 50 genes that ACMG recommended should be tested in any patient undergoing genetic screening in a clinical setting, with the information automatically returned to patients. After a storm of pushback, ACMG backed off, saying patients should be allowed to opt out of receiving such findings.

One of the co-authors of the original ACMG list now calls the automatic return of incidental findings from research projects "a bridge too far." "Even though the number of people who did not want the results back is very small," says Robert Nussbaum, chief medical officer at Invitae, a medical diagnostics company, "do we want to 'ride roughshod' over [them]? The answer to that is probably no." ■



New Yorkers, like those at this Brooklyn mobile pharmacy, can get $100 for getting vaccinated.

PHOTO: SPENCER PLATT/GETTY IMAGES

COVID-19

Studies probe how payouts affect U.S. vaccination rates

Money can sometimes tip the balance for the hesitant

By **Anil Oza**

What does it take to convince someone to get vaccinated? U.S. officials all the way up to President Joe Biden are hoping cash will do the trick.

In the past few months, as COVID-19 vaccination rates slowed across the United States and the Delta coronavirus variant spread, companies and local governments have offered creative incentives for people to get their shots, including free doughnuts, baseball tickets, and cars. Nearly 20 states now entice those hesitant to get vaccine with a lottery. And the White House last week called on state and local governments to directly pay $100 to anyone willing to get a first dose—an incentive New York City began to offer on 30 July.

Behavioral researchers and other scientists are closely watching, and in a few cases helping set up, these financial incentives to see how well they work. Early data on vaccine lotteries, for example, suggest they can convince a modest number of people teetering on the edge of committing to a shot. "Incentives of various different types, including lotteries, may be one effective strategy," says Dena Gromet, a psychologist at the University of Pennsylvania who helped set up Philadelphia's vaccination lottery, which has a grand prize of $50,000. "But we are likely going to have to rely on a multiprong approach to be able to really make a dent in vaccination going forward."

Directly giving Americans money to get a vaccine hasn't been tested before, but studies in Africa and Latin America have shown cash payments prod individuals to make use of health services, including pediatrician visits, nutrition consultations, and educational sessions, according to Amanda Glassman, a public health expert at the Center for Global Development. One study, conducted in Nigeria, found that conditional cash transfers increased childhood measles vaccination rates by 27%.

Although there are few studies of the effectiveness of lotteries as a medical incentive, some economists suggest they appeal to unvaccinated people who are more tolerant of risks. A lottery could also capitalize on a phenomenon called "probability neglect," where individuals will irrationally interpret probabilities in their favor. "Certain people just underestimate, say, the risks of a disease like COVID-19 for themselves, or spreading it to others, and also simultaneously would overweight, or be overoptimistic about their odds of winning the lottery," says economist Jeremy West of the University of California (UC), Santa Cruz, a co-author of a recent working paper about Ohio's vaccine lottery.

A lottery may also be more cost-effective than direct cash payments. West and his colleagues calculated that Ohio's $5.6 million lottery has enticed 80,000 residents to get vaccinated—a figure they arrived at by comparing the state's actual vaccination data with a simulation incorporating rates for demographically similar states that did not implement a lottery. That amounted to $68 for each new person vaccinated—a bargain compared with a $100 payment.

West also estimates that the lottery payout saved Ohio $66 million in intensive care unit costs. "Meaning, it's a no-brainer to do. But it's also not going to be the cure-all for the pandemic, or for vaccination hesitancy," West says. Three other recent analyses of the Ohio lottery have come to similar conclusions, with estimates of new vaccinations produced ranging from no significant change to 100,000.

Cash payments and lotteries may change cost-benefit calculations for individuals. Although COVID-19 vaccines are free to everyone in the United States, many people face indirect costs, from transportation expenses to a vaccine provider to taking time off from work for the shot or subsequent side effects. For the poor or those who perceive they are at lower risk for severe disease, that may be enough to put off vaccinations. "The lotteries aren't really about changing beliefs [about vaccine safety and need]. They are intended to increase the perceived benefit of receiving the vaccine and motivate indecisive people to go get the shot," says Oberlin College economist Maggie Brehm, who is a co-author of another study of Ohio's vaccine lottery.

Political scientist Lynn Vavreck of UC Los Angeles notes that cash incentives can only go so far. "We are not trying to get everyone—we are trying to get the people on the margin, the next set of people who can imagine themselves [getting vaccinated] under some circumstance," she says.

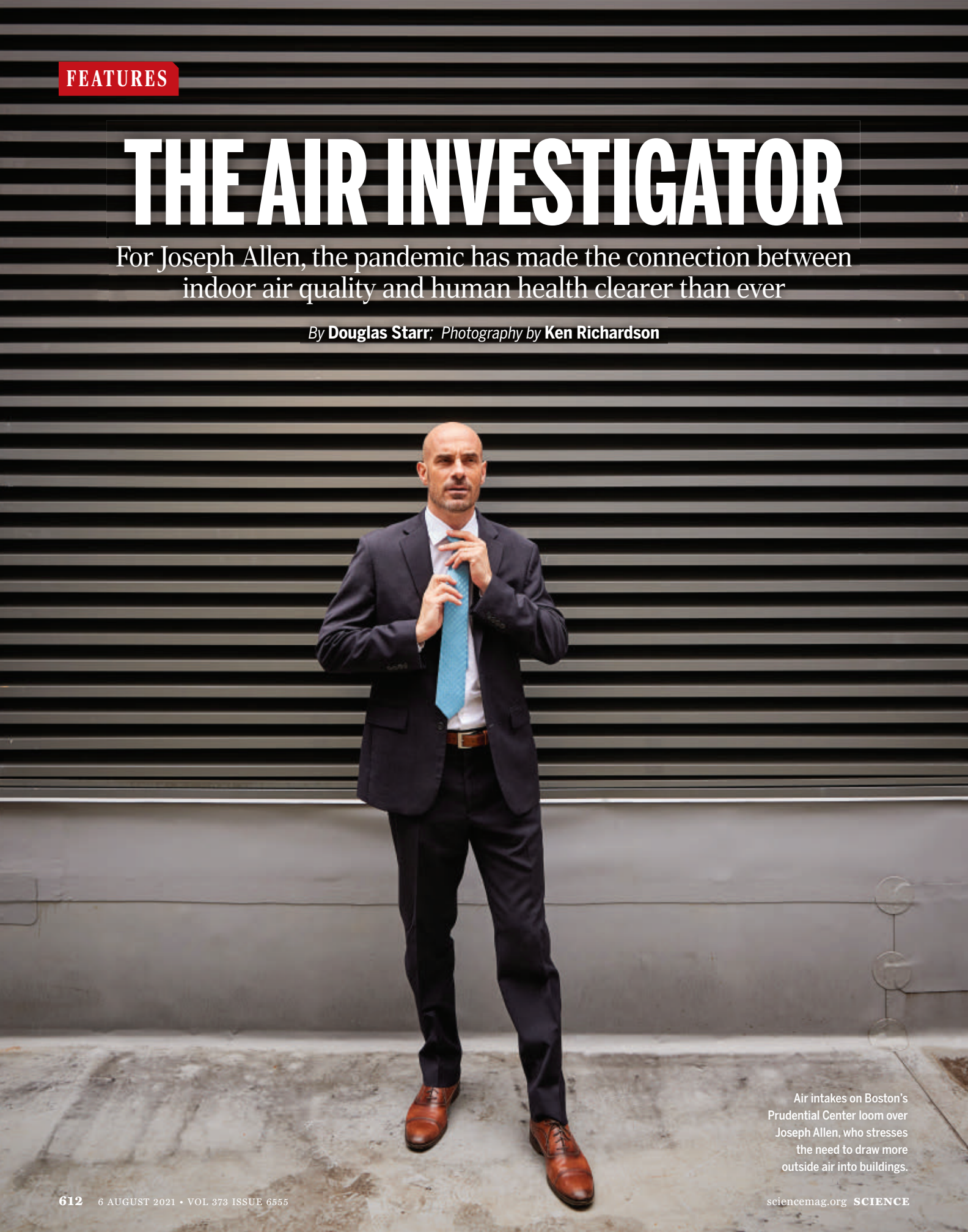
Glassman and others raise the concern that financial incentives could set a bad precedent, turning vaccination or other public health precautions into something done for selfish reasons rather than altruistic ones. "People should be motivated by intrinsic motivation to do this thing for pro-social reasons," she says.

But Glassman also doesn't see the middle of a pandemic as the time to debate the moral fiber of the country. "How could we not test every tool in the armament to make sure that people get vaccinated?" she asks. ■

THE AIR INVESTIGATOR

For Joseph Allen, the pandemic has made the connection between indoor air quality and human health clearer than ever

By **Douglas Starr**; *Photography by* **Ken Richardson**



Air intakes on Boston's Prudential Center loom over Joseph Allen, who stresses the need to draw more outside air into buildings.

Joseph Allen runs a major public health research project at Harvard University, probing how indoor air quality affects human health and cognition. He consults with companies on ventilation and air filtration, and during the pandemic he became a prominent voice on public health, writing dozens of op-eds criticizing early guidance from health authorities and debunking misconceptions about how the virus spreads. But none of it would have happened if he hadn't washed out as an FBI recruit.

The son of a New York City homicide detective who opened his own investigative agency, Allen spent his teens and 20s helping with the family business. He did surveillance, undercover work, computer forensics, and skiptrace—tracking down people who left town to avoid alimony. Eventually he took over the agency, leading investigations and supervising eight agents.

"I enjoyed the work and thought it was challenging," Allen recalls. But part of him always wanted to be a scientist. He majored in environmental science at Boston College, and in his late 20s, still torn, he began to apply to graduate school even as he started the process to become an FBI agent. After 2 years of interviews and testing, the last step was a routine polygraph test. He failed the first round—the trick questions he was asked were so obvious that he could not take them seriously. So FBI flew in one of its toughest examiners from Iraq—a hulking, jackbooted guy who got right in Allen's face, screaming that he knew he was lying. But Allen kept cool, and after a while, the interrogator stormed out and slammed the door.

"I thought he would come back in the room and say, 'Congratulations,' cause I'm thinking I'm crushing it," Allen recalls. "But they failed me because they said I employed countermeasures." FBI apparently didn't want an agent who couldn't be unnerved by a polygraph test. And that solved Allen's career dilemma. "I guarantee I'm the only public health student ever to fail an FBI lie detector polygraph in the morning and start graduate school a few hours later," Allen says. But his investigative instincts never left him.

A tall, athletic-looking man with a bald head and stylish stubble, Allen directs the Healthy Buildings Program at Harvard's T.H. Chan School of Public Health, where he studies the effects of toxic gases emitted from furniture, carpets, and paints; stale air; and high levels of carbon dioxide. Years of studies by Allen and others have shown poorly circulated air in buildings impairs our ability to think clearly and creatively. Considering that we spend more than 90% of our lifetimes indoors, those findings have implications for personal well-being—and for businesses concerned about their bottom line.

"Joe has always had a unique understanding of this range of domains—from how buildings work, to environmental exposure assessment, to making connections with health outcomes," says Brent Stephens, chair of the Department of Civil, Architectural, and Environmental Engineering at the Illinois Institute of Technology. "There's not a tremendous number of people in this world that have worked on that whole spectrum."

When the COVID-19 pandemic arrived, the previously esoteric field of indoor air quality suddenly became the focus of widespread concern. Like many of his colleagues, Allen jumped into the fray, advising school systems, police departments, entertainment companies, the Boston Symphony, and a host of other entities on how to make their indoor air healthier, during the pandemic and afterward.

"COVID really changed the conversation," says Matt Murray, vice president of leasing at Boston Properties, the largest publicly traded developer in the United States and one of Allen's consulting clients. Before the pandemic, the company would have to explain to bored executives why they should pay attention to indoor air. "Now, the CEOs are all saying, 'What filters do you use? How you process the air you bring into the workspace?'" Murray says. "And we're ready for those conversations because we've been working with Joe."

> **"The idea of a healthy building has been made too complicated."**
>
> **Joseph Allen and John Macomber**, Harvard University

AFTER HE FAILED his FBI exam, Allen became a different kind of sleuth. For his doctoral thesis at the Boston University School of Public Health, he investigated toxic flame-retardant chemicals released into the air by furniture, and found they were nearly ubiquitous. (The chemicals were later banned.) After graduation he got a job with a consulting firm, where he investigated problems such as toxic emissions from drywall and outbreaks of Legionnaires' disease, which is caused by bacteria that grow in plumbing and become aerosolized by ventilation systems, showers, or even flushed toilets. Those investigations introduced him to "sick building syndrome," a problem first identified in the 1970s in which the occupants experience fatigue, itchy eyes, headaches, and other symptoms. Exactly what causes these ailments isn't clear, but exposure to contaminated air is a likely culprit. Allen became convinced that the building you work in can have more impact on your health than your doctor.

In 2014, Allen accepted a position at Harvard, where he soon turned his attention to how the indoor environment can affect people's cognitive abilities. Many of us have struggled to pay attention during a long staff meeting in a stuffy conference room. Research by Allen and others suggests that lassitude may not be due solely to boredom, but also to the carbon dioxide (CO_2)-rich conference room air.

Ever since the energy shocks of the 1970s, buildings in the United States have been made as airtight and energy-efficient as possible. The result was a buildup of toxic volatile organic compounds (VOCs) and exhaled CO_2. "Green building standards" introduced in the late '90s focused on reducing toxic materials and making buildings healthier as well as more sustainable, but they didn't prioritize indoor air quality and ultimately did little to improve it.

In a multiyear series of experiments, Allen and his team have investigated the consequences. In the first study, published in 2015, they had 24 white-collar volunteers spend six working days in environmentally controlled office spaces at Syracuse University's Total Indoor Environmental Quality Laboratory. On various days the experimenters would alter ventilation rates and levels of CO_2 and VOCs. Each afternoon the volunteers were tested on their ability to think analytically and react to a crisis. (One test, for example put the volunteer in the role of a small-town mayor trying to react to an emergency.) All tests were double-blind: Neither the volunteers nor the study personnel knew that day's environmental conditions.

The results were dramatic. When volunteers worked in well-ventilated conditions (which lowered the levels of CO_2 and VOCs), they scored 61% higher than when they worked in typical office building conditions. When they worked in the cleanest conditions, with even lower CO_2 levels and higher ventilation rates, their scores climbed 101%.

To find out whether the results held up in the real world, Allen and his team recruited 109 volunteers from 10 office buildings across the United States. Six had been renovated to create better heat and humidity control, improve ventilation, and lower the use of toxic materials. Four had not. Allen's team gave each office worker a Fitbit-like bracelet to record heart rate, skin temperature, sleep patterns, and other physiological signs of well-being. Workers also completed a survey each day about how comfortable they felt and whether they experienced symptoms such as drowsiness or headaches. At the end of the

week, they took the cognitive tests. Workers in the buildings with good ventilation and lower levels of indoor pollution scored 26.4% higher than those in the unimproved buildings. They also reported sleeping better and experiencing fewer "sick building" symptoms.

"This is really important, interesting work," says Elliott Gall, an indoor air scientist at Portland State University. "It's a great example of the kind of interdisciplinary work [that explores] the complexity of indoor air and how it affects us."

Over time, Allen came to see businesspeople as natural allies who could act on his public health findings faster than government officials. He teamed up with John Macomber, a Harvard Business School lecturer and former CEO of one of the largest construction companies in New England. Macomber was impressed with Allen's research suggesting a tiny sacrifice in energy efficiency through improved ventilation could increase a business's bottom line by as much as 10% by decreasing absenteeism and boosting worker productivity. "I realized we've been missing the boat," Macomber says. "We're chasing pennies on energy when there's thousands of dollars in productivity issues."

Allen and Macomber consulted with companies and spoke at corporate conferences, making the economic case for improving ventilation and filtration as well as adjusting lighting, temperature, and humidity. "The idea of a healthy building has been made too complicated," they wrote in a book they co-authored, *Healthy Buildings: How Indoor Spaces Drive Performance and Productivity*. "There are just a handful of things we need to do to make a building healthier."

Allen's group continued to investigate how the indoor environment affects our mental state. They found that airline pilots exposed to CO_2 levels common in cockpits did worse on Federal Aviation Administration–mandated emergency response tests than when they breathed better air. They showed that during a heat wave, students who lived in non–air-conditioned dorms had slower reaction times and poorer problem-solving skills than those with air conditioning. They showed that bringing plants and views of nature into the workplace can lower office workers' heart rate, blood pressure, and other physiological indicators of stress.

In 2019, Allen's team embarked on an ambitious international project to examine the long-term impacts of indoor air quality by tracking the physical and cognitive health of more than 300 office workers in 43 buildings in six countries over the course of 1 year. They mailed each worker a wristband to monitor their physiology and a small sensor to continuously measure levels of fine particulates and CO_2 in their workspace. At predetermined times and levels of CO_2 and particulates, the program pinged each worker's smartphone with a quiz to test reaction time and cognitive function. The studies showed that in offices across the world, poor ventilation, CO_2, and particulates (which carry VOCs) conspire to significantly impair cognitive function.

WHEN THE first reports of the new coronavirus emerged from Wuhan, China, in January 2020, Allen realized his years researching air quality and disease transmission in indoor environments had new relevance. "Even though the virus was novel, there are elements in all this that feel quite familiar," he says. "It doesn't matter if it's a radiological hazard, biological hazard, or chemical hazard. We know how to assess the risk and put in appropriate controls."

Early in the pandemic, experts at the World Health Organization (WHO) and the U.S. Centers for Disease Control and Prevention (CDC) latched onto the idea that the virus was spread by large, exhaled droplets that float for a short time and then settle on surfaces. But scientists specializing in aerosols knew airborne viruses are more likely to ride on tinier particles exhaled when people breathe, sneeze, cough, or talk. Smaller than 5 microns, these particles can travel across a room and linger in indoor air for hours.

"One of our biggest frustrations over the past year is that we knew enough to act early on."
Joseph Allen, Harvard University

Aerosol experts such as Lidia Morawska of Australia's Queensland University of Technology, Gardens Point; Donald Milton of the University of Maryland, College Park; and Linsey Marr of Virginia Polytechnic Institute and State University argued that the focus on larger droplets had led to wrong-headed guidance about washing packages with bleach, staying 2 meters apart—even outdoors—and other forms of what some researchers called "hygiene theater." They urged policies that emphasized indoor mask wearing and less draconian regulations for people outdoors, where the virus would quickly disperse.

Allen signed on to that fight, collaborating with aerosol experts and public health researchers on scientific papers and bringing their case to the public. "He's a really good public communicator," says Marr, who credits Allen with helping her work get the attention it deserves.

"One of our biggest frustrations over the past year is that we knew enough to act early on," Allen says. "Even by late January 2020 we knew that airborne transmission of aerosols was not only likely, but probable." Waiting for proof made no sense. "This was a pandemic, an all-in moment, so why wouldn't we have immediately deployed every strategy that could have helped?" Those strategies, Allen knew from his research, include bringing more outside air into chronically underventilated buildings and using higher efficiency filters in ventilation units.

He started waking up at 4 a.m. to write opinion pieces. His first two, in the *Financial Times* and *The New York Times*, argued that buildings, if properly ventilated, could be formidable weapons in the fight against the virus. He and his team had measured air flow in schoolrooms under various conditions, and Allen explained that schools could easily be made safe by opening windows and buying the kind of high-efficiency particulate air purifiers sold at local home stores.

So much of his writing involved correcting misimpressions that he felt he was playing editorial whack-a-mole. No, he argued, you do not have to wipe your groceries with bleach. No, you do not have to avoid exercising outside. No, schools do not need to install costly air-purifying systems.

In July 2020, Morawska and Milton wrote an open letter to WHO—a full-throated appeal to recognize the importance of aerosol transmission. Allen was among 237 other scientists in 32 countries who signed on to the letter, which urged greater emphasis on indoor air quality. But WHO continued to downplay the importance of aerosol transmission. "I think some of the reluctance was that if [health authorities] say a disease is airborne, we'd have to provide N95 masks for every health worker and have negative pressure rooms in every hospital, which wasn't possible," Marr says.

Later that month, Allen and his Harvard colleague Parham Azimi published a study in the *Proceedings of the National Academy of Sciences* that used computer modeling to reconstruct the spread of the COVID-19 outbreak on the *Diamond Princess* cruise ship. By mapping the ship's ventilation system and the locations of people who came down with the disease, they showed that only aerosols, not larger droplets, could have traveled the necessary distances through the ducts. Similar findings emerged from studies by Marr, Morawska, and others.

Finally, in early May, after a series of persuasive papers in major journals, WHO and CDC acknowledged that the virus was transmitted primarily by fine aerosols. (Even then, the agencies issued no major announcements but simply changed the wording on their websites.) Since then, CDC has gone further,

Something in the air

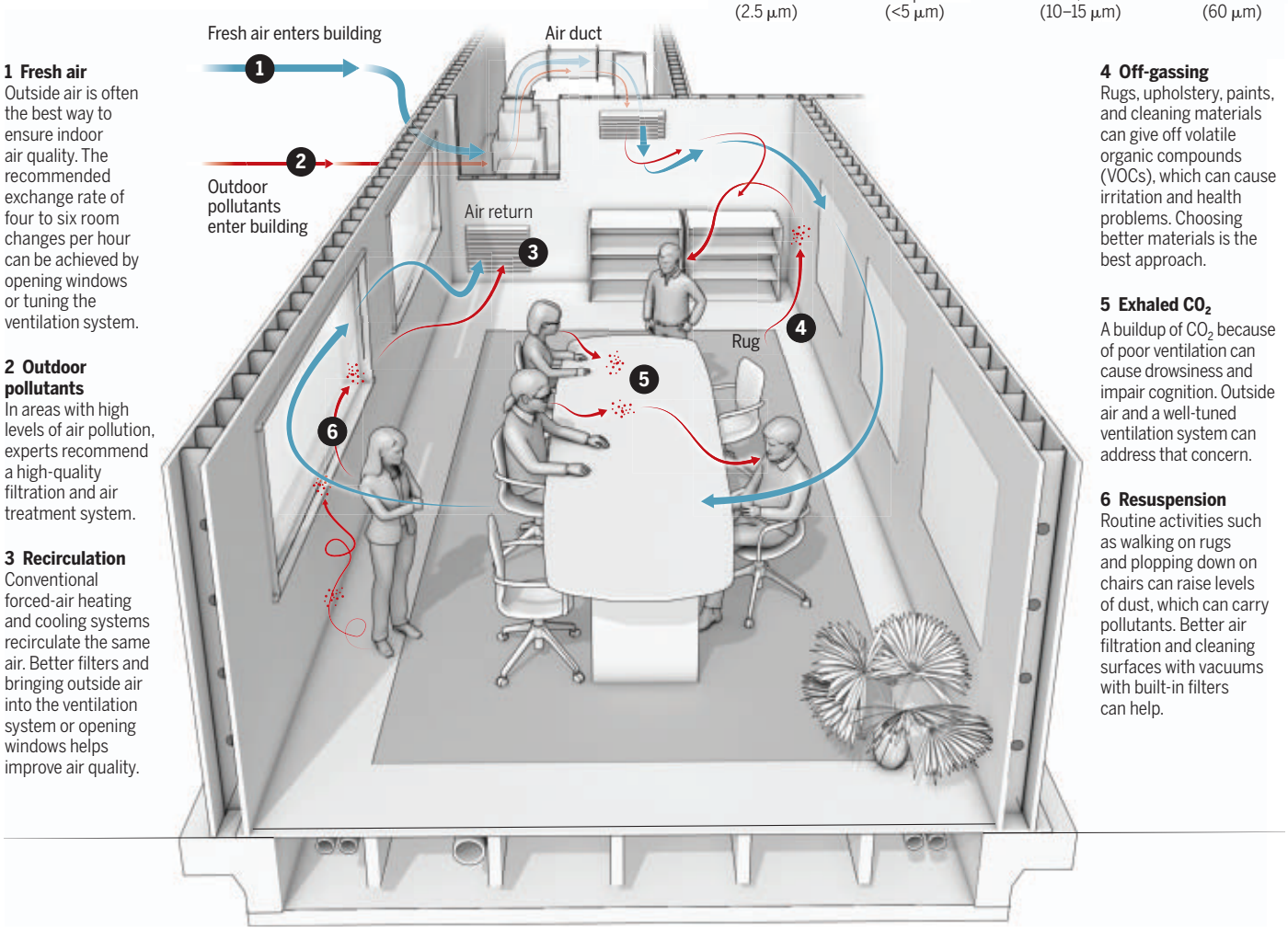
Many sources of indoor air pollution can affect human health and cognition. These include particles and gases emitted by furniture and building materials, as well as carbon dioxide (CO_2) exhaled by a building's occupants. Choosing better materials and improving ventilation, filtration, and air processing can help make buildings healthier.

Size matters

Of the many particles found in indoor air, exhaled particles smaller than 5 micrometers (μm) have become a focus during the COVID-19 pandemic because they can linger in the air and transmit disease.



Dust (2.5 μm) · Aerosol particles (<5 μm) · Pollen, molds (10–15 μm) · Grain of salt (60 μm)



1 Fresh air
Outside air is often the best way to ensure indoor air quality. The recommended exchange rate of four to six room changes per hour can be achieved by opening windows or tuning the ventilation system.

2 Outdoor pollutants
In areas with high levels of air pollution, experts recommend a high-quality filtration and air treatment system.

3 Recirculation
Conventional forced-air heating and cooling systems recirculate the same air. Better filters and bringing outside air into the ventilation system or opening windows helps improve air quality.

4 Off-gassing
Rugs, upholstery, paints, and cleaning materials can give off volatile organic compounds (VOCs), which can cause irritation and health problems. Choosing better materials is the best approach.

5 Exhaled CO_2
A buildup of CO_2 because of poor ventilation can cause drowsiness and impair cognition. Outside air and a well-tuned ventilation system can address that concern.

6 Resuspension
Routine activities such as walking on rugs and plopping down on chairs can raise levels of dust, which can carry pollutants. Better air filtration and cleaning surfaces with vacuums with built-in filters can help.

GRAPHIC: C. BICKEL/*SCIENCE*

issuing recommendations for reopening schools that stress the importance of good ventilation in addition to vaccinations.

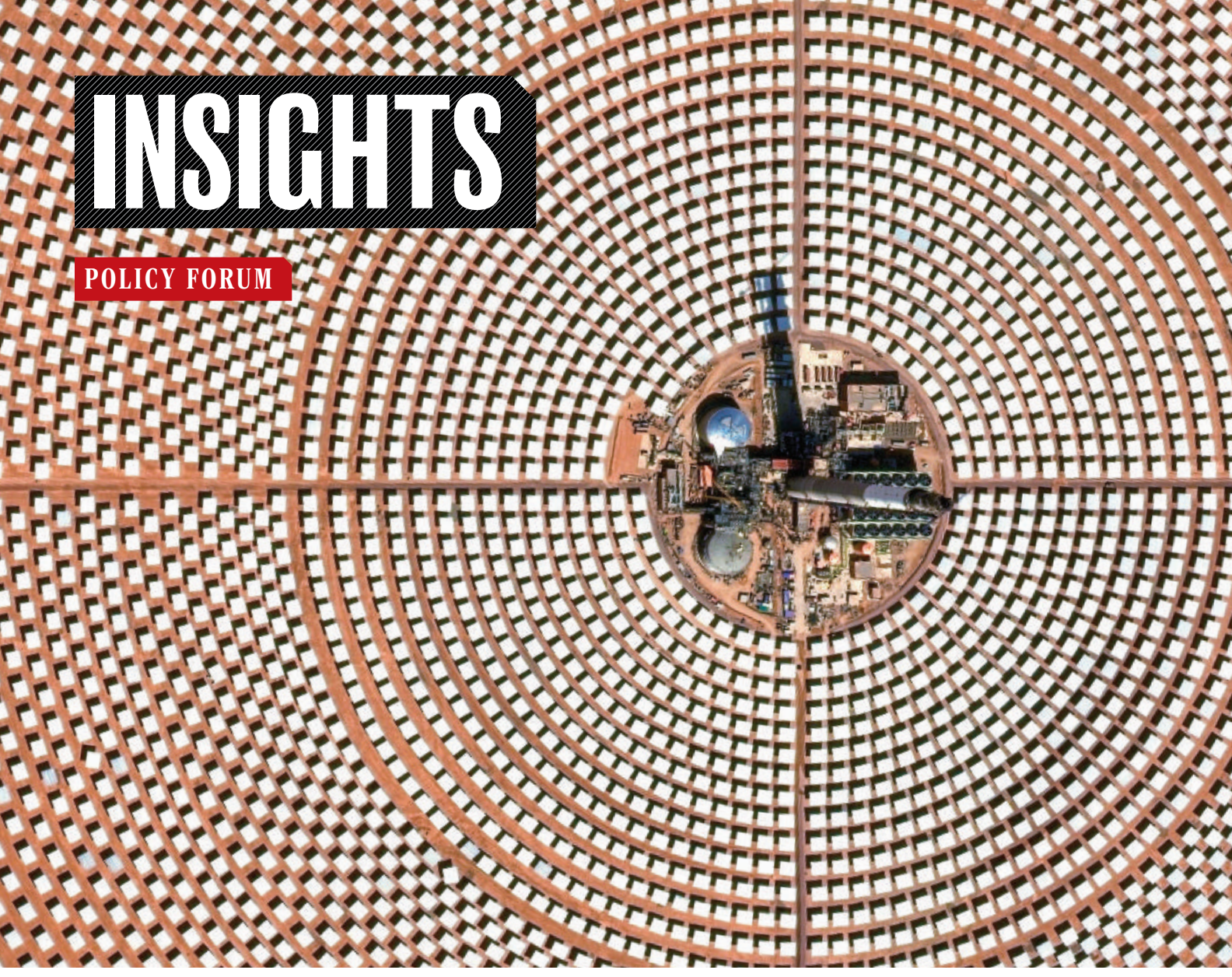
Meanwhile, Allen and his colleagues at the Harvard Healthy Buildings Program created a website with a comprehensive guide to maintaining proper ventilation in schools, homes, and businesses. The website advises building managers to bring in as much outside air as possible—a room air exchange rate of four to six times per hour, more than double the rate in a typical office or school building. In buildings that recirculate interior air, managers should upgrade to hospital-grade MERV 13 filters, which remove up to 90% of particles 2.5 microns or smaller, rather than the typical MERV 8, which can remove as little as 20%.

The new focus on indoor air quality could help hasten the end of the current pandemic and perhaps even help forestall the next one. It may bring broader changes as well. Businesspeople are recognizing the value of improving indoor air to create better working conditions and add value to their properties. "What we're seeing with some parts of the market—notably the higher end, real estate investment trusts, owners of multiple office buildings or apartments—is they're thinking really hard about competing" in the postpandemic market, Macomber says. "And one way to compete is to have healthier buildings."

Allen predicts that the new availability of cheap personal air quality monitors will quicken that competition and heighten people's awareness of the indoor environment. Previously, the only way to assess indoor air quality was to hire an expensive consultant. Now, with monitors available online for less than a couple of hundred dollars, any office worker or hotel guest can quickly monitor CO_2; some devices even detect VOCs. If consumers post results on websites like Yelp, businesses would be forced to pay attention. (Indeed, some building owners already boast about air quality in advertisements.)

"I think there's going to be a fundamental rebalancing in terms of how we think about indoor spaces," Allen says. "I think that people won't tolerate sick buildings, where you feel tired, your eyes itch, you have a headache, or you're stuffed into a closetlike office with no windows." It's one lasting positive from the pandemic. "That era is over," Allen says. "Rightly so, and good riddance." ■

Douglas Starr is a journalist in the Boston area.



SUSTAINABLE ENERGY

An action agenda for Africa's electricity sector

Modernization and expansion require heightened efforts

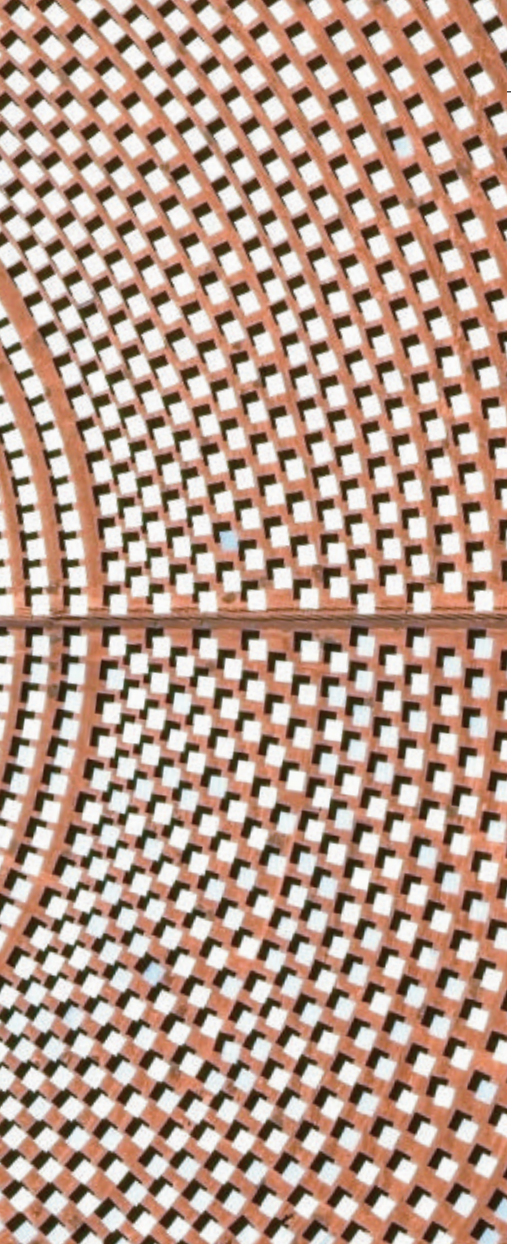
By **Daniel Puig[1], Magda Moner-Girona[2], Daniel M. Kammen[3], Yacob Mulugetta[4], Atef Marzouk[5], Maximilian Jarrett[6], Yohannes Hailu[7], Nebojša Nakićenović[8]**

To meet the needs of a growing population in a manner that is socially equitable, economically viable, and environmentally sustainable, Africa's electricity sector will require a major transformation (*1*). It has already undergone some important changes over the past decade. Efforts to expand access to electricity have proceeded at a slightly faster pace than anticipated 10 years ago. In parallel, the deployment of renewable energy technologies has progressed apace, despite new discoveries of natural gas across the continent and favored by the volatility of oil prices. Nonetheless, the expansion and modernization of Africa's electricity sector need heightened efforts, as evidenced by current electrification rates, generation-capacity levels, and security-of-supply indicators. We identify a suite of actions that, if implemented, would put Africa's electricity sector on track to sharply increase electrification rates across the continent while securing long-term access to affordable cleaner energy, and reducing greenhouse-gas emissions and emissions of local-air pollutants (see the figure) (*2*).

SUPPLY-SIDE INCENTIVES, DEMAND-SIDE SUBSIDIES

Despite progress with rural electrification, at least 250 million people in Africa cannot afford electricity, a gap that the COVID-19 global health pandemic has widened by about 80 million people who have fallen into extreme poverty (*3*). A combination of supply-side incentives and demand-side subsidies can go a long way toward expanding

PHOTO: DIGITALGLOBE/GETTY IMAGES



With a capacity of 580 MW, Morocco's Ouarzazate solar power station is one of the world's largest solar thermal plants.

electricity markets, by breaking down the vicious circle whereby the affordability gap limits electricity supply options.

Supply-side incentives refers to the combination of risk-guarantee schemes and so-called blending instruments. Through international finance, risk-guarantee schemes cover private lenders against the risk of a government failing to perform its contractual obligations and, in some instances, against expropriation, civil unrest, and other imponderables. Blending instruments are public-private partnerships through which grants are used to attract larger volumes of private and, in some cases, public finance. The combination of risk-guarantee schemes and blending instruments is set to play a central role in expanding electrification in Africa (*4*). However, it is increasingly clear that coordination among the various partners is indispensable for this combination of schemes to work. Here, coordination refers to the ability of the various actors to align eligibility and implementation requisites across their respective portfolios.

Given this precondition and the magnitude of the energy access problem in Africa, a continent-wide risk-guarantee scheme should be established, ideally by a combination of African and other multilateral lending institutions. Such an integrated approach, through which overall savings can outweigh risk premia (*5*), could be articulated under the aegis of the African Single Electricity Market, launched in early February 2021 with the main goal of harmonizing regulatory and technical aspects of electricity generation, transmission, and distribution across the continent.

Demand-side subsidies refer to direct or indirect payments to secure electricity access to the poorest households in Africa. This kind of subsidy can take the form of a reduced rate, calculated by using 12-month or shorter rolling-average electricity-consumption data, for example, and may be complemented with cross-subsidies. Although the experience with these types of subsidies remains limited, initiatives such as the European Union's GET. Invest program and the World Bank's Africa Renewable Energy and Access program are putting the concept on the policy agenda. These initiatives suggest that, when combined with supply-side incentives, demand-side subsidies can help close the affordability gap referred to above by serving the poorest households. African governments and bilateral and multilateral lenders should adopt and scale up this kind of approach. In doing so, they should be mindful of the need to rely on highly disaggregated estimates of affordability to underpin the design of demand-side subsidy programs, because affordability levels can vary greatly, even within small (one square kilometer–sized) areas (*6*).

Moving from successful pilot projects to transformative off-grid rural-electrification programs at the continental scale requires a concerted effort by all types of program financiers. Such an effort would be best articulated around a set of good practice requirements, reflecting the considerations outlined above and adopted by African governments and bilateral and multilateral lenders alike.

THE GREEN ENERGY–INFORMATION TECHNOLOGY INNOVATION FRONTIER

The electricity sector is in the midst of a major transformation, as utilities' primary role is gradually becoming "the management of energy supplied by independent power producers, rather than building and owning capacity themselves" (*7*). Such transformation is driven by digitalization, which makes it possible to "identify who needs energy and deliver it at the right time, in the right place, and at the lowest cost" (*8*). Worldwide, digitalization in the power sector is forecast to become a US$ 64 billion market by 2025, with smart-meters alone accounting for US$ 26 billion in the same year (*7*).

For Africa's electricity sector, digitalization entails three main opportunities (*8*). First, reduce the need for additional investments in electricity infrastructure: Through preset financial incentives, digital monitoring and data analysis tools make it possible to adjust electricity consumption during peak-demand and off-peak periods, accommodate variability in supply, and limit network congestion. Second, reduce electricity costs for consumers, by facilitating the integration of distributed electricity generation (typically, household-level solar photovoltaic panels and storage facilities) into local grids. Third, help manage the uncertainties in supply that are inherent to the variability of renewable energy–powered electricity generation and electricity supplied by interconnected mini-grids.

To reap the full benefits of digitalization, a pan-African cross-sectoral digital road map, such as the European Union's Digital Single Market Strategy, needs to be developed, consistent with existing guidelines and mechanisms for secure data sharing (*8*). At the national level, governments would benefit from adopting policies that are conducive to the expansion of renewable energy markets, and of technological innovation, notably with regard to the development of digital start-ups—namely, the programmers, vendors, and certifying companies that are needed to put digitalization to work (*7*). When measured against 21 criteria related to the domestic renewable energy sector and digital business fabric, the African countries that are best placed to embrace digitalization are Nigeria, Egypt, Kenya, Ethiopia, and Ghana (*7*).

LOCAL CONTENT REQUIREMENTS

In Africa, financial incentives for renewable energy–powered electricity generation have had modest success. Feed-in tariffs, through which governments set the price of an electricity-generation project, and auctions, which are based on competitive bidding by industry to determine the price, are the two main financial incentives used. Whereas feed-in tariffs are becoming less relevant given the

[1]Technical University of Denmark, Copenhagen, Denmark. [2]European Commission Joint Research Centre, Ispra, Italy. [3]University of California, Berkeley, CA, USA. [4]University College London, London, UK. [5]African Union Commission, Addis Ababa, Ethiopia. [6]International Energy Agency, Paris, France. [7]United Nations Economic Commission for Africa, Addis Ababa, Ethiopia. [8]International Institute for Applied Systems Analysis, Laxenburg, Austria. Email: dapu@dtu.dk

fall in cost of renewable energy–powered technologies, auctions suffer from limited competition in the electricity market and the lack of domestic expertise, South Africa being the notable exception. Coherent policies that are stable over the long term, not least with regard to renewable energy targets, independent electricity-market regulators, and, as relevant, measures to reform monopolies, are indispensable for financial incentives to deliver at their fullest, as they have done in other parts of the world.

Irrespective of the type of financial incentive used, most electricity projects in Africa are undertaken by foreign developers, notably European, Chinese, and United States companies, owing to their experience and, especially, their ability to secure financing. As a result, African governments have introduced different types of so-called local-content requirements, namely obligations concerning local employment, procurement of local goods and services, and the transfer of technologies and know-how, to which foreign investors have to abide. In countries such as Kenya and Nigeria, these requirements are defined through quantitative targets, whereas in other countries, such as Uganda and Zambia, they take the form of qualitative goals.

Empirical evidence, from Africa and elsewhere (*9*), reveals that there is ample scope for improving the effectiveness of local-content requirements. Two design shortcomings are especially common in Africa: the mismatch between requirements and actual domestic capacities, which deters and delays foreign investors, and the lack of sustained demand, which limits the ability of local providers to plan and undertake capacity expansions.

Experience from other natural-resource–based sectors shows that, to increase the effectiveness of local-content requirements, national governments in countries with limited domestic capacities can focus on a small number of promising subsectors, as opposed to a broad-based approach targeting all aspects of electricity generation, transmission, and distribution. For the subsectors chosen, a certification system can help ensure the required quality standards. National governments in countries with stronger capacities can take a more ambitious approach, structured around introducing legislative provisions and governance arrangements for local-content requirements; engaging both foreign investors and local providers in the definition of the requirements; and monitoring and reporting on the extent to which the requirements are met. For both types of countries, it is advisable to tighten requirements only gradually while continuously adjusting ambition levels to reflect technology and labor market dynamics.

POWER POOLS

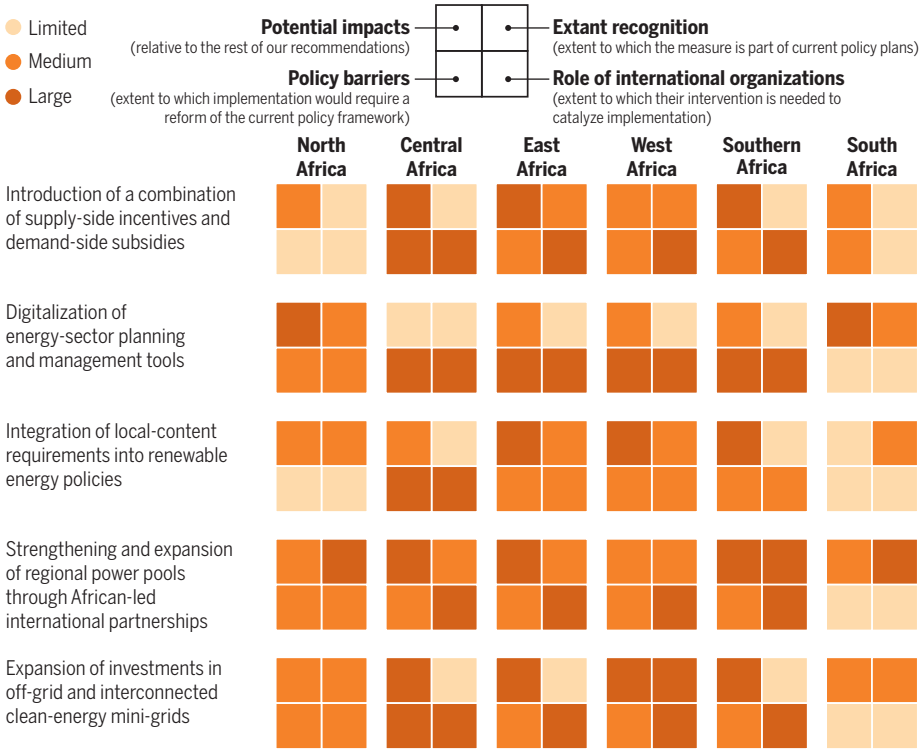
Power pooling, through cross-border trade in electric power, helps reduce electricity bills and enhances the reliability of electricity supply. Regional power pools, based increasingly on renewable energy supplies, are now possible across most of the African continent (*10*). Nonetheless, additional efforts are needed to reap the full benefits of power pooling (*11*). Notwithstanding shared challenges, notably underinvestment, the nature of the additional efforts needed varies across African regions.

South Africa is the main electricity producer for the Southern African power pool, facilitated by the Southern African Development Community (SADC). Given the challenges that the country is increasingly facing to meet its domestic demand for electricity, and the sharp decreases in cost of solar, wind, and energy storage, the case for relying on solar and wind energy–powered electricity generation becomes stronger in the region. Yet, at present, for both renewable energy and electric-power transmission, many of the investment discussions in the SADC region focus on large dams, which have been the technology of choice for decades. Concentrating solar power, a technology that generates electricity from the heat obtained by concentrating solar energy (in contrast to converting solar energy directly into electricity, as photovoltaic systems do), is already being deployed in South Africa, owing to its much larger generation capacity compared to photovoltaic system–based installations and its ability to store large amounts of electricity. Concentrating solar power technology can help shift the balance away from hydropower and toward solar energy, but only to the extent that stronger financial incentives are in place, compared to those introduced thus far (*1*).

To date, the members of the Maghreb Electricity Committee (COMELEC), Northern Africa's power pool, have only engaged in cross-border trade with the Iberian Peninsula, across the Mediterranean Sea (Spain currently exports electricity to Morocco). As concentrating solar power in Morocco develops,

Enabling the clean energy revolution in Africa

Power-sector reform, energy access, and investing in Sustainable Development Goal 7 each build economic opportunities. Color-coded assessment is based on the authors' collective judgment.



North Africa: Algeria, Egypt, Libya, Mauritania, Morocco, and Tunisia; West Africa: Benin, Burkina Faso, Cabo Verde, Côte d'Ivoire, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Niger, Nigeria, Senegal, Sierra Leone, and Togo; East Africa: Burundi, Comoros, Djibouti, Eritrea, Ethiopia, Kenya, Rwanda, Seychelles, Somalia, South Sudan, Sudan, Tanzania, and Uganda; Central Africa: Cameroon, Central African Republic, Chad, Congo, Democratic Republic of Congo, Equatorial Guinea, and Gabon; Southern Africa: Angola, Botswana, Eswatini, Lesotho, Madagascar, Malawi, Mauritius, Mozambique, Namibia, São Tomé and Príncipe, Zambia, and Zimbabwe; South Africa: South Africa. This regional breakdown is used by the United Nations Economic Commission for Africa. The designations employed do not imply the expression of any opinion whatsoever on the part of the authors concerning the legal status of any country, territory, or area, or of its authorities, or concerning the delimitation of its boundaries.

GRAPHIC: N. DESAI/*SCIENCE* BASED ON PUIG *ET AL.*

the country plans to export electricity to Spain and possibly Portugal. Tunisia and Egypt are planning similar export arrangements (with Italy and Greece, respectively). Against this background, COMELEC has pledged to launch, in 2025, a common electricity market for its five members. A planned interconnection between Morocco and Mauritania stands as the first milestone toward that goal. At present, efforts to develop intra-regional trade are hampered by insufficiently liberalized electricity markets and limited tradition of cross-border infrastructure investments in the region, which are both issues on which COMELEC has influence.

Both the Eastern Africa Power Pool (EAPP) and the West African Power Pool (WAPP) originate from preexisting cross-border arrangements aimed at promoting cooperation on energy issues. In both regions, cooperation thus far has been limited to bilateral agreements, such as the lines linking Kenya with Ethiopia and Ghana with Burkina Faso. Nonetheless, efforts are underway in both regions to harmonize regulations, contractual terms, and tariff levels applicable to electricity transmission, with a view to expanding power pooling beyond these bilateral agreements. In the EAPP region, this goal is hampered by the lack of an updated development plan, reflecting changes in assumptions about costs and the expected impacts of newly implemented policies, because such a plan is a precondition for raising the funding required to establish a regional power pool. In the WAPP region, utility near-bankruptcy is a key hurdle to intra-regional trade, which calls for the regulated entry of foreign investment.

The Central African Power Pool (CAPP) remains underdeveloped. Poverty and other developmental challenges in the region limit the size of the electricity market, thus inflating prices. Compounding this challenge, weak governance arrangements result in insufficient competition in the sector, leading to underinvestment and ultimately thwarting any intra-regional power-pooling ambitions in the short term. In light of these challenges, progress with regional power pooling should build on gradually expanding bilateral electricity-trade agreements involving the Democratic Republic of Congo, owing to its high hydropower potential, drawing on support by international lenders.

MINI-GRIDS

In moderately populated areas, where both grid extension and deployment of a relatively large number of stand-alone electricity-generation systems would be prohibitively expensive, off-grid mini-grids are the most economical electrification option in most cases (*12*). In light of strained public-sector budgets, scaled-up deployment of off-grid mini-grids depends on expanded and sustained private-sector investment. Two approaches to mobilizing private-sector finance can bolster off-grid mini-grid deployment rates, as evidenced by the financial model used to set up the latest generation of off-grid mini-grids.

First, so-called third-generation mini-grids, which combine photovoltaic solar systems and batteries with or without a back-up diesel-powered electricity generator, require less than 2 weeks of scheduled maintenance per year. Such a high level of reliability makes it possible to incentivize off-grid mini-grid deployment through performance-based subsidies (*13*). For example, with World Bank backing, Nigeria's rural electrification agency pays off-grid mini-grid developers US$ 350 per connection, provided that the customer has had a steady supply of power for at least 3 months. Similarly, the reliability of third-generation mini-grids allows developers to offer customers a contract that includes, in addition to the electricity connection, the option to purchase income-generating appliances, such as machines for welding, milling, and rice hulling, thus increasing deployment rates (*13*).

Second, interconnected mini-grids can generate and distribute electricity to customers that are not connected to the main grid, while selling excess power to the main grid, as found in Nigeria, or they can choose to sell to the main grid all the electricity generated, as do some mini-grids in Tanzania. Electricity consumers benefit from interconnected mini-grids through increased reliability of supply and, in most instances, lower prices (*14*). Interconnected mini-grid development is hampered by both regulation, which is unclear at best and in some cases acts as a deterrent to interconnections, and technical standards, which may differ between mini-grids and the main grid. Overcoming the barriers to interconnected mini-grid development requires national governments to clarify licensing procedures and tariff regulations and ultimately establish unambiguous tariff levels for the various interconnection options, a set of tasks that can be facilitated by the International Renewable Energy Agency.

A PIVOTAL ROLE

Implementing our recommendations requires careful consideration of the policy options that may be most appropriate in each situation. As is the case with most aspects of public policy, power, agency, and politics will have an impact on the extent to which the policy options chosen are compatible with societal goals related to environmental quality, employment, and equity, among others (*15*). Specifically, the choice of policy options will be influenced by two sets of policy drivers that, in a developmental context, are especially powerful. The first is prevailing national-level incumbency issues (namely, resistance to change by incumbents in the energy sector) and information asymmetries among different stakeholders in a country, which invariably punish prospective new entrants in the energy sector. The second is unduly rigid priorities and procedures on the part of bilateral and multilateral lenders.

Against this background, we contend that the Sustainable Energy for All initiative has a pivotal role to play in ensuring that the policy choices that will shape the future of Africa's electricity sector are not captured by the two sets of policy drivers referred to above. As an international organization that is independent from, but works in partnership with, both the United Nations and its member countries, the Sustainable Energy for All initiative can act as the watchdog that upholds environmental and developmental concerns, notably with regard to promoting renewable energy technologies and championing local employment. It can play this role under the aegis of the newly launched African Single Electricity Market and by capitalizing on United Nations–sponsored high-level consultations, such as the High-level Dialogue on Energy 2021, which will take place in New York in September of this year. The opportunity is ripe, and leadership has to be up to the mark. ■

REFERENCES AND NOTES

1. IEA, *Africa Energy Outlook 2019* (International Energy Agency, Paris, 2019).
2. C. E. Casillas, D. M. Kammen, *Science* **330**, 1181 (2010).
3. IEA, *World Energy Outlook 2020* (International Energy Agency, Paris, 2020).
4. IEA, *World Energy Investment 2020* (International Energy Agency, Paris, 2020).
5. D. Matthäus, M. Mehling, *Joule* **4**, 2627 (2020).
6. S. Szabó *et al.*, *Sci. Rep.* **11**, 3226 (2021).
7. BNEF, *Digitalization of Energy Systems* (Bloomberg Energy Finance, New York, 2017).
8. IEA, *Digitalisation and Energy* (International Energy Agency, Paris, 2017).
9. B. Probst, V. Anatolitis, A. Kontoleon, L. D. Anadón, *Nat. Energy* **5**, 842 (2020).
10. J. P. Carvallo, B. J. Shaw, N. I. Avila, D. M. Kammen, *Environ. Sci. Technol.* **51**, 10232 (2017).
11. RM, *African Power Pools Outlook 2018-2019* (Research and Markets, Dublin, 2019).
12. P. Alstone, D. Gershenson, D. M. Kammen, *Nat. Clim. Chang.* **5**, 305 (2015).
13. ESMAP, *Mini Grids for Half a Billion People: Market Outlook and Handbook for Decision Makers* (The World Bank's Energy Sector Management Assistance Program, Washington, DC, 2019).
14. J.-P. Carvallo, J. Taneja, D. Callaway, D. M. Kammen, *Proc. IEEE* **107**, 1906 (2019).
15. K. S. Rogge, K. Reichardt, *Res. Policy* **45**, 1620 (2016).

ACKNOWLEDGMENTS

D.M.K. acknowledges support from the Zaffaroni and Karsten Family Foundations and the James and Katherine Lau Chair in Sustainability.

10.1126/science.abh1975



ORGANISMAL BIOLOGY

Learning to move in the real world

Leaping squirrels show that locomotion entails perceiving and innovating possibilities for action from moment to moment

By **Karen E. Adolph**[1] *and* **Jesse W. Young**[2]

Comparative research reveals extraordinary animal athleticism—mites lift >1000 times their body weight, mantis shrimp strike their prey with the force of a bullet, and peregrine falcons dive toward prey at 335 mph (539.13 km/hour) (*1*). Even human babies travel the distance of eight football fields per hour during free play (*2*). However, movement in the real world is not about being the strongest, fastest, or most active. Rather, effective action is a moment-to-moment process of matching the current status of the body to features of the environment (*3*). Locomotion—like other actions—must be tailored to local conditions. On page 697 of this issue, Hunt *et al.* (*4*) provide an elegant demonstration of the creativity of functional movement, showing that wild squirrels tune their leaps to branch bendiness and target distance, even inventing ingenious maneuvers when required.

Matching behavior to local conditions—that is, perceiving "affordances" for action—involves perception-action coupling (*5*). Animals must generate perceptual information about which actions are possible and then select appropriate actions from this set of possibilities (*2*, *3*). The arboreal canopy is an ideal natural laboratory for studies of perception-action coupling because support diameters vary from tiny twigs to massive boughs, with more than three orders of magnitude of variation in branch compliance (*6*). Moreover, mistakes can have serious consequences—falling injuries account for most limb-bone fractures in free-ranging primates (*7*). Arboreal animals must avoid errors or quickly correct them.

Hunt *et al.* created an outdoor obstacle course and trained wild squirrels to jump from cantilevered perches to cross gaps of varying distances. Launch perches varied in compliance, requiring squirrels to negotiate a critical trade-off: Moving toward the end of the perch would shorten leaping distance but compromise stability and force production; staying closer to the base would ensure a secure launching platform but at the cost of increased gap distance. Squirrels launched closer to the base of the perch, which suggests that support compliance is a critical factor in arboreal locomotion (*8*). Notably, squirrels also demonstrated the creativity of functional movement. After learning the leaping task, squirrels encountered new adjustments to launch-perch compliance and gap distance, necessitating moment-to-moment modifications in behavior to ensure gap-crossing success. Squirrels innovated new strategies—parkour-like jumping maneuvers off the back wall of the apparatus when launching impulse was insufficient to cross the gap, and front and back flips to grasp the landing perch when their leaps over- or undershot the target.

Squirrels are not alone in the precision and creativity of their locomotion. Every animal must perceive and exploit affordances for locomotion under variable conditions (*3*, *5*). Bats and iguanas alter locomotor forces to compensate for increased body mass associated with feeding and pregnancy (*9*, *10*). Running guinea fowl adjust limb postures within a single step to maintain stability after an unexpected drop (*11*). Likewise, human infants gauge affordances with exquisite precision and invent new locomotor strategies on the fly (e.g., sliding down steep slopes or high drop-offs on their bottoms, backward feetfirst, or headfirst like Superman) (*2*, *3*).

So how do animals learn to gauge and adjust their movements? Hunt *et al.* suggest that trial-and-error learning governs affor-

[1]Department of Psychology, New York University, New York, NY, USA. [2]Department of Anatomy and Neurobiology, Northeast Ohio Medical University, Rootstown, OH, USA. Email: karen.adolph@nyu.edu; jwyoung@neomed.edu

PHOTO: LAWRENCE WANG, DOMINIC PANG, SEBASTIAN LEE, JEREMY SNOWDEN, TINA KUANG, FRANK ALIAGA AUQUI



Wild squirrels adjust their leaps in real-time to the mechanical properties of the environment, elegantly demonstrating the creativity of functional movement in the real world.

dance perception in the course of a single session. However, adult squirrels have had a lifetime of learning, so development must be a critical factor. During development, new affordances emerge as animals' bodies, skills, and effective environments change (*2*). Human infants can grow up to 2 cm in a single day (*12*). One week, babies are crawlers; the next, they are walkers (*13*)—yesterday, objects on the coffee table were out of sight and beyond reach; today, they are accessible (*2*). Thus, learning occurs in the context of development, and the flux of body growth and motor-skill acquisition ensures that infants do not learn fixed solutions. Indeed, static solutions would be maladaptive in a continually changing ecosystem. Instead, infants "learn to learn." They learn to detect information for affordances at each moment to determine which actions are possible with their current body and skills in a given environment. Learning amid development results in perception-action coupling that is sufficiently flexible to scale up to the novelty and variability of action in the real world (*2*, *3*).

Human perceptual-motor development is an iterative process, where experience moving in a variable environment generates perception of new affordances that, in turn, facilitates new experiences (*2*). Human infants move to learn while they are learning to move. Likely, infant squirrels and other arboreal animals show similar calibration and creativity as they learn to navigate the canopy, particularly because misperceptions can prove fatal. Future work should consider the ontogeny of perception-action coupling in natural habitats.

Just as movement in the real world requires flexibility and creativity, researchers studying natural locomotion must be as ingenious as their animal subjects. The trick is to capture movement in all its complexity while retaining sufficient experimental control and measurement fidelity (*2*). The study of Hunt *et al.* is a beautiful example. Their unexpected results elucidate what every homeowner knows: Squirrels are clever acrobats when navigating complex environments. ■

REFERENCES AND NOTES

1. D. J. Irschick, T. E. Higham, *Animal Athletes: An Ecological and Evolutionary Approach* (Oxford Univ. Press, 2016).
2. K. E. Adolph, *Hum. Development* **63**, 180 (2019).
3. K. E. Adolph, S. R. Robinson, in *Cognitive Processes*, L. S. Liben, U. Mueller, Eds., vol. 2 of *Handbook of Child Psychology and Developmental Science* (Wiley, 2015).
4. N. H. Hunt, J. Jinn, L. F. Jacobs, R. J. Full, *Science* **373**, 697 (2021).
5. J. J. Gibson, *The Ecological Approach to Visual Perception* (Houghton Mifflin, 1979).
6. N. T. Dunham, A. McNamara, L. Shapiro, T. Hieronymus, J. W. Young, *Am. J. Phys. Anthropol.* **167**, 569 (2018).
7. N. C. Lovell, *Am. J. Phys. Anthropol.* **34**, 117 (1991).
8. J. W. Young, B. M. Stricklen, B. A. Chadwell, *J. Exp. Biol.* **219**, 2659 (2016).
9. J. Iriarte-Diaz, D. K. Riskin, K. S. Breuer, S. M. Swartz, *PLOS ONE* **7**, e36665 (2012).
10. J. Scales, M. Butler, *Integr. Comp. Biol.* **47**, 285 (2007).
11. M. A. Daley, A. A. Biewener, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15681 (2006).
12. M. Lampl, J. D. Veldhuis, M. L. Johnson, *Science* **258**, 801 (1992).
13. K. E. Adolph, S. R. Robinson, J. W. Young, F. Gill-Alvarez, *Psychol. Rev.* **115**, 527 (2008).

ACKNOWLEDGMENTS

K.E.A. was supported by National Institute of Child Health and Human Development grant R01HD033486 and Defense Advanced Research Projects Agency grant N66001-19-2-4035. J.W.Y. was supported by National Science Foundation grant BCS-1921135.

10.1126/science.abj6733

STRUCTURAL BIOLOGY

Maturation of HIV-1

Structural transformation of HIV-1 matrix during virus maturation promotes infection

By **Yuta Hikichi** *and* **Eric O. Freed**

HIV-1 particle assembly is driven by the Gag polyprotein precursor, which contains several structural and functional domains that engage in protein-protein, protein-lipid, and protein-RNA interactions during virion assembly. Concomitant with virus release from an infected cell, the viral protease cleaves the Gag precursor to liberate the mature Gag proteins, triggering a morphological transformation of the virus particle, called maturation. The matrix (MA) domain plays key roles in directing Gag to the plasma membrane of the host cell and in the incorporation of the viral envelope glycoprotein (Env) complex into the assembling particle. On page 700 of this issue, Qu *et al.* (*1*) report the cryo–electron tomography (cryo-ET) structure of MA in both immature and mature particles. The results provide important insights into HIV-1 assembly and maturation and the role that MA plays in these processes. These findings may suggest new antiviral strategies that target MA.

The Gag polyprotein precursor contains MA, capsid (CA), nucleocapsid (NC), and p6 domains (*2*, *3*). The NC domain is flanked by two spacer peptides, SP1 and SP2. The structure of the intact HIV-1 Gag precursor has been challenging to determine because of its large size and flexibility. However, structures are available for individual mature Gag proteins and for some domains of the Gag precursor—notably, CA. As an isolated protein, MA folds into a highly globular structure (*4*, *5*). A myristic acid moiety covalently linked to the amino terminus of MA anchors Gag in the lipid bilayer of the virion, and a highly basic region of MA interacts electrostatically with the acidic headgroup of the host cell plasma mem-

Virus-Cell Interaction Section, HIV Dynamics and Replication Program, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, MD, USA. Email: efreed@nih.gov

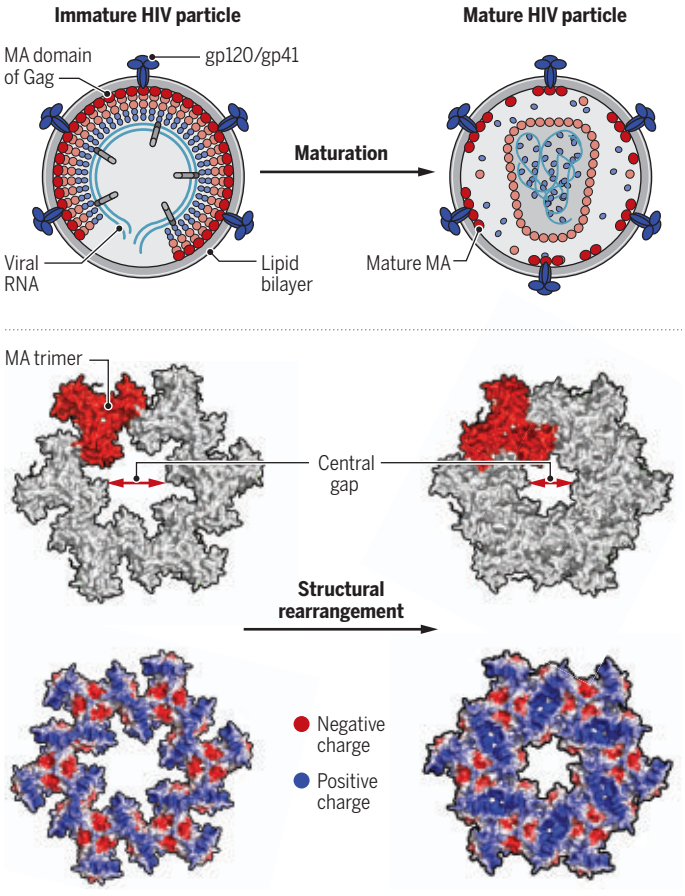
brane lipid, phosphatidylinositol 4,5-bisphosphate [$PI(4,5)P_2$] (*6*, *7*). The interaction between the MA domain of Gag and $PI(4,5)P_2$ targets Gag to the inner leaflet of the plasma membrane, where HIV-1 assembly takes place. The structural changes that take place within Gag and Env in the newly released virus particle during the maturation step are essential for entry into and productive infection of new cells.

HIV-1 MA was shown to crystallize as a trimer (*5*) and assemble into a hexameric lattice of trimers on an artificial $PI(4,5)P_2$–containing membrane monolayer (*8*). Although CA forms a hexagonal lattice in both the immature particle and the mature CA core, the orientation and intersubunit interactions of CA differ in these two contexts (*9*). In contrast to CA, the arrangement of MA in HIV-1 virions has until now been largely refractory to high-resolution structural analysis because of the curved and flexible nature of the viral membrane. Defining the structure of the MA lattice in virions is key to understanding how MA functions in Gag targeting, Env incorporation, and viral infection.

To address this key gap in our knowledge of HIV-1 structure, Qu *et al.* demonstrate that MA assembles into hexamers of trimers in both immature and mature virions, with the trimeric structure similar to that of the previously reported crystallized MA trimer (*5*) (see the figure). However, the arrangement of MA in the immature and mature lattice is very different, demonstrating that the MA lattice, like the CA lattice (*9*), undergoes a rearrangement on particle maturation. A hole in the MA lattice is located at the center of the hexamer of trimers. This hole, which has been hypothesized to be a docking site for the cytoplasmic tail of the viral transmembrane Env glycoprotein gp41, is lined with basic residues in the immature particle but neutral residues in the mature virion. MA mutations that impair Env incorporation are not exposed to the central holes of the immature MA lattice but rather, in some cases, are located at trimer interfaces, which is consistent with a role for MA trimerization in Env incorporation (*10*, *11*).

Maturation of the matrix lattice

The membrane-bound HIV-1 matrix (MA) domain assembles into a hexamer of trimers in both immature and mature virions. The arrangement and intertrimer interactions of the MA lattice differ in immature and mature particles. The central gap in the MA lattice is hypothesized to interact with the envelope glycoprotein (Env, comprising gp41 and gp120) cytoplasmic tail (double-headed arrows). At the bottom is shown the electrostatic surface potentials.



The results of Qu *et al.* also indicate changes in MA-lipid interaction. Specifically, in the immature particle, both acyl chains of $PI(4,5)P_2$ are embedded in the membrane bilayer, whereas in the mature particle, a density consistent with an acyl chain, possibly derived from $PI(4,5)P_2$, is embedded into a proposed acyl chain binding groove in MA (*7*). This study demonstrates that particle maturation changes both the protein-protein and protein-lipid interaction properties of MA, presumably priming the virion for the subsequent infection process, possibly by changing the dynamic properties of the viral membrane.

Because gp41 could not be resolved in the current analysis, it is unclear how the gp41 cytoplasmic tail is accommodated by the MA lattice, a key step in Env incorporation during particle assembly. Moreover, it is unknown how the conformational changes in MA that accompany particle maturation affect the structure and dynamics of Env. Previous studies have shown that Env in immature virions is fusion-incompetent and that cleavage events downstream of SP1 can activate fusogenicity, enabling the virus to enter target cells (*12*, *13*). Qu *et al.* likewise show that cleavage of Gag downstream of SP1 is sufficient to trigger structural rearrangement of the MA lattice. The suppression of fusion observed in immature particles requires the gp41 cytoplasmic tail (*12*, *13*). Env trimers cluster in a maturation-dependent fashion on HIV-1 particles, and again, this clustering is dependent on the gp41 cytoplasmic tail (*14*).

Together, these findings suggest that structural rearrangement of the MA lattice during maturation may activate Env fusogenicity by changing the conformation, mobility, and/or clustering of Env trimers, allowing virus entry. Because it has been reported that virus maturation affects the stiffness of HIV-1 particles (*15*), what role do the structural rearrangements of MA, and the resulting extrusion of lipid acyl chains from the viral membrane, play in this shift in rigidity? Further understanding HIV-1 assembly and maturation, as provided by the study of Qu *et al.*, will potentially reveal new therapeutic targets for controlling this pathogen. ■

REFERENCES AND NOTES

1. K. Qu *et al.*, *Science* **373**, 700 (2021).
2. E. O. Freed, *Nat. Rev. Microbiol.* **13**, 484 (2015).
3. W. I. Sundquist, H.-G. Kräusslich, *Cold Spring Harb. Perspect. Med.* **2**, a006924 (2012).
4. M. A. Massiah *et al.*, *J. Mol. Biol.* **244**, 198 (1994).
5. C. P. Hill, D. Worthylake, D. P. Bancroft, A. M. Christensen, W. I. Sundquist, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 3099 (1996).
6. A. Ono, S. D. Ablan, S. J. Lockett, K. Nagashima, E. O. Freed, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14889 (2004).
7. J. S. Saad *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11364 (2006).
8. A. Alfadhli, R. L. Barklis, E. Barklis, *Virology* **387**, 466 (2009).
9. F. K. Schur *et al.*, *Nature* **517**, 505 (2015).
10. P. R. Tedbury, S. D. Ablan, E. O. Freed, *PLOS Pathog.* **9**, e1003739 (2013).
11. P. R. Tedbury, M. Novikova, S. D. Ablan, E. O. Freed, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E182 (2016).
12. D. J. Wyma *et al.*, *J. Virol.* **78**, 3429 (2004).
13. T. Murakami, S. Ablan, E. O. Freed, Y. Tanaka, *J. Virol.* **78**, 1026 (2004).
14. J. Chojnacki *et al.*, *Science* **338**, 524 (2012).
15. N. Kol *et al.*, *Biophys. J.* **92**, 1777 (2007).

10.1126/science.abj9075

GRAPHIC: C. BICKEL/*SCIENCE*

Therapy based on functional RNA elements

Fragments of long noncoding RNAs show potential in treating a metabolic disorder in mice

By **Rotem Ben-Tov Perry**[1,2] and **Igor Ulitsky**[1,2]

Over the past several years, advances in RNA sequencing have led to an increased appreciation of the prevalence and function of noncoding RNAs, including long noncoding RNAs (lncRNAs). These are typically expressed in a tissue-specific manner in healthy tissues and are often dysregulated in disease, making them potential biomarkers and therapeutic targets. On page 662 of this issue, Li *et al.* (*1*) reveal the biological importance of a lncRNA in an inherited metabolic disorder called phenylketonuria (PKU) and demonstrate in mice that a molecule that mimics the functional region of this lncRNA is a promising therapeutic. This discovery suggests that short lncRNA fragments could overcome some of the challenges faced by other RNA therapeutic modalities.

RNA-based and RNA-targeting therapeutics have many advantages: They are cost-effective, are relatively simple to manufacture, can target otherwise undruggable pathways, and have demonstrated success in the treatment of several diseases. Although RNA therapeutics have a long and bumpy history, advances in the generation, purification, and cellular delivery of short oligonucleotides and long RNAs have led to regulatory approval of several RNA-focused therapies, including the much-celebrated messenger RNA (mRNA)–based COVID-19 vaccines.

The human genome encodes a large number of RNA molecules that do not encode functional proteins, including tens of thousands that are classified as lncRNAs (*2*). lncRNAs and mRNAs are virtually identical at the molecular level, although lncRNA production is typically much more tissue specific. Also, lncRNA genes evolve much faster than protein-coding ones (*3*). lncRNAs have diverse roles, including in gene regulation and as scaffolds for macromolecular assemblies. Some lncRNAs function in cis—that is, in the vicinity of their site of transcription—whereas others are trans-acting, and their function is not affected by their production site within the genome. Because lncRNAs are expressed in a cell-, tissue-, developmental stage–, or disease-specific manner, their modulation could have substantial, but focal, consequences, which are expected to be well tolerated. However, the progress in elucidating their functions and causally linking genetic changes in lncRNA loci to disease has been slow.

Antisense oligonucleotides (ASOs) are currently the most common approach for therapeutic targeting of RNAs. These are single-stranded oligonucleotides that base pair with a target RNA and can either lead to target degradation or alter target RNA structure and/or its ability to interact with other factors. Chemical modifications of ASOs make them highly stable and able to permeate cells, and considerable progress has been made in the improvement of their pharmacological properties, allowing development of effective therapeutics such as nusinersen for spinal muscular atrophy (*4*). However, the limited sequence conservation of lncRNAs between human and mouse poses a substantial challenge, because many human lncRNAs do not have recognizable mouse orthologs (*3*). For those that are conserved, it is often impossible to find an ASO sequence that will recognize both the human and the mouse sequences, which substantially complicates preclinical drug development.

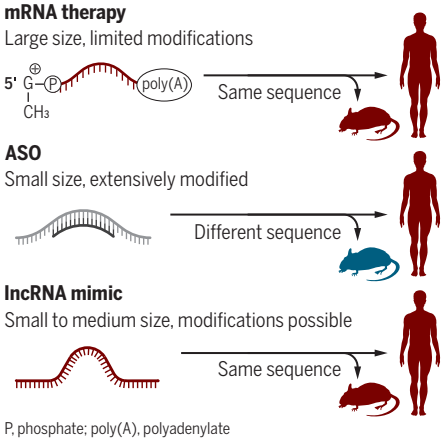
In other cases, increased lncRNA expression is sought, either because the lncRNA is mutated in a disease or because an increase in its concentration carries benefits. One conceptual challenge is that for lncRNAs that function in cis, exogenous delivery to the entire cell will likely not sufficiently increase their concentration at the target locus and may hence remain inconsequential. In any case, a major challenge is the delivery of a large RNA molecule. This can be potentially overcome by identifying and using a functionally active fragment of the full lncRNA. Such a functional element can be a region in the lncRNA molecule that is responsible for interacting with other factors, possibly resulting in changes to their abundance or activity.

For example, the lncRNA *Nron* (noncoding repressor of NFAT) was identified in mice as a critical suppressor of bone resorption, which is a pathological mechanism in osteoporosis (*5*). Delivery of full-length *Nron* using a bone-resorption surface-targeting nucleic acid delivery system inhibits bone resorption but causes side effects in mice, including splenomegaly, probably because of a strong immune response to the delivered RNA. However, the delivery of just the conserved functional motif of *Nron*, which binds the E3 ubiquitin ligase cullin-4B, effectively reversed bone loss in mice without any obvious side effects, indicating its potential translational use in osteoporosis (*5*).

Li *et al.* developed a therapeutic strategy based on the activity of the *HULC* (hepatocellular carcinoma up-regulated long non-coding RNA) lncRNA which, as they demonstrate, increases the activity of phenylalanine hydroxylase (PAH), which is mutated in PKU. They used lncRNA mimics containing a short fragment of *HULC* sequence that is tagged with an *N*-acetylgalactosamine (GalNAc) moiety that facilitates delivery to hepatocytes. Two different lncRNAs, *Pair* and *HULC*, perform this function in mouse and human liver, respectively, yet both were able to function equivalently in cells from both species, and the mimics of the functional region in human *HULC* were effective in vivo at improving PAH function in the mouse liver, with-

RNA-focused therapeutics

RNA-focused therapeutics—such as messenger RNAs (mRNAs), antisense oligonucleotides (ASOs), and long noncoding RNA (lncRNA) mimics—differ in size and whether chemical modifications can be introduced. Their function in humans and mice also varies, which affects preclinical development.



P, phosphate; poly(A), polyadenylate

GRAPHIC: A. MASTIN/*SCIENCE*

[1]Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel. [2]Department of Molecular Neuroscience, Weizmann Institute of Science, Rehovot 76100, Israel. Email: igor.ulitsky@weizmann.ac.il

out any detectable adverse effects on liver or kidney function.

The use of mimics of lncRNA functional motifs to treat human disease has several advantages compared with other approaches (see the figure). In contrast to therapeutic mRNAs, which need to be translated by ribosomes, and similarly to ASOs, lncRNA mimics can be extensively modified, which can facilitate high in vivo stability and decrease immunogenicity. They can also be easily tagged with organ-targeting peptides for tissue-specific distribution. Functional RNA motifs often do not have strict sequence requirements, which allows flexibility in designing lncRNA mimics and minimizing undesired activities, such as triggering antiviral pathways that recognize different RNA modalities. Because endogenous lncRNA activities are often tissue specific, there is, in principle, a relatively low potential for toxicity. Lastly, as exemplified by Li *et al.*, functional elements can have conserved functions even if their sequences are entirely different, and so the same element can be equivalently active in humans and mice, overcoming a major challenge for ASOs.

"...short [long noncoding RNA] fragments could overcome some of the challenges faced by other RNA therapeutic modalities."

Several hurdles still need to be overcome before lncRNAs or fragments thereof realize their full therapeutic potential. Perhaps most important is the need for advances in the methods to deliver RNA molecules to specific tissues and cell types (as nanoparticles or through other vehicles), which will also benefit therapeutic mRNAs and ASOs (*6*). The repertoire of lncRNAs whose biology is properly understood and linked to specific pathological states also needs to be expanded. Lastly, for as long as the delivery of full-length lncRNAs remains a challenge, new approaches will be needed in computational and/or experimental identification of lncRNA functional domains and of minimal backbones that will facilitate stability and desired subcellular localization. ■

REFERENCES AND NOTES

1. Y. Li *et al.*, *Science* **373**, 662 (2021).
2. M. K. Iyer *et al.*, *Nat. Genet.* **47**, 199 (2015).
3. I. Ulitsky, *Nat. Rev. Genet.* **17**, 601 (2016).
4. X. Shen, D. R. Corey, *Nucleic Acids Res.* **46**, 1584 (2018).
5. F. Jin *et al.*, *Nat. Commun.* **12**, 3319 (2021).
6. M. D. Buschmann *et al.*, *Vaccines* **9**, 65 (2021).

10.1126/science.abj7969

NEURODEGENERATION

Treatments for Alzheimer's disease emerge

Anti-amyloid immunotherapies will provide the first disease-modifying therapeutics

By **Dennis J. Selkoe**

Few of life's experiences evoke greater apprehension than a diagnosis of Alzheimer's disease (AD). Virtually unknown to the public until the 1980s, it is alone among the 10 most common fatal diseases of developed nations in lacking a disease-modifying treatment. AD affects people of all ethnicities; in the United States, African Americans have twice the prevalence of European Americans (*1*). The cumulative financial cost to society of late-life dementias (of which AD comprises ~60%) is estimated to exceed those of heart disease and cancer (*2*). This dismal reality may now be changing. The properties of the key proteins comprising the amyloid plaques [amyloid-β (Aβ)] and neurofibrillary tangles (tau) that define the neuropathology of AD have been identified. Coupled with extensive genetic studies, a sequence of lesion formation in brain networks serving memory and cognition is suggested. Antibodies that target these proteins are in advanced trials, and aducamumab, which clears Aβ, was recently approved, though not without controversy.

Through longitudinal analyses of humans with rare, causative mutations in APP (the Aβ precursor protein) and presenilin (the catalytic subunit of γ-secretase, which cleaves APP to generate Aβ), it has become clear that biochemical alterations in the brain begin at least two decades before cognitive symptoms develop. During this long presymptomatic interval, extracellular accumulation of the self-aggregating Aβ42 peptide into initially soluble oligomers and then increasingly large polymers and insoluble fibrils is accompanied by binding of the oligomers to the plasma membranes of microglia, astrocytes, and myriad neurites and synapses (see the figure). Although this amyloid hypothesis of AD is often drawn linearly for simplicity (*3*), many of the changes likely arise in temporal proximity (*4*).

Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. Email: dselkoe@bwh.harvard.edu

Genome-wide association studies in typical late-onset AD (i.e., after age 65) have converged on risk alleles in diverse genes mediating cholesterol and lipid regulation, synaptic network functions, and especially microgliosis (altered microglia) and neuroinflammation. The most potent genetic risk factor is the apolipoprotein E (*APOE*) ϵ4 variant: Heterozygosity raises AD risk 2- to 5-fold, and homozygosity increases it >5- to 10-fold. Its pathogenic mechanism appears to involve decreased glial-mediated clearance of Aβ from the brain's extracellular space, leading to more amyloid in cerebral plaques and microvessels (*5*). In mice, the APOE4 protein can also promote tau-mediated neurodegeneration and glial activation, both in the presence and absence of amyloid (*6*). Some other AD genetic risk factors have likewise been linked to enhanced Aβ deposition and/or the macrophage and microglial reaction to it.

Two decades ago, theories about AD pathogenesis seemed divided over the primacy of amyloid versus tau deposition. This false dichotomy has been supplanted by a growing consensus that Aβ aggregation in the brain [indicated by declines in soluble Aβ monomers in cerebrospinal fluid (CSF) and accrual of insoluble plaques seen on amyloid-PET (positron emission tomography) scans] begins early in people destined to develop AD and is followed by glia-mediated inflammation and the accumulation and spread of tau tangles in brain regions that serve cognition (*7, 8*). Rising amounts of extracellular Aβ lead to aggregates, including soluble oligomers, that appear to enhance the accrual of tau tangles and altered neurites beyond the medial temporal lobe, where these lesions are often present in older people without AD. Such tau accumulation and spread in the brain, perhaps via neuron-to-neuron connections, seems necessary for the development of cognitive symptoms in AD (*9*). In APP transgenic mice, deletion of the gene that encodes tau does not alter amyloid plaques but significantly lessens their behavioral consequences. Thus, Aβ oligomerization appears to initiate AD neuropathology, leading to altered tau in neurites and cell bodies as

well as microgliosis and blood monocyte infiltration into the brain.

The failure to reach primary and secondary outcomes in numerous trials of potentially AD-modifying agents may be explained in one or more ways: failure of the agent to achieve robust and selective target engagement in the brain; initiating treatment at a clinical stage that is too advanced to be effective; underpowered trials; adverse side effects on cognition; and faulty trial execution. The precise reasons differ among the unsuccessful trials to date. But a few recent trials appear to have met their primary endpoints or come close to them and have also achieved some secondary endpoints.

The clearest evidence of disease modification so far has come from secondary biomarker endpoints, principally a substantial decrease in amyloid plaques over 18 months, as measured by amyloid-PET. For example, this occurred in the two phase 3 trials of aducanumab, an Aβ monoclonal antibody that was approved by the US Food and Drug Administration (FDA) on 7 June 2021. Additional biomarker changes included a decrease in the elevated CSF concentration of phosphorylated tau protein and a reduction of brain tau-PET signal, but these outcomes were only measured in a small minority of aducanumab recipients. Although the marked decrease in amyloid deposits can be viewed as biological evidence of disease modification, this was accompanied by a decidedly mixed outcome on cognitive testing, with one aducanumab trial (EMERGE, NCT02484547) meeting its prespecified primary and secondary endpoints at the highest dose, whereas the other (ENGAGE, NCT02477800) did not achieve them. Although differences in cumulative dosing and uneven trial execution have been offered as explanations for this discrepancy, an FDA advisory committee was unconvinced and voted against approval. Nonetheless, the FDA granted an "accelerated approval," citing robust amyloid lowering across both trials and an expectation that this should lead to less cognitive decline. It also required that a confirmatory trial be performed while marketing commences.

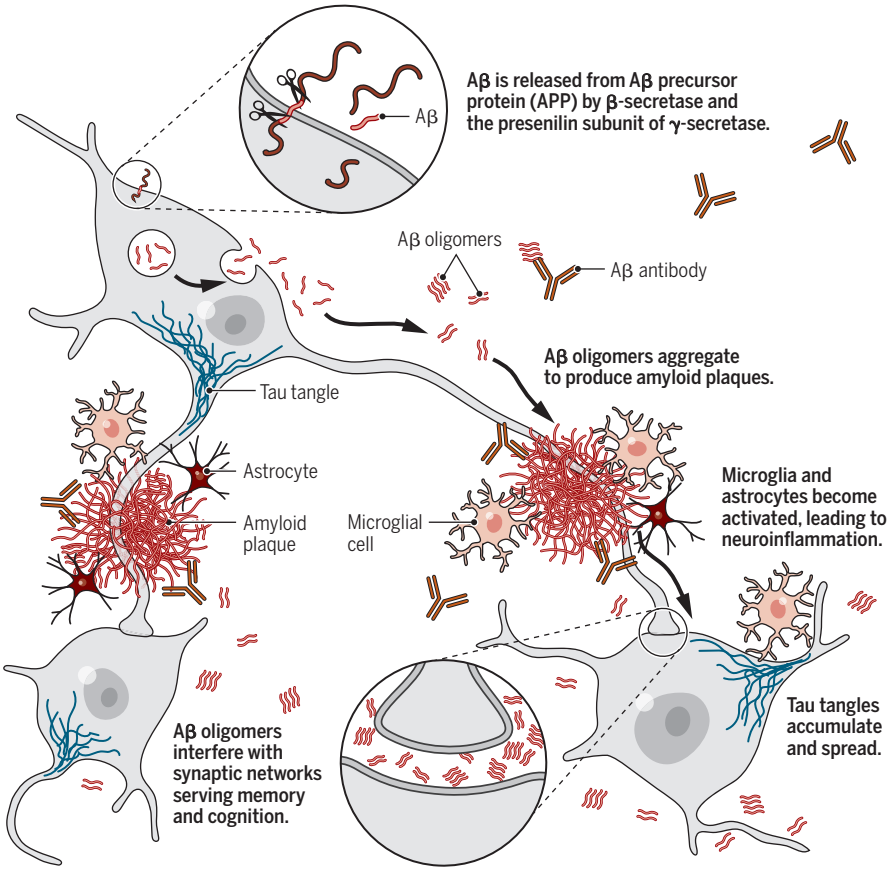
The controversy over aducanumab should be considered in the context of other recent AD immunotherapy trials. A large phase 2 trial of the monoclonal antibody lecanemab, designed to bind and clear Aβ protofibrils and oligomers, achieved its primary and secondary endpoints, including substantial amyloid plaque lowering and significantly less cognitive decline (*10*), and has advanced to phase 3 (NCT03887455). Another Aβ monoclonal antibody, gantenerumab, produced amyloid plaque reductions in phase 2 with less cognitive decline (*11*) and is in phase 3 (NCT03443973). Moreover, the antibody donanemab, which targets a low-abundance but aggregation-prone variant of Aβ with a modified amino terminus containing pyroglutamate-3, was recently shown in a moderate-sized phase 2 trial to markedly lower amyloid burden, accompanied by significant slowing of decline in psychometric tests and daily activities (*12*). Notably, donanemab conferred its cognitive effects in patients with relatively low tau burdens at trial entry (as judged by tau-PET), not in those with higher tau concentrations. Stratifying patients by tau burden was wise and could be used in future anti-Aβ trials. Unfortunately, however, both tau-PET and amyloid-PET only quantify fibrillar deposits, not soluble oligomers that appear to be responsible for neurotoxicity.

These four antibodies against Aβ unambiguously clear amyloid deposits from brain regions that are important for cognition, and this effect is accompanied by a variable 20 to 40% slowing of cognitive decline in 18-month trials. Collectively, these data represent the closest the AD field has come to a disease-modifying approach. So far, cognitive benefits are modest, and the challenge of assessing their clinical meaningfulness for patients and caregivers remains. But this challenge has been experienced in other chronic diseases, e.g., the controversy over the initial limited benefits of the antiretroviral drug zidovudine for HIV and AIDS when it was first approved (*13*).

Disease-modifying agents for AD are expected to slow cognitive decline more effectively the longer—and earlier—they are given. Indeed, treating amyloid-positive individuals in the presymptomatic period is more likely to be efficacious. In mouse models of AD, early treatment with aducanumab reduced Aβ deposition and downstream neuropathology later in life. Gaining real-world experience with a first, albeit modest, treatment should encourage development of more potent second-generation agents.

Drug targets for Alzheimer's disease

Alzheimer's disease neuropathology includes extracellular amyloid plaques containing myriad amyloid- β (Aβ) oligomers and intraneuronal tangles containing phosphorylated tau. Microglia and astrocytes become activated, leading to neuroinflammation and the spread of neuropathology. Antibodies to Aβ, administered intravascularly, can clear amyloid plaques.



GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

Overnight, managing an untreatable, ultimately fatal disease has been converted into the complex challenge of offering treatment plans to myriad AD patients. Surprisingly, the indication on the aducanumab label initially read "Alzheimer's disease," but after facing criticism, the FDA soon changed that to mild cognitive impairment and mild AD, mirroring the entry criteria for the phase 3 trials. AD clinicians will likely also require evidence of amyloid pathology. The latter can be established through amyloid-PET imaging, but this is not widely accessible, so CSF profiling will be relied upon to document the characteristic decrease in Aβ42 monomers and increase in phospho-tau that has long been used to confirm AD.

A special challenge to clinicians will be considering amyloid-positive patients who are more impaired than those in the trials for treatment. AD practices offering aducanumab should establish transparent guidelines for patient eligibility, hopefully with limited variation among sites. The drug label specifies dose and infusion intervals, but criteria for how long to treat patients will evolve as any slowing of cognitive decline becomes apparent. The practical challenges of an infusible therapeutic will lead to subcutaneous formulations that can be administered at home. Parenthetically, the slower-release subcutaneous route may lessen the occurrence of the key adverse effect of antibodies against Aβ: focal cerebral edema (ARIA-E), which is self-limited and asymptomatic in three-quarters of those who develop it and may be a sign of amyloid clearance or an inflammatory response at local vessels. Occasional microhemorrhages (ARIA-H) developed in a minority of those aducanumab recipients who had ARIA-E, and these appeared to be asymptomatic. The initial price of aducanumab (~$56,000/year) is very high and will need to be covered by insurance or national health care providers. These and other challenges in the march to implement the first approved AD therapeutic require thoughtful planning and resourcefulness, but this is just the process that patients and caregivers have long awaited.

"Disease-modifying agents for [Alzheimer's disease] are expected to slow cognitive decline the longer—and earlier—they are given."

A key advance has been the emergence of blood tests that can detect AD neuropathology. Plasma assays for certain fragments (*14*) and phospho-epitopes (*15*) of tau appear particularly promising, because tau alteration follows Aβ accumulation in those who develop AD symptoms. Comparison of various tau and Aβ plasma assays for their sensitivity in diagnosing AD and monitoring progression is needed. Accelerating the development of plasma biomarkers is critical to meet the challenge of screening innumerable patients globally for eligibility for AD-modifying agents.

Additional therapeutic approaches are crucial. Among small-molecule approaches, β-secretase inhibitors have been thwarted by mechanism-based side effects, although lower doses are being considered. An understudied class is the γ-secretase modulators that allosterically alter the conformation of presenilin and thereby shift APP processing from longer, amyloidogenic forms (Aβ42, Aβ43) to shorter, anti-amyloidogenic forms (Aβ37, Aβ38). Beyond Aβ, effort is focused on slowing tau accumulation, e.g., by immunotherapy or antisense oligonucleotides. Modulating the pathological responses of macrophages and microglia is of great interest, given the strong genetic evidence for their involvement in AD. Nonpharmacological approaches toward preventing AD must also be pursued, including exercise, sleep hygiene, a Mediterranean diet, and intellectual and social enrichment.

For many chronic diseases, the initial therapeutic compounds have limited efficacy and are often steadily replaced by more effective drugs. The emerging immunotherapeutics slow the AD biological process but confer modest clinical benefit. The approval of aducanumab may provide a proof of concept that can be rapidly improved upon. It may also enable combination treatments, as is typical in chronic diseases. In therapeutics, as in life, one must walk before one can run. ■

REFERENCES AND NOTES

1. K. B. Rajan, J. Weuve, L. L. Barnes, R. S. Wilson, D. A. Evans, *Alzheimers Dement*. **15**, 1 (2019).
2. M. D. Hurd, P. Martorell, A. Delavande, K. J. Mullen, K. M. Langa, *N. Engl. J. Med*. **368**, 1326 (2013).
3. D. J. Selkoe, J. Hardy, *EMBO Mol. Med*. **8**, 595 (2016).
4. B. De Strooper, E. Karran, *Cell* **164**, 603 (2016).
5. J. M. Castellano *et al*., *Sci. Transl. Med*. **3**, 89ra57 (2011).
6. Y. Shi *et al*., *Nature* **549**, 523 (2017).
7. R. J. Bateman *et al*., *N. Engl. J. Med*. **367**, 795 (2012).
8. C. R. Jack Jr. *et al*., *Lancet Neurol*. **12**, 207 (2013).
9. J. S. Sanchez *et al*., *Sci. Transl. Med*. **13**, eabc0655 (2021).
10. C. J. Swanson *et al*., *Alzheimers Res. Ther*. **13**, 80 (2021).
11. R. Doody, *J. Prev. Alzheimers Dis*. **4**, 264 (2017).
12. M. Mintun *et al*., *N. Engl. J. Med*. **384**, 1691 (2021).
13. M. A. Fischl *et al*., *N. Engl. J. Med*. **317**, 185 (1987).
14. J. P. Chhatwal *et al*., *Nat. Commun*. **11**, 6024 (2020).
15. S. Janelidze *et al*., *Nat. Med*. **26**, 379 (2020).

ACKNOWLEDGMENTS

D.J.S. is a director of and consultant for Prothena Biosciences.

10.1126/science.abi6401

NEURODEGENERATION

Phenotyping Alzheimer's disease with blood tests

Affordable and accessible tools could be used for screening and therapy monitoring

By **Kaj Blennow**

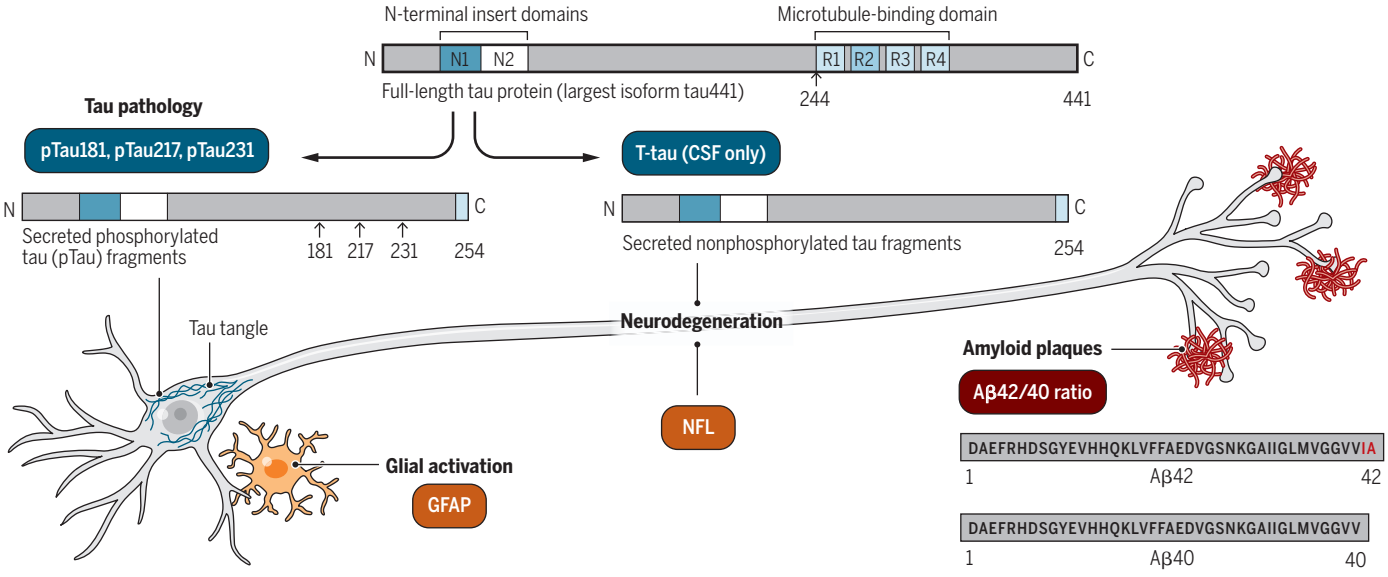
Alzheimer's disease (AD) is characterized by brain protein aggregates of amyloid-β (Aβ) and phosphorylated tau (pTau) that become plaques and tangles, and dystrophic neurites surrounding the plaques, which are accompanied by downstream neurodegeneration. These protein changes can be used as biomarkers detected through positron emission tomography (PET) imaging and in cerebrospinal fluid (CSF), allowing for ATN (amyloid, tau, and neurodegeneration) classification of patients. This phenotyping has become standard in AD clinical trials to overcome the high misclassification rate (20 to 30%) for clinical criteria and also enables enrollment of preclinical AD patients. The recent approval of the first disease-modifying anti-amyloid immunotherapy, aducanumab, for AD will generate a need for widely accessible and inexpensive biomarkers for ATN classification of patients with cognitive complaints. Technological advances have also overcome the challenges of measuring the extraordinarily low amounts of brain-derived proteins in blood samples, and recent studies indicate that AD blood tests may soon be possible.

The Aβ42 variant of Aβ is aggregation-prone and is deposited in plaques in the brains of people with AD, whereas the shorter Aβ40 isoform is by far the most abundant Aβ peptide (see the figure). Thus, as AD progresses and Aβ42 forms plaques, its concentration in the CSF and blood is reduced. Ascertaining the ratio of Aβ42 and Aβ40 concentrations in the CSF is known to adjust for between-in-

Clinical Neurochemistry Laboratory, Institute of Neuroscience and Physiology, the Sahlgrenska Academy at Gothenburg University, Mölndal, Sweden.
Email: kaj.blennow@neuro.gu.se

Biomarkers of Alzheimer's disease

Low amyloid-β (Aβ) 42/40 isoform ratio is associated with brain amyloidosis, and several phosphorylated tau (pTau) fragments increase with tau pathology; both are specific blood biomarkers for Alzheimer's disease (AD). Among neurodegeneration biomarkers, neurofilament light (NFL) is modestly increased in AD, and total tau (T-tau) is markedly increased only in cerebrospinal fluid (CSF), and not blood, in AD. Glial fibrillary acidic protein (GFAP) is a candidate blood biomarker for astrocytic activation, to indicate neuroinflammation.



dividual differences in "total" Aβ production, thereby increasing concordance with amyloid PET imaging to detect brain amyloidosis. Applying the same principle for blood plasma Aβ, immunoprecipitation–mass spectrometry (IP-MS) measures of plasma Aβ42/Aβ40 ratio can reach an accuracy exceeding 90% to identify brain amyloidosis (*1*). A population-based study of 441 asymptomatic elderly individuals indicates that IP-MS plasma Aβ can identify those who are amyloid PET-positive with high accuracy (*2*).

The question then arises whether plasma Aβ detection can replace PET or CSF tests for brain amyloidosis. A potential issue is that Aβ is produced not only in the brain but also in platelets and peripheral tissues, which will obscure the central nervous system–derived Aβ signal in plasma. Consequently, in amyloid PET-positive cases, plasma Aβ42/Aβ40 ratio is only ~10% lower than in individuals without brain amyloidosis, whereas it is more than 40% lower in CSF (*3*). This leads to an overlap that introduces challenges to robustly classify individuals as being either amyloid positive or negative, especially in those with Aβ42/Aβ40 ratios close to the cut-off for normality. Algorithms combining plasma Aβ42/Aβ40 ratio with the ϵ4 variant of apolipoprotein E (*APOE*), which is the major AD risk gene, and age (the main risk factor for AD) increase accuracy in detecting brain amyloidosis by 2 to 6% (*2*, *3*). However, merging biomarker data with genetic risk and aging may cause confusion because some younger *APOE*-ϵ4 noncarriers with low plasma Aβ42/Aβ40 will be misclassified as amyloid negative by the algorithm, whereas a proportion of older individuals with homozygous *APOE*-ϵ4 but normal plasma Aβ42/Aβ40 will be wrongly classified as amyloid positive.

Tau protein is truncated into amino-terminal to mid-domain fragments before being secreted in blood plasma and CSF (*4*). CSF pTau has long been used as an AD-specific biomarker. A major breakthrough is the use of new ultrasensitive methods that allow for quantification of pTau in blood plasma, with high concentrations occurring in AD (*5*). Of 321 patients and controls, high plasma concentrations of pTau181 fragments were associated with brain tau pathology as measured by PET (*6*). Similar results were subsequently presented for other pTau species, including pTau217 (*7*) and pTau231 (*8*).

The findings of very high accuracy of plasma pTau217 in the ability to discriminate AD from other neurodegenerative disorders (*7*) and IP-MS data showing a higher magnitude of increase and better association with amyloid plaques by PET of plasma pTau217 than of pTau181 (*4*) suggest that there may be diagnostic or pathophysiological differences between pTau species, but this remains a matter of debate. Nonetheless, these pTau blood biomarkers all show high concordance with AD pathology at autopsy, with accuracies in differentiating AD from non-AD dementia cases up to 99% for pTau231 (*8*).

However, these studies are based on different analytical methods and cohorts. In an attempt to directly compare these pTau species, a study of 381 participants employing digital immunoassays for pTau181, pTau217, and pTau231 found strong correlations with amounts of pTau species in CSF. Moreover, although the fold change was highest for pTau217, the accuracy in identifying amyloid PET positivity was very high for all pTau species (*9*), suggesting that differences are not meaningful. A study of two large cohorts of 883 individuals with cognitive symptoms also showed high accuracy (90 to 91%) of both plasma pTau181 and pTau217 to predict clinical progression to AD dementia in algorithms that include memory and executive function tests and *APOE* genotyping (*10*). Overall, plasma pTau biomarkers fulfill many requirements for a clinically useful AD test, with a high fold change in AD (between two to four times higher in AD than non-AD controls across studies), and an increase early in the AD continuum (even preclinically), an association with amyloid-associated tau pathophysiology and tangle burden in the brain, and an increase specifically found in AD but not in other types of dementia.

The findings of an early increase in plasma pTau fragments in patients with evidence of amyloid plaques, but not tau abnormalities, by PET imaging may be interpreted as a neuronal response to Aβ aggregates that gives rise to increased pTau secretion into CSF and blood plasma. However, findings in biomarker studies are only associations and may not directly reveal causal relationships. For example, plasma pTau231 shows a 10- to 15-fold increase within 24 hours after acute traumatic brain injury, especially evident in younger patients (who are unlikely to have amyloid or tau pathology) (*11*).

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

Total tau (T-tau), referring to any tau variant or fragment regardless of phosphorylation, and other brain proteins such as glial fibrillary acidic protein (GFAP) also increase in blood plasma, hypothetically mediated by a trauma-induced compromise of the blood-brain barrier, with release of proteins pre-existing in the extracellular space. Even if different mechanisms operate in specific disorders, further research is needed to understand the mechanisms underlying the increase in plasma pTau in AD.

In the search for blood biomarkers of neurodegeneration, it has become evident that in contrast to CSF, where T-tau is markedly increased in AD, T-tau does not work as a biomarker of AD neurodegeneration in blood. Instead, another axonal protein, neurofilament light (NFL), has been evaluated as a substitute AD neurodegeneration biomarker, even though it is not involved in AD pathogenesis. Plasma NFL concentrations correlate well with CSF concentrations, supporting that it reflects brain pathophysiology. But high amounts are found in a wide variety of neurodegenerative disorders, so this biomarker lacks specificity. Nevertheless, plasma NFL, which shows a modest increase in AD, predicts both cognitive deterioration and rate of neurodegeneration as measured by atrophy on brain imaging. Notably, both plasma and CSF NFL concentrations increase in cognitively unimpaired people with autosomal dominant AD 7 years before symptom onset (*12*), so this may be a good biomarker for predicting AD.

Another candidate AD blood biomarker includes the astrocytic protein GFAP, which is markedly increased in AD. Plasma GFAP distinguishes amyloid PET-positive and -negative cognitively normal elderly with high accuracy (*13*), and may serve as a blood biomarker for glial activation and neuroinflammation.

Despite both rapid and robust reductions in amyloid PET ligand binding after treatment with Aβ immunotherapies (indicative of drug target engagement), effects on cognitive outcomes have been less evident. Therefore, biomarker evidence for downstream effects on reducing tau pathology and neurodegeneration is important to support disease-modifying effects by this class of drugs. Given that in most clinical trials only a small percentage of enrolled patients undergo repeat lumbar puncture for CSF testing, blood biomarkers could play an important role to accomplish this. Data from other areas of clinical neuroscience show that children with spinal muscular atrophy have a marked increase in CSF NFL, but treatment with the antisense oligonucleotide drug nusinersen results in a successive reduction of NFL concentrations in CSF with normalization after ~7 months, and the reduction correlates with clinical improvements (*14*). Similar, but less pronounced, reductions of plasma NFL are seen with disease-modifying treatments in multiple sclerosis patients. These findings may serve as proof of concept for the usefulness of plasma NFL in identifying downstream drug effects on neurodegeneration. Target engagement for the anti-Aβ drug, aducanumab, was demonstrated in 2017, with dose-dependent reductions on amyloid PET (*15*), but to date there are no reports of effects on blood biomarkers of neurodegeneration (or tau pathology) from any Aβ immunotherapy trial.

Current studies of blood AD biomarkers come exclusively from cohorts at highly specialized research centers. Thus, further clinical validation is needed, specifically on the diagnostic accuracy of the AD blood biomarkers, alone or in combination, in consecutive patient populations at memory clinics and in primary care settings. In addition, because plasma pTau increases progressively with tau pathology in the brain and more advanced clinical stage, more data are needed on the accuracy of plasma pTau biomarkers to identify individuals with preclinical or early symptoms who will go on to develop AD. Moreover, studies comparing plasma pTau species in the same cohorts and using the same technology are needed to understand if there are pathophysiological differences across the pTau epitopes. Current assays are research grade, and full analytical validation of methods is needed to achieve accurate and comparable results between laboratories, as well as global efforts to develop certified reference materials to achieve harmonization across assay platforms. Transferring the blood tests to fully automated platforms would also help to streamline these procedures and to establish these blood tests as clinically useful tools. Lastly, to make blood biomarkers attractive substitutes for imaging, costs need to be substantially lower than costs for the PET scans.

REFERENCES AND NOTES

1. A. Nakamura *et al.*, *Nature* **554**, 249 (2018).
2. A. Keshavan *et al.*, *Brain* **144**, 434 (2021).
3. S. E. Schindler *et al.*, *Neurology* **93**, 17 (2019).
4. N. R. Barthélemy, K. Horie, C. Sato, R. J. Bateman, *J. Exp. Med.* **217**, e20200861 (2020).
5. M. M. Mielke *et al.*, *Alzheimers Dement.* **14**, 989 (2018).
6. T. K. Karikari *et al.*, *Lancet Neurol.* **19**, 422 (2020).
7. S. Palmqvist *et al.*, *JAMA* **324**, 772 (2020).
8. N. J. Ashton *et al.*, *Acta Neuropathol.* **141**, 709 (2021).
9. M. Suárez-Calvet *et al.*, *EMBO Mol. Med.* **12**, e12921 (2020).
10. S. Palmqvist *et al.*, *Nat. Med.* **27**, 1034 (2021).
11. R. Rubenstein *et al.*, *JAMA Neurol.* **74**, 1063 (2017).
12. O. Preische *et al.*, *Nat. Med.* **25**, 277 (2019).
13. P. Chatterjee *et al.*, *Transl. Psychiatry* **11**, 27 (2021).
14. B. Olsson *et al.*, *J. Neurol.* **266**, 2129 (2019).
15. J. Sevigny *et al.*, *Nature* **546**, 564 (2017).

ACKNOWLEDGMENTS

K.B. has consulted for Axon, Biogen, Lilly, and Roche Diagnostics and is cofounder of Brain Biomarker Solutions in Gothenburg AB.

10.1126/science.abi5208

NANOFLUIDICS

Bioinspired nanofluidic iontronics

Electrolytes in planar nanochannels are predicted to function as nanofluidic memristors

By **Yaqi Hou**[1] *and* **Xu Hou**[1,2]

In digital computing, functions such as processing and memory require separate components wired together for electronic conduction. Neurons, the functional units equivalent to processor and memory areas, are integrated in the brain and transmit signals through ionic and neurotransmitter conduction. Inspired by the energy-efficient computation architectures from biological systems, on page 687 of this issue, Robin *et al.* (*1*) used theory and simulations to predict that two-dimensional (2D) nanofluidic channels can show nonlinear conduction and function as memory-effect transistors. By incorporating two nanofluidic memristors in an elementary circuit that refers to Hodgkin and Huxley's model (*2*), the neuromorphic responses of emitting voltage spikes were reproduced in simulations of experimental devices.

Although digital computing can execute artificial intelligence (AI) tasks, the very large number of electronic components needed to process information and transmit data leads to intensive energy consumption. A shift in computation from calculation and storage to pattern recognition and other AI tasks also drives the need for more energy-efficient architectures. In conventional electronics, state switching in semiconductors is executed by changing the population of charged electrons and holes, and their drifting and filling behaviors are invariant in time and space once the wires in circuits are connected. In biological systems, ions are the charge carriers. The action potential across cell membranes to transmit data based on the time- and voltage-variant ionic conductivity modulation of ion

[1]State Key Laboratory of Physical Chemistry of Solid Surfaces, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China. [2]Research Institute for Soft Matter and Biomimetics, College of Physical Science and Technology, Xiamen University, Xiamen 361005, China. Email: houx@xmu.edu.cn

channels leads to coordination and precise timing of physiological outputs. These outputs range from elementary muscle contractions to high-level mental activities.

Despite the dynamic interactions between electrons and ions being a source of various electronic and ionic functionalities in biology, chemistry, and physics, electronics and ionics have developed with different priorities (*3*). In electronics, one goal is to reduce component size of devices in integrated circuits. In ionics, the focus is on realizing sophisticated control of dissolved ions through varieties of materials and structures. The aim is to identify characteristic ionic signals that could interface with and control biological processes in the complex aqueous environment.

Iontronics has emerged as a tool for signal processing that combines electronic properties with ionic conductivities. In some of the first transistors, electrolytes were used to control the current flow of semiconductors. The analogous description of the semiconductors and electrolyte solutions led to the development of charge-selective structures that acted as p- or n-type semiconductors, which led to the development of current rectifiers in aqueous solution in the 1950s (*4*). Since then, various materials and structures have been explored to show the diode-like rectification phenomena of electrolytic solutions. The transport of ions passing through nanometer-sized biological channels accounts for a wide array of physiological processes as well as the study of fluids under nanoscale confinement (*5*–*7*). The development of micro- and nanofabrication technologies in the semiconductor industries has provided incisive experimental toolkits for nanofluidics, as well as instruments that can be used for direct imaging and characterization.

Guided by the structural variety of ion channels, nanoconfinement with different geometries, both theoretically and experimentally, has resulted in several types of ionic and molecular transport that can be used in iontronics (*6*, *7*). For 1D nanoconfinement, asymmetric designs in geometries and inner-surface properties can reproduce diode-like ion rectification (*8*). Biological nanochannels are deformable and dynamic, and curvature-tunable carbon nanotubes have recently been obtained for regulating ionic rectifications in real time (*9*).

2D materials—such as graphene, hexagonal boron nitride, and molybdenum disulfide—provide a route to experimentally accessible 2D nanoconfinement (*10*). Robin *et al.* suggested that compared with 1D confinement, planar confinement expands translational degrees of freedom for ionic transport and would lead to greater ionic correlation times and potential memory effects. In their theoretical framework, the monolayer electrolyte confined between two graphite layers displays rather slow dynamics that allows for self-association of ions into clusters (ion pairs and polyelectrolyte chains of ion pairs) under an oscillating electric field. This association is stronger with divalent versus monovalent cations (Ca^{2+} versus Na^{+}) and decreases conductivity. The time needed to form clusters and to dissociate them into conductive free ions in response to voltage changes results in the nonlinear conduction that forms the basis for memristive effects. Thus, in the memristor voltage loops, applying the voltage results in the more linear curves closer to Ohm's law, during which free ions self-associate into clusters. Upon returning to zero voltage, a bigger drop in current is observed because the low conductivity clusters need time to break apart. These 2D-channel–based devices can be used in circuits that generate voltage spike trains analogous to those generated by biological neurons (see the figure).

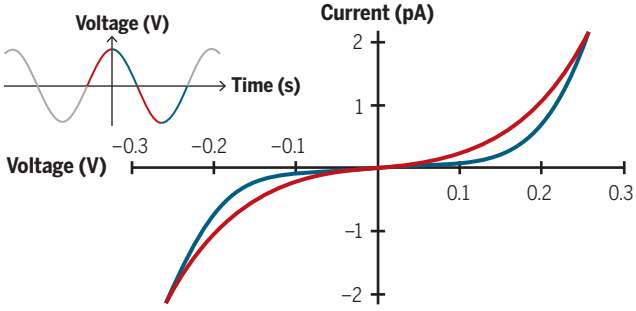
Progress in ion-based detection signal processing based on iontronics could have implications for interfacing devices with neural systems. Such devices could have compatible signals with neurons, which could enable lower power operation, and would be compatible with aqueous physiological environments. The theoretical work of Robin *et al.* should help in the development of wearable or implantable iontronic devices or even neuronal-computer interfaces. ■

Ionic voltage spike trains

Robin *et al.* reproduced the voltage spike trains of the Hodgkin-Huxley neuron model in a simulation of two-dimensional nanofluidic circuits.

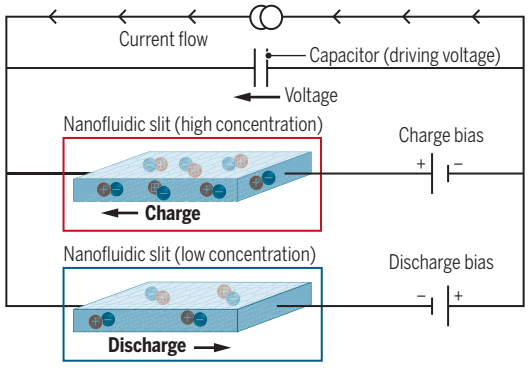
Nanofluidic memristors

Aqueous ions between graphene sheets show memristive responses. The resistance rises upon applying voltage (red) and decreases upon returning to zero voltage (blue). In this model, the voltage U is driven over time t as $U_0\cos(2\pi ft)$, where $U_0 = 0.26$ V and frequency $f = 160$ Hz.



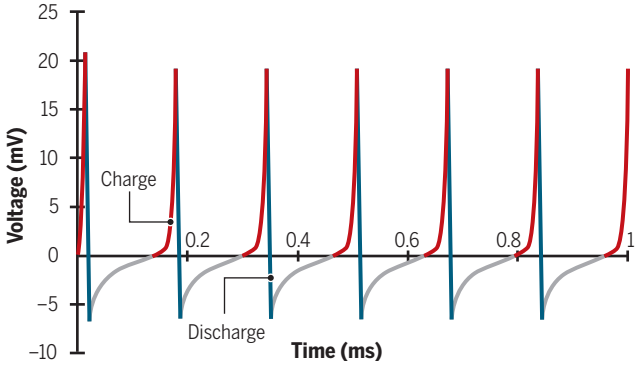
Modeling a neuronal circuit

The simulated nanofluidic device recapitulates the neuronal Hodgkin-Huxley model of axonal voltage spikes.



Spike trains

The charging memristor drives the main rise of the positive voltage spikes (red), and the discharging memristor drives the recovery (blue).



GRAPHIC: K. FRANKLIN/*SCIENCE*

REFERENCES AND NOTES

1. P. Robin, N. Kavokine, L. Bocquet, *Science* **373**, 687 (2021).
2. A. L. Hodgkin, A. F. Huxley, *J. Physiol.* **117**, 500 (1952).
3. S. Z. Bisri, S. Shimizu, M. Nakano, Y. Iwasa, *Adv. Mater.* **29**, 1607054 (2017).
4. H. Chun, T. D. Chung, *Annu. Rev. Anal. Chem.* **8**, 441 (2015).
5. L. Bocquet, *Nat. Mater.* **19**, 254 (2020).
6. M. Wang, Y. Hou, L. Yu, X. Hou, *Nano Lett.* **20**, 6937 (2020).
7. J. Zhong *et al.*, *Acc. Chem. Res.* **53**, 347 (2020).
8. X. Hou, H. Zhang, L. Jiang, *Angew. Chem. Int. Ed.* **51**, 5296 (2012).
9. M. Wang *et al.*, *Adv. Mater.* **31**, 1805130 (2019).
10. A. Esfandiar *et al.*, *Science* **358**, 511 (2017).

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (project no. 2018YFA0209500) and the National Natural Science Foundation of China (52025132 and 21975209).

10.1126/science.abj0437



BOOKS *et al.*

Air pollution from dumps like this one outside Delhi directly harms poor waste pickers, argue the authors.

PHOTO: RAVI CHOUDHARY/*HINDUSTAN TIMES*/GETTY IMAGES

HEALTH AND MEDICINE

To move beyond the paradigms of Western medicine

A pair of authors aspire to decolonize medicine through liminality, fugitivity, and abolition

By **Chandra L. Ford**

In *Inflamed,* physician Rupa Marya and scholar Raj Patel use the concept of inflammation to probe the many ills currently plaguing humanity and the planet. The text moves uneasily between inflammation as a cause of illness, inflammation as a signal that something is wrong, and inflammation as a framework for examining European and American colonialism, which Marya and Patel consider the root cause of contemporary inequities. Drawing primarily on knowledge the authors have gathered about Indigenous cultures over the years, *Inflamed* presents evidence that alternatives to Western capitalist biomedical ways of relating and knowing are possible. Those alternatives, the authors maintain, are preferable because their foundations lie not in extraction and exploitation but in connection and healing.

The authors' focus on inflammation is timely, as a growing number of studies seek to understand the mechanisms by which it causes diseases. Their focus on social injustice—in particular, on incarceration, environmental harm, and racism—is also timely, as these issues became visible during the first year of the COVID-19 pandemic.



Inflamed: Deep Medicine and the Anatomy of Injustice
Rupa Marya and Raj Patel
Farrar, Straus and Giroux, 2021. 496 pp.

The book's simple title, its length, and its linear organization all suggest an encyclopedic textbook. As the authors explain, however, *Inflamed* is intended to be "a subversive political anatomical survey, taking individual bodily systems and showing how ancient and modern science connects that body to the web of life through histories and relationships of power."

Marya and Patel trace contemporary biomedical thinking to its origins in European and American empires. In doing so, they aptly reveal how this knowledge is socially constructed, exposing the power investments it serves and the racism, misogyny, and colonialism that remain embedded within it.

The authors' liberal use of metaphor, simile, analogy, and other literary strategies adds lyricism to the writing, but at times, these techniques oversimplify. For example, citing racially oriented genetics research, the authors assert, "Social oppression preconditions the bodies of Black people in the United States," a statement that lacks nuance and reinforces racial essentialism.

Throughout the book, Marya and Patel highlight interactions with Indigenous people. Chief Caleen Sisk, they note, "spoke to us from her home on the remaining Winnemem Wintu land, a 42-acre…trailer park in Shasta County." And Marya asks Gregg Castro, "a culture bearer for the Ramaytush Ohlone tribe," for "his blessing and permission to practice my medicine in his homelands."

The Indigenous persons they cite are the most authoritative experts on the topic, but Marya and Patel have chosen not to include any of them as coauthors. As Indigenous scholar Eve Tuck and her coauthor K. Wayne Yang have explained, such an approach enables non-Indigenous people to embrace Indigenous cultures without acknowledging their own role in colonial projects (*1*).

In urging a (re)turn to Indigenous knowledge, *Inflamed* sometimes reproduces the hierarchies it seeks to dismantle. For instance, Patel desires "to wear nation like language, speaking many and being able to move through them." However, this fluidity requires national boundaries (and those within them) to be fixed in order to traverse them.

Given the scope and complexity of *Inflamed*, the book's conclusion seems truncated and superficial. Citing eminent American studies scholars Fred Moten, Jack Halberstam, and Ruth Wilson Gilmore, the authors call for "direct action to evade the structures and logic that have come with liberal settler colonialism" and for the abolition of all systems of incarceration. However, the book provides little substantive guidance on how to do so and does not explain how those actions will block the routes to inflammation identified in the previous eight chapters. These considerations echo tensions articulated by founding critical race theorist Derrick Bell about the role of the lawyer in prioritizing the interests of client versus the interests of the movement for equity for which the client is representative (*2*).

Inflamed offers evidence that activists can use in their struggles against capitalist biomedicine, and it grants clinicians and scientists permission to exchange the perpetual drive toward self-advancement, ownership (of bodies and of knowledge), and profit for the opportunity to become engaged in deep medicine. To all readers, it offers creative, aspirational visions of individual, societal, and planetary healing. ■

The reviewer is at the Center for the Study of Racism, Social Justice and Health, Department of Community Health Sciences, UCLA Fielding School of Public Health, Los Angeles, CA 90095, USA. Email: clford@ucla.edu

REFERENCES AND NOTES

1. E. Tuck, K. W. Yang, *Decolonization: Indigeneity, Education & Society* **1**, 1 (2012).
2. D. A. Bell Jr., *Yale Law J.* **85**, 470 (1976).

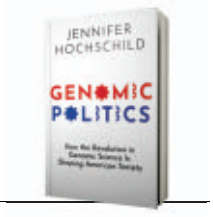
10.1126/science.abj6550

SCIENCE AND SOCIETY

Governing the genome

A political scientist proposes a framework for understanding genomic policy debates

Genomic Politics: How the Revolution in Genomic Science Is Shaping American Society
Jennifer L. Hochschild
Oxford University Press, 2021. 336 pp.



By **Jessica Blatt**

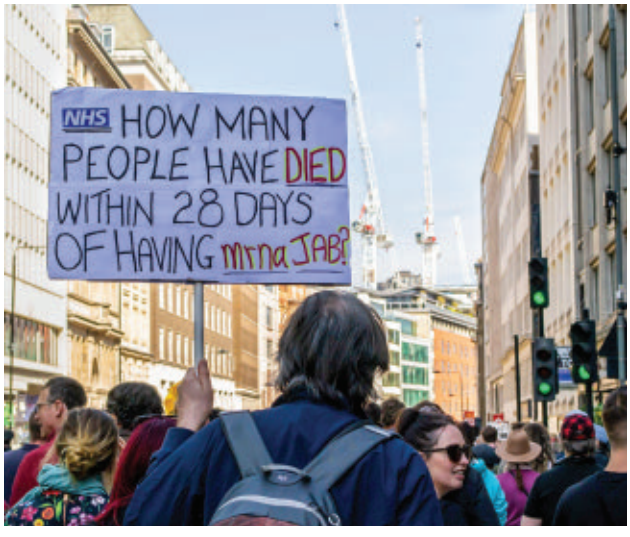
In the decades since the Human Genome Project was launched in 1990, genomic science has opened a dizzying array of scientific, medical, and commercial possibilities. Some of its applications have generated visible public controversy. Perhaps the most salient now is the furor over COVID-19 vaccinations, with vocal activists feeding vaccine hesitancy by insisting (mistakenly) that groundbreaking mRNA vaccines can change recipients' DNA.

Other controversies are more substantive, if often less well publicized. For example, abortion opponents and some disability rights activists have raised alarms that prenatal genetic testing could have eugenic effects, while critics of popular ancestry testing products and "racially specific medicine," such as BiDil, a heart medication marketed to African Americans, argue that these products commodify misleading ideas about the biological reality of race, with pernicious social effects and, in the case of pharmaceuticals, potentially dangerous consequences for patient health.

As Harvard University political scientist Jennifer Hochschild points out in her new book, *Genomic Politics: How the Revolution in Genomic Science Is Shaping American Society*, these controversies, and others like them, often make for strange bedfellows. Just as some conservative Trump supporters and liberal wellness devotees may come together over their opposition to vaccines, religious conservatives and queer theorists may both oppose the idea that homosexuality is inborn, for example.

The book's greatest strength is its series of concise, elegantly written overviews of a range of positions on these and other issues, including the use of DNA databases by law enforcement and various medical and research applications of gene editing technologies. The questions raised by these debates are fascinating, and Hochschild is adept at using participants' own words to bring out the complexity and nuances involved.

Beyond simply exploring the range of issues and attitudes raised by various applications of genomic science, Hochschild seeks to make sense of this welter of attitudes by sorting experts and survey respondents into four categories according to their attitudes about genomics: "enthusiasts," for whom genomics has both great explanatory power and technological promise; "skeptics," who see it as powerful but potentially dangerous; the "hopeful," who may see some enthusiasts' claims as overblown but still believe that genomics research can be beneficial; and "rejectors," who, for religious, ethical, or other reasons, disavow genomics research entirely. The idea is that this exercise will lend clarity to questions about whether, how much, and by whom genomic science ought to be regulated. In the end, however, the framework has limited explanatory power, and readers are left without a clear sense of how it might relate to ongoing governance debates.



False beliefs about mRNA vaccines threaten COVID-19 vaccination efforts.

Early in the book and again at the end, Hochschild poses the question "Is genomics a thing?" That is, does genomics present a coherent set of issues that are likely to invoke a more or less coherent set of responses? Her answer is yes, on the evidence that people's responses to questions about various aspects of genomics can be sorted into her proposed framework. But it seems more plausible that this is an artifact of the framework's generality and most people's unfamiliarity with the range of issues involved rather than some underlying consistency in attitudes toward genomics research. After all, one could legitimately hold privacy or racial justice concerns about law enforcement use of DNA databases yet wholeheartedly support certain medical applications of genomic research. Likewise, one might have deep religious objections to cloning or gene editing yet happily spring for a commercial ancestry testing kit without necessarily seeing those issues as connected in any way.

Hochschild's framework, moreover, reduces genomic politics to people's attitudes about genomics. But this is a rather thin conception of politics that sidesteps what, to many readers, will be the far more interesting political issues that the book vividly evokes, such as the political economy of health care or the power imbalances between research subjects and scientists.

In the book's last chapter, Hochschild comes out as a genomic "enthusiast." This will have been apparent all along to the careful reader, who will have noted her uncritical repetition of errors by fellow enthusiasts—assertions by geneticists who conflate the empirically valid concept of "population" with the dubious one of biological "race," for example—and her dismissive attitude toward some skeptics' concerns.

By her own account, however, Hochschild's enthusiasm waxes and wanes depending on the domain. This, too, would seem to suggest that the avenues of research and technological, cultural, and policy changes made possible by advances in genomic science are only becoming more complex and resistant to tidy organization, despite Hochschild's interesting and learned, but ultimately limited, attempt to do so. ■

The reviewer is at the Department of Politics and Human Rights, Marymount Manhattan College, New York, NY 10021, USA. Email: jblatt@mmm.edu

10.1126/science.abj5409

PHOTO: JESSICA GIRVAN/ALAMY STOCK PHOTO



US state policies put gray wolf populations at risk.

PHOTO: DENNIS FAST/VWPICS/UNIVERSAL IMAGES GROUP/GETTY IMAGES

Edited by **Jennifer Sills**

Restore protected status for gray wolves

In January, the US Fish and Wildlife Service terminated Endangered Species Act protections for gray wolves (*1*), with the caveat that a status review to determine whether relisting is warranted could be prompted "if a change in State law or management objectives would significantly increase the threat to the wolf population." Since the decision, Idaho and Montana have sanctioned the killing of 90% of their wolves (*2*). Those promoting these massive statewide hunts argue that culling wolves is necessary to protect the livelihoods of ranchers from the depredations of livestock. However, these policies put wolves, humans, and ecosystems at risk.

Counterintuitively, analyses of lethal wolf control programs indicate that killing wolves may disrupt wolf social structures, leading to more, not fewer, livestock deaths (*3*, *4*). In addition, wolves are responsible for at most 1 to 2% of unwanted livestock deaths (*5*). The economic cost of livestock losses attributed to wolves is far outweighed by the economic benefits that wolves provide by controlling deer populations. Smaller deer populations cause fewer deer-automobile collisions, saving human lives and preventing injuries (*6*). Supporters of lethal wolf control also claim that the approved quotas for hunting will not imperil the wolves because their reproduction can easily make up for the losses. However, the hunting limits approved in Montana and Idaho are more than four times the rates of human-caused mortality witnessed in recent decades (*1*) and are reminiscent of the historical massacres that extirpated wolves.

The delisting decision (*1*) was based on the observation that wolf population numbers had exceeded the numerical goals laid out in the 1987 recovery plan, specifically the target of more than 10 breeding pairs of wolves in each of three designated management areas. The problem is that these goals are far too low—something that has been noted as an issue not just for wolves but also as a broader problem affecting many species listed in the Endangered Species Act (*7*, *8*). In addition, the population target for wolves does not account for the added threat of climate change, the extent of which was not well understood when the recovery plan was written.

Wolves are a keystone predator (*9*, *10*). For keystone species (such as wolves, beavers, and sea otters) that engineer wild ecosystems, recovery goals should take into account not just the bare minimum survival of the species but also the restoration of the structuring role these species play in their ecosystems (*11*). Recovery goals that are set too low undermine the original intent of the Endangered Species Act. We urge the Department of the Interior to relist gray wolves and to examine recovery goals for all species in light of climate change and essential ecosystem functions.

P. Kareiva[1]*, J. A. Estes[2], M. Marvier[3]
[1]Aquarium of the Pacific, Long Beach, CA 90802, USA. [2]Ecology and Evolutionary Biology Department, University of California, Santa Cruz, Santa Cruz, CA 94060, USA. [3]Department of Environmental Studies and Sciences, Santa Clara University, Santa Clara, CA 95053, USA.
*Corresponding author. Email: pkareiva@lbaop.org

REFERENCES AND NOTES

1. Fish and Wildlife Service," Removing the gray wolf (Canis lupus) from the list of endangered and threatened wildlife, Rule effective January 4, 2021," *Federal Register* **85**, 69778 (2020).
2. M. Beck, "Mountain West lawmakers take aim at wolves," (Boise State Public Radio, 2021); www.boisestatepublicradio.org/environment/2021-04-22/mountain-west-lawmakers-take-aim-at-wolves.
3. R. B. Wielgus, K. A. Peebles, *PLOS One* **9**, e113505 (2014).
4. F. J. Santiago-Avila, A. M. Cornman, A. Treves, *PLOS One* **13**, e0189729 (2018).
5. The Humane Society of the United States, "Government data confirm that wolves have a negligible effect on U.S. cattle & sheep industries" (2019); www.humanesociety.org/sites/default/files/docs/HSUS-Wolf-Livestock-6.Mar_.19Final.pdf.
6. J. L. Raynor, C. A. Grainger, D. P. Parker, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023251118 (2021).
7. T. H. Tear, J. M. Scott, P. H. Hayward, B. Griffith, *Science* **262**, 976 (1993).
8. K. A. Pawluk, C. H. Fox, C. N. Service, E. H. Stredulinsky, H. M. Bryan, *PLOS One* **14**, e0224021 (2019).
9. T. D. Gable, S. M. Johnson-Bice, A. T. Homkes, S. K. Windels, J. K. Bump, *Sci. Adv.* **6**, eabc5439 (2020).
10. C. Eisenberg, *The Wolf's Tooth: Keystone Predators, Trophic Cascades, and Biodiversity* (Island Press, Washington, DC, 2013).
11. M. E. Soulé, J. A. Estes, J. Berger, C. M. Del Rio, *Conserv. Biol.* **17**, 1238 (2003).

10.1126/science.abk2278

Brazil can protect sharks worldwide

In the past few years, some countries have prohibited the removal of shark fins in an effort to protect shark populations (*1*). However, sharks remain at risk from this practice as long as other countries drive up demand by buying the finless shark carcasses for a cheap price (*2*–*4*). Brazil, the world's largest importer of shark meat (*3*, *4*), imports finless shark carcasses and steaks from countries that are involved in the fin trade, such as China and Spain, and from Uruguay, which exports processed shark meat (*3*, *4*). The growing shark meat trade in Brazil can be attributed in part to attractive prices, but the demand is complicated by mislabeled products; Brazilians are often unaware that they are eating sharks (*2*, *5*). Brazil has the power to disrupt the global shark market, but it will require policies that limit shark meat imports as well as an effort to provide consumers with accurate information.

In Brazil, although protected sharks cannot be legally marketed by local fishers or entrepreneurs, they can be imported without any restrictions. Moreover, it is mandatory to provide information for the proper labeling only if the imported frozen fish is in the Salmonidae family (which includes salmon and trout) or the Gadidae family (which includes cod and haddock) (*6*). Instead of being packaged with proper labeling, Brazil's shark meat is sold as unidentifiable carcasses or in pieces marked as *cação*, a deliberately ambiguous name used for multiple species (*7*). Despite a growing debate about shark mislabeling among nongovernmental organizations and academic communities (*8*), no government measures have been implemented. As a result, the consumers in Brazil remain unaware that they are purchasing shark meat and contributing to the decline of vulnerable shark species.

Sharks are facing irreversible population reductions, and action is urgently needed worldwide (*9*), especially in Brazil (*10*). As a first step, Brazil's government should widely disseminate the fact that *cação* may refer to shark meat. The country should require all domestic and imported products to be labeled with their scientific names throughout the supply chain, ensuring accurate monitoring of the species in the system and allowing consumers to decide whether to eat a species at risk of extinction. As a result of such changes, demand would likely decrease, limiting the market for sharks with illegally removed fins. Brazil could also protect sharks worldwide by prohibiting the importation of species on Brazil's National Red List (*11*). Because of Brazil's outsize role in global shark trade, these changes could vastly improve conservation efforts.

Bianca S. Rangel[1,2]*, Rodrigo Barreto[3], Nathalie Gil[2], Anna Del Mar[2], Carolina Castro[2]
[1]Departamento de Fisiologia, Instituto de Biociências, Universidade de São Paulo, Cidade Universitária, São Paulo, SP, Brazil. [2]Sea Shepherd Brasil, Porto Alegre, RS, Brazil. [3]Centro Nacional de Pesquisa e Conservação da Biodiversidade Marinha do Sudeste e Sul do Brasil, Instituto Chico Mendes de Conservação da Biodiversidade, Itajaí, Brazil.
*Corresponding author.
Email: biarangel.sharks@gmail.com

REFERENCES AND NOTES

1. L. Biery, D. Pauly, *J. Fish Biol.* **80**, 1643 (2012).
2. R. R. Barreto *et al.*, *Mar. Pol.* **85**, 114 (2017).
3. F. Dent, S. Clarke, "State of the global market for shark products" (FAO Fisheries and Aquaculture technical paper no. 590, Rome, 2015).
4. S. Niedermueller *et al.*, "The shark and ray meat network: A deep dive into a global affair 2021" (World Wildlife Fund Mediterranean Marine Initiative, Rome, 2021).
5. H. Bornatowski *et al.*, *Ethnobiol. Lett.* **6**, 196 (2015).
6. Government of Brazil, "Instrução normativa n. 53, de 1º de setembro de 2020" (Diário Oficial da República Federativa do Brasil, Brasília, DF, 2020); www.in.gov.br/en/web/dou/-/instrucao-normativa-n-53-de-1-de-setembro-de-2020-275906964 [in Portuguese].
7. H. Bornatowski *et al.*, *Science* **340**, 923 (2013).
8. M. Alvarenga *et al.*, *Biol. Conserv.* **257**, 109119 (2021).
9. N. Pacoureau *et al.*, *Nature*, **589**, 567 (2021).
10. R. R. Barreto *et al.*, *Conserv. Biol.* **30**, 792 (2016).
11. Chico Mendes Institute for Biodiversity Conservation, "Livro vermelho da fauna brasileira ameaçada de extinção: Volume VI–Peixes" (2018) [in Portuguese].

10.1126/science.abj9634

Call for protection of scatter-hoarding rodents

Mutualistic interactions between seed dispersers and tree species are important to both animal and plant populations. Scatter-hoarding rodents benefit plant species by carrying seeds to areas with more space for growth and increasing the number of seeds that germinate and grow. Forest fragmentation has compromised scatter-hoarding rodent habitats, causing population declines worldwide. In turn, disruption of dispersal mutualisms may predispose some trees to population decline or even local extinction (*1*). However, compared with large mammals, scatter-hoarding rodents have been largely overlooked in conservation plans.

In Central American neotropical forests, the ground-dwelling acouchies and agoutis that disperse and scatter-hoard large-seeded tropical endemic tree species, including the arara nut-tree (*Joannesia princeps*) and palms such as *Astrocaryum standleyanum* and *A. aculeatissimum*, are sensitive to and threatened by forest fragmentation (*2*). In North America, the largest kangaroo rat, *Dipodomys ingens*, an endangered species that plays a crucial role in scattering the seeds of various plants, has suffered substantial population decline and now is restricted to 3% of its historical range (*3*). The scatter-hoarding Edwards's long-tailed giant rat (*Leopoldamys edwardsi*) disappeared after forest fragmentation compromised its habitat in southern Thailand (*4*) and has been lost in Hong Kong (*5*). Small-bodied scatter-hoarding rats have also become extinct as a result of fragmentation (*6*).

Scatter-hoarding squirrels, which are widely distributed across the world and include the Siberian flying squirrel (*Pteromys volans*), northern flying squirrel (*Glaucomys sabrinus*), eastern gray squirrel (*Sciurus carolinensis*), eastern fox squirrel (*S. niger*), Eurasian red squirrel (*S. vulgaris*), and American red squirrel (*Tamiasciurus hudsonicus*), are rarely found in the small forest fragments that remain after the loss of larger habitats (*7*–*10*). Eight species in the Sciurid family have declined as a result of deforestation in Singapore since the early 19th century (*11*). Smith's bush squirrel (*Paraxerus cepapi*), the sole secondary seed disperser of marula fruit (*Sclerocarya birrea*), is endangered in the African savanna (*12*).

Disrupting the patterns of scatter-hoarding rodents and the plants they support puts entire ecosystems at risk. Given the crucial role these species play in regions across the world, local governments and forestry administrations worldwide should implement conservation measures to preserve and restore large, unfragmented areas to protect them.

Sijie Yi[1] and Xianfeng Yi[2]*
[1]College of Life and Environmental Sciences, University of Exeter, Exeter, EX4 4PY, UK. [2]College of Life Sciences, Qufu Normal University, Qufu, 273165, China.
*Corresponding author.
Email: 20190021@qfnu.edu.cn

REFERENCES AND NOTES

1. M. Galetti *et al.*, *Bot. J. Linn. Soc.* **151**, 141 (2006).
2. P. Mittelman *et al.*, *Biol. Rev*. 10.1111/brv.12761 (2021).
3. I. V. Widick, W. T. Bean, *Divers. Distribut*. **25**, 1074 (2019).
4. A. J. Lynam, I. Billick, *Biol. Conserv*. **91**, 191 (1999).
5. K. Chung, R. T. Corlett, *Biodiv. Conserv.* **15**, 4521 (2006).
6. C. P. Kofron, F. X. Villablanca, *J. Fish Wildl. Manage.* **7**, 237 (2016).
7. J. L. Koprowski, *Anim. Conserv.* **8**, 369 (2005).
8. A. Garner *et al.*, *Conserv. Genet.* **6**, 759 (2005).
9. S. Hanski, *Ecography* **26**, 641 (2003).
10. L. E. Ritchie *et al.*, *Forest Ecol. Manage.* **257**, 1920 (2009).
11. R. C. H. Teo, S. Rajathurai, *Garden Bull. Singapore* **49**, 353 (1997).
12. J. J. Midgley *et al.*, *J. Trop. Ecol.* **28**, 227 (2012).

10.1126/science.abk0347

PRIZE ESSAY

NEUROMODULATION

Neurorobotics for neurorehabilitation

Peripheral neuromodulation boosts health and cognitive benefits in leg amputees

By **Stanisa Raspopovic**

Advances in peripheral nervous system (PNS) interfacing present a promising alternative to traditional neuromodulation (*1*), particularly for individuals with upper-limb amputations (*2*–*7*). Implanted electrodes have been shown to diminish phantom limb pain (PLP) in such subjects and enable close-to-natural touch sensations. Individuals with peripheral neural stimulators are even able to control the amount of force exerted by a prosthesis and to discern among objects with different compliances and shapes with prosthesis (*2*–*5*). Successive studies have shown the long-term utility of these technologies (*6*, *7*).

Having successfully achieved functional stimulation and chronic biocompatibility through neural interfaces, the focus of neuromodulation research is shifting toward achieving optimal design and policy of use (*8*). There is great variety in electrode geometries, stimulating contact numbers, and placement within the nervous system (see the figure) as well as in possible stimulation protocols. Optimization will not be achieved with brute force but requires the development of computational models (*9*, *10*) capable of exploiting the knowledge that has been accumulated on this topic.

SHIFTING THE FOCUS TOWARD THE LOWER LIMBS

Until recently, most research has focused on hand amputees, neglecting the clinical reality that four out of five amputees have lower-limb loss. Subjects with lower-limb amputation frequently do not engage fully in everyday activities because they are afraid of falls and do not perceive the prosthesis as part of their body (low "embodiment"). Such individuals often report poor satisfaction with their prostheses, citing the prosthesis as an excessive weight, despite prosthetic limbs typically being less than half the weight of a natural limb (*11*). They also tend to have reduced mobility (*12*), which can induce a sedentary lifestyle that promotes disease development and hinders reinsertion into society. PLP is also common and is poorly managed with current medications (*13*). Additionally, those with lower-limb amputations face substantially higher metabolic costs while walking, resulting in an increased risk of heart attack compared with the general population (*14*).

We have pioneered a human-machine system that translates prosthetic sensors' readouts into "language" understandable by the nervous system, using a detailed computational model (*9*, *10*) that indicates an optimal number of implants for the targeted nerve. Compared with traditional frequency variation, the model favors current amplitude modulation, for increasing efficiency of mapping and closed-loop stimulation (*15*).

Placement of the cable connecting implants to the stimulator is a longstanding problem and a frequent cause of failure in implantable technologies during clinical testing (*7*, *8*). We addressed this issue during surgical preparation by implementing a release loop and stabilization within the fascia tissue graft with cables embedded in the middle (*14*).

REGAINING THE FUNCTIONALITY

We developed a "sensing leg," for lower-limb amputees, by connecting sensors from the prosthetic knee and under the foot to the residual PNS (see the figure). An effective connection was achieved by equipping a microprocessor-controlled prosthesis with a purposely developed sensorized insole. An external controller that communicates wirelessly with the "sensorized prosthesis" proportionally transduces the readout of the insole and knee sensors into stimulation parameters. The stimulator then injects the current into the intraneural electrodes, eliciting sensations from the missing lower limb. The whole process ran at a delay unperceivable to the user, enabling real-time neuromodulation dependent on leg status.

This intervention enabled recovery of rich leg and foot perceptions (such as touch, proprioception, and both simultaneously). Users were able to recognize when the prosthetic leg—physically disconnected from their body and communicating wirelessly with the implants—was touched over different foot positions, flexed, or both. Users were also able to avoid a substantially higher number of stumbles when walking over obstacles, while wearing glasses that blind their lower field of vision, than when not exploiting the restored feedback.

Climbing stairs is often a challenge for above-knee amputees, resulting in very slow motion and considerable fatigue. When neuromodulation restored limb perception, their mobility substantially increased (*15*).

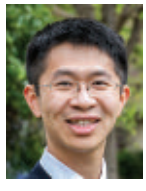
After these laboratory tests, volunteers stepped outside into a more natural environment. Because of the fully portable neuromodulating system, their confidence was



GRAND PRIZE WINNER

Stanisa Raspopovic

Stanisa Raspopovic received undergraduate degrees from the University of Pisa and a PhD from Scuola Superiore Sant'Anna, Italy. After completing his postdoctoral fellowship at EPFL, he started his laboratory in the Department of Health Science and Technology at ETH Zürich in 2018. His research focuses on deep understanding of nervous system interaction with electric field through computational modeling, design of sensory neuroprostheses, and bioelectronics solutions and the investigation of human interaction with these.



FINALIST

Weijian Yang

Weijian Yang received his undergraduate degree from Peking University and a PhD from the University of California, Berkeley. After completing his postdoctoral fellowship at Columbia University, he started his laboratory in the Department of Electrical and Computer Engineering at the University of California, Davis in 2017. His research aims to develop advanced optical methods and neurotechnologies to interrogate and modulate brain activity, with a goal to understand how neural circuits organize and function and how behaviors emerge from neuronal activity.

www.sciencemag.org/content/373/6555/635

PHOTOS: (LEFT TO RIGHT) COURTESY OF STANISA RASPOPOVIC. COURTESY OF WEIJIAN YANG

increased, and subjects were able to walk with increased speed over what would normally be challenging sandy terrain. At the same time, volunteers' metabolic consumption was diminished when sensory neurofeedback was switched on. The decreased energy expenditure when using neuromodulation could potentially limit cardiovascular system fatigue—a tremendously important health benefit for lower-limb amputees (*14*).

MODULATING THE PNS TO ADDRESS CENTRAL ISSUES

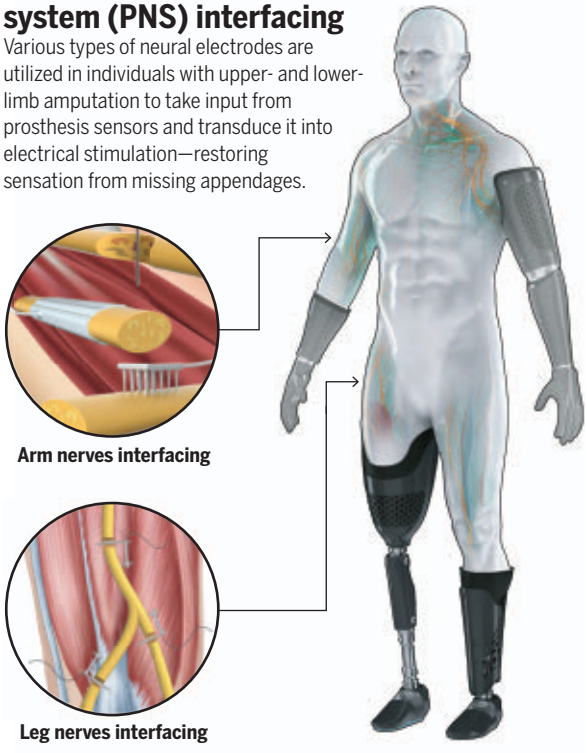
When the implantable system was used in "neuro-pacemaker modality"—stimulating the nerve without connection to the prosthesis—a reduction in PLP was observed. Through precise somatotopic stimulation, we have evoked pleasant, close-to-natural sensations within regions of referred pain. By contrast, commercial stimulation devices mainly deliver prefixed and often ineffective patterns of stimuli, which do not elicit physiologically plausible sensations and fail to deliver effective relief (*16*), and spinal cord stimulators involve the induction of paresthesia (an uncomfortable tingling), which does not always completely relieve the pain (*17*). Meanwhile, neurostimulators that directly target the peripheral nerve deliver either nonselective stimulation or induce an analgesic nerve block, both of which have considerable drawbacks (*18*). Our neuromodulation pain treatment represents a real advance with respect to existing treatments, in that we restore naturalistic percepts, revitalizing the physiological pathway for sensations. Beside the imminent pain relief, this potentially induces beneficial long-term neuroplastic changes at the central nervous system (CNS) level, offering not only an analgesic but also a "curative" effect.

As a consequence of the restoration of physiologically plausible sensations, subjects experienced ("embodied") the prosthesis similar to a real limb. Embodiment is typically measured in "nonfunctional" scenarios [such as rubber-hand experiments (*19*)]. We were able to measure an objective functional embodiment (*20*) increase during our experiments with the bionic leg, with and without feedback (*15*). Increased neural embodiment decreased weight perception (*11*)—a subjective percept influenced by cognitive processes. Brain cognitive load, measured with electroencephalography, also decreased while walking with the neuroprosthesis and performing a dual task (*15*).

Neuroengineering Laboratory, Institute for Robotics and Intelligent Systems, Department of Health Sciences and Technology, ETH Zürich, 8092 Zürich, Switzerland.
Email: stanisa.raspopovic@hest.ethz.ch

Different neurotechnologies for the peripheral nervous system (PNS) interfacing

Various types of neural electrodes are utilized in individuals with upper- and lower-limb amputation to take input from prosthesis sensors and transduce it into electrical stimulation—restoring sensation from missing appendages.



GRAPHIC: C. BICKEL/*SCIENCE* BASED ON S. RASPOPOVIC

UNVEILING NEUROROBOTICS MECHANISMS

Neuromodulation triggered by a robotic device influences sensorimotor strategies employed by users, by means of its integration into their "traditional" nervous system. To better understand underlying mechanisms, we measured gait features of leg amputees during motor tasks of different difficulty while using the neuroprosthesis. They performed an easy task (walking over ground) and a challenging task (ascending and descending stairs) while gait and neurostimulating parameters were collected. The neuroprosthesis reshaped subjects' legs' kinematics toward a more physiological gait owing to sensorimotor strategies that allowed users to intuitively exploit various features of the neural code during different tasks (*21*). These strategies included different temporal order, or spatial usage of stimulation channels, resulting in simple but robust intuitively integrated neural codes for different motor behaviors. In a hypothetical scenario, which required a leg amputee to simulate driving a conventional car, we demonstrated a finer pressure estimation from the prosthesis, suggesting that even a simple neural code could effectively improve wearable neuromodulating devices.

These studies not only provided clear evidence of the benefit of neuromodulation for lower-limb amputees but also provided insights into fundamental mechanisms of supraspinal integration of the restored sensory modalities. Even with only a limited restoration of sensations from foot and knee, the CNS was able to successfully integrate and exploit this information. Analogous findings were observed in animals that compensated for lack of a single sensory modality through supraspinal structures (*22*).

TOWARD WIDESPREAD USE

The health benefits achievable from neuromodulation are of paramount importance to millions of impaired individuals. Because the economic cost of such technologies remains considerable, it is important to emphasize the accompanying benefits, which could eliminate the need for treatments related to pain or cardiovascular problems (*1*).

Together with pioneering results in neuromodulating treatment for neuropathy, the described research presents a conceptually new framework for neuroprosthetic device design, implementation, and testing. This iterative framework consists of (i) developing a deep understanding of the problem through models and experiments, (ii) influencing the device design, and (iii) a meticulously planned clinical testing phase. Multifaceted validation of experiments—including functional, emotional, and cognitive outcomes—feeds back to increase our knowledge and further optimize design. Model-based, deep understanding of the effects of neuromodulation could benefit future projects in the emerging field of bioelectronic medicine (*23*, *24*). ■

REFERENCES AND NOTES

1. S. Raspopovic, *Science* **370**, 290 (2020).
2. S. Raspopovic *et al.*, *Sci. Transl. Med.* **6**, 222ra19 (2014).
3. J. A. George *et al.*, *Sci. Robot.* **4**, eaax2352 (2019).
4. D. W. Tan *et al.*, *Sci. Transl. Med.* **6**, 257ra138 (2014).
5. C. M. Oddo *et al.*, *eLife* **5**, e09148 (2016).
6. M. Ortiz-Catalan *et al.*, *N. Engl. J. Med.* **382**, 1732 (2020).
7. F. M. Petrini *et al.*, *Ann. Neurol.* **85**, 137 (2019).
8. S. Raspopovic *et al.*, *Nat. Mater.* **20**, 925 (2021).
9. S. Raspopovic *et al.*, *Proc. IEEE* **105**, 34 (2017).
10. M. Zelechowski *et al.*, *J. Neuroeng. Rehabil.* **17**, 24 (2020).
11. G. Preatoni *et al.*, *Curr. Biol.* **31**, 1065 (2021).
12. L. Nolan *et al.*, *Gait Posture* **17**, 142 (2003).
13. H. Flor, *Lancet Neurol.* **1**, 182 (2002).
14. F. M. Petrini *et al.*, *Nat. Med.* **25**, 1356 (2019).
15. F. M. Petrini *et al.*, *Sci. Transl. Med.* **11**, eaav8939 (2019).
16. C. Richardson, J. Kulkarni, *J. Pain Res.* **10**, 1861 (2017).
17. K. Kumar *et al.*, *Surg. Neurol.* **50**, 110, discussion 120 (1998).
18. Y. A. Patel, R. J. Butera, *J. Neural Eng.* **15**, 031002 (2018).
19. M. Botvinick, J. Cohen, *Nature* **391**, 756 (1998).
20. G. Rognini *et al.*, *J. Neurol. Neurosurg. Psychiatry* **90**, 833 (2019).
21. G. Valle *et al.*, *Sci. Adv.* **7**, eabd8354 (2021).
22. S. Rossignol *et al.*, *Physiol. Rev.* **86**, 89 (2006).
23. M. A. Hamza *et al.*, *Diabetes Care* **23**, 365 (2000).
24. L. V. Borovikova *et al.*, *Nature* **405**, 458 (2000).

10.1126/science.abj5259

PRIZE ESSAY

NEUROMODULATION

Manipulating neuronal circuits, in concert

Two-photon holographic optogenetics enables precise modulation of brain activity

By **Weijian Yang**

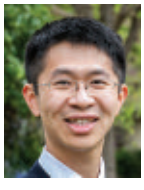
Perception and behavior emerge from the coordinated and orchestrated activity of neurons in brain circuits. Individual, functionally coherent neurons form ensembles, which become the building blocks of large-scale circuitry to drive the brain machinery (*1*). The ability to modulate brain activity in a spatiotemporal pattern with high specificity (millisecond time scale and cellular resolution) has great implications for interfacing with this sophisticated machinery. In fundamental science, it provides a powerful tool to dissect neuronal circuits in very fine detail and study causality among neural activity, circuit dynamics, and behavior (*2*–*6*). In translational medicine, it plays an important role in treating brain disorders (*7*, *8*) and holds great promise to become a new tool for precision medicine.

Electrical stimulation is the most mature approach to modulating brain activity. However, penetrating electrodes are highly invasive, and there is a lack of spatial specificity in the targeted brain regions. Therefore, neurons in a large brain volume are indiscriminately stimulated simultaneously regardless of their individual function in the brain circuit and resultant link to behavior. Such low spatial specificity and the associated unspecific off-target effects not only limit the application of these approaches in studying brain circuits, but also pose concerns with regard to overall efficacy and side effects in clinical therapy (*9*, *10*).

Optical methods, particularly when coupled with optogenetics (*11*, *12*), offer a new approach to modulating brain activity with cell-type specificity. Although high spatial specificity can be achieved in two-dimensional (2D) samples, such as thin brain slices, early use of optogenetics in living brains faced the same challenges as electrical stimulation: The dispersed light failed to distinguish individual cells in a 3D volume and instead stimulated all neurons together. Two-photon light resolves the problem of spatial specificity and achieves cellular resolution. Borrowing this technique from laser scanning microscopy, two-photon optogenetics sequentially stimulates neurons one by one (*13*, *14*). Although specificity is high, sequential single-cell stimulation fails to mimic intrinsic activity patterns in the brain, where multiple neurons can fire action potentials simultaneously.

Metaphorically speaking, manipulating neurons in a circuit is akin to pressing the keys of a piano keyboard. Photostimulating neurons one at a time is like playing the piano with a single finger, which would fail to produce a rhythmic and melodious concert piece. To modulate neural activity in a coordinated manner, it is necessary to simultaneously stimulate an ensemble of neurons, distributed in a 3D brain volume, with cellular resolution—as if playing the piano with all 10 fingers. We are among the first to tackle this challenge in vivo (*15*) and demonstrate the power of optogenetics in studying the link between neural activity and behavior (*2*) (see the figure).

Department of Electrical and Computer Engineering, University of California, Davis, Davis, CA 95616, USA. Email: wejyang@ucdavis.edu



FINALIST

Weijian Yang

Weijian Yang received his undergraduate degree from Peking University and a PhD from the University of California, Berkeley. After completing his postdoctoral fellowship at Columbia University, he started his laboratory in the Department of Electrical and Computer Engineering at the University of California, Davis in 2017. His research aims to develop advanced optical methods and neurotechnologies to interrogate and modulate brain activity, with a goal to understand how neural circuits organize and function and how behaviors emerge from neuronal activity. **www.sciencemag.org/content/373/6555/635**

TWO-PHOTON HOLOGRAPHIC OPTOGENETICS

Leveraging the computer-generated hologram, we encoded the 3D spatial information of the targeted neurons into the phase hologram using a spatial light modulator, to develop two-photon 3D holographic techniques for precision optogenetics (*15*). By projecting a holographic light pattern, which contains beamlets focused on the target neurons in a mouse brain, we can precisely modulate the activity of neuronal ensembles. Two-photon excitation ensures that the light can penetrate deep into the scattering tissue and stimulate the target neurons distributed in a 3D volume with excellent specificity. To maximize the number of neurons that can be stimulated at once without imposing high doses of light (and thus heat) on the brain, we increased the two-photon excitation efficiency by adapting a low–repetition rate femtosecond laser. This allowed us to simultaneously stimulate a large group of neurons (>50) with a minimum amount of light power (a few milliwatts per neuron). By rapidly switching the holograms (millisecond time scale), we can stimulate different groups of neurons with high temporal specificity. Our two-photon holographic optogenetics approach thus enables modulating the neuronal activity in a desired spatiotemporal pattern.

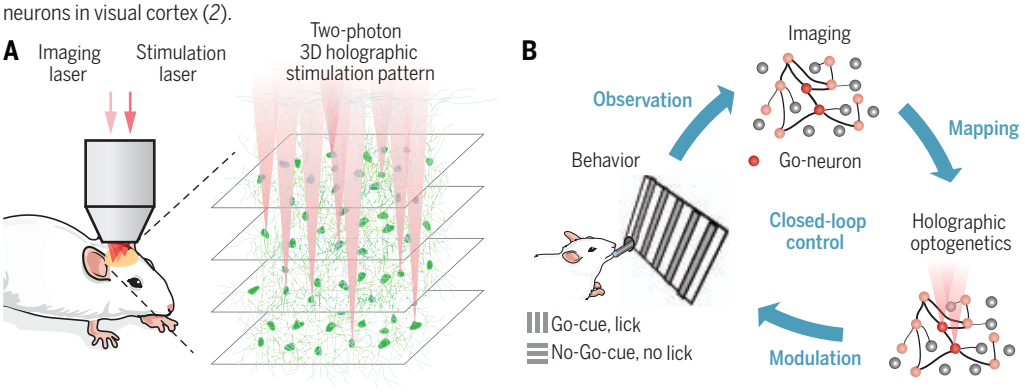
ALL-OPTICAL INTERROGATION OF THE NEURONAL ACTIVITY

To expertly manipulate brain circuitry, we needed a 3D neuronal map. We built a dual-path microscope with two different lasers, integrating two-photon high-speed volumetric calcium imaging with two-photon holographic optogenetics (*15*). The imaging path was equipped with an electrically tunable lens for fast 3D imaging (*15*, *16*), and the optogenetics path was equipped with a spatial light modulator to generate the 3D photostimulation pattern. To avoid cross-talk between imaging and optogenetics, we selected indicators and opsins with distinct light excitation spectra: calcium indicator GCaMP6 (*17*) for imaging and opsin C1V1 (*18*) for optogenetics. Using this dual-path microscope, we were among the first to demonstrate simultaneous 3D imaging and holographic photostimulation of cortical activity in awake mice (see the figure). Such an all-optical setup allowed us to precisely stimulate an arbitrary group of neurons while monitoring the response of the circuit, and thus enabled closed-loop control of brain activity, all with high temporal specificity and cellular resolution across a large 3D brain volume.

PHOTO: COURTESY OF WEIJIAN YANG

Two-photon holographic optogenetics

(**A**) Schematics of simultaneous two-photon volumetric calcium imaging and two-photon 3D holographic patterned photostimulation in a mouse brain. A user-defined group of neurons can be stimulated simultaneously with high spatiotemporal specificity. (**B**) Closed-loop control of neuronal activity and behavior. The neuronal circuit imaged during animal behavior provides a map to modulate the brain through two-photon holographic optogenetics. We demonstrated mouse performance in a Go/No-Go visual discrimination task can be enhanced by photoactivating only two core ensemble neurons in visual cortex (*2*).



GRAPHIC: H. BISHOP/*SCIENCE* BASED ON W. YANG

MODULATING BEHAVIOR BY TRIGGERING NEURONAL ENSEMBLE ACTIVITY

Understanding the role of neuronal ensembles could lead to new insight into how behaviors emerge as well as innovative therapies for brain diseases (*8*). Using our all-optical method, we studied the causal link between ensemble activity and behavior and demonstrated an efficient approach to modulate behavior (*2*) (see the figure).

We designed a Go/No-Go visual discrimination task, in which two orientations of drifting gratings were randomly displayed and the mouse discriminated between them by licking a waterspout. We hypothesized that modulation of ensemble activity could affect behavior. We holographically photoactivated a random group of unspecific neurons in the mouse visual cortex during the task. Unsurprisingly, the resultant "noise" in the visual cortex decreased task performance. We then asked if directed neuronal modulation could improve the task outcome. Using a machine learning algorithm, we extracted the neuronal ensembles and the core ensemble neurons related to the "Go-cue" of the visual stimuli in the visual cortex. Surprisingly, holographic photoactivation of only two core ensemble neurons during the Go-cue could enhance task performance (*2*).

Through imaging, we observed that the activation of core ensemble neurons drove widespread recruitment of other neurons within the ensemble. Such a pattern completion mechanism, potentially involving recurrent neural networks, eventually amplified the activation effect of core ensemble neurons and ultimately modulated the behavioral outcome. This effect was so pronounced that the holographic activation could elicit mouse licking associated with the Go-cue even when the Go-cue was not physically presented (*2*).

Compared to previous approaches whereby large brain regions are stimulated at once, either electrically or through single-photon optogenetics, our holographic approach provides much greater specificity and efficiency. Not only does our study prove the functional and behavioral relevance of neuronal ensembles and provide a direct illustration of pattern completion, but the ability to precisely write information into the brain to trigger behavior opens a new avenue in precision medicine to correct the pathophysiology of mental disorders (*8*).

AN OUTLOOK OF PRECISION OPTOGENETICS

The invention of optogenetics has given neuroscientists a new tool to modulate cell-type–specific neuronal activity. Our in vivo two-photon holography technique has brought optogenetics into a new, precision era. Today and in the near future, 4D spatiotemporal modulation patterns, which parallel the intrinsic physiology of the neural system, could be applied to elicit recurrent activity and recruit downstream activity and behavior.

We have demonstrated the triggering of visually guided behavior through two-photon holographic optogenetics in the mouse visual cortex (*2*), and others have applied this technique to the mouse hippocampus (*4*), to drive spatial behavior. In other animal models such as the larval zebrafish (*6*), the method has been used to elicit motor behavior. In each case, activation of only a small number of neurons was able to modulate animal behavior.

In addition to studying circuit causality, two-photon holographic optogenetics is an ideal tool to induce network plasticity through Hebbian plasticity (*19*). By repeatedly photostimulating a group of neurons, we demonstrated that functional connectivity increased in a subset of these neurons (*20*). When pairing the holographic photoactivation of a behaviorally unspecific ensemble with a behavioral reward, it was shown that the animal could learn to associate the ensemble activation with the award (*3, 21, 22*). Such findings suggest that two-photon holographic optogenetics could be used to reprogram the brain and create an artificial link between neuronal activity and cognitive states. This result has tremendous translational importance and could potentially be used to reestablish brain functions of a damaged region in a new region.

The past *3* years have witnessed a new wave of findings enabled by two-photon holographic optogenetics in awake mice. Much could be done to further exploit its potential, particularly in translational medicine. In our noninvasive demonstrations, target neurons were confined to the cortical layers. The ability to target deep brain regions with a noninvasive or minimally invasive approach will greatly broaden its application. The closed-loop, real-time control of imaging, optogenetics, and monitoring of behavior could potentially create a new type of brain machine interface. As the first type of precise brain modulation modality, we envision that two-photon holographic optogenetics (*15, 23–27*) will continue to play a pivotal role in both fundamental neuroscience and translational medicine. ■

REFERENCES AND NOTES

1. R. Yuste, *Nat. Rev. Neurosci.* **16**, 487 (2015).
2. L. Carrillo-Reid *et al.*, *Cell* **178**, 447 (2019).
3. J. H. Marshel *et al.*, *Science* **365**, eaaw5202 (2019).
4. N. T. M. Robinson *et al.*, *Cell* **183**, 1586 (2020).
5. K. Daie, K. Svoboda, S. Druckmann, *Nat. Neurosci.* **24**, 259 (2021).
6. M. dal Maschio *et al.*, *Neuron* **94**, 774 (2017).
7. A. M. Lozano *et al.*, *Nat. Rev. Neurol.* **15**, 148 (2019).
8. L. Carrillo-Reid, W. Yang, J. E. Kang Miller, D. S. Peterka, R. Yuste, *Annu. Rev. Biophys.* **46**, 271 (2017).
9. D. Cyron, *Front. Integr. Neurosci.* **10**, 17 (2016).
10. M. Z. Zarzycki, I. Domitrz, *Acta Neuropsychiatr.* **32**, 57 (2020).
11. G. Nagel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13940 (2003).
12. E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, K. Deisseroth, *Nat. Neurosci.* **8**, 1263 (2005).
13. R. Prakash *et al.*, *Nat. Methods* **9**, 1171 (2012).
14. J. P. Rickgauer *et al.*, *Nat. Neurosci.* **17**, 1816 (2014).
15. W. Yang *et al.*, *eLife* **7**, e32671 (2018).
16. S. Han, W. Yang, R. Yuste, *Cell Rep.* **27**, 2229 (2019).
17. T. W. Chen *et al.*, *Nature* **499**, 295 (2013).
18. O. Yizhar *et al.*, *Nature* **477**, 171 (2011).
19. D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory* (Wiley, 1949).
20. L. Carrillo-Reid *et al.*, *Science* **353**, 691 (2016).
21. H. W. Dalgleish *et al.*, *eLife* **9**, e58889 (2020).
22. J. V. Gill *et al.*, *Neuron* **108**, 382 (2020).
23. W. Yang, R. Yuste, *Curr. Opin. Neurobiol.* **50**, 211 (2018).
24. A. M. Packer *et al.*, *Nat. Methods* **12**, 140 (2015).
25. A. R. Mardinly *et al.*, *Nat. Neurosci.* **21**, 881 (2018).
26. A. Forli *et al.*, *Cell Rep.* **22**, 3087 (2018).
27. I.-W. Chen *et al.*, *J. Neurosci.* **39**, 3484 (2019).

10.1126/science.abj5260

The Most ADVANCED Patch Clamp System Available Today

Visit
www.dpatch.com
to learn more



SUTTER INSTRUMENT

+1.415.883.0128
info@sutter.com
www.sutter.com

Where Science Gets Social.

AAAS.ORG/COMMUNITY







AAAS' Member Community is a one-stop destination for scientists and STEM enthusiasts alike. It's "Where Science Gets Social": a community where facts matter, ideas are big and there's always a reason to come hang out, share, discuss and explore.

Member COMMUNITY
AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



PLANT GENOMICS

An a-maize-ing set of genomes

Maize is an important crop that is cultivated worldwide. As maize spread across the world, selection for local environments resulted in variation, but the impact on differences between the genome has not been quantified. By producing high-quality genomic sequences of the 26 lines used in the maize nested association mapping panel, Hufford *et al*. map important traits and demonstrate the diversity of maize. Examining RNA and methylation of genes across accessions, the authors identified a core set of maize genes. Beyond this core set, comparative analysis across lines identified high levels of variation in the total set of genes, the maize pan-genome. The value of this resource was further exemplified by mapping quantitative traits of interest, including those related to pathogen resistance. —LMZ *Science*, abg5289, this issue p. 655

Maize varieties share a common set of core genes, but a genomic survey reveals the diversity of genes in different inbred lines.

PHOTO: PHILIPPE PSAILA/SCIENCE SOURCE

CHEMICAL KINETICS

Spectral fingerprint of stabilized •QOOH

Carbon-centered radicals containing the hydroperoxy group, commonly denoted as •QOOH, are elusive but are among the most critical intermediate species for kinetic modeling of hydrocarbon oxidation in various atmospheric and combustion processes. Their direct experimental observation is a long-standing challenge, with only one successful previous attempt. Using a combination of infrared activation spectroscopy and ultraviolet laser–induced fluorescence detection, Hansen *et al*. directly characterized the vibrational structure of a •QOOH intermediate in isobutane oxidation, collisionally stabilized and isolated, and followed its dissociative evolution under infrared activation with time and energy resolution. High-level electronic structure calculations revealed an important role of heavy-atom tunneling in this process. —YS *Science*, abj0412, this issue p. 679

IMMUNOMETABOLISM

Chemerin counters cold thermogenesis

Conversion of white adipose cells in fat depots into glucose-burning, beige adipose cells contributes to the maintenance of normal energy homeostasis and the prevention of obesity. Chemerin, an obesity-associated adipokine, inhibits the cold-induced "beiging" of adipocytes. Lin *et al.* found that genetic deletion of chemerin or its receptor, CMKLR1, in mice increased cold-induced interleukin-33 production by adipocytes, thereby promoting beige fat formation. Mice that had CMKLR1 conditionally removed in adipocytes were less vulnerable than control mice to the development of obesity after consuming a high-fat diet. These findings point to adipocyte-expressed CMKLR1 as a potential target for pharmaceutical approaches aimed at treating obesity and linked metabolic phenotypes. —IW *Sci. Immunol.* **6**, eabg9698 (2021).

NEUROSCIENCE

Arresting fear and anxiety

Opioid receptors signal through two kinds of downstream partners, G proteins and β-arrestins. Because many side effects associated with opioid use are mediated by β-arrestins, opioids that bias signaling toward G protein–mediated pathways are preferred for treating pain. However, Ko *et al*. found that β-arrestin–biased drugs may have potential for treating fear and anxiety. Natural and synthetic opioids altered anxiety-like and conditioned fear–related behaviors in mice through G protein– and β-arrestin–specific signaling in multiple brain regions, indicating distinct and overlapping functions for β-arrestin isoforms. —LKF *Sci. Signal.* **14**, eaba0245 (2021).

QUANTUM SENSING

Quantum enhanced sensing

Harnessing quantum mechanical effects is expected to provide an advantage over classical sensing technology. By entangling the center-of-mass motional state of approximately 150 ions trapped in a two-dimensional Coulomb crystal with their collective spin state, Gilmore *et al*. demonstrate

a quantum-enhanced measurement sensitivity of displacement and electric field. Such enhanced sensitivity could, for instance, find application in probing proposed weak interactions between dark matter and normal matter, as well as enhancing gravitational wave detection. —ISO

Science, abi5226, this issue p. 673

METALLURGY

Locking structure to high temperature

Because of atomic diffusion, metal alloys with nanometer-sized crystal grains do not retain their structure at high temperature. Xu *et al*. found that a minimum-interface structure allows a supersaturated aluminum-magnesium alloy to be retained at temperatures higher than the melting point. This system is known for high atomic diffusivity, highlighting the importance of the underlying interface structure. These observations have implications for designing structural alloys for high-temperature applications. —BG

Science, abh0700, this issue p. 683

COOLING TEXTILES

A durable way to keep cool

The fibers that make up textiles can be augmented to change how they interact with thermal radiation, but the resulting materials often are not durable. Zeng *et al*. developed a multilayer metafabric composed of a titanium oxide polylactic acid composite laminated with a polytetrafluoroethylene layer. This combination creates a textile that has passive radiative cooling properties with good mechanical properties and scalability. The textile can be made into clothes or car covers and keeps a person or a car much cooler than other fabrics. —BG

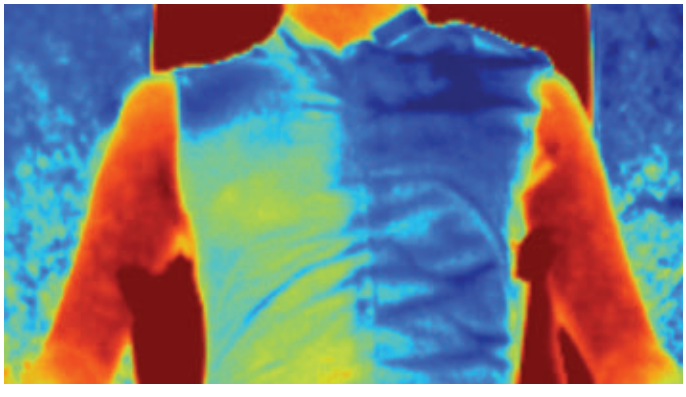
Science, abi5484, this issue p. 692

STRUCTURAL BIOLOGY

Structural changes in HIV maturation

Nascent HIV particles assemble at the plasma membrane of an infected cell and bud into a membrane-enveloped, immature virion. Assembly and budding are driven by a polyprotein called Gag, which consists of a matrix domain (MA) that is recruited to the plasma membrane, a capsid domain (CA) responsible for self-assembly, and a nucleocapsid domain (NC) that recruits the viral RNA genome. Gag cleavage results in a structural rearrangement that produces the mature virion. Qu *et al*. imaged mature and immature HIV particles by electron tomography and focused in on the MA domain (see the Perspective by Hikichi and Freed). They found that MA rearranges between two distinct hexameric lattices, and mature MA modulates the viral membrane by binding to a membrane lipid. This finding suggests that MA may play functional roles in the mature virion. —VV

Science, abe6821, this issue p. 700; see also abj9075, p. 621



An infrared camera reveals the cooling effect of a shirt made with a composite metafabric (right) versus normal fabric (left).

IN OTHER JOURNALS

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

PHYSIOLOGY

Winter feast of feces

Cold temperatures and greatly reduced resources are challenging for small mammalian species. Pikas are rabbit relatives that live at high altitudes and survive through winter without recourse to hibernation at temperatures that often dip below −30°C. Speakman *et al.* found that plateau pikas, *Ochotona curzoniae*, reduced their daily energy expenditure over winter by nearly 30% through thyroid-regulated lowered body temperature and reduced activity. However, if domestic yaks were present, pikas were able to save more energy by consuming yak feces. This cross-species coprophagy resulted in convergence of yak and pika gut microbiota and allowed the small mammals to thrive in the presence of their large, supposed competitors. —SNV

Proc. Natl. Acad. Sci. U.S.A. **118**, e2100707118 (2021).



Plateau pikas survive the Tibetan winter without hibernating.

RNA FOLDING

RNA knots provide resistance

Exoribonuclease enzymes are recruited to defend cells against viruses by degrading viral RNA. Some viruses, including flaviviruses such as Zika virus, have evolved RNA that is resistant to exoribonuclease. Structural studies show that the resistant RNAs contain a knot-like structure. To find out how the knot makes the RNA exoribonuclease resistant, Zhao and Woodside measured the forces needed to unfold single resistant RNA molecules. They mapped a series of structural transitions and show how the RNA must fold in the correct order to give a very stable knotted structure. Alternate structures with less-stable knot-like structures do not resist degradation. Targeting key intermediates on the path to the degradation-resistant structure could be developed into a therapeutic strategy. —VV

Nat. Chem. Biol. 10.1038/s41589-021-00829-z (2021).

CHEMICAL PHYSICS

How water affects aldehyde oxidation

Although water is a copiously studied solvent, its role in mediating gas phase reactions remains uncertain and hard to determine precisely. Neeman *et al.* reexamined a previous report that water complexation accelerates the gas phase reaction of acetaldehyde with hydroxyl radicals at 60 Kelvin. Their experiments show that water instead appears to slow the reaction down, and they suggest that the earlier work may have been skewed by aldehyde dimerization at high concentration. Accompanying theory predicts that, despite lowering

CREDITS (LEFT TO RIGHT) ZENG *ET AL.*; (PHOTO) STAFFAN WIDSTRAND/MINDEN PICTURES



the energy barrier for hydrogen abstraction from $CH_3CHO(H_2O)$, water complexation diminishes the dipolar attraction of the reactants. —JSY

J. Chem. Phys. **155**, 034306 (2021).

QUANTUM NETWORKS

Getting practical with quantum security

Current communication platforms, exchange of data, and online financial transactions are secured by classical encryption methods. The advent of quantum computers is expected to replace these protocols at some point. Quantum key distribution (QKD) relies on quantum mechanics to ensure that security is maintained, but the operation of the quantum components is manufacturer dependent, making it difficult for different systems to talk to each other. De Marco *et al.* close that communication gap by developing a multirate, multiprotocol QKD transmitter. The operating regime can be adjusted in real time by changing the driving signal, thereby allowing it to communicate with different receiving devices. Such operational flexibility will be useful for securing what will be a multiuser, multivendor quantum network environment. —ISO

Optica **8**, 911 (2021).

CHEMICAL REACTIONS

Chemical reactivity from local temperature

In recent years, density functional theory (DFT) has become truly accurate, with uncertainties comparable to typical uncertainties in many experiments. This should be leading to a complete transformation of chemistry but so far it has not. DFT calculations yield results that cannot be used directly because they must be translated into a special language to point out their chemical relevance. Although chemical reactions are too numerous to count, there are only a limited number of reactivity principles that could be applicable in general. Guo *et al.* show that the concept of local temperature from the context of conceptual DFT can be used to predict regioselectivity for nucleophilic and electrophilic attacks and to develop the corresponding descriptors of chemical reactivity. —YS

J. Phys. Chem. Lett. **12**, 5623 (2021).

CELL BIOLOGY

Recycling the recycling machinery

During phagocytosis, phagocytic cells physically engulf particles from the extracellular milieu and enclose them in phagosomes. These phagosomes fuse with lysosomes and their contents are degraded; however, lysosomal enzymes cannot digest particulate matter. How can a cell avoid building up a junkyard of phagolysosomes replete with indigestible material and maintain a pool of lysosomes ready for phagocytosis and degradation of further cargos? Lancaster *et al.* studied macrophages as they internalized and degraded bacteria. They found that phagolysosomes separated into small vesicles and tubules by cytoskeletal interactions and clathrin, such that small lysosomes devoid of indigestible content were regenerated with the help of newly synthesized components. This process allows the macrophages to cope with successive rounds of phagocytosis. —SMH

J. Cell Biol. **220**, e202005072 (2021).

MICROBIOTA

Commensal responses go retro

Vertebrate immune systems must be able to process and respond to myriad intrakingdom and interkingdom interactions by commensals and pathogens. Counterintuitively, Lima-Junior *et al.* found that colonization of mouse skin by the bacterium *Staphylococcus epidermidis* initiates an antiviral immune program. Sensing of *S. epidermis* by keratinocytes induced the expression of endogenous retroviruses that could be detected by the host's cGAS–STING signaling pathway. This pathway promotes the tissue repair function of commensal-specific T cells. Antiretroviral treatment impairs repair processes, whereas a high-fat diet enhances endogenous retrovirus expression, leading to excess skin inflammation. Thus, the immune system integrates the endogenous virome and exogenous microbiome into a fruitful alliance. —STS

Cell **184**, P3794 (2021).

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

NANOMATERIALS

Advances in colloidal quantum dots

The confinement found in colloidal semiconductor quantum dots enables the design of materials with tunable properties. García de Arquer *et al*. review the recent advances in methods for synthesis and surface functionalization of quantum dots that enable fine tuning of their optical, chemical, and electrical properties. These important developments have driven the commercialization of display and lighting applications and provide promising developments in the related fields of lasing and sensing. —MSL

Science, aaz8541, this issue p. 640

CORONAVIRUS

SARS-CoV-2 from alpha to epsilon

As battles to contain the COVID-19 pandemic continue, attention is focused on emerging variants of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus that have been deemed variants of concern because they are resistant to antibodies elicited by infection or vaccination or they increase transmissibility or disease severity. Three papers used functional and structural studies to explore how mutations in the viral spike protein affect its ability to infect host cells and to evade host immunity. Gobeil *et al*. looked at a variant spike protein involved in transmission between minks and humans, as well as the B1.1.7 (alpha), B.1.351 (beta), and P1 (gamma) spike variants; Cai *et al*. focused on the alpha and beta variants; and McCallum *et al*. discuss the properties of the spike protein from the B1.1.427/B.1.429 (epsilon) variant. Together, these papers show a balance among mutations that enhance stability, those that increase binding to the human receptor ACE2, and those that confer resistance to neutralizing antibodies. —VV

Science, abi6226, abi9745, abi7994, this issue p. 641, p. 642, p. 648

PHENYLKETONURIA

RNA solution for a genetic problem

Phenylketonuria is a classic example of the benefit of newborn metabolic screening: It is a single-gene disease that can be detected at birth, and its neurological effects can be prevented by dietary therapy. Unfortunately, this is not always straightforward because the disease-causing mutations in phenylalanine hydroxylase vary between patients and affect the severity of the phenotype, such that some patients' symptoms do not fully respond to the available interventions. Li *et al*. identified two long noncoding RNAs, one in mice and one in humans, that interact with phenylalanine hydroxylase and modulate its function (see the Perspective by Ben-Tov Perry and Ulitsky). Administration of modified RNAs mimicking their effects ameliorated the disease phenotype in mouse models of phenylketonuria. —YN

Science, aba4991, this issue p. 662; see also abj7969, p. 623

NANOFLUIDICS

Confined flow effects

Most memory resistor ("memristor") systems use electrons as the charge carrier, but it may also be possible to use ionic carriers, similar to the way that neurons work. Robin *et al*. studied an aqueous electrolyte confined into a pseudo two-dimensional gap between two graphite layers (see the Perspective by Hou and Hou). The authors observed a current–voltage relation that exhibits hysteresis, and the conductance depends on the history of the system, also known as the memresistor effect. Using simulations of their system, they can model the emission of voltage spikes characteristic of neuromorphic activity. —MSL

Science, abf7923, this issue p. 687; see also abj0437, p. 628

BIOMECHANICS

Squirrel parkour

Every day, there are acrobatic extravaganzas going on above our heads. Squirrels navigate remarkably complex and unpredictable environments as they leap from branch to branch, and mistakes can be fatal. These feats require a complex combination of evolved biomechanical adaptations and learned behaviors. Hunt *et al*. characterized the integration of these features in a series of experiments with free-living fox squirrels (see the Perspective by Adolph and Young). They found that the squirrels' remarkable and consistent success was due to a combination of learned impulse generation when assessing the balance between distance and branch flexibility and the addition of innovative leaps and landings in the face of increasingly difficult challenges. —SNV

Science, abe5753, this issue p. 697; see also abj6733, p. 620

INFECTIOUS DISEASE

Banishing bad bacteria

Group A *Streptococcus* is one cause of the flesh-eating disease necrotizing fasciitis. Anand *et al*. show that a branch of the unfolded protein response is responsible for the infection-induced increase in host cell asparagine production and release known to help this bacterium infiltrate connective and soft tissue. Conversely, using inhibitors of the unfolded protein response, or the endoplasmic reticulum stress that induces it, resulted in restricted bacterial access to asparagine. Treatment with either inhibitor protected mice against group A streptococcal infection in prophylactic or therapeutic models of necrotizing fasciitis, suggesting a strategy to treat this potentially deadly disease. —CAC

Sci. Transl. Med. **13**, eabd7465 (2021).

NEURODEGENERATION

Therapeutics in sight?

The development of drugs to treat Alzheimer's disease has largely been disappointing until the recent approval of aducanumab. Although not without controversy, particularly given the cost and lack of consistent clinical outcomes, the approval of aducanumab highlights important issues in drug development for complex chronic diseases. In a Perspective, Selkoe discusses the challenges raised by the recent approval and the implications for the pipeline of drugs in advanced clinical trials that may soon be evaluated for approval. Whether these trials will offer more consistent clinical responses is unknown, but what is clear is that there may finally be disease-modifying drugs, at least for some patients, to begin to tackle the devastating effects of Alzheimer's disease. —GKA

Science, abi6401, this issue p. 624

NEURODEGENERATION

Blood tests for Alzheimer's disease

The progression of Alzheimer's disease is accompanied by certain changes in the brain, especially the formation of amyloid plaques and tau tangles. Whether these changes are causative of the disease remains a matter for debate, but they can be used diagnostically, usually through positron emission tomography imaging or by measuring amyloid-β or tau protein concentrations in the cerebrospinal fluid. However, such tests are highly invasive.

Given the progress in therapeutics for Alzheimer's disease, there is a need for simple blood tests that can be used in clinical trials to assess therapeutic efficacy and to ensure accurate diagnosis for subsequent treatment. In a Perspective, Blennow discusses the progress in developing blood tests for Alzheimer's disease and the hurdles that need to be overcome. —GKA

Science, abi5208, this issue p. 626

REVIEW SUMMARY

NANOMATERIALS

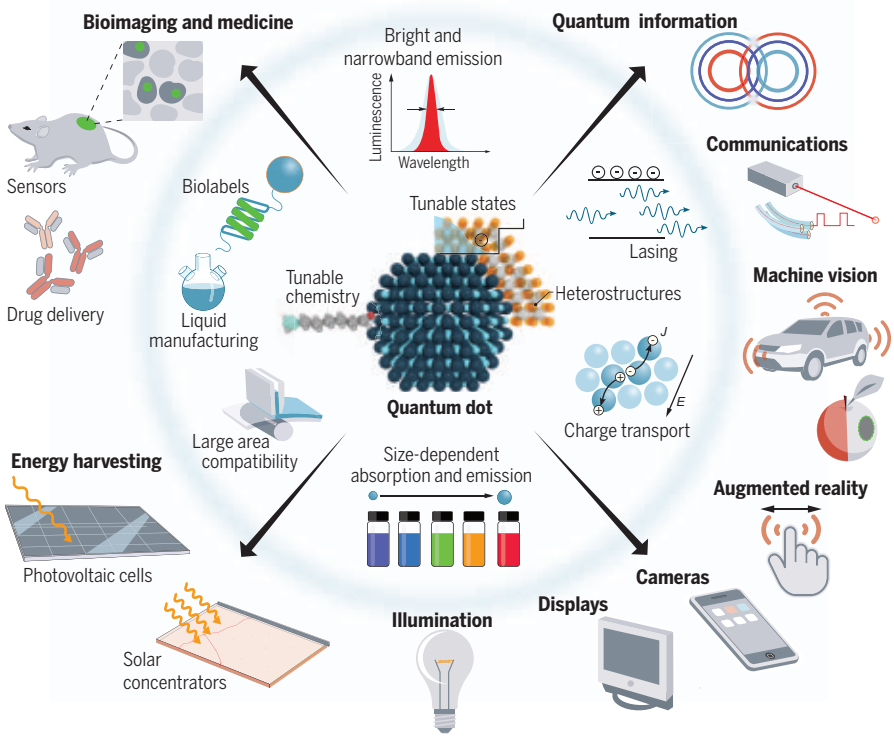
Semiconductor quantum dots: Technological progress and future challenges

F. Pelayo García de Arquer, Dmitri V. Talapin, Victor I. Klimov, Yasuhiko Arakawa, Manfred Bayer, Edward H. Sargent*

BACKGROUND: Semiconductor materials feature optical and electronic properties that can be engineered through their composition and crystal structure. The use of semiconductors such as silicon gallium arsenide sparked technologies from computers and mobile phones to lasers and satellites. Semiconductor quantum dots (QDs) offer an additional lever: Because their size is reduced to the nanometer scale in all three dimensions, the restricted electron motion leads to a discrete atom-like electronic structure and size-dependent energy levels. This enables the design of nanomaterials with widely tunable light absorption, bright emission of pure colors, control over electronic transport, and a wide tuning of chemical and physical functions because of their large surface-to-volume ratio.

ADVANCES: The bright and narrowband light emission of semiconductor QDs, tunable across the visible and near-infrared spectrum, is attractive to realize more efficient displays with purer colors. QDs are engineered compositionally and structurally to manipulate energy states and charge interactions, leading to optical gain and lasing, relevant to light emission across visible and infrared wavelengths and fiberoptic communication. Their tunable surface chemistry allows application as optical labels in bio-imaging, made possible by tethering QDs with proteins and antibodies. The manipulation of QD surfaces with capping molecules that have different chemical and physical functions can be tailored to program their assembly into semiconducting solids, increasing conductivity and enabling the transduction of photonic and chemical stimuli into electrical signals. Optoelectronic devices such as transistors and photodetectors lead to cameras sensitive to visible and infrared light. Highly crystalline QDs can be grown epitaxially on judiciously chosen substrates by using high-temperature and vacuum conditions, and their use has led to commercially viable high-performance lasers. The advent of colloidal QDs, which can be fabricated and processed in solution at mild conditions, enabled large-area manufacturing and widened the scope of QD application to markets such as consumer electronics and photovoltaics.

OUTLOOK: From a chemistry perspective, further advances in QD fabrication are needed to sustain and improve desired chemical and optoelectronic properties and to do so with high reproducibility. This entails the use of inexpensive synthesis methods and precursors that are able to retain laboratory-scale QD properties to market-relevant volumes. A better understanding of the yet-incomplete picture of QD surfaces, atomic arrangement, and metastable character is needed to drive further progress. From a regulatory perspective, added attention is needed to achieve high-quality materials that do not rely on heavy metals such as Cd, Pb, and Hg. The role of nanostructuring in toxicity and life cycle analysis for each application is increasingly important. From a materials and photophysics perspective, exciting opportunities remain in the understanding and harnessing of electrons in highly confined materials, bridging the gap between mature epitaxial QDs and still-up-and-coming colloidal QDs. The yet-imperfect quality of the latter—a price paid today in exchange for their ease of manufacture—remains a central challenge and must be addressed to achieve further-increased performance in devices. From a device perspective, colloidal QD manufacturing must advance to translate from laboratory-scale to large-area applications such as roll-to-roll and inkjet printing. Photocatalysis, in which light is used to drive chemical transformations, is an emerging field in which QDs are of interest. Quantum information technologies, which rely on the transduction of coherent light and electrons, bring new challenges and opportunities to exploit quantum confinement effects. Moving forward, opportunities remain in the design of QD-enabled new device architectures. ■



Semiconductor quantum dot technologies. QDs feature widely tunable optical, electrical, chemical, and physical properties. Their applications span energy harvesting, illumination, displays, cameras, sensors, communication and information technology, biology, and medicine. These nanostructures have been exploited to realize efficient lasers, displays, biotags, and solar harvesting devices available in the market and are emerging in photovoltaics, sensing, and quantum information.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: ted.sargent@utoronto.ca
Cite this article as F. P. García de Arquer *et al.*, *Science* **373**, eaaz8541 (2021). DOI: 10.1126/science.aaz8541

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aaz8541

REVIEW

NANOMATERIALS

Semiconductor quantum dots: Technological progress and future challenges

F. Pelayo García de Arquer[1,2], Dmitri V. Talapin[3], Victor I. Klimov[4], Yasuhiko Arakawa[5], Manfred Bayer[6], Edward H. Sargent[1]*

In quantum-confined semiconductor nanostructures, electrons exhibit distinctive behavior compared with that in bulk solids. This enables the design of materials with tunable chemical, physical, electrical, and optical properties. Zero-dimensional semiconductor quantum dots (QDs) offer strong light absorption and bright narrowband emission across the visible and infrared wavelengths and have been engineered to exhibit optical gain and lasing. These properties are of interest for imaging, solar energy harvesting, displays, and communications. Here, we offer an overview of advances in the synthesis and understanding of QD nanomaterials, with a focus on colloidal QDs, and discuss their prospects in technologies such as displays and lighting, lasers, sensing, electronics, solar energy conversion, photocatalysis, and quantum information.

The electronic and optical properties of conventional bulk semiconductors are determined by materials composition, crystal structure, and intentional and unintentional impurities (dopants). Advances in layer-by-layer crystal growth techniques such as molecular beam epitaxy (MBE) and metal organic chemical vapor deposition (MOCVD) enabled the realization of highly crystalline Si and III-V (for example, GaAs, InP, and GaN) semiconductors (*1*) with widely tunable optoelectronic properties. Quantum-confined structures exhibit size-dependent electronic properties, leading to extra degrees of tunability compared with bulk semiconductors and additional levers in the design of materials and devices.

Quantum confinement emerges when electrons are constrained to a domain comparable with their de Broglie wavelength. Quantum-confined structures are classified as two-dimensional (2D)—in which electrons are free to move in two directions—1D, and 0D. These include quantum wells, quantum wires, and quantum dots (QDs), respectively. In QDs, electrons and holes exhibit a discrete (quantized), atomic-like density of states (DOS) (Fig. 1A) (*2*). As QDs become smaller, quantum confinement increases the effective bandgap, leading to a blue shift of the absorption and emission spectra. An electron excited across the bandgap experiences strong interactions with the remaining valence band hole. Coulomb attraction and spin-exchange coupling produce strongly confined electron-hole pairs (excitons). Under high excitation levels, multiple excitons populate a QD. The close proximity between charge carriers in QDs leads to enhanced many-body phenomena that affect their electronic and optoelectronic properties (*3*).

QDs were first realized experimentally as glass-embedded particles (Fig. 1B) (*4*) and, shortly after that, as chemically synthesized colloidal nanocrystals (*5*, *6*). Independently, the concept of 3D quantum confinement in semiconductor nanostructures for lasing applications was presented (*7*) and demonstrated by combining a quantum well potential with high magnetic fields (*8*). The resulting structures were labeled 3D quantum wells or quantum well boxes.

Fabrication and assembly of QDs

Two main strategies exist to fabricate QDs: physical vacuum-based methods and wet-chemical approaches.

Top-down physical fabrication relies on lithography or milling to define a nanometer-sized volume in an existing semiconductor (Fig. 1C). In bottom-up techniques, QD growth occurs through assembly of atomic or molecular building blocks and is driven by built-in strain [Stranski-Krastanov (S-K) growth mode]. MBE and MOCVD have enabled the realization of high-quality epitaxial QDs (eQDs) prepared on top of a crystalline substrate (Fig. 1D). S-K growth was used to achieve In(Ga)As/GaAs eQDs (*9*–*11*). Alternative droplet epitaxial growth of eQDs occurs through sequential deposition of group III and V atoms without using lattice-mismatch, offering a path toward strain-free eQDs that is yet to be fulfilled (*12*). eQDs have been applied in areas such as optical fiber communications (as laser sources), military night-vision cameras, and aerospace (for example, optoelectronic circuits and ultrahigh-efficiency solar cells) (*13*).

The chemical solution-phase fabrication of colloidal QDs (cQDs) is an approach distinct to physical vacuum-based epitaxy. Modern cQD syntheses can be traced to the colloidal method introduced in 1993 (*14*). The synthetic methodology for cQDs evolved from early work on arrested precipitation of inside small aqueous micelles (*15*) toward reactions between molecular precursors in organic solvents at mild temperatures (100° to 350°C) (Fig. 1E). The nucleation and growth of cQDs is controlled by surfactant molecules (ligands) that bind dynamically to cQD surfaces. Judicious selection of precursors and surfactants, as well as manipulation of reaction temperature and duration, enable precise control over the stoichiometry, size, and shape of the cQD. Colloidal approaches have been successfully applied to grow cQDs of II-VI (*14*), III-V (*16*), IV-VI (*17*), and group IV (*18*–*20*) semiconductors and, more recently, metal halide perovskites ($CsPbX_3$; X = I, Br, or Cl) (*21*).

The quality of cQDs is determined by the crystalline perfection of their cores, the completeness of surface passivation, and uniformity in size and shape. High monodispersity is crucial to retaining the near-discrete character of the DoS for an ensemble of cQDs. Exploration of the ample synthetic parameter space has led to a continued improvement in cQD syntheses, leading to absorption linewidths approaching the homogeneous (single QD) limit (*14*, *22*, *23*). Doping of cQDs provides an additional avenue to tune their DOS and type of majority charge carrier (*24*–*26*).

The surface ligands in cQDs—typically bulky organic molecules, such as oleic acid and oleylamine—introduce repulsive forces between cQDs dispersed in a solvent, rendering them colloidally stable (*27*). The deposition of cQDs onto solid substrates can lead to either glassy or partially ordered QD films, determined by nanocrystal monodispersity, solvent drying kinetics, and the interaction among surface ligands (*28*–*32*).

Solution-based cQD deposition techniques are scalable and well-suited for the realization of large-area devices. cQD thin-film fabrication is compatible with high-throughput manufacturing and with a variety of substrates, which facilitates integration with platforms such as silicon electronics, plastic circuits, fiber optics, and fabrics.

QDs: From engineered functionalities to applications

Size-dependent bandgap

The bandgap of QDs (E_g) can be size-tuned across a wide range of energies from the

[1]Department of Electrical and Computer Engineering, University of Toronto, 35 St. George Street, Toronto, ON M5S 1A4, Canada. [2]ICFO–Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, Barcelona 08860, Spain. [3]Department of Chemistry, University of Chicago, Chicago, IL 60637, USA. [4]Chemistry Division, C-PCS, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. [5]University of Tokyo, Meguro, Tokyo 153-8505, Japan. [6]Technische Universitat Dortmund, 44221 Dortmund, Germany.
*Corresponding author. Email: ted.sargent@utoronto.ca

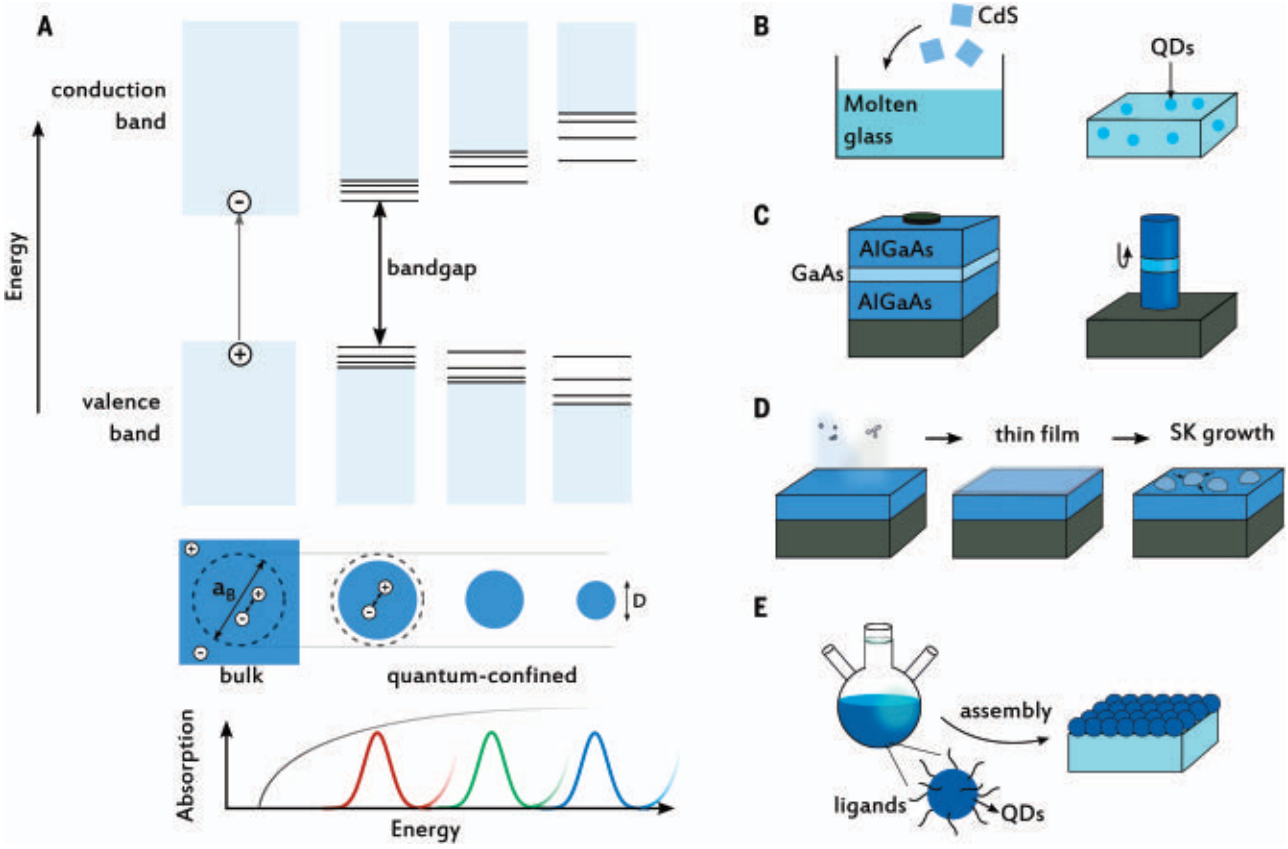


Fig. 1. Quantum confinement and fabrication of QD materials. (**A**) Quantum confinement, leading to size-dependent optical and electrical properties that are distinct from those of parental bulk solids, occurs when the spatial extent of electronic wave functions is smaller than the Bohr exciton diameter (a_B). *D*, QD diameter. (**B** to **E**) Examples of fabrication techniques of 0D semiconductor nanostructures, including early demonstrations of high-temperature precipitation in molten glasses; top-down lithography; nucleation and growth of nano-islands through epitaxial layer-by-layer deposition by using molecular beam epitaxy (MBE); and solution-based, low-to-moderate temperature colloidal synthesis.

ultraviolet (UV) to the infrared (IR) (Fig. 1A). The lower-bound on E_g is set by the bandgap of the parental bulk solid. Control over E_g is central to QD-based technologies, affecting solar energy harvesting, lighting and displays, lasers for telecommunications, sensing, metrology, imaging, and medical diagnostics.

Narrow linewidth, bright emission

In highly monodisperse cQD samples, the discrete, atomic-like structure of electronic states leads to a narrow ensemble emission linewidth of 20 to 80 meV at room temperature [defined as a full width at half maximum (FWHM)], which approaches a single-dot linewidth (*23*, *33*). This enables the high color purity needed to reach target performance in next-generation displays (*34*). Best cQD samples also achieve near-unity photoluminescence quantum yield (PLQY; the number of emitted photons per absorbed photon) (*35*). The narrowband and bright emission of the cQDs has been exploited in commercial televisions and displays. These features are also of interest for luminescent solar concentrators (LSCs), devices that act as large-area sunlight collectors for PV modules (*36*).

In lasing, the discrete character of QD electronic states is an important advantage. The sharp DOS concentrates oscillator strength into the desired ground-state transition, while a wide separation between quantized energy levels inhibits thermal depopulation of the emitting band-edge states (*7*). Present-day commercial QD lasers are realized by using S-K grown In(Ga)As/GaAs eQDs. cQD-based lasers are still under active development (*37*).

Tunable surface chemistry

QDs feature a large surface-to-volume ratio, making them sensitive to their environment. In cQDs, the surfaces are typically terminated with molecules or ions with different morphologies and functional groups (*27*, *38*). This offers a route to manipulate cQD interactions with their environment. cQDs can be tethered to proteins, antibodies, or other biologic species and used as optically addressable biolabels (*39*).

"Surface programming" offers an additional tool for manipulating energy levels and driving the assembly of cQDs into conductive semiconductor solids. Ligand exchange strategies seek to replace bulky molecules with shorter, more conductive ligands to enhance interdot coupling and facilitate charge transport (*27*, *40*). This process can be carried out while the cQDs are in a liquid solution or after they had been assembled into a solid film. It is necessary that these steps do not distort electronic surface passivation so as to prevent formation of intra-gap states that would compromise electronic and optical properties (*41*).

Tunable charge transport

The ability of cQD assemblies to pass current is determined by the ability of charge carriers to cross interparticle barriers (*42*). cQD solids exhibit modest charge carrier mobilities (typically below 10^{-1} cm^2 V^{-1} s^{-1}) compared with Si or epitaxial III-Vs semiconductors (10^2 to 10^3 cm^2 V^{-1} s^{-1}). Most cQD solids exhibit a complex interplay between carrier confinement, cQD interfacial properties, and electronic coupling. Enhanced cQD coupling for higher mobility is often accompanied by the increase

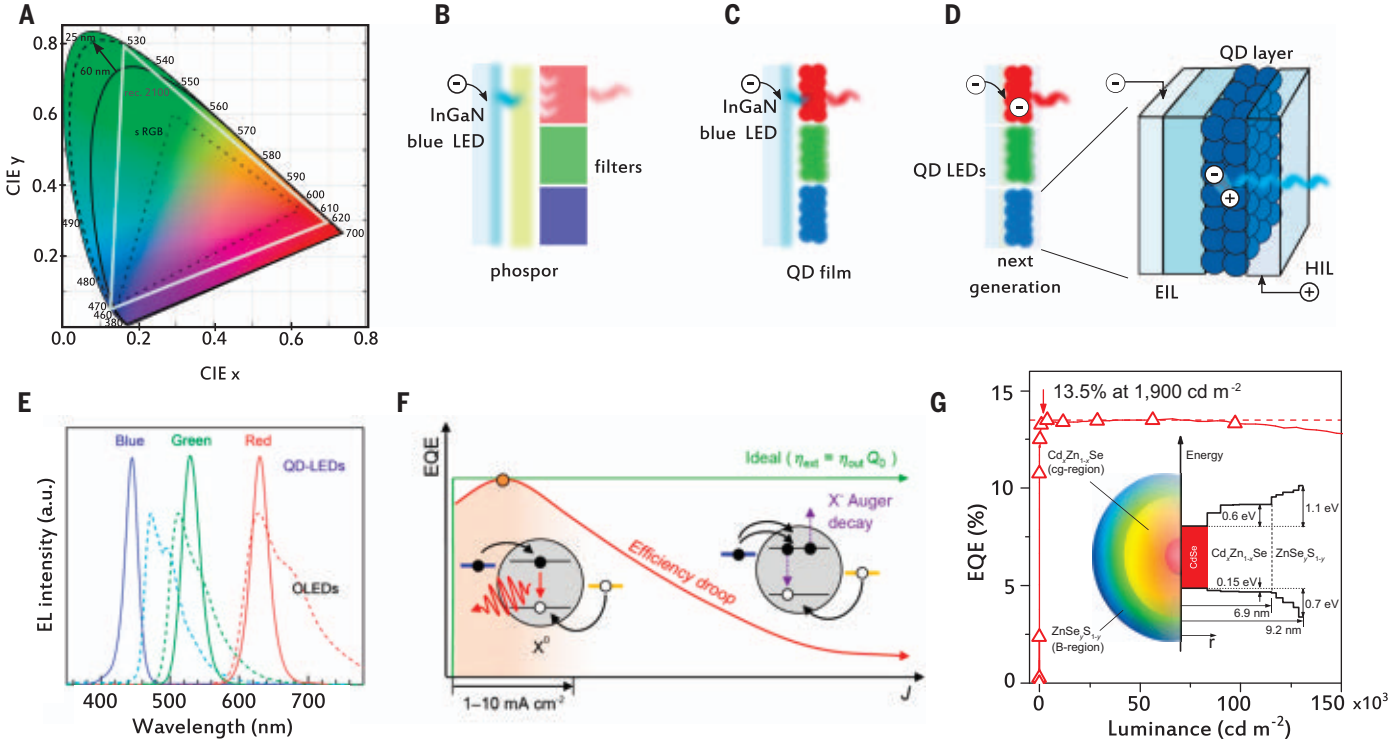


Fig. 2. QD materials for displays and lighting. (**A**) In displays, red, green, and blue colors are mixed to obtain entire color palette. Chromaticity diagrams are used to quantify the quality of color for displays. The triangles represent the color gamut standard for current (sRGB) and next-generation displays (Rec. 2100) (*34*). (**B** to **D**) Different QD-enabled display technologies in which QDs are excited either optically or electrically. EIL and HIL are electron and hole injection layers, respectively. (**E**) Illustration of high color purity of QD emission (solid lines) compared with those of OLEDs emitting at similar wavelengths (dashed lines) (*48*). (**F**) Efficiency droop in QD-LEDs. EQE roll-off at higher current densities (*J*) is typically a result of imbalanced electron and hole injections, which leads to accumulation of long-lived uncompensated charges and associated carrier losses due to nonradiative Auger recombination. (**G**) A virtually droop-free EQE of ~13.5% maintained up to high luminance levels of ~0.15 Mcd m^{-2} is realized by using cQDs with suppressed Auger recombination, achieved by grading the cQD composition (inset) (*52*).

of intra-gap tail states that shrink the electronic gap of the cQD solid. The passivating layers, added to preclude cQD fusion and suppress intra-gap states, often introduce charge-transport barriers between adjacent dots. By controlling these competing trends, mobilities up to 10 cm^2 V^{-1} s^{-1} for one type of carrier have been realized (*43*, *44*). Recently, encouraging progress has been achieved for cQD solids showing balanced charge transport for both polarities (*n* and *p*) that preserve discrete quantum-confined electronic states (*45*).

QD technologies and challenges

QD materials for displays and lighting

The narrowband cQD emission represents a competitive advantage compared with other semiconductors for the generation of pure colors, a requisite for next-generation displays (Fig. 2A). Displays can use cQDs either as color-converting phosphors excited by light-emitting diodes (LEDs) or as active electroluminescent materials directly driven by an applied bias. In the first mode (Fig. 2B), a polymer composite containing red- and green-emitting nanocrystals is combined with a backlight unit frame of blue InGaN LEDs, incorporated into a liquid-crystal display architecture as an RGB (red-green-blue) backlight (*46*). This approach offers improved color gamut and reduced light losses during color filtering as compared with those of traditional white LED backlight approaches. In another scheme, patterned cQDs are used as a photoactive material, absorbing short-wavelength blue light and re-emitting light of longer-wavelength blue, green, and red colors (Fig. 2C). This eliminates the need for separate color filters, removing color cross-talk (*47*); reduces the number of layers in the device stack; enhances the viewing angle; and increases the light output and device efficiency.

In cQD-based electroluminescent structures, cQDs are used to implement RGB LEDs that are addressed electrically (Fig. 2D). This approach can help reduce screen thickness, enhance dynamic range, improve black-color rendering, and increase viewing angle and frame rates. Compared with organic LEDs (OLEDs), cQD-based LEDs offer narrower emission linewidths (<30 nm versus >60 nm for OLEDs) (Fig. 2E) and correspondingly higher color purity, as needed to meet Rec. 2100 color gamut specifications (*34*, *48*).

In QD-LEDs, a cQD active layer is sandwiched between electron and hole injecting layers (Fig. 2D). An important LED characteristic is the external quantum efficiency (EQE)—the ratio of the number of emitted photons to the number of injected electrons. The prerequisites of high EQEs are a high PLQY and good balance between electron and hole injection currents to avoid CQD charging because the formation of charged excitons promotes nonradiative Auger recombination (*49*, *50*). During Auger decay, the electron-hole recombination energy is released not as a photon but instead transferred to the resident charge carrier (Fig. 2F). Auger recombination has been identified as at least one of the reasons for EQE droop—a decrease in device efficiency with increasing current density. This creates problems even in the case of standard displays operating at low-to-moderate brightness (500 to 1000 cd m^{-2}); and becomes a serious challenge for outdoor systems whose brightness should be comparable with or greater than that of natural sunlight (5000 cd m^{-2}).

The droop problem has been tackled with both device optimization and cQD structure control. Compositionally graded cQD multishell heterostructures have been shown to impede Auger recombination because of creation of a "smooth" confinement potential that suppresses

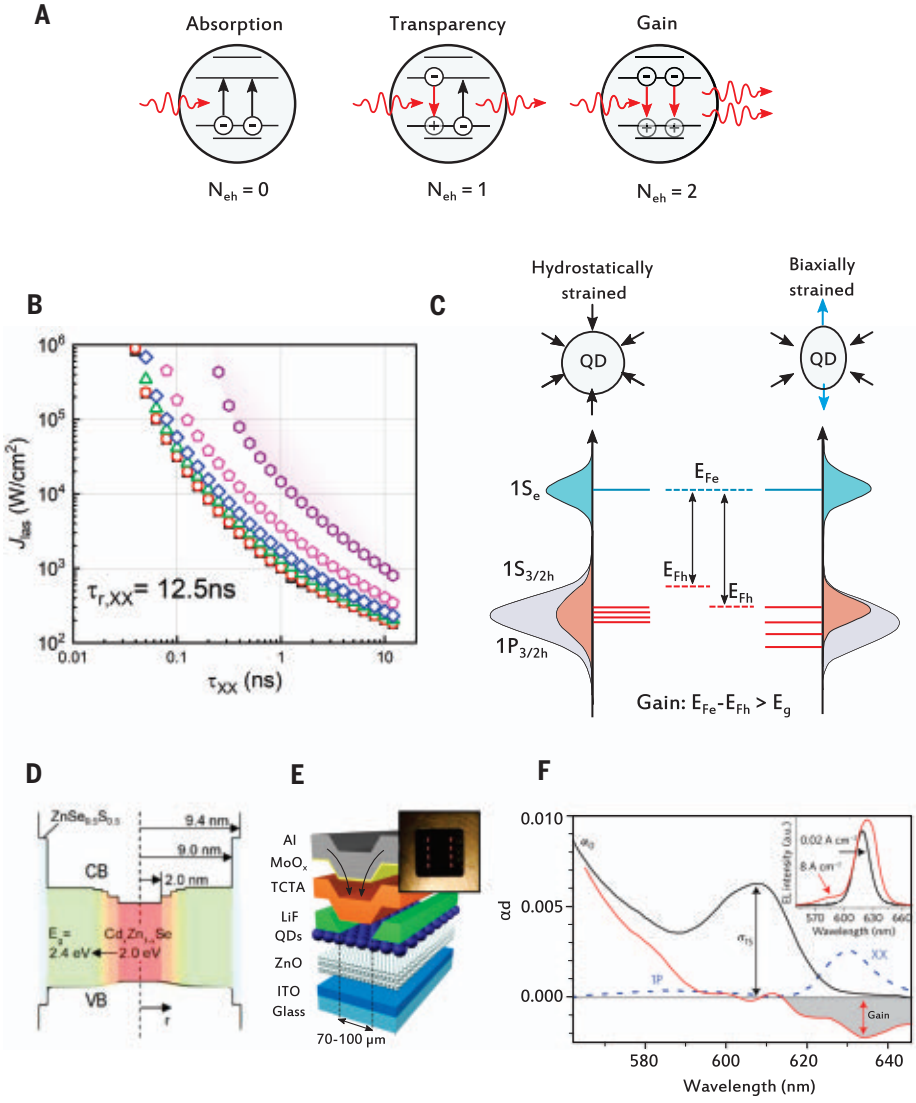


Fig. 3. Principles of QD lasing and recent advances. (**A**) Optical gain in QDs originates from biexcitons; hence, suppression of Auger recombination is critical to realize lasing, especially in the case of cw optical and dc electrical pumping (*63*). (**B**) Modeling indicates strong dependence of the cw lasing threshold on biexciton lifetime ($\tau_{XX} = \tau_{r,XX}\tau_{A,XX}/(\tau_{r,XX} + \tau_{A,XX})$). Here, $\tau_{r,XX}$ and $\tau_{A,XX}$ are, respectively, the radiative and the Auger lifetime of a biexciton (*65*). (**C**) Increasing splitting between the light and heavy hole states owing to biaxial strain leads to the reduction of the optical gain threshold, which facilitates the realization of cw lasing (*69*). (**D**) Continuously graded cQDs (cg-cQDs) exhibit strong suppression of Auger decay, which increases the biexciton emission efficiency and extends optical gain lifetimes. (**E**) A current-focusing LED architecture helps boost the current density to levels sufficient for achieving population inversion in a cQD active layer. (**F**) Light amplification by using cg-cQDs incorporated into a current-focusing LED (*51*).

the intragap transition involving the energy-accepting carrier (*51*). This enabled red-emitting LEDs with high droop-free EQEs up to ~100,000 cd m^{-2} (Fig. 2G) (*52*). cQD surface modifications have been also pursued to facilitate balanced charge injection, enabling green-emitting LEDs with brightness >400,000 cd m^{-2} (*53*). In addition to improving EQEs and brightness, good charge balance helps reduce heat generation by suppressing Auger decay, which is essential to maintaining high EQE stability and extending device longevity (*54*). Substantial strides in cQD-LED performance have led to EQEs near the limit defined by the light extraction efficiency from a high-index semiconductor medium, exceeding 20% for red and green colors and 18% for blue (*55*–*57*).

The realization of efficient IR cQD LEDs—desired for technologies such as optical telecommunications, biological imaging, and chemical sensing—had previously been hindered by difficulty in obtaining simultaneously high-IR PLQY and highly efficient, balanced charge injection (*58*). The advent of cQD-in-perovskite solids—cQDs embedded in a metal halide perovskite—provided an avenue to resolving these problems (*59*). These materials take advantage of ambipolar charge transport of perovskites and spectral tunability of the cQDs. The cQDs and the perovskite matrix form a defect-free epitaxial junction with a band alignment that can be tuned to facilitate carrier injection into the cQDs. This materials platform led to high-performance IR LEDs with good power conversion efficiencies (PCEs) and high brightness (*60*). The use of quantum-confined perovskite matrices, in which charges were injected as excitons, led to improved charge balance and helped further increase the EQE and the brightness (*61*). An alternative QD-in-matrix approach used inorganic bulk-heterojunction solids implemented by using percolated networks of PbS and ZnO nanocrystals to achieve a PCE of 9.3% (*62*).

QD lasers

Semiconductor lasers are sources of coherent light applied in numerous technologies, including optical communications, on-chip interconnects, digital projection systems, manufacturing, surgical instruments, metrology, and emerging quantum information technologies. Lasing requires population inversion in which the occupancy of a higher-energy state of the emitting transition exceeds that of a lower-energy state. For QDs with twofold-degenerate electron and hole band-edge states, the onset of population inversion and optical gain occurs when the average number of electron-hole pairs per-dot is one ($\langle N_{eh} \rangle = 1$) (*63*). This corresponds to the regime of optical transparency or optical-gain threshold when absorption and stimulated emission exactly compensate each other (Fig. 3A).

To enact optical gain, at least a fraction of the QDs in the sample must contain two or more excitons, implying that optical amplification in QD media relies on biexcitons and other higher-order multiexcitons. This greatly complicates the realization lasing because of the extremely fast deactivation of optical gain through nonradiative Auger recombination (*63*, *64*). Fast Auger decay represents an especially serious obstacle for realizing continuous-wave (cw) lasing. In particular, in the case of small-size standard (nonengineered) cQDs with <100-ps Auger lifetimes, it leads to prohibitively high lasing thresholds of around 10^5 to 10^6 W cm^{-2} (Fig. 3B) (*65*). As a result, most experimental demonstrations of cQD lasing used pulsed optical excitation (*37*).

The efforts to suppress Auger recombination explored "giant" CdSe/CdS cQDs with a thick shell to increase the spatial extent of electronic wave functions (*66*) and interfacial alloying to realize a "smooth" confinement potential (*67*, *68*). These approaches enabled a considerable reduction of lasing thresholds in

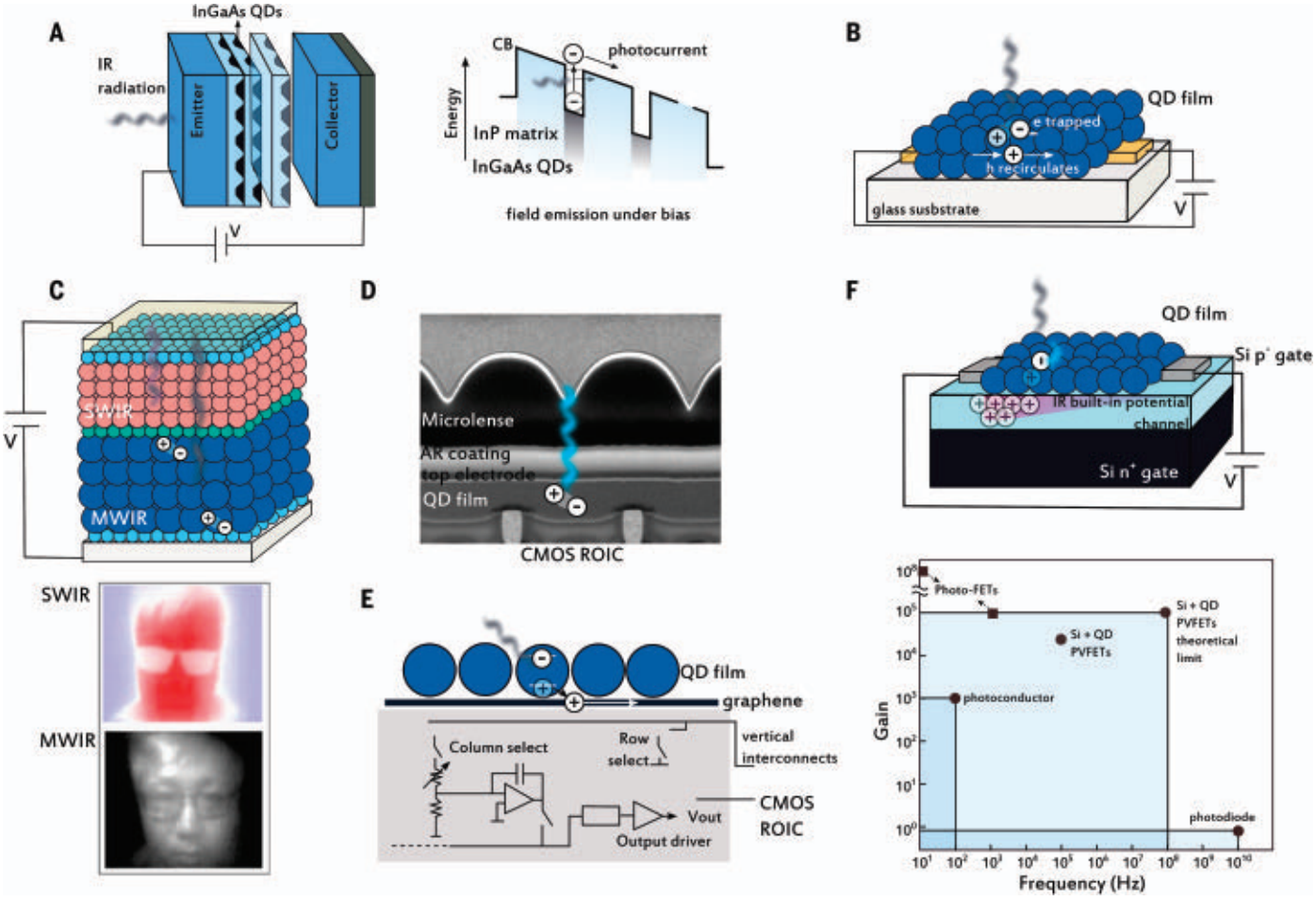


Fig. 4. QD materials for sensing. (**A**) Early QD photodetectors relied on field emission from III-V eQDs embedded into a wider-bandgap InP matrix where charge transport occurred. (**B**) The programmed surface modification and assembly of cQDs led to photoconductive cQD solids with high SWIR sensitivity (*80*). (**C**) Photodetection at the MWIR is challenging and typically achieved with HgTe cQDs. Shown here, a cQD back-to-back photodiode enables monolithic detection of SWIR and MWIR bands (*85*). (**D**) cQDs have been incorporated into Si-based imaging technologies, enabling new sensing architectures such as top-surface photodetection: cQDs sensitize a Si CMOS readout-integrated-circuit (ROIC) (*93*). (**E**) The combination of cQDs with high-mobility materials such as graphene can provide large photoconductive gains. The integration of this sensing platform into CMOS has been recently reported (*95*). (**F**) cQDs have also been monolithically integrated with Si in a photovoltage transistor modulating charge transport in Si by use of an IR cQD-generated photovoltage (*96*).

the pulsed regime (*65*) and allowed for demonstrating cw lasing with pump powers of a few kilowatts per square centimeter (*69*). The latter study also took advantage of strong biaxial strain realized by means of epitaxial facet-selective growth of CdS on CdSe cores. This increased the energy splitting between valence-band states and helped impede thermal depopulation of the band-edge hole level involved in lasing (Fig. 3C).

The development of compositionally graded core-multishell cQDs have stimulated further advances in the lasing field (Fig. 3D) (*37*, *51*). The electronic band alignment in these structures increases the spatial overlap between electron and hole wave functions and thus boosts the emission rate compared with quasi–type II CdSe/CdS systems. The radially graded composition also leads to strong suppression of Auger decay, which extends optical gain lifetimes. These properties have been exploited to demonstrate electrically driven optical gain in "current-focusing" LEDs (Fig. 3, E and F) and to realize dual-function devices with an integrated optical cavity that operated as an optically pumped laser and a standard LED (*70*). These are major milestones on the path to electrically pumped cQD laser diodes.

Whereas cQD lasers are still at the stage of exploratory devices, lasers based on eQDs have already reached technological maturity (*71*). In particular, record-low thresholds (*72*) and high operating temperatures of up to 220°C (*73*) have been achieved by using III-V eQDs. eQD lasers have become a key component in silicon photonics, optical interconnects, telecommunication, and data centers (*74*).

QD materials for optical sensing and imaging

The wide spectral tunability of QD absorption, combined with good mobility of photogenerated charge carriers, make them competitive candidates for optical sensor applications for wavelengths beyond the Si bandgap. Early QD sensors relied on InGaAs eQDs grown on InP, in which photoexcited electrons were emitted into the InP transport matrix with the help of an electric field (Fig. 4A) (*75*). eQD IR detectors were then considered a compelling sensing platform of IR light (*76*), demonstrating low dark current and high sensitivity and speed (*77*, *78*).

The high cost/performance tradeoff of existing IR sensors based on bulk narrow-gap semiconductors opened additional opportunities for cQDs, which have the benefit of readier integration with Si chips. Initial cQD IR sensors consisted of PbS cQDs embedded in a polymer matrix, in which a type II heterojunction provided selective charge circulation (*79*). Surface modification of cQDs enabled electrically coupled cQD solids with increased IR sensitivity

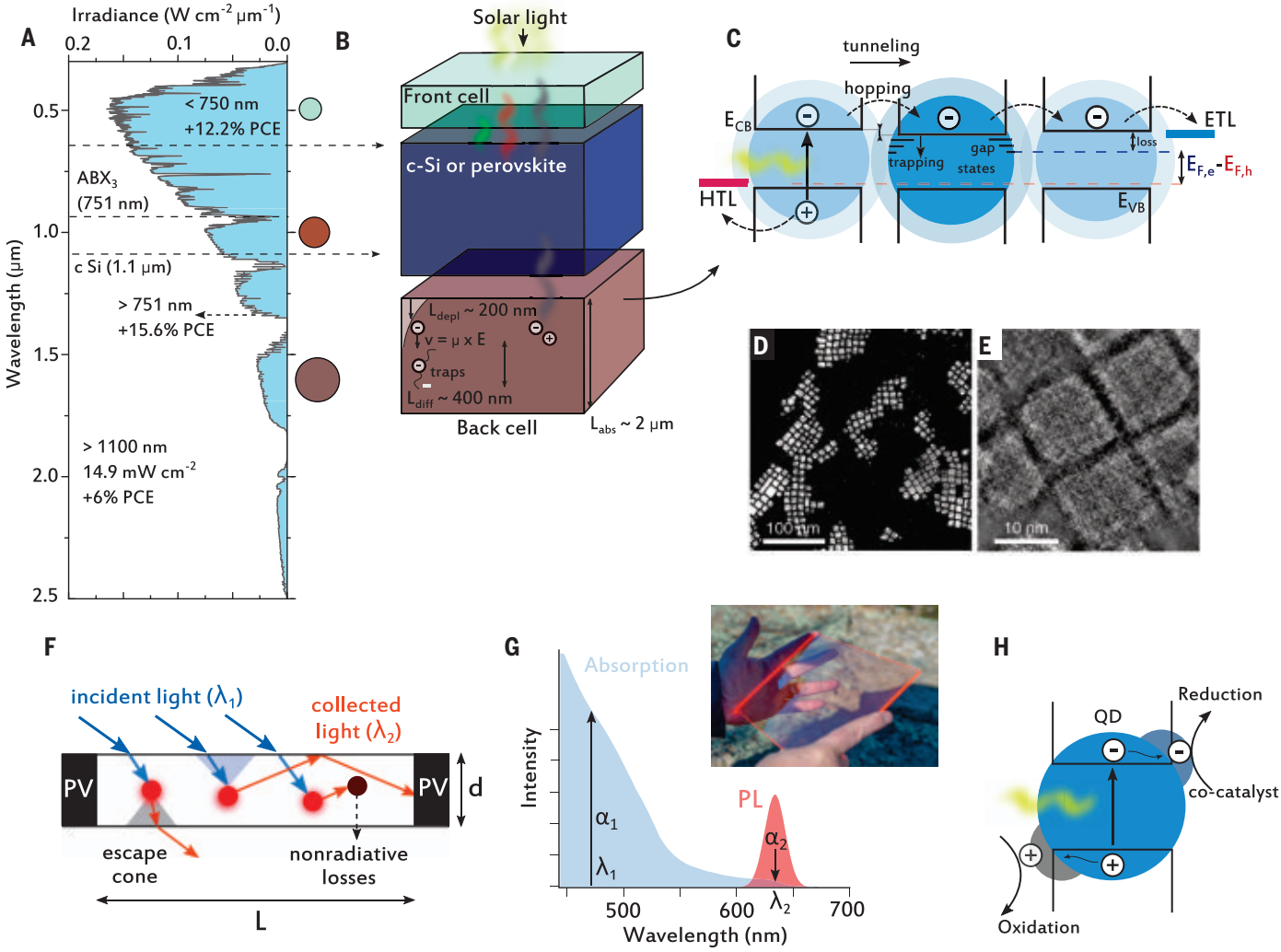


Fig. 5. QD semiconductors for solar energy harvesting. (**A**) Solar irradiance spectrum and cumulative power. (**B**) cQD absorption can be tuned across the entire solar spectrum, which is of interest for lightweight standalone solar cells or tandem cells to increase the net PCE of other technologies such as perovskites (additional +15.6% PCE) and cSi (+6% PCE). (**C**) Efficient photovoltaic operation in QD solids requires minimization of defects pining quasi-Fermi level splitting and open-circuit voltage and sufficient QD coupling to facilitate charge transport to electron (ETL) and hole (HTL) transport electrodes. (**D** and **E**) Perovskite cQDs have emerged as a strategy to achieve high-quality perovskite solids with stabilized composition and record PCE (*113*). (**F**) In an LSC, cQDs absorb incident sunlight and then reemit lower-energy photons that are captured by means of total internal reflection in an LSC slab and waveguided toward edge-mounted solar cells. (**G**) The LSC performance can be enhanced by increasing the LSC quality factor (Q_{LSC}), which is defined as the ratio of the absorption coefficients for incident (α_1) and reemitted (α_2) photons ($Q_{LSC} = \alpha_1/\alpha_2$). [Inset courtesy of UbiQD.] (**H**) cQDs are candidates for photo- and electrocatalysts to store solar energy in the form of chemical bonds.

(Fig. 4B) (*80*). These devices exhibited high photoconductivity and specific detectivities (D^*; noise-equivalent power normalized to measurement bandwidth and device area) comparable with those of III-V technologies. The high photoconductive gain—while one type of photocarrier is trapped, the other recirculates many times leading to gain—came with the downside of high dark currents and time response limited to approximately milliseconds.

Unlike photoconductors, photodiodes rely on the collection of minority carriers, opening the door to lower dark currents and faster time response. Advances in surface chemistry enabled sensitive short-wavelength IR (SWIR) cQD-based photodiodes with 3-MHz 3-dB bandwidths but no gain (*81*).

Sensing of mid-IR wavelengths—of interest for spectroscopy, gas and health monitoring, aerospace, and thermal imaging—is more challenging. As the semiconductor bandgap decreases, it becomes closer to background radiation and transport energy barriers. This increases dark current and noise and makes charge collection by using conventional electronics more difficult. The materials choice for IR sensing is reduced to few candidates in the HgCdTe family, InSb, and III-V epitaxial quantum-well structures exploiting intraband transitions that demand complex fabrication and low temperature for efficient operation. Compared with other technologies, such as bulk HgCdTe photodiodes or Schottky photoemissive sensors, cQD-based IR sensors offer the prospect of lower dark currents, higher-temperature operation, and higher detectivity.

Improvements in colloidal synthesis of HgTe cQDs, doping, and surface engineering led to the demonstration of mid-IR sensors with room-temperature photoresponse beyond 5 μm (*82*–*84*). Multispectral IR detectors have been demonstrated by using two opposing

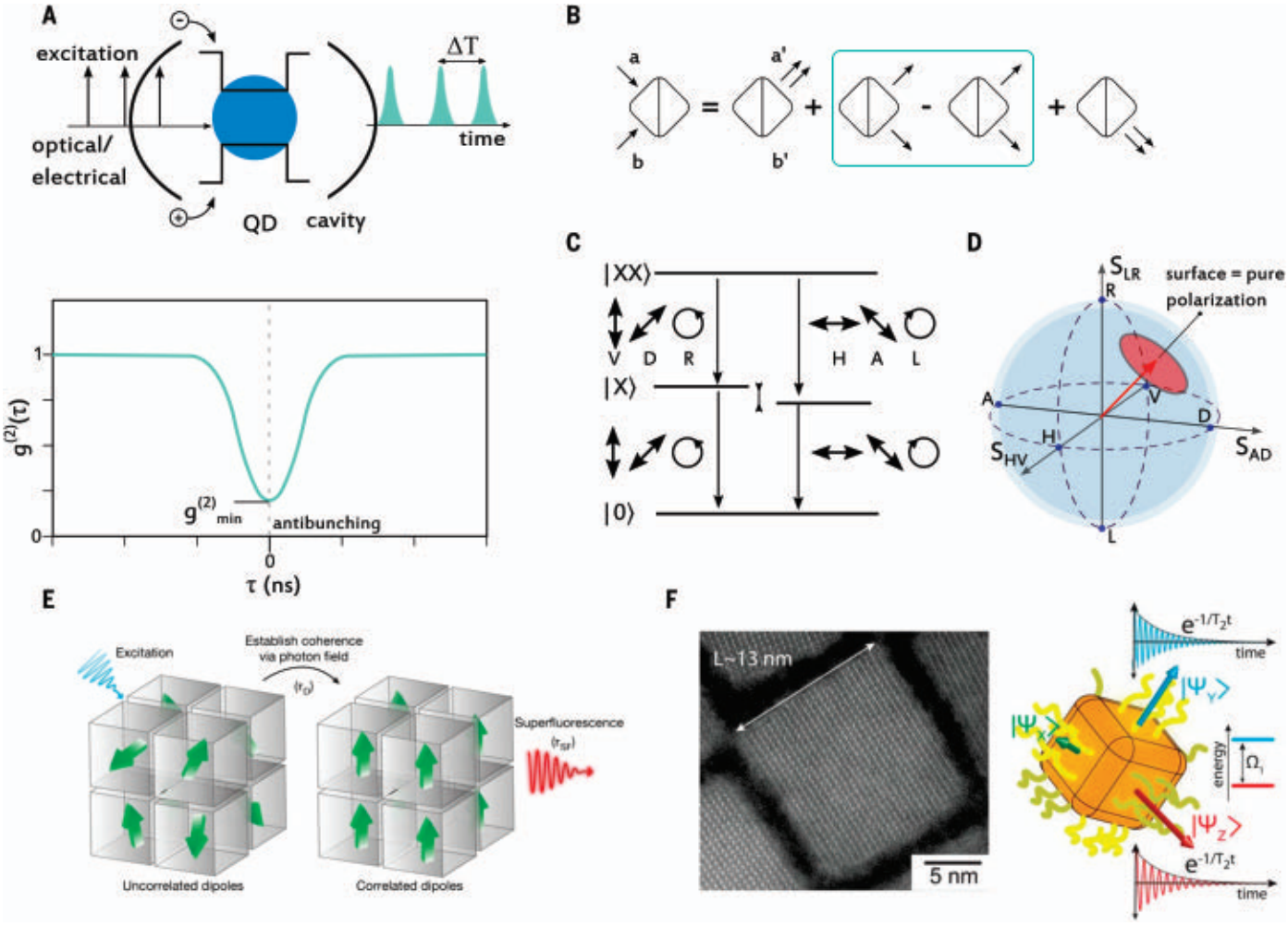


Fig. 6. QD materials for quantum light generation. (**A**) (Top) A single-photon source based on a QD coupled to a resonant cavity. (Bottom) The second-order intensity correlation function (g^2) exhibits an "antibunching dip" at time zero, a signature of high single-photon purity. (**B**) If two indistinguishable photons simultaneously reach a beam splitter, they leave it along the same optical path as a photon pair. This experiment is used to assess the degree of photon indistinguishability. (**C**) Generation of polarization-entangled photon pairs through a "radiative cascade" produced by decay of a biexciton. (**D**) The degree of entanglement can be assessed with polarization tomography. Coherent interaction would lead to states on the surface of the Poincare sphere. (**E**) Superfluorescence from colloidal $CsPbX_3$ ensembles (*169*). (**F**) Single-photon emission from $CsPbX_3$ QDs has been achieved with coherence times of tens of picoseconds (*155*).

Schottky junctions of cQDs with different bandgaps (Fig. 4C) (*85*). IR intraband photodetectors have also been demonstrated with doped cQDs (*86*–*88*) or mixtures of different cQDs (*89*).

Solution-processing of cQDs opened the door to their integration with incumbent visible sensor platforms such as Si complementary metal-oxide semiconductor (CMOS) readout electronics. In standard CMOS-based cameras, readout circuitry and photoactive elements coexist within the same layer, limiting the photodetection area (fill-factor) to ~30%. Top-surface integration of cQDs onto CMOS chips led to improved sensors with 100% fill-factor (Fig. 4D). The direct bandgap of cQDs also allowed for thinner photoactive films (~600 nm), reducing signal crossover from adjacent pixels compared with silicon back-surface–illuminated sensors.

The integration of cQDs with materials that have high mobilities, such as metal halide perovskites, and 2D materials such as graphene and transition-metal chalcogenides enabled decoupling sensitization (light absorption) and charge transport (*90*). In QD-in-perovskite solids, it has been shown that one can combine the field emission of photoexcited charges from PbS cQDs into the perovskite matrix, along with ensuing charge transport within the higher-mobility perovskite matrix (*91*). cQDs and 2D materials have been incorporated in a phototransistor architecture (Fig. 4E). As photoexcited electron-hole pairs in the cQD layer split, one of the charges drifts into the 2D transport layer, where it recirculates multiple times before recombination, leading to photoconductive gain (*92*). This mode of operation enables larger photoconductive gain than conventional cQD solids (because of faster charge recirculation) without compromising time response (*93*). However, it requires special integration with current CMOS technologies (*94*). CMOS sensor arrays modified to include graphene and PbS cQDs have achieved high on-chip detectivity from the visible to the IR (Fig. 4E) (*95*).

cQDs have also been integrated into Si transistors, modulating Si transport by means of a photovoltage effect (*96*). Light absorption within the PbS cQD layer leads to a modification of the depletion region in the Si layer, controlling transport in Si and extending its photoresponse up to 1500 nm. A comparison of performance

characteristics for different sensor architectures points to a favorable combination of gain and bandwidth in these devices (Fig. 4F).

The highly tunable emission of cQDs was exploited in alternative sensing schemes. Upconversion devices transduce lower-energy photons (such as IR) into visible light, which is then directly imaged (*97*, *98*); down-conversion devices convert high-energy radiation into lower-energy photons, which is of interest for detection of radiation such as x-rays. $CsPbBr_3$ perovskite cQDs have been used to realize inexpensive flexible scintillators with high performance (*99*). cQD tunability has also been exploited to realize fast spectrometers (*100*).

QD materials for solar energy harvesting

Solar energy harvesting is widely achieved by using photovoltaic Si modules, with impressive levelized costs below 2¢/kilowatt-hour (kWh) in well-isolated geographic regions, thus approaching grid parity. The high crystallinity of eQDs offers a path toward high-efficiency multijunction solar cells. cQDs can be deployed in single-junction lightweight solar cells or in tandem configurations to complement existing PV technologies (Fig. 5, A and B) and used as the photoactive material in LSCs and photocatalytic systems.

cQD solar cells

In cQD photovoltaic devices, the energy of photogenerated excitons is harvested in the form of electrons and holes, which are collected and used to generate an electric current. Early cQD solar cells relied on a dye-sensitized configuration as organic dye replacements. Electron-hole pairs were harvested by means of redox reactions and charge transport facilitated through TiO_2 and electrolyte. Initial solid-state cQD PV devices used PbS cQDs embedded in a conducting polymer to enable charge transport (*101*).

Improvements in charge transport in cQD solids opened the door to cQD active layers (Fig. 6C) (*102*). This was first implemented by using a Schottky junction configuration, in which the built-in field at the cQD-metal interface facilitated charge extraction. This was followed by other architectures such as the depleted heterojunction, which maximized field-assisted charge collection (*103*).

Higher-performance cQD solar cells would still require efficient charge extraction at near-flatband, maximum–power point conditions in which charge transport is diffusion-based rather than field-assisted. Because the diffusion length (L_{diff}) of minority carriers was shorter than the length required to maximize light absorption, this led to an absorption-extraction compromise. Extended L_{diff} required fewer defects in the cQD solid, to increase carrier lifetime and mobility, and avoidance of quasi-Fermi level and open-circuit voltage (V_{oc}) pinning—also curtailed by cQD polydispersity and disorder (*104*, *105*).

Advances in cQD quality and PV PCE have been achieved through improvements in synthesis; ligand exchanges relying on small metal halides to improve cQD coupling, protect against oxidation, and reduce the density of defects (*106*); doping and surface dipole engineering (*107*); device architectures (*103*); and light management (*108*).

The implementation of cQD bulk heterojunction (BHJs)—percolating paths of cQDs forming a type II heterojunction (*109*)—offered a route to extend L_{diff} by separating electron and hole transport, reducing recombination. One challenge is the retention of the desired cQD configuration once the final solid is assembled. BHJs are also attractive to reduce V_{oc} deficit through DOS modification and remote passivation (*62*).

The introduction of single-step solution-phase ligand exchanges enabled to preserve cQD monodispersity in the final solid, leading to a more uniform energy landscape, reduced bandtail states, and improved charge transport, resulting in higher V_{oc} and PCE > 11% (*110*). Further improvements in cQD order and coupling achieved through perovskite bridging have been shown to boost PCE to >14% (*111*).

Advances in device architecture have led in parallel to similar PCEs. These exploited the combination of PbS cQDs with small-molecule organic semiconductor layers to complement cQD absorption and enhance charge extraction (*112*).

Rapid progress has been made in recent years using metal halide perovskite cQDs (Fig. 5, D and E). The use of presynthesized high-quality $CsPbX_3$ cQDs as the precursor to perovskite solids—as opposed to an on-substrate perovskite crystallization—has enabled control over perovskite phase stability. The lower density of electronic defects in perovskite cQD solids compared with PbS led to PCEs that reached 16.6% (*113*–*115*).

The tunable bandgap of cQDs can be exploited to augment the PCE of other PV architectures by harvesting IR light. This strategy can lead to up to +6 and +12% additional PCE points when combined with Si or perovskites, respectively (Fig. 5A). To date, PbS cQD:perovskite tandems have achieved a 24.7% PCE when combined in a four-terminal configuration (*116*).

Alternative strategies to improve PV performance include solar spectra reshaping, in which cQDs absorb and reemit light at a region of interest. $CsPbX_3$ cQDs doped with Yb^{3+} ions can efficiently absorb blue light and reemit in near-IR, with quantum efficiency approaching 200% (*117*).

Luminescent solar concentrators

LSCs are light-management devices envisioned as large-area sunlight collectors for building-integrated solar cells (*118*). In LSCs, light is absorbed by fluorophores embedded in an optically transparent slab. Excited fluorophores reemit lower-energy photons, which are guided by means of total internal reflection to slab edges to be collected by PV modules (Fig. 5F). If the light-collecting area of the LSC is greater than the area of its edges, the output photon flux density (ϕ_{out}) can exceed the incident flux density (ϕ_{in}), concentrating light. In contrast to traditional lens- and mirror-based concentrators, LSCs can operate equally efficiently for direct and diffuse light, making them well suited as large-area sunlight collectors for building-integrated PVs installed as solar windows and solar sidings (*119*).

An important performance-limiting factor of LSCs is light reabsorption by the fluorophores themselves, which restrict the maximum device size. The interplay between sunlight-harvesting ability and losses due to reabsorption can be quantified by an LSC quality factor ($Q_{LSC} = \alpha_1/\alpha_2$), which is defined as the ratio of the absorption coefficients for the harvested (α_1) and the reemitted (α_2) light (Fig. 5G). The maximum concentration factor ($C = \phi_{out}/\phi_{out}$) obtained in the large-area limit is approximately equal to Q_{LSC}, highlighting the importance of achieving $Q_{LSC} \gg 1$, which can be realized with engineered cQDs (*120*). An additional requirement is spectral matching between fluorophore's emission and PV absorption. In the case of Si PVs, this implies that for optimal LSC operation, cQDs should combine efficient NIR emission and a spectrally displaced NIR absorption onset.

Initial efforts to tackle reabsorption explored giant CdSe/CdS cQDs. A thick CdS shell served as a light-harvesting antenna that funneled photogenerated carriers into a small emitting CdSe core (*121*). Because the bandgap of the core is smaller than that of the shell, reemitted light is not attenuated by absorption arising from a large-volume shell. This approach enables Q_{LSC} of more than 100 and leads to high concentration factors exceeding ~60 (*120*, *122*).

The amount of sunlight absorbed by giant CdSe/CdS cQDs is limited by the high bandgap of the shell. This problem has been addressed by using narrower-gap I-III-VI $CuInSe_xS_{2-x}$ cQDs (*123*, *124*). In addition to improved sunlight harvesting, these structures exhibit low reabsorption because of a peculiar light emission mechanism that involves an intragap hole-like state (*125*). Further improvements in the LSC efficiency have been obtained by exploiting spectral splitting in tandem devices implemented by using a combination of II-VI and I-III-VI cQDs (*126*).

Following principles exploited in LSCs, cQDs can be used for spectral reshaping of incident sunlight for applications in agriculture to match the absorption of photoactive molecules and thereby boosting crop growth. This approach

is presently being tested for greenhouse and indoor farming (*127*).

Photo- and electrocatalysis

Storage of renewable energy as chemical bonds—for example, transforming greenhouse gases or pollutants into fuels and chemical feedstock—is a path toward carbon-neutral energy systems (*128*). In this context, cQD materials might enable photon-to-chemical energy conversion across the solar spectrum, combining benefits of heterogeneous and homogeneous catalysis (*129*). cQDs can be used as standalone photocatalyst or as the sensitizing agent of metal catalytic sites.

In a photocatalytic system, excited electron-hole pairs in a semiconductor are directed from conduction and valence bands to catalytic sites to drive reduction and oxidation reactions, respectively (Fig. 5H). The cQD large surface-to-volume ratio offers a path to increased reaction rates. Strain and defect engineering was shown to increase cQD photocatalytic activity (*130*). Energy positioning is interesting to match the potentials of different reactions and control selectivity and to drive electrons and holes toward different reaction sites, minimizing product recombination. cQD surfaces can be manipulated to influence the interaction between solvents and adsorbates (*131*).

cQDs have been used in different photocatalytic systems, such as H_2/O_2 evolution (*132*) and CO_2 reduction (*133*), lignocellulose (*134*) and plastic (*135*) reforming, ammonia generation (*136*), and water purification (*137*). cQDs have also been implemented in hybrid strategies to sensitize living bio-organisms, enabling the production of CO_2-upgraded feedstock (*138*).

Carbon-based cQDs are an attractive route to realize metal-free photocatalysts. Their optical and electronic properties can be widely manipulated to control their size, shape, and doping (*139*). Their chemical inertness brings benefits in aqueous reactions at extreme pH conditions. Quantum-confined transition-metal dichalcogenides have also shown a path for water splitting and CO_2 reduction (*140*), among other reactions.

In water-supported reactions, single-junction photocatalytic systems must overcome the energy gap and overpotentials required to split water (>1.23 eV) and outcompete product recombination. This can be achieved through two-step excitation and Z-scheme energy transfer in cQD heterojunctions and using tandem photoelectrochemical (PEC) systems, in which redox reactions take place separately at the photocathode (anode). The design of photoelectrodes follows the same logic as that of photovoltaic systems but brings added challenges of cocatalyst integration and more demanding chemical stability. Further progress in PEC performance and system cost are still required for this technology to have impact commercially (*141*).

Pure electrochemical systems have the advantage of separate optimization of PV and electrocatalyst modules. cQD heterostructures might enable a design platform to tailor the physical, chemical, and electronic properties of catalysts—which undergo extensive surface reconstruction through oxidation and reduction from their initial configuration as they are operated.

QD materials for quantum light generation

The development of QD technology for quantum computers and quantum communication is of growing interest. Quantum technologies are desirable for fast computation and secure communication (*142*). The artificial atom-like features of QDs triggered efforts to use them as quantum technology hardware, leveraging potential advantages such as ease of miniaturization, scalability, and integration.

Although the coherence properties of QD quantum bits are superior to higher-dimensional semiconductors (*143*–*148*), they so far have remained considerably behind those of atoms or other solid-state systems such as defect centers. For optically active excitons in self-assembled QDs, the coherence time can be as long as nanoseconds; for optically inactive excitons, it may reach microseconds, similar to spins.

QDs are attractive as quantum light sources, providing emission of single as well as entangled photons with high fidelity (*149*). Excellent performance parameters have been achieved, mostly by using eQD structures thus far. A key factor in that respect is the "silencing" of the quantum emitter environment—for example, by suppressing lattice vibrations and charge fluctuations. The first can be achieved with cryogenic cooling, whereas the second requires high material quality and separation from surfaces and surfactants located therein.

Single-photon sources

A QD in a high-quality optical resonator cavity is the basic unit of a single-photon source (Fig. 6A) (*150*). After tailored pulsed excitation, the QD will ideally emit one and only one photon, which is called antibunching. The quality of antibunching can be characterized by measuring the second-order correlation function $g^{(2)}(\tau = 0)$, which should reach zero in case of perfect operation because it gives the probability of detecting simultaneously two photons (Fig. 6A, bottom). Over the years, the continuous improvement of In(Ga)As/GaAs eQD materials (*151*, *152*) has led to the suppression of $g^{(2)}(\tau = 0)$ to less than 10^{-4} (*153*). High-brightness In(Ga)As/InAs eQD single-photon sources have been accomplished by using optical (*154*, *155*) or electrical (*156*) excitation with record-high operation frequencies of up to 1 GHz. Single-photon sources are usually operated at cryogenic temperatures, but operation at room temperature or above is desired for practical quantum integrated circuit systems. On the basis of the large biexciton binding energy (>60 meV) realized in GaN eQDs embedded in a GaN/AlGaN nanowire, single-photon emission was achieved at 350 K (*157*).

In addition, photons emitted during a sequence of pulsed excitation should be indistinguishable (*158*), a requirement also for photon-based quantum simulators and computers that exploit Fock number states with a well-defined number of identical photons. Indistinguishability concerns their energy, polarization, and spatiotemporal mode structure and can be tested for two photons by means of a "which-path" experiment (Fig. 6B): When two indistinguishable photons simultaneously reach a 1:1 beam splitter, they leave this splitter only in pairs. This is due to destructive interference of the transition amplitudes for the two photons leaving the beam splitter through separate ports (*159*). Photon indistinguishability exceeding 98% has been achieved for eQD devices (*160*).

Sources of entangled photons

QDs can generate entangled photons (for example, by using their polarization degree of freedom through the biexciton cascade recombination) given a sufficiently small splitting of the polarized exciton states so that the photons cannot be distinguished through their energies (Fig. 6C) (*161*). The biexciton is a zero-angular-momentum eigenstate, so the polarizations of the two photons from the biexciton decay cancel. On the other hand, each photon can have either of the two possible complementary polarizations, leading to entanglement (Fig. 6D) (*162*–*165*). The fidelity of entanglement generation has already exceeded 98% (*166*). Efforts have been also made to generate cluster states with a photon number exceeding two (*167*).

Remaining goals include further enhancement of photon source and detector performance parameters and integration into nanophotonic circuits. Currently, operation has been limited to cryogenic temperatures. Another priority to achieve greater impact is the demonstration of similarly high-fidelity performance in the fiber-based telecommunication wavelength range, which is around 1.5 μm.

Colloidal QDs for quantum light generation

QD quantum light sources have largely relied on eQDs, even though one of the very first demonstrations of antibunching involving QDs used colloidal nanocrystals (*168*). Ordered superlattices of $CsPbX_3$ perovskite nanocrystals have been shown to achieve superfluorescence light generation (*169*), potentially opening the door to the implementation of multiphoton entangled quantum light sources (Fig. 6E).

Individual $CsPbX_3$ QDs have been used to demonstrate highly tunable single-photon sources (*155*) that, combined with their solution processability, results in a compelling material platform to serve as the building blocks of next-generation quantum light sources (Fig. 6F).

Challenges and outlook

Furthering technological impact of QDs will require continued advances on multiple fronts, including QD synthesis and assembly, integration with existing technological platforms, and the development of effective QD-specific device designs.

From a synthetic perspective, the realization of high-quality cQD materials based on solution-chemistry syntheses that transition toward widely available, inexpensive precursors and solvents is of increasing importance (*170*, *171*). Scaling up synthesis processes is required to meet the large-volume material demands of markets in consumer electronics, wearable devices, displays, and energy generation and storage. Broad adoption of cQD devices will put additional pressure on reducing the manufacturing costs of highly monodisperse cQDs.

The cost to volume and availability of precursors have a strong impact on the final cost as production scales up. The available body of knowledge on cQD synthesis suggests that it should be possible to switch from currently used costly trimethylsilyl-based chalcogenide and pnictide precursors with poor atom economy (*172*) to simpler species, such as H_2S and PH_3. Other approaches, such as the use of amine complexes or thiourea, offer promising cQD quality and a path for greener synthesis with lower projected costs (*170*). Sustainable large-scale synthesis should consider recycling the large volume of organic solvents used during synthesis and cQD assembly to decrease their cost and associated carbon footprint (*171*). The implementation of continuous-flow automated syntheses is expected to bring down production costs substantially. Generally, cQD synthesis will benefit from the development of quantitative kinetic models based on the mechanistic understanding of each reaction step. Recent developments in machine learning and artificial intelligence can be implemented to increase the predictive power of these models (*173*).

The regulation of heavy metals such as Pb, Cd, and Hg, among others, requires advancing the synthesis, processes, and performance of more widely accepted cQD materials. Widespread technological adoption brings additional constraints and challenges. Many groundbreaking fundamental studies have used CdSe cQDs and Cd-based core shell structures. In terms of color purity and other performance metrics, CdSe cQDs are superior to cQDs of InP, $CuInSe_{2-x}S_x$, and other less toxic semiconductors. However, legislative regulations, additional manufacturing and transportation costs, and environmental concerns have required displays to switch to InP cQDs. One can anticipate a similar trend with other emerging cQD technologies. Substantial progress has been achieved by using Cu-, Bi-, Sn-, Sb-, and In-based cQDs in display and energy-harvesting applications (*174*–*176*). Accelerated materials discovery is also expected to play an increasing role in these areas (*177*). Another important direction for cQD synthesis will be the development of new routes toward III-V cQDs beyond currently available InP, InAs, and InSb. The record-performing eQD devices all used Ga-based materials such as GaAs and GaN, which are very difficult to synthesize with current solution methods.

Long-term material stability is an additional metric that needs careful study when proof-of-concept demonstrations are successful. In general, nanomaterials are metastable with respect to bulk crystals. This raises an important problem of their morphological and chemical stability during operation—often at elevated temperatures. An increased understanding of sintering and grain growth in materials composed of <10-nm semiconductor grains is needed (*178*). Generally, covalent semiconductors show fewer tendencies to electromigration, ion diffusion, and sintering compared with those of ionic compounds. Lattice-matched cQD-in-perovskite systems have been shown to extend the lifetime of the cQDs and perovskite matrix and enhance their properties (*179*), representing a promising path forward.

From a manufacturing perspective, more efforts are needed to demonstrate reproducible cQD devices with high performance by using techniques such as roll-to-roll and spray casting. Patterning of cQD layers is an important step for manufacturing almost any electronic or optoelectronic device containing cQDs. The development of material- and process-tailored patterning methods will be critical for cQD technologies (*180*). Inkjet printing colloidal solutions offers exciting opportunities for device manufacturing, but patterns with <10-μm features will require different, typically lithographic methods, such as direct optical lithography of functional nanomaterials (*181*). The translation of existing cQD surface chemistries—mostly developed at smaller scales or optimized for other deposition processes—to these manufacturing conditions is not trivial (*103*). The integration of cQD technologies in manufacturing processes of products with increasing complexity may bring additional challenges regarding cQD stability.

At the device level, the evaluation of performance metrics at conditions relevant to the final application and standardized stress tests combined with material degradation studies are needed to advance in the reliability of cQD devices.

Challenges remain on the path to commercialization of cQD LEDs. So far, the best-performing devices still rely on Cd or Pb, which are highly toxic heavy metals. Another challenge is the need for pixelation of cQDs toward high-resolution, multicolor-LED–based displays. Different variations of photolithographic and printing techniques have been explored for this purpose but have yet to fully meet resolution, fidelity, and throughput requirements (*182*). LEDs based on cQDs must target further increases in power efficiency and stability at relevant operating currents and brightness. This is especially the case for UV, blue, and infrared LEDs, whose performance is lower than that for green and red colors.

The demonstration of electrically driven cQD laser diodes is an important present objective. Further advances in the design and implementation of complex cQD heterostructures are expected to help accomplish this goal. Interesting opportunities are associated with cQD-specific recent lasing concepts that do not require biexcitons to achieve optical gain. These include single-exciton gain (*183*) and charged-exciton gain (*184*–*186*). Promising results have been also obtained with emerging nanocrystalline materials such as $CsPbX_3$ perovskite cQDs (*187*) and quasi-2D nanoplatelets (*188*, *189*). Recently, cQD optical gain and amplified spontaneous emission were realized within the IR telecom window by using PbS cQDs (*190*).

Challenges also remain in the area of eQD lasers. Improvements in eQD fabrication (S-K growth) are needed to increase the in-plane density of QDs and number of layers in the stack without compromising materials quality and QD uniformity. These are expected to provide with extended high-temperature stability of the threshold current, increased modulation bandwidth, and reduced noise features that are important for quantum light applications. Lasers that use droplet eQDs have yet to be demonstrated.

In addition, it is interesting to develop eQD lasers that are based on other materials. In particular, the realization of efficient ultraviolet GaN eQD lasers would open the door to manufacturing and lithography applications. The fabrication of eQD lasers on CMOS-compatible silicon-on-insulator substrates is also an important unresolved challenge.

In photodetection, deepened understanding of cQD assembly, charge transport, intragap states, and interfaces are needed to improve the sensitivity of sensors operating at the SWIR and beyond (*191*). Mobilities beyond 10 $cm^2 V^{-1} s^{-1}$ and low dielectric constants are desired to meet the demands of fast sensing (<1 ns) for time-of-flight 3D imaging and light detection and ranging (LIDAR)—which is of

interest to the expanding autonomous driving industry (*93*), machine vision, and augmented reality for consumer electronics. Efforts to improve device architecture (*92*), charge mobility, and trap density in cQD films are necessary to improve D^* and response time (*192*).

Attention is still needed to increase the PCE and stability of cQD solar cells. Improvements in diffusion length need to be pursued together with strategies to ensure long-term chemical and mechanical compatibility with other elements in tandem devices. Because IR harvesting requires larger-diameter cQDs that have a different facet distribution, different surface chemistry strategies are needed relative to those used for smaller cQDs (*193*). Multiple exciton generation or carrier multiplication, in which one high-energy photon ($E > 2E_g$) generates multiple electron-hole pairs (*194*), remains a path to be fulfilled for achieving PCEs that exceed a traditional Shockley-Queisser limit. Long-term stability at maximum power point and under temperature and humidity, which is now standard in perovskite PV reports (*195*), needs to be consistently reported to evaluate the impact of this technology and guide its further development.

In the context of eQDs, the realization of devices that achieve high uniformity, high density, and multiple layers has the prospect to enable ultrahigh-efficiency solar cells with near 80% PCE by using multiple intermediate levels (*196*).

The largest cQD-based LSCs are yet within a few tens of square centimeters (*197*). An important challenge is the scale-up of these proof-of-principle structures to meter-sized devices that would be of relevance for building-integrated sunlight collectors. This will require further CQD developments to improve the LSC-relevant parameters (such as PLQY and Q_{LSC}), especially in the case of NIR emitters. Other challenges include the elimination of propagation losses in the LSC waveguide and the enhancement of device stability under outdoor conditions.

The use of cQDs in photo- and electrochemistry is still in its early stages. It is still to be seen which cQD properties can be exploited in practical systems, demanding long-term stability at harsh electrochemical conditions.

In the application of QDs to nonclassical quantum light sources such as single-photon and entangled-photon sources, high-precision position and wavelength control and integration to a quantum circuit are important issues. Nonclassical quantum light sources are positioned as indispensable devices for quantum key distribution and quantum repeater systems for quantum communication. Emerging distributed quantum computer systems (edge quantum computing systems) will be mutually connected by quantum networks. In that case, quantum communication based on on-demand and highly efficient nonclassical quantum light sources will play large roles.

In addition, quantum bits (qubits) based on semiconductor QDs are one of key candidates for realizing quantum computer systems. Up to now, these have been realized by using QDs under depletion effect induced by electric fields. eQDs relying on group IV such as silicon or germanium are emerging as next-generation qubits. The application of QDs for quantum sensing and metrology could also enable QDs to replace diamond nitrogen-vacancy centers and carbon nanotubes.

REFERENCES AND NOTES

1. A. Y. Cho, J. R. Arthur, Molecular beam epitaxy. *Prog. Solid State Chem.* **10**, 157–191 (1975). doi: 10.1016/0079-6786(75)90005-9
2. A. P. Alivisatos, A. L. Harris, N. J. Levinos, M. L. Steigerwald, L. E. Brus, Electronic states of semiconductor clusters: Homogeneous and inhomogeneous broadening of the optical spectrum. *J. Chem. Phys.* **89**, 4001–4011 (1988). doi: 10.1063/1.454833
3. V. I. Klimov, Multicarrier interactions in semiconductor nanocrystals in relation to the phenomena of Auger recombination and carrier multiplication. *Annu. Rev. Condens. Matter Phys.* **5**, 285–316 (2014). doi: 10.1146/annurev-conmatphys-031113-133900
4. A. I. Ekimov, A. A. Onushchenko, Quantum size effect in three-dimensional microscopic semiconductor crystals. *JETP Lett.* **34**, 345 (1981).
5. R. Rossetti, S. Nakahara, L. E. Brus, Quantum size effects in the redox potentials, resonance Raman spectra, and electronic spectra of CdS crystallites in aqueous solution. *J. Chem. Phys.* **79**, 1086–1088 (1983). doi: 10.1063/1.445834
6. L. Brus, Electronic wave functions in semiconductor clusters: Experiment and theory. *J. Phys. Chem.* **90**, 2555–2560 (1986). doi: 10.1021/j100403a003
7. Y. Arakawa, H. Sakaki, Multidimensional quantum well laser and temperature dependence of its threshold current. *Appl. Phys. Lett.* **40**, 939–941 (1982). doi: 10.1063/1.92959
8. Y. Arakawa, H. Sakaki, M. Nishioka, H. Okamoto, N. Miura, Spontaneous emission characteristics of quantum well lasers in strong magnetic fields–An approach to quantum-well-box light source–. *Jpn. J. Appl. Phys.* **22**, L804–L806 (1983). doi: 10.1143/JJAP.22.L804
9. L. Goldstein, F. Glas, J. Y. Marzin, M. N. Charasse, G. Le Roux, Growth by molecular beam epitaxy and characterization of InAs/GaAs strained-layer superlattices. *Appl. Phys. Lett.* **47**, 1099–1101 (1985). doi: 10.1063/1.96342
10. D. Leonard, M. Krishnamurthy, C. M. Reaves, S. P. Denbaars, P. M. Petroff, Direct formation of quantum-sized dots from uniform coherent islands of InGaAs on GaAs surfaces. *Appl. Phys. Lett.* **63**, 3203–3205 (1993). doi: 10.1063/1.110199
11. J. Oshinowo, M. Nishioka, S. Ishida, Y. Arakawa, Highly uniform InGaAs/GaAs quantum dots (~15 nm) by metalorganic chemical vapor deposition. *Appl. Phys. Lett.* **65**, 1421–1423 (1994). doi: 10.1063/1.112070
12. M. Gurioli, Z. Wang, A. Rastelli, T. Kuroda, S. Sanguinetti, Droplet epitaxy of semiconductor nanostructures for quantum photonic devices. *Nat. Mater.* **18**, 799–810 (2019). doi: 10.1038/s41563-019-0355-y; pmid: 31086322
13. M. Bayer, Bridging two worlds: Colloidal versus epitaxial quantum dots. *Ann. Phys.* **531**, 1900039 (2019). doi: 10.1002/andp.201900039
14. C. B. Murray, D. J. Norris, M. G. Bawendi, Synthesis and characterization of nearly monodisperse CdE (E = sulfur, selenium, tellurium) semiconductor nanocrystallites. *J. Am. Chem. Soc.* **115**, 8706–8715 (1993). doi: 10.1021/ja00072a025
15. M. L. Steigerwald, L. E. Brus, Semiconductor crystallites: A class of large molecules. *Acc. Chem. Res.* **23**, 183–188 (1990). doi: 10.1021/ar00174a003
16. O. I. Micic, C. J. Curtis, K. M. Jones, J. R. Sprague, A. J. Nozik, Synthesis and characterization of InP quantum dots. *J. Phys. Chem.* **98**, 4966–4969 (1994). doi: 10.1021/j100070a004
17. F. W. Wise, Lead salt quantum dots: The limit of strong quantum confinement. *Acc. Chem. Res.* **33**, 773–780 (2000). doi: 10.1021/ar970220q; pmid: 11087314
18. R. A. Bley, S. M. Kauzlarich, A low-temperature solution phase route for the synthesis of silicon nanoclusters. *J. Am. Chem. Soc.* **118**, 12461–12462 (1996). doi: 10.1021/ja962787s
19. D. D. Vaughn 2nd, R. E. Schaak, Synthesis, properties and applications of colloidal germanium and germanium-based nanomaterials. *Chem. Soc. Rev.* **42**, 2861–2879 (2013). doi: 10.1039/C2CS35364D; pmid: 23128962
20. F. Wang *et al.*, One-step synthesis of highly luminescent carbon dots in noncoordinating solvents. *Chem. Mater.* **22**, 4528–4530 (2010). doi: 10.1021/cm101350u
21. L. Protesescu *et al.*, Nanocrystals of cesium lead halide perovskites ($CsPbX_3$, X = Cl, Br, and I): Novel optoelectronic materials showing bright emission with wide color gamut. *Nano Lett.* **15**, 3692–3696 (2015). doi: 10.1021/nl5048779; pmid: 25633588
22. C. B. Murray, C. R. Kagan, M. G. Bawendi, Synthesis and characterization of monodisperse nanocrystals and close-packed nanocrystal assemblies. *Annu. Rev. Mater. Sci.* **30**, 545–610 (2000). doi: 10.1146/annurev.matsci.30.1.545
23. M. C. Weidman, M. E. Beck, R. S. Hoffman, F. Prins, W. A. Tisdale, Monodisperse, air-stable PbS nanocrystals via precursor stoichiometry control. *ACS Nano* **8**, 6363–6371 (2014). doi: 10.1021/nn5018654; pmid: 24840645
24. D. J. Norris, A. L. Efros, S. C. Erwin, Doped nanocrystals. *Science* **319**, 1776–1779 (2008). doi: 10.1126/science.1143802; pmid: 18369131
25. D. Mocatta *et al.*, Heavily doped semiconductor nanocrystal quantum dots. *Science* **332**, 77–81 (2011). doi: 10.1126/science.1196321; pmid: 21454783
26. D. Yu, C. Wang, P. Guyot-Sionnest, n-Type conducting CdSe nanocrystal solids. *Science* **300**, 1277–1280 (2003). doi: 10.1126/science.1084424; pmid: 12764194
27. M. A. Boles, D. Ling, T. Hyeon, D. V. Talapin, The surface science of nanocrystals. *Nat. Mater.* **15**, 141–153 (2016). doi: 10.1038/nmat4526; pmid: 26796733
28. M. C. Weidman, D.-M. Smilgies, W. A. Tisdale, Kinetics of the self-assembly of nanocrystal superlattices measured by real-time in situ x-ray scattering. *Nat. Mater.* **15**, 775–781 (2016). doi: 10.1038/nmat4600; pmid: 26998914
29. E. V. Shevchenko, D. V. Talapin, N. A. Kotov, S. O'Brien, C. B. Murray, Structural diversity in binary nanoparticle superlattices. *Nature* **439**, 55–59 (2006). doi: 10.1038/nature04414; pmid: 16397494
30. C. B. Murray, C. R. Kagan, M. G. Bawendi, Self-organization of CdSe nanocrystallites into three-dimensional quantum dot superlattices. *Science* **270**, 1335–1338 (1995). doi: 10.1126/science.270.5240.1335
31. M. P. Boneschanscher *et al.*, Long-range orientation and atomic attachment of nanocrystals in 2D honeycomb superlattices. *Science* **344**, 1377–1380 (2014). doi: 10.1126/science.1252642; pmid: 24948734
32. Y. Nagaoka *et al.*, Superstructures generated from truncated tetrahedral quantum dots. *Nature* **561**, 378–382 (2018). doi: 10.1038/s41586-018-0512-5; pmid: 30232427
33. X. Jin *et al.*, Cation exchange assisted synthesis of ZnCdSe/ZnSe quantum dots with narrow emission line widths and near-unity photoluminescence quantum yields. *Chem. Commun.* **56**, 6130–6133 (2020). doi: 10.1039/D0CC01302A; pmid: 32364191
34. BT Series, 2100: Image parameter values for high dynamic range television for use in production and international programme exchange; www.itu.int/rec/R-REC-BT.2100-2-201807-I/en.
35. D. A. Hanifi *et al.*, Redefining near-unity luminescence in quantum dots with photothermal threshold quantum yield. *Science* **363**, 1199–1202 (2019). doi: 10.1126/science.aat3803; pmid: 30872520
36. F. Meinardi, F. Bruni, S. Brovelli, Luminescent solar concentrators for building-integrated photovoltaics. *Nat. Rev. Mater.* **2**, 17072 (2017). doi: 10.1038/natrevmats.2017.72
37. Y.-S. Park, J. Roh, B. T. Diroll, R. D. Schaller, V. I. Klimov, Colloidal quantum dot lasers. *Nat. Rev. Mater.* **6**, 382–401 (2021). doi: 10.1038/s41578-020-00274-9
38. J. Owen, The coordination chemistry of nanocrystal surfaces. *Science* **347**, 615–616 (2015). doi: 10.1126/science.1259924; pmid: 25657234
39. I. L. Medintz, H. T. Uyeda, E. R. Goldman, H. Mattoussi, Quantum dot bioconjugates for imaging, labelling and sensing. *Nat. Mater.* **4**, 435–446 (2005). doi: 10.1038/nmat1390; pmid: 15928695
40. N. Yazdani *et al.*, Charge transport in semiconductors assembled from nanocrystal quantum dots. *Nat. Commun.* **11**, 2852 (2020). doi: 10.1038/s41467-020-16560-7; pmid: 32503965

41. C. Giansante, I. Infante, Surface traps in colloidal quantum dots: A combined experimental and theoretical perspective. *J. Phys. Chem. Lett.* **8**, 5209–5215 (2017). doi: 10.1021/acs.jpclett.7b02193; pmid: 28972763
42. N. Yazdani *et al*., Charge transport in semiconductors assembled from nanocrystals. *Nat. Commun.* 10.1038/s41467-020-16560-7 (2019).
43. D. V. Talapin, C. B. Murray, PbSe nanocrystal solids for n- and p-channel thin film field-effect transistors. *Science* **310**, 86–89 (2005). doi: 10.1126/science.1116703; pmid: 16210533
44. J.-S. Lee, M. V. Kovalenko, J. Huang, D. S. Chung, D. V. Talapin, Band-like transport, high electron mobility and high photoconductivity in all-inorganic nanocrystal arrays. *Nat. Nanotechnol.* **6**, 348–352 (2011). doi: 10.1038/nnano.2011.46; pmid: 21516091
45. X. Lan *et al*., Quantum dot solids showing state-resolved band-like transport. *Nat. Mater.* **19**, 323–329 (2020). doi: 10.1038/s41563-019-0582-2; pmid: 31988516
46. H. Chen, J. He, S.-T. Wu, Recent advances on quantum-dot-enhanced liquid-crystal displays. *IEEE J. Sel. Top. Quantum Electron.* **23**, 1–11 (2017).
47. B.-L. Huang *et al*., Color converting film with quantum-dots for the liquid crystal displays based on inkjet printing. *IEEE Photonics J.* **11**, 1–9 (2019). doi: 10.1109/JPHOT.2019.2911308
48. W. K. Bae, S. Brovelli, V. I. Klimov, Spectroscopic insights into the performance of quantum dot light-emitting diodes. *MRS Bull.* **38**, 721–730 (2013). doi: 10.1557/mrs.2013.182
49. J. M. Caruge, J. E. Halpert, V. Wood, V. Bulović, M. G. Bawendi, Colloidal quantum-dot light-emitting diodes with metal-oxide charge transport layers. *Nat. Photonics* **2**, 247–250 (2008). doi: 10.1038/nphoton.2008.34
50. B. S. Mashford *et al*., High-efficiency quantum-dot light-emitting devices with enhanced charge injection. *Nat. Photonics* **7**, 407–412 (2013). doi: 10.1038/nphoton.2013.70
51. J. Lim, Y.-S. Park, V. I. Klimov, Optical gain in colloidal quantum dots achieved with direct-current electrical pumping. *Nat. Mater.* **17**, 42–49 (2018). doi: 10.1038/nmat5011; pmid: 29180770
52. J. Lim, Y.-S. Park, K. Wu, H. J. Yun, V. I. Klimov, Droop-free colloidal quantum dot light-emitting diodes. *Nano Lett.* **18**, 6645–6653 (2018). doi: 10.1021/acs.nanolett.8b03457; pmid: 30198267
53. X. Li *et al*., Bright colloidal quantum dot light-emitting diodes enabled by efficient chlorination. *Nat. Photonics* **12**, 159–164 (2018). doi: 10.1038/s41566-018-0105-8
54. X. Dai, Y. Deng, X. Peng, Y. Jin, Quantum-Dot Light-Emitting Diodes for large-area displays: Towards the dawn of commercialization. *Adv. Mater.* **29**, 1607022 (2017). doi: 10.1002/adma.201607022; pmid: 28256780
55. X. Dai *et al*., Solution-processed, high-performance light-emitting diodes based on quantum dots. *Nature* **515**, 96–99 (2014). doi: 10.1038/nature13829; pmid: 25363773
56. H. Shen *et al*., Visible quantum dot light-emitting diodes with simultaneous high brightness and efficiency. *Nat. Photonics* **13**, 192–197 (2019). doi: 10.1038/s41566-019-0364-z
57. O. Wang *et al*., High-efficiency, deep blue $ZnCdS/Cd_xZn_{1-x}S$/ZnS quantum-dot-light-emitting devices with an EQE exceeding 18. *Nanoscale* **10**, 5650–5657 (2018). doi: 10.1039/C7NR09175C; pmid: 29528343
58. D. Bozyigit, O. Yarema, V. Wood, Origins of low quantum efficiencies in quantum dot LEDs. *Adv. Funct. Mater.* **23**, 3024–3029 (2013). doi: 10.1002/adfm.201203191
59. Z. Ning *et al*., Quantum-dot-in-perovskite solids. *Nature* **523**, 324–328 (2015). doi: 10.1038/nature14563; pmid: 26178963
60. X. Gong *et al*., Highly efficient quantum dot near-infrared light-emitting diodes. *Nat. Photonics* **10**, 253–257 (2016). doi: 10.1038/nphoton.2016.11
61. L. Gao *et al*., Efficient near-infrared light-emitting diodes based on quantum dots in layered perovskite. *Nat. Photonics* **14**, 227–233 (2020). doi: 10.1038/s41566-019-0577-1
62. S. Pradhan *et al*., High-efficiency colloidal quantum dot infrared light-emitting diodes via engineering at the supra-nanocrystalline level. *Nat. Nanotechnol.* **14**, 72–79 (2019). doi: 10.1038/s41565-018-0312-y; pmid: 30510279
63. V. I. Klimov *et al*., Optical gain and stimulated emission in nanocrystal quantum dots. *Science* **290**, 314–317 (2000). doi: 10.1126/science.290.5490.314; pmid: 11030645
64. V. I. Klimov, A. A. Mikhailovsky, D. W. McBranch, C. A. Leatherdale, M. G. Bawendi, Quantization of multiparticle Auger rates in semiconductor quantum dots. *Science* **287**, 1011–1013 (2000). doi: 10.1126/science.287.5455.1011; pmid: 10669406
65. Y.-S. Park, W. K. Bae, T. Baker, J. Lim, V. I. Klimov, Effect of auger recombination on lasing in heterostructured quantum dots with engineered core/shell interfaces. *Nano Lett.* **15**, 7319–7328 (2015). doi: 10.1021/acs.nanolett.5b02595; pmid: 26397312
66. F. García-Santamaría *et al*., Suppressed auger recombination in "giant" nanocrystals boosts optical gain performance. *Nano Lett.* **9**, 3482–3488 (2009). doi: 10.1021/nl901681d; pmid: 19505082
67. G. E. Cragg, A. L. Efros, Suppression of auger processes in confined structures. *Nano Lett.* **10**, 313–317 (2010). doi: 10.1021/nl903592h; pmid: 20017564
68. W. K. Bae *et al*., Controlled alloying of the core-shell interface in CdSe/CdS quantum dots for suppression of Auger recombination. *ACS Nano* **7**, 3411–3419 (2013). doi: 10.1021/nn4002825; pmid: 23521208
69. F. Fan *et al*., Continuous-wave lasing in colloidal quantum dot solids enabled by facet-selective epitaxy. *Nature* **544**, 75–79 (2017). doi: 10.1038/nature21424; pmid: 28321128
70. J. Roh, Y.-S. Park, J. Lim, V. I. Klimov, Optically pumped colloidal-quantum-dot lasing in LED-like devices with an integrated optical cavity. *Nat. Commun.* **11**, 271 (2020). doi: 10.1038/s41467-019-14014-3; pmid: 31937771
71. M. Grundmann, The present status of quantum dot lasers. *Phys. E* **5**, 167–184 (1999). doi: 10.1016/S1386-9477(99)00041-7
72. K. Otsubo *et al*., Temperature-insensitive eye-opening under 10-Gb/s modulation of 1.3-μm P-doped quantum-dot lasers without current adjustments. *Jpn. J. Appl. Phys.* **43** (No. 8B), L1124–L1126 (2004). doi: 10.1143/JJAP.43.L1124
73. T. Kageyama *et al*., in 2011 Conference on Lasers and Electro-Optics Europe and 12th European Quantum Electronics Conference (CLEO EUROPE/EQEC) (IEEE, 2011); http://ieeexplore.ieee.org/document/5943701.
74. Y. Arakawa, T. Nakamura, J. Kwoen, Quantum dot lasers for silicon photonics. *Semicond. Semimet.* **101**, 91–138 (2019). doi: 10.1016/bs.semsem.2019.07.007
75. H. Lim, S. Tsao, W. Zhang, M. Razeghi, High-performance InAs quantum-dot infrared photodetectors grown on InP substrate operating at room temperature. *Appl. Phys. Lett.* **90**, 131112 (2007). doi: 10.1063/1.2719160
76. P. Martyniuk, A. Rogalski, Quantum-dot infrared photodetectors: Status and outlook. *Prog. Quantum Electron.* **32**, 89–120 (2008). doi: 10.1016/j.pquantelec.2008.07.001
77. B. Chen *et al*., Low dark current high gain InAs quantum dot avalanche photodiodes monolithically grown on Si. *ACS Photonics* **7**, 528–533 (2020). doi: 10.1021/acsphotonics.9b01709
78. A. Ren, L. Yuan, H. Xu, J. Wu, Z. Wang, Recent progress of III–V quantum dot infrared photodetectors on silicon. *J. Mater. Chem. C Mater. Opt. Electron. Devices* **7**, 14441–14453 (2019). doi: 10.1039/C9TC05738B
79. R. Saran, R. J. Curry, Lead sulphide nanocrystal photodetector technologies. *Nat. Photonics* **10**, 81–92 (2016). doi: 10.1038/nphoton.2015.280
80. G. Konstantatos *et al*., Ultrasensitive solution-cast quantum dot photodetectors. *Nature* **442**, 180–183 (2006). doi: 10.1038/nature04855; pmid: 16838017
81. J. P. Clifford *et al*., Fast, sensitive and spectrally tuneable colloidal-quantum-dot photodetectors. *Nat. Nanotechnol.* **4**, 40–44 (2009). doi: 10.1038/nnano.2008.313; pmid: 19119281
82. C. Livache, B. Martinez, N. Goubet, J. Ramade, E. Lhuillier, Road map for nanocrystal based infrared photodetectors. *Front Chem.* **6**, 575 (2018). doi: 10.3389/fchem.2018.00575; pmid: 30547026
83. M. M. Ackerman, X. Tang, P. Guyot-Sionnest, Fast and sensitive colloidal quantum dot mid-wave infrared photodetectors. *ACS Nano* **12**, 7264–7271 (2018). doi: 10.1021/acsnano.8b03425; pmid: 29975502
84. S. Keuleyan, E. Lhuillier, V. Brajuskovic, P. Guyot-Sionnest, Mid-infrared HgTe colloidal quantum dot photodetectors. *Nat. Photonics* **5**, 489–493 (2011). doi: 10.1038/nphoton.2011.142
85. X. Tang, M. M. Ackerman, M. Chen, P. Guyot-Sionnest, Dual-band infrared imaging using stacked colloidal quantum dot photodiodes. *Nat. Photonics* **13**, 277–282 (2019). doi: 10.1038/s41566-019-0362-1
86. Z. Deng, K. S. Jeong, P. Guyot-Sionnest, Colloidal quantum dots intraband photodetectors. *ACS Nano* **8**, 11707–11714 (2014). doi: 10.1021/nn505092a; pmid: 25343383
87. J. Kim, D. Choi, K. S. Jeong, Self-doped colloidal semiconductor nanocrystals with intraband transitions in steady state. *Chem. Commun.* **54**, 8435–8445 (2018). doi: 10.1039/C8CC02488J; pmid: 29972153
88. I. Ramiro *et al*., Mid- and long-wave infrared optoelectronics via intraband transitions in PbS colloidal quantum dots. *Nano Lett.* **20**, 1003–1008 (2020). doi: 10.1021/acs.nanolett.9b04130; pmid: 31934762
89. C. Livache *et al*., A colloidal quantum dot infrared photodetector and its use for intraband detection. *Nat. Commun.* **10**, 2125 (2019). doi: 10.1038/s41467-019-10170-8; pmid: 31073132
90. G. Konstantatos *et al*., Hybrid graphene-quantum dot phototransistors with ultrahigh gain. *Nat. Nanotechnol.* **7**, 363–368 (2012). doi: 10.1038/nnano.2012.60; pmid: 22562036
91. F. P. García de Arquer *et al*., Field-emission from quantum-dot-in-perovskite solids. *Nat. Commun.* **8**, 14757 (2017). doi: 10.1038/ncomms14757; pmid: 28337981
92. D. Kufer, G. Konstantatos, Photo-FETs: Phototransistors enabled by 2D and 0D nanomaterials. *ACS Photonics* **3**, 2197–2210 (2016). doi: 10.1021/acsphotonics.6b00391
93. F. P. García de Arquer, A. Armin, P. Meredith, E. H. Sargent, Solution-processed semiconductors for next-generation photodetectors. *Nat. Rev. Mater.* **2**, 16100 (2017). doi: 10.1038/natrevmats.2016.100
94. G. Konstantatos, Current status and technological prospect of photodetectors based on two-dimensional materials. *Nat. Commun.* **9**, 5266 (2018). doi: 10.1038/s41467-018-07643-7; pmid: 30531824
95. S. Goossens *et al*., Broadband image sensor array based on graphene–CMOS integration. *Nat. Photonics* **11**, 366–371 (2017). doi: 10.1038/nphoton.2017.75
96. V. Adinolfi, E. H. Sargent, Photovoltage field-effect transistors. *Nature* **542**, 324–327 (2017). doi: 10.1038/nature21050; pmid: 28178236
97. D. Y. Kim *et al*., PbSe nanocrystal-based infrared-to-visible up-conversion device. *Nano Lett.* **11**, 2109–2113 (2011). doi: 10.1021/nl200704h; pmid: 21504192
98. W. Zhou *et al*., Solution-processed upconversion photodetectors based on quantum dots. *Nat. Electron.* **3**, 251–258 (2020). doi: 10.1038/s41928-020-0388-x
99. Q. Chen *et al*., All-inorganic perovskite nanocrystal scintillators. *Nature* **561**, 88–93 (2018). doi: 10.1038/s41586-018-0451-1; pmid: 30150772
100. J. Bao, M. G. Bawendi, A colloidal quantum dot spectrometer. *Nature* **523**, 67–70 (2015). doi: 10.1038/nature14576; pmid: 26135449
101. S. A. McDonald *et al*., Solution-processed PbS quantum dot infrared photodetectors and photovoltaics. *Nat. Mater.* **4**, 138–142 (2005). doi: 10.1038/nmat1299; pmid: 15640806
102. I. Gur, N. A. Fromer, M. L. Geier, A. P. Alivisatos, Air-stable all-inorganic nanocrystal solar cells processed from solution. *Science* **310**, 462–465 (2005). doi: 10.1126/science.1117908; pmid: 16239470
103. A. R. Kirmani, J. M. Luther, M. Abolhasani, A. Amassian, Colloidal quantum dot photovoltaics: Current progress and path to gigawatt scale enabled by smart manufacturing. *ACS Energy Lett.* **5**, 3069–3100 (2020). doi: 10.1021/acsenergylett.0c01453
104. S. Pradhan, A. Stavrinadis, S. Gupta, S. Christodoulou, G. Konstantatos, Breaking the open-circuit voltage deficit floor in PbS quantum dot solar cells through synergistic ligand and architecture engineering. *ACS Energy Lett.* **2**, 1444–1449 (2017). doi: 10.1021/acsenergylett.7b00244
105. C.-H. M. Chuang *et al*., Open-circuit voltage deficit, radiative sub-bandgap states, and prospects in quantum dot solar cells. *Nano Lett.* **15**, 3286–3294 (2015). doi: 10.1021/acs.nanolett.5b00513; pmid: 25927871
106. S. Kahmann, M. A. Loi, Trap states in lead chalcogenide colloidal quantum dots - Origin, impact, and remedies. *Appl. Phys. Rev.* **7**, 041305 (2020). doi: 10.1063/5.0019800
107. C.-H. M. Chuang, P. R. Brown, V. Bulović, M. G. Bawendi, Improved performance and stability in quantum dot solar cells through band alignment engineering. *Nat. Mater.* **13**, 796–801 (2014). doi: 10.1038/nmat3984; pmid: 24859641
108. T. Kim, X. Jin, J. H. Song, S. Jeong, T. Park, Efficiency limit of colloidal quantum dot solar cells: Effect of optical interference on active layer absorption. *ACS Energy Lett.* **5**, 248–251 (2020). doi: 10.1021/acsenergylett.9b02504
109. A. K. Rath *et al*., Solution-processed inorganic bulk nano-heterojunctions and their application to solar cells. *Nat. Photonics* **6**, 529–534 (2012). doi: 10.1038/nphoton.2012.139
110. M. Liu *et al*., Hybrid organic-inorganic inks flatten the energy landscape in colloidal quantum dot solids. *Nat. Mater.* **16**, 258–263 (2017). doi: 10.1038/nmat4800; pmid: 27842072

111. B. Sun *et al*., Monolayer perovskite bridges enable strong quantum dot coupling for efficient solar cells. *Joule* **4**, 1542–1556 (2020). doi: 10.1016/j.joule.2020.05.011
112. S.-W. Baek *et al*., Efficient hybrid colloidal quantum dot/organic solar cells mediated by near-infrared sensitizing small molecules. *Nat. Energy* **4**, 969–976 (2019). doi: 10.1038/s41560-019-0492-1
113. A. Swarnkar *et al*., Quantum dot-induced phase stabilization of α-CsPbI3 perovskite for high-efficiency photovoltaics. *Science* **354**, 92–95 (2016). doi: 10.1126/science.aag2700; pmid: 27846497
114. M. Hao *et al*., Ligand-assisted cation-exchange engineering for high-efficiency colloidal $Cs_{1-x}FA_xPbI_3$ quantum dot solar cells with reduced phase segregation. *Nat. Energy* **5**, 79–88 (2020). doi: 10.1038/s41560-019-0535-7
115. J. Yuan *et al*., Metal halide perovskites in quantum dot solar cells: Progress and prospects. *Joule* **4**, 1160–1185 (2020). doi: 10.1016/j.joule.2020.04.006
116. B. Chen *et al*., Enhanced optical path and electron diffusion length enable high-efficiency perovskite tandems. *Nat. Commun.* **11**, 1257 (2020). doi: 10.1038/s41467-020-15077-3; pmid: 32152324
117. M. J. Crane, D. M. Kroupa, D. R. Gamelin, Detailed-balance analysis of Yb^{3+}:$CsPb(Cl_{1-x}Br_x)_3$ quantum-cutting layers for high-efficiency photovoltaics under real-world conditions. *Energy Environ. Sci.* **12**, 2486–2495 (2019). doi: 10.1039/C9EE01493D
118. W. H. Weber, J. Lambe, Luminescent greenhouse collector for solar radiation. *Appl. Opt.* **15**, 2299–2300 (1976). doi: 10.1364/AO.15.002299; pmid: 20165383
119. M. G. Debije, P. P. C. Verbunt, Thirty years of luminescent solar concentrator research: Solar energy for the built environment. *Adv. Energy Mater.* **2**, 12–35 (2012). doi: 10.1002/aenm.201100554
120. H.-J. Song *et al*., Performance limits of luminescent solar concentrators tested with seed/quantum-well quantum dots in a selective-reflector-based optical cavity. *Nano Lett.* **18**, 395–404 (2018). doi: 10.1021/acs.nanolett.7b04263; pmid: 29226688
121. F. Meinardi *et al*., Large-area luminescent solar concentrators based on 'Stokes-shift-engineered' nanocrystals in a mass-polymerized PMMA matrix. *Nat. Photonics* **8**, 392–399 (2014). doi: 10.1038/nphoton.2014.54
122. N. D. Bronstein *et al*., Quantum dot luminescent concentrator cavity exhibiting 30-fold concentration. *ACS Photonics* **2**, 1576–1583 (2015). doi: 10.1021/acsphotonics.5b00334
123. F. Meinardi *et al*., Highly efficient large-area colourless luminescent solar concentrators using heavy-metal-free colloidal quantum dots. *Nat. Nanotechnol.* **10**, 878–885 (2015). doi: 10.1038/nnano.2015.178; pmid: 26301902
124. M. R. Bergren *et al*., High-performance $CuInS_2$ quantum dot laminated glass luminescent solar concentrators for windows. *ACS Energy Lett.* **3**, 520–525 (2018). doi: 10.1021/acsenergylett.7b01346
125. A. S. Fuhr *et al*., Light emission mechanisms in $CuInS_2$ quantum dots evaluated by spectral electrochemistry. *ACS Photonics* **4**, 2425–2435 (2017). doi: 10.1021/acsphotonics.7b00560
126. K. Wu, H. Li, V. I. Klimov, Tandem luminescent solar concentrators based on engineered quantum dots. *Nat. Photonics* **12**, 105–110 (2018). doi: 10.1038/s41566-017-0070-7
127. N. S. Makarov *et al*., Fiber-coupled luminescent concentrators for medical diagnostics, agriculture, and telecommunications. *ACS Nano* **13**, 9112–9121 (2019). doi: 10.1021/acsnano.9b03335; pmid: 31291097
128. L. Trahey *et al*., Energy storage emerging: A perspective from the Joint Center for Energy Storage Research. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12550–12557 (2020). doi: 10.1073/pnas.1821672117; pmid: 32513683
129. X.-B. Li, C.-H. Tung, L.-Z. Wu, Semiconducting quantum dots for artificial photosynthesis. *Nat. Rev. Chem.* **2**, 160–173 (2018). doi: 10.1038/s41570-018-0024-8
130. J. Wang *et al*., Enabling visible-light-driven selective CO_2 reduction by doping quantum dots: Trapping electrons and suppressing H_2 evolution. *Angew. Chem. Int. Ed.* **57**, 16447–16451 (2018). doi: 10.1002/anie.201810550; pmid: 30350910
131. E. A. Weiss, Designing the surfaces of semiconductor quantum dots for colloidal photocatalysis. *ACS Energy Lett.* **2**, 1005–1013 (2017). doi: 10.1021/acsenergylett.7b00061
132. S. W. Boettcher, T. E. Mallouk, F. E. Osterloh, Themed issue on water splitting and photocatalysis. *J. Mater. Chem. A Mater. Energy Sustain.* **4**, 2764–2765 (2016). doi: 10.1039/C6TA90014C
133. H. Wu, X. Li, C. Tung, L. Wu, Semiconductor quantum dots: An emerging candidate for CO_2 photoreduction. *Adv. Mater.* **31**, 1900709 (2019). doi: 10.1002/adma.201900709
134. D. W. Wakerley *et al*., Solar-driven reforming of lignocellulose to H_2 with a CdS/CdO_x photocatalyst. *Nat. Energy* **2**, 17021 (2017). doi: 10.1038/nenergy.2017.21
135. T. Uekert, M. F. Kuehnel, D. W. Wakerley, E. Reisner, Plastic waste as a feedstock for solar-driven H_2 generation. *Energy Environ. Sci.* **11**, 2853–2857 (2018). doi: 10.1039/C8EE01408F
136. X. Chen, N. Li, Z. Kong, W.-J. Ong, X. Zhao, Photocatalytic fixation of nitrogen to ammonia: State-of-the-art advancements and future prospects. *Mater. Horiz.* **5**, 9–27 (2018). doi: 10.1039/C7MH00557A
137. U. A. Rani, L. Y. Ng, C. Y. Ng, E. Mahmoudi, A review of carbon quantum dots and their applications in wastewater treatment. *Adv. Colloid Interface Sci.* **278**, 102124 (2020). doi: 10.1016/j.cis.2020.102124; pmid: 32142942
138. K. K. Sakimoto, A. B. Wong, P. Yang, Self-photosensitization of nonphotosynthetic bacteria for solar-to-chemical production. *Science* **351**, 74–77 (2016). doi: 10.1126/science.aad3317; pmid: 26721997
139. H. Luo, Q. Guo, P. Á. Szilágyi, A. B. Jorge, M. M. Titirici, Carbon dots in solar-to-hydrogen conversion. *Trends Chem.* **2**, 623–637 (2020). doi: 10.1016/j.trechm.2020.04.007
140. H. Jung *et al*., Highly efficient and stable CO_2 reduction photocatalyst with a hierarchical structure of mesoporous TiO_2 on 3D graphene with few-layered MoS_2. *ACS Sustain. Chem.& Eng.* **6**, 5718–5724 (2018). doi: 10.1021/acssuschemeng.8b00002
141. T. Takata, K. Domen, Particulate photocatalysts for water splitting: Recent advances and future prospects. *ACS Energy Lett.* **4**, 542–549 (2019). doi: 10.1021/acsenergylett.8b02209
142. C. Anton, K. S. Ranade, Eds., *Quantum Technology, From Research to Application* (National Academy of Science and Engineering, 2015); https://en.acatech.de/publication/perspektiven-der-quantentechnologien.
143. P. Borri *et al*., Ultralong dephasing time in InGaAs quantum dots. *Phys. Rev. Lett.* **87**, 157401 (2001). doi: 10.1103/PhysRevLett.87.157401; pmid: 11580725
144. T. H. Stievater *et al*., Rabi oscillations of excitons in single quantum dots. *Phys. Rev. Lett.* **87**, 133603 (2001). doi: 10.1103/PhysRevLett.87.133603; pmid: 11580588
145. A. Greilich *et al*., Mode locking of electron spin coherences in singly charged quantum dots. *Science* **313**, 341–345 (2006). doi: 10.1126/science.1128215; pmid: 16857937
146. J. R. Petta *et al*., Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science* **309**, 2180–2184 (2005). doi: 10.1126/science.1116955; pmid: 16141370
147. D. Brunner *et al*., A coherent single-hole spin in a semiconductor. *Science* **325**, 70–72 (2009). doi: 10.1126/science.1173684; pmid: 19574387
148. K. De Greve *et al*., Ultrafast coherent control and suppressed nuclear feedback of a single quantum dot hole qubit. *Nat. Phys.* **7**, 872–878 (2011). doi: 10.1038/nphys2078
149. P. Michler, *Quantum Dots for Quantum Information Technologies* (Springer, 2017).
150. P. Michler *et al*., A quantum dot single-photon turnstile device. *Science* **290**, 2282–2285 (2000). doi: 10.1126/science.290.5500.2282; pmid: 11125136
151. N. Somaschi *et al*., Near-optimal single-photon sources in the solid state. *Nat. Photonics* **10**, 340–345 (2016). doi: 10.1038/nphoton.2016.23
152. S. Unsleber *et al*., Highly indistinguishable on-demand resonance fluorescence photons from a deterministic quantum dot micropillar device with 74% extraction efficiency. *Opt. Express* **24**, 8539–8546 (2016). doi: 10.1364/OE.24.008539; pmid: 27137291
153. L. Schweickert *et al*., On-demand generation of background-free single photons from a solid-state source. *Appl. Phys. Lett.* **112**, 093106 (2018). doi: 10.1063/1.5020038
154. O. Gazzano *et al*., Bright solid-state sources of indistinguishable single photons. *Nat. Commun.* **4**, 1425 (2013). doi: 10.1038/ncomms2434; pmid: 23385570
155. H. Utzat *et al*., Coherent single-photon emission from colloidal lead halide perovskite quantum dots. *Science* **363**, 1068–1072 (2019). doi: 10.1126/science.aau7392; pmid: 30792359
156. X. Lin *et al*., Electrically-driven single-photon sources based on colloidal quantum dots with near-optimal antibunching at room temperature. *Nat. Commun.* **8**, 1132 (2017). doi: 10.1038/s41467-017-01379-6; pmid: 29070867
157. M. J. Holmes, S. Kako, K. Choi, M. Arita, Y. Arakawa, Single photons from a hot solid-state emitter at 350 K. *ACS Photonics* **3**, 543–546 (2016). doi: 10.1021/acsphotonics.6b00112
158. C. Santori, D. Fattal, J. Vučković, G. S. Solomon, Y. Yamamoto, Indistinguishable photons from a single-photon device. *Nature* **419**, 594–597 (2002). doi: 10.1038/nature01086; pmid: 12374958
159. K. Toyoda, R. Hiji, A. Noguchi, S. Urabe, Hong-Ou-Mandel interference of two phonons in trapped ions. *Nature* **527**, 74–77 (2015). doi: 10.1038/nature15735; pmid: 26536958
160. X. Ding *et al*., On-demand single photons with high extraction efficiency and near-unity indistinguishability from a resonantly driven quantum dot in a micropillar. *Phys. Rev. Lett.* **116**, 020401 (2016). doi: 10.1103/PhysRevLett.116.020401; pmid: 26824530
161. O. Benson, C. Santori, M. Pelton, Y. Yamamoto, Regulated and entangled photons from a single quantum Dot. *Phys. Rev. Lett.* **84**, 2513–2516 (2000). doi: 10.1103/PhysRevLett.84.2513; pmid: 11018923
162. N. Akopian *et al*., Entangled photon pairs from semiconductor quantum dots. *Phys. Rev. Lett.* **96**, 130501 (2006). doi: 10.1103/PhysRevLett.96.130501; pmid: 16711973
163. R. J. Young *et al*., Improved fidelity of triggered entangled photons from single quantum dots. *New J. Phys.* **8**, 29–29 (2006). doi: 10.1088/1367-2630/8/2/029
164. M. Müller, S. Bounouar, K. D. Jöns, M. Glässl, P. Michler, On-demand generation of indistinguishable polarization-entangled photon pairs. *Nat. Photonics* **8**, 224–228 (2014). doi: 10.1038/nphoton.2013.377
165. D. Huber *et al*., Strain-tunable GaAs quantum dot: A nearly dephasing-free source of entangled photon pairs on demand. *Phys. Rev. Lett.* **121**, 033902 (2018). doi: 10.1103/PhysRevLett.121.033902; pmid: 30085806
166. M. Müller *et al*., Quantum-dot single-photon sources for entanglement enhanced interferometry. *Phys. Rev. Lett.* **118**, 257402 (2017). doi: 10.1103/PhysRevLett.118.257402; pmid: 28696738
167. I. Schwartz *et al*., Deterministic generation of a cluster state of entangled photons. *Science* **354**, 434–437 (2016). doi: 10.1126/science.aah4758; pmid: 27608669
168. P. Michler *et al*., Quantum correlation among photons from a single quantum dot at room temperature. *Nature* **406**, 968–970 (2000). doi: 10.1038/35023100; pmid: 10984045
169. G. Rainò *et al*., Superfluorescence from lead halide perovskite quantum dot superlattices. *Nature* **563**, 671–675 (2018). doi: 10.1038/s41586-018-0683-0; pmid: 30405237
170. M. P. Hendricks, M. P. Campos, G. T. Cleveland, I. Jen-La Plante, J. S. Owen, A tunable library of substituted thiourea precursors to metal sulfide nanocrystals. *Science* **348**, 1226–1230 (2015). doi: 10.1126/science.aaa2951; pmid: 26068846
171. J. Jean *et al*., Synthesis cost dictates the commercial viability of lead sulfide and perovskite quantum dot photovoltaics. *Energy Environ. Sci.* **11**, 2295–2305 (2018). doi: 10.1039/C8EE01348A
172. Y. Pu, F. Cai, D. Wang, J.-X. Wang, J.-F. Chen, Colloidal synthesis of semiconductor quantum dots toward large-scale production: A review. *Ind. Eng. Chem. Res.* **57**, 1790–1802 (2018). doi: 10.1021/acs.iecr.7b04836
173. O. Voznyy *et al*., Machine learning accelerates discovery of optimal colloidal quantum dot synthesis. *ACS Nano* **13**, 11122–11128 (2019). doi: 10.1021/acsnano.9b03864; pmid: 31539477
174. Y.-H. Won *et al*., Highly efficient and stable InP/ZnSe/ZnS quantum dot light-emitting diodes. *Nature* **575**, 634–638 (2019). doi: 10.1038/s41586-019-1771-5; pmid: 31776489
175. J. Du, R. Singh, I. Fedin, A. S. Fuhr, V. I. Klimov, Spectroscopic insights into high defect tolerance of $Zn:CuInSe_2$ quantum-dot-sensitized solar cells. *Nat. Energy* **5**, 409–417 (2020). doi: 10.1038/s41560-020-0617-6
176. M. Bernechea *et al*., Solution-processed solar cells based on environmentally friendly $AgBiS_2$ nanocrystals. *Nat. Photonics* **10**, 521–525 (2016). doi: 10.1038/nphoton.2016.108
177. D. P. Tabor *et al*., Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* **3**, 5–20 (2018). doi: 10.1038/s41578-018-0005-z
178. W. Walravens *et al*., Setting carriers free: Healing faulty interfaces promotes delocalization and transport in nanocrystal solids. *ACS Nano* **13**, 12774–12786 (2019). doi: 10.1021/acsnano.9b04757; pmid: 31693334
179. M. Liu *et al*., Lattice anchoring stabilizes solution-processed semiconductors. *Nature* **570**, 96–101 (2019). doi: 10.1038/s41586-019-1239-7; pmid: 31118515

180. J. Yang, M. K. Choi, D.-H. Kim, T. Hyeon, Designed Assembly and Integration of Colloidal Nanocrystals for Device Applications. *Adv. Mater.* **28**, 1176–1207 (2016). doi: 10.1002/adma.201502851; pmid: 26707709
181. Y. Wang, I. Fedin, H. Zhang, D. V. Talapin, Direct optical lithography of functional inorganic nanomaterials. *Science* **357**, 385–388 (2017). doi: 10.1126/science.aan2958; pmid: 28751606
182. J. Yang *et al*., Toward full-color electroluminescent quantum dot displays. *Nano Lett.* **21**, 26–33 (2021). doi: 10.1021/acs.nanolett.0c03939; pmid: 33258610
183. V. I. Klimov *et al*., Single-exciton optical gain in semiconductor nanocrystals. *Nature* **447**, 441–446 (2007). doi: 10.1038/nature05839; pmid: 17522678
184. K. Wu, Y.-S. Park, J. Lim, V. I. Klimov, Towards zero-threshold optical gain using charged semiconductor quantum dots. *Nat. Nanotechnol.* **12**, 1140–1147 (2017). doi: 10.1038/nnano.2017.189; pmid: 29035399
185. O. V. Kozlov *et al*., Sub-single-exciton lasing using charged quantum dots coupled to a distributed feedback cavity. *Science* **365**, 672–675 (2019). doi: 10.1126/science.aax3489; pmid: 31416959
186. C. Wang, L. Brian, Wehrenberg, Chui Y. Woo, Philippe Guyot-Sionnest, Light Emission and Amplification in Charged CdSe Quantum Dots (2004), doi:.doi: 10.1021/JP0489830
187. S. Yakunin *et al*., Low-threshold amplified spontaneous emission and lasing from colloidal nanocrystals of caesium lead halide perovskites. *Nat. Commun.* **6**, 8056 (2015). doi: 10.1038/ncomms9056; pmid: 26290056
188. C. She *et al*., Red, yellow, green, and blue amplified spontaneous emission and lasing using colloidal CdSe nanoplatelets. *ACS Nano* **9**, 9475–9485 (2015). doi: 10.1021/acsnano.5b02509; pmid: 26302368
189. B. Guzelturk, M. Pelton, M. Olutas, H. V. Demir, Giant modal gain coefficients in colloidal II-VI nanoplatelets. *Nano Lett.* **19**, 277–282 (2019). doi: 10.1021/acs.nanolett.8b03891; pmid: 30539638
190. S. Christodoulou *et al*., Single-exciton gain and stimulated emission across the infrared telecom band from robust heavily doped PbS colloidal quantum dots. *Nano Lett.* **20**, 5909–5915 (2020). doi: 10.1021/acs.nanolett.0c01859; pmid: 32662655
191. R. H. Gilmore *et al*., Epitaxial dimers and auger-assisted detrapping in PbS quantum dot solids. *Matter* **1**, 250–265 (2019). doi: 10.1016/j.matt.2019.05.015
192. M. Chen *et al*., High carrier mobility in HgTe quantum dot solids improves mid-IR photodetectors. *ACS Photonics* **6**, 2358–2365 (2019). doi: 10.1021/acsphotonics.9b01050
193. S. Zheng, J. Chen, E. M. J. Johansson, X. Zhang, PbS colloidal quantum dot inks for infrared solar cells. *iScience* **23**, 101753 (2020). doi: 10.1016/j.isci.2020.101753; pmid: 33241199
194. R. D. Schaller, V. I. Klimov, High efficiency carrier multiplication in PbSe nanocrystals: Implications for solar energy conversion. *Phys. Rev. Lett.* **92**, 186601 (2004). doi: 10.1103/PhysRevLett.92.186601; pmid: 15169518
195. M. V. Khenkin *et al*., Consensus statement for stability assessment and reporting for perovskite photovoltaics based on ISOS procedures. *Nat. Energy* **5**, 35–49 (2020). doi: 10.1038/s41560-019-0529-5
196. A. Luque, A. Martí, C. Stanley, Understanding intermediate-band solar cells. *Nat. Photonics* **6**, 146–152 (2012). doi: 10.1038/nphoton.2012.1
197. H. Li, K. Wu, J. Lim, H.-J. Song, V. I. Klimov, Doctor-blade deposition of quantum dots onto standard window glass for low-loss large-area luminescent solar concentrators. *Nat. Energy* **1**, 16157 (2016). doi: 10.1038/nenergy.2016.157

ACKNOWLEDGMENTS

Funding: F.P.G.d.A. and E.H.S. acknowledge support from the Canada Research Chair. F.P.G.d.A. was also supported by the Ministry of Economy and Competitiveness of Spain through the "Severo Ochoa" program for Centres of Excellence in R&D (SE5- 0522), Fundació Privada Cellex, Fundació Privada Mir-Puig, and the Generalitat de Catalunya through the CERCA program. D.V.T. was supported by the US Department of Defense (DOD) Air Force Office of Scientific Research under grant FA9550-18-1-0099 and MICCoM, as part of the Computational Materials Sciences Program funded by the US Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, through Argonne National Laboratory, under contract DE-AC02-06CH11357. V.I.K. acknowledges support from the Solar Photochemistry Program of the Chemical Sciences, Biosciences and Geosciences Division, Office of Basic Energy Sciences, Office of Science, US Department of Energy (overview of photophysical aspects of cQDs relevant to solar energy conversion) and the Laboratory Directed Research and Development program of Los Alamos National Laboratory under project 20200213DR (overview of studies of cQD lasers and LEDs). Y.A. acknowledges funding from JSPS KAKENHI Grant-in-Aid for Specially Promoted Research (15H05700). M.B. acknowledges the support from the Deutsche Forschungsgemeinschaft in the frame of the ICRC TRR 160 (projects A1 and B1) and by Mercur Foundation (grant Pe-2019-0022).

10.1126/science.aaz8541

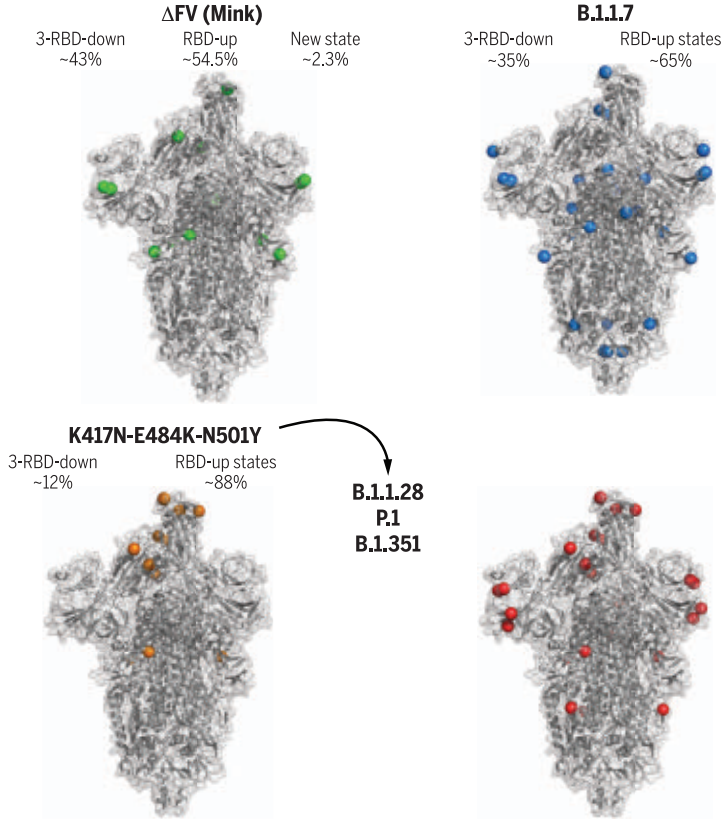
RESEARCH ARTICLE SUMMARY

CORONAVIRUS

Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity

Sophie M.-C. Gobeil, Katarzyna Janowska, Shana McDowell, Katayoun Mansouri, Robert Parks, Victoria Stalls, Megan F. Kopp, Kartik Manne, Dapeng Li, Kevin Wiehe, Kevin O. Saunders, Robert J. Edwards, Bette Korber, Barton F. Haynes, Rory Henderson*, Priyamvada Acharya*

INTRODUCTION: Variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have been circulating worldwide since the beginning of the pandemic. Some are termed Variants of Concern (VOC) because they show evidence for increased transmissibility, higher disease severity, resistance to neutralizing antibodies elicited by current vaccines or from previous infection, reduced efficacy of treatments, or failure of diagnostic detection methods. VOCs accumulate mutations in the spike (S) glycoprotein. Some VOCs that arose independently in different geographical locations show identical changes, implying convergent evolution and selective advantages of the acquired variations. A set of three amino acid substitutions in the receptor-binding domain (RBD)—Lys^{417} → Asn (K417N), Glu^{484} → Lys (E484K), and Asn^{501} → Tyr (N501Y)—occurred in the B.1.1.28 and B.1.351 lineages that originated in Brazil and South Africa, respectively. The P.1 lineage that branched off B.1.1.28 harbored a Lys^{417} → Thr (K417T) substitution while retaining the E484K and N501Y changes. The E484K substitution has attracted attention as a result of its location within the epitope of many potent neutralizing antibodies. The N501Y substitution also occurred in the B.1.1.7 variant that originated in the UK and was implicated in increased receptor binding and higher transmissibility of the variant. The B.1.1.7 variant, in turn, shares the His^{69}/Val^{70} spike deletion mutation with spike from a variant that was implicated in transmission between humans and minks (ΔFVI).



Cryo-EM structures of SARS-CoV-2 spike ectodomains. Naturally occurring amino acid variations are represented by colored spheres. Spike mutations from a mink-associated (ΔFV) (top left), B.1.1.7 (top right), B.1.351 (bottom right), and a spike with three RBD mutations (bottom left) are shown. Relative proportions of the RBD down and up populations are indicated for each. The three amino acid substitutions in the RBD—K417N/T, E484K, and N501Y—were found in the B.1.1.28 variant and are shared with the P.1 and B.1.351 lineages.

RATIONALE: Global sequencing initiatives and in vitro neutralization and antibody binding assays have rapidly provided critical and timely information on the VOCs. Here, by combining cryo–electron microscopy (cryo-EM) structural determination with binding assays and computational analyses on the variant spikes, we sought to visualize the impact of the amino acid substitutions on spike conformation to understand how these changes affect their biological function.

RESULTS: We measured angiotensin-converting enzyme 2 (ACE2) receptor and antibody binding for 19 SARS-CoV-2 S ectodomain constructs harboring amino acid changes found in circulating variants. These included a variant involved in interspecies SARS-CoV-2 transmission between humans and minks, as well as several VOCs including the B.1.1.7, B.1.1.28/P.1, and B.1.351 variants. Consistent with published neutralization data, B.1.1.7 showed decreased binding to N-terminal domain (NTD)–directed antibodies, whereas P.1 and B.1.351 showed reduced binding to both NTD- and RBD-directed antibodies. All variants showed increased binding to ACE2, which was mediated by higher propensity for RBD-up states, and affinity-enhancing mutations in the RBD. We observed spike instability in the mink-associated variant, highlighted by the presence of a population in the cryo-EM dataset with missing density for the S1 subunit of one protomer. Modulation of contacts between the SD1 and HR1 regions led to increased RBD-up states of the B.1.1.7 spike, with the protein stability maintained by a balance of stabilizing and destabilizing mutations. A local destabilizing effect of the RBD E484K mutation was implicated in resistance of the B.1.1.28/P.1 and B.1.351 variants to some potent RBD-directed neutralizing antibodies.

CONCLUSION: Our study revealed details of how amino acid substitutions affect spike conformation in circulating SARS-CoV-2 VOCs. We define communication networks that modulate spike allostery and show that the S protein uses different mechanisms to converge upon similar solutions for altering the RBD up/down positioning. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: rory.henderson@duke.edu (R.H.); priyamvada.acharya@duke.edu (P.A.)
This is an open-access article distributed under the terms of the Creative Commons Attribution license (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Cite this article as S. M.-C. Gobeil *et al.*, *Science* 373, eabi6226 (2021). DOI: 10.1126/science.abi6226

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abi6226

RESEARCH ARTICLE

CORONAVIRUS

Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity

Sophie M.-C. Gobeil[1], Katarzyna Janowska[1], Shana McDowell[1], Katayoun Mansouri[1], Robert Parks[1], Victoria Stalls[1], Megan F. Kopp[1], Kartik Manne[1], Dapeng Li[1], Kevin Wiehe[1,2], Kevin O. Saunders[1,3,4,5], Robert J. Edwards[1,2], Bette Korber[6], Barton F. Haynes[1,2,5], Rory Henderson[1,2]*, Priyamvada Acharya[1,3,7]*

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants with multiple spike mutations enable increased transmission and antibody resistance. We combined cryo–electron microscopy (cryo-EM), binding, and computational analyses to study variant spikes, including one that was involved in transmission between minks and humans, and others that originated and spread in human populations. All variants showed increased angiotensin-converting enzyme 2 (ACE2) receptor binding and increased propensity for receptor binding domain (RBD)–up states. While adaptation to mink resulted in spike destabilization, the B.1.1.7 (UK) spike balanced stabilizing and destabilizing mutations. A local destabilizing effect of the RBD E484K mutation was implicated in resistance of the B.1.1.28/P.1 (Brazil) and B.1.351 (South Africa) variants to neutralizing antibodies. Our studies revealed allosteric effects of mutations and mechanistic differences that drive either interspecies transmission or escape from antibody neutralization.

The emergence of rapidly spreading variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent for COVID-19, threatens to prolong an already devastating pandemic. Some variants have exhibited resistance in in vitro assays to neutralization by antibodies (Abs) and plasma from convalescent or vaccinated individuals, raising concerns that their resistance may reduce the efficiency of current vaccines (*1*, *2*) (www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html). Additionally, SARS-CoV-2 transmission between humans and animals has been observed in mink farms, leading to culling of large mink populations in Denmark and other countries to prevent establishment of a non-human reservoir of SARS-CoV-2 variants (*3*). Changes in the spike (S) glycoprotein (*4*, *5*) in these variants are under scrutiny because the S protein has a central role in engaging the angiotensin-converting enzyme 2 (ACE2) receptor to mediate cellular entry (*6*) and is a dominant target of neutralizing antibodies (nAbs) elicited by either vaccination or natural infection (*7*, *8*).

[1]Duke Human Vaccine Institute, Durham, NC 27710, USA. [2]Department of Medicine, Duke University, Durham, NC 27710, USA. [3]Department of Surgery, Duke University, Durham, NC 27710, USA. [4]Department of Molecular Genetics and Microbiology, Duke University, Durham, NC 27710, USA. [5]Department of Immunology, Duke University, Durham, NC 27710, USA. [6]Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. [7]Department of Biochemistry, Duke University, Durham, NC 27710, USA.
*Corresponding author. Email: rory.henderson@duke.edu (R.H.); priyamvada.acharya@duke.edu (P.A.)

The prefusion SARS-CoV-2 S trimer is composed of S1 and S2 subunits separated by a furin cleavage site (Fig. 1). The S1 subunit contains the N-terminal domain (NTD), ACE2 receptor binding domain (RBD), and two subdomains (SD1 and SD2). The NTD and RBD are dominant targets for nAbs (*9*–*12*). The RBD transitions between a "closed" ("down") receptor-inaccessible conformation and an "open" ("up") conformation that allows binding to the ACE2 receptor (*13*–*15*). Variations in distal regions of the S protein can have allosteric effects on RBD up/down disposition (*16*–*20*), with SD1 and SD2 playing essential roles in modulating spike allostery (*16*). Whereas the S1 subunit shows large motions, the prefusion S2 remains mostly invariant. The S2 subunit contains a TMPRSS2 cleavage site (S2′), followed by the fusion peptide (FP), heptad repeat 1 (HR1), central helix (CH), connector domain (CD), heptad repeat 2 (HR2), transmembrane domain (TM), and cytoplasmic tail (CT) (Fig. 1). After binding ACE2 receptor, and following proteolysis at the furin and TMPRSS2 cleavage sites, the spike undergoes large conformational changes leading to cellular entry (*6*, *21*–*23*).

The autumn of 2020 was marked by the appearance of several fast-spreading SARS-CoV-2 variants with S protein variations accumulating in the background of the Asp^{614} → Gly (D614G) substitution (*24*). Some amino acid substitutions recur in variants that originated independently in different geographical locations, suggesting convergent evolution and selective advantages of these changes. Here, we determined the structures of S protein variants and measured the binding of these variants to ACE2 and Abs.. These include a variant that was implicated in SARS-CoV-2 transmission between humans and minks (*25*) and a few that originated and spread in human populations. Three RBD substitutions—Lys^{417} → Asn (K417N), Glu^{484} → Lys (E484K), and Asn^{501} → Tyr (N501Y)—occurred in the B.1.1.28 and the B.1.351 lineages that originated in Brazil and South Africa, respectively. The P.1 lineage, which branched off from B.1.1.28, incorporated a Lys^{417} → Thr (K417T) change and retained the E484K and N501Y substitutions. The N501Y substitution also occurred in the B.1.1.7 variant that originated in the UK (*26*–*31*). Our studies revealed different residue interaction networks in the variant spikes that converge on similar solutions for altering spike conformation and RBD up/down positioning. These findings elucidate the structural mechanisms underlying the effects of spike mutations on transmissibility and immune evasion.

Binding of SARS-CoV-2 S protein variants to ACE2 receptor and antibodies

We used the previously described S-GSAS-D614G S ectodomain as a template here (Fig. 1 and table S1) (*16*) (referred to as "D614G spike" hereafter). This template includes SARS-CoV-2 S residues 1 to 1208, an Arg-Arg-Ala-Arg (RRAR) to Gly-Ser-Ala-Ser (GSAS) substitution that renders the furin cleavage site inactive, and a foldon trimerization motif at the spike C terminus, followed by a C-terminal TwinStrep tag. All purified S proteins showed similar migration profiles upon SDS–polyacrylamide gel electrophoresis (PAGE) and size exclusion chromatography (SEC), with high-quality spike preparations confirmed by negative-stain electron microscopy (NSEM) (fig. S1) (*32*).

We used surface plasmon resonance (SPR) and enzyme-linked immunosorbent assay (ELISA) to measure spike binding to the ACE2 receptor ectodomain and to Abs (Fig. 1, figs. S2 to S4, and table S2). Abs included RBD-directed, potent nAbs DH1041 and DH1043, whose epitopes overlap with the ACE2 binding site; RBD-directed highly cross-reactive nAb DH1047, which neutralizes SARS-CoV-1, SARS-CoV-2, and bat CoVs; NTD-directed nAbs DH1050.1 and DH1050.2, which bind an antigenic supersite; NTD-directed non-neutralizing Ab (nnAb) DH1052; fusion peptide–directed cross-reactive Ab DH1058; and S2 glycan cluster–directed nnAb 2G12 (fig. S4) (*9*, *33*–*37*). All variants bound ACE2 at higher levels relative to the D614G spike (Fig. 1C and figs. S2 and S3), with S-GSAS-B.1.1.7 ("B.1.1.7 spike") displaying the greatest increase. DH1047 showed similar binding levels to all spike variants (Fig. 1D and figs. S2 and S3), consistent with neutralization of B.1.1.7 and B.1.351 by DH1047 (*34*). The RBD-directed nAb DH1041 showed similar binding levels to the B.1.1.7 and D614G spikes, consistent with its neutralization of the B.1.1.7 pseudovirus

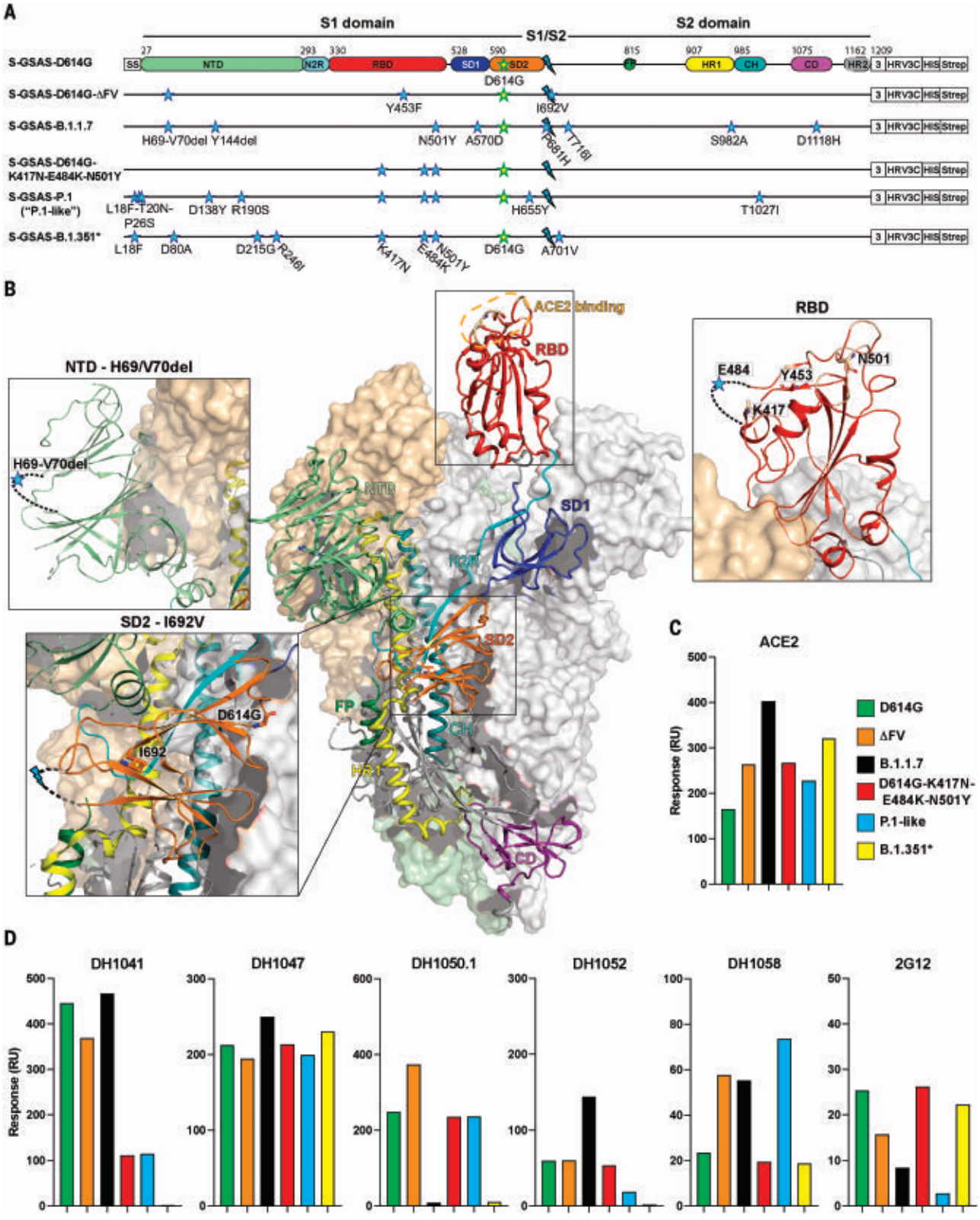


Fig. 1. SARS-CoV-2 spike (S) protein ectodomains for characterizing structures and antigenicity of S protein variants. (**A**) Domain architecture of the SARS-CoV-2 spike protomer. The S1 subunit contains a signal sequence (SS), the NTD (N-terminal domain, pale green), N2R (NTD-to-RBD linker, cyan), RBD (receptor binding domain, red), and SD1 and SD2 (subdomains 1 and 2, dark blue and orange). The S2 subunit contains the FP (fusion peptide, dark green), HR1 (heptad repeat 1, yellow), CH (central helix, teal), CD (connector domain, purple), and HR2 (heptad repeat 2, gray) regions. The transmembrane domain (TM) and cytoplasmic tail (CT) have been truncated and replaced by a foldon trimerization sequence (3), an HRV3C cleavage site (HRV3C), a His-tag (HIS), and a strep-tag (Strep). The D614G mutation (yellow star with green outline) is in SD2. The S1/S2 furin cleavage site (RRAR)

has been mutated to GSAS (blue lightning). The substitutions in each variant are indicated by blue stars. *A few ectodomain constructs were prepared on the B.1.351 spike backbone; these differed in their NTD mutations (see table S1). Binding data for the other constructs, including the one representing the dominant circulating form (L18F, D80A, D215G, Δ242-244, K417N, E484K, N501Y, D614G, A701V), are shown in figs. S2 and S3. The construct shown here was used for determining the cryo-EM structure (Fig. 6). The "P.1-like" spike was prepared in the P.1 backbone but retained the K417N RBD substitution (instead of the K417T in the P.1 spike; see table S1). (**B**) Representation of the trimeric SARS-CoV-2 spike ectodomain in a prefusion conformation with one RBD up (PDB ID 7KDL). The S1 subunit on an RBD-down protomer is shown as a pale orange molecular surface; the S2 subunit is shown in pale green. The subdomains on an RBD-up protomer are colored according to (A) on a ribbon diagram. Each inset corresponds to the spike regions harboring mutations included in this study. (**C** and **D**) Binding of ACE2 (C) and of RBD-directed antibodies DH1041 and DH1047, NTD-directed antibodies DH1050.1 and DH1052, and S2-directed antibodies DH1058 and 2G12 (D) to spike variants measured by SPR. Data are representative of two independent experiments.

(*38*). The S-GSAS-D614G-K417-E484K-N501Y (the "triple mutant spike") showed reduced binding to RBD-directed nAbs DH1041 and DH1043. These results are consistent with the inability of class 2 RBD–binding Abs, where the E484K substitution occurs within the epitope, to neutralize variants that harbor the E484K substitution (*2*).

We tested several variants in the B.1.351 spike backbone (Fig. 1, figs. S2 and S3, and table S1). We found that the commonly occurring 242–244 deletion, and a rare Arg^{246} → Ile substitution that is included in some reagent panels and candidate vaccines (*39*), can each affect binding of not only NTD-directed Abs, but also RBD-directed Abs DH1041 and DH1043. Whereas binding of NTD-directed nAbs DH1050.1 and DH1050.2 to B.1.1.7 and B.1.351 spikes was markedly reduced, their binding to the triple mutant spike and S-GSAS-P.1 (or "P.1-like spike") remained unchanged. This is consistent with neutralization data, where mAbs 5-24 and 4-8 (which target the same antigenic supersite as DH1050.1) lost activity against B.1.351 but neutralized P.1 (*40*).

In summary, our binding data are consistent with biological data obtained in in vitro neutralization assays, thus establishing that our SARS-CoV-2 S ectodomain constructs are an effective mimic of native spikes and supporting their use for studying structural changes due to amino acid substitutions in spike variants.

Structural analysis of mink-associated "cluster 5" spike mutations

Spillover of SARS-CoV-2 from humans to minks, and then from minks to humans, was first reported in April 2020 in the Netherlands and subsequently independently reported in Denmark, Spain, Italy, the United States, Sweden, and Greece (*25*). Five S mutations were observed in a variant named "cluster 5"; these included a His^{69}/Val^{70} NTD deletion (ΔH69/V70), RBD Tyr^{453} → Phe (Y453F) substitution, SD2 Ile^{692} → Val (I692V) substitution, and Met^{1229} → Ile (M1229I) in the TM. To understand how these affect spike conformations, we determined cryo-EM structures of S-GSAS-D614G-ΔFV ("ΔFV spike"), which included all but the TM M1229I substitution (Fig. 1, A and B, and table S1). We identified four 3-RBD-down populations, which we named 3D-1, 3D-2, 3D-3, and 3D-4 (PDB 7LWL, 7LWI, 7LWK, and 7LWJ, respectively) (Fig. 2A), refined to overall resolutions of 2.8 to 3.2 Å; three 1-RBD-up populations, which we named 1U-1, 1U-2, and 1U-3 (PDB 7LWM, 7LWN, and 7LWO, respectively), refined to resolutions of 2.8 to 2.9 Å; and one 2-RBD-up population (2U; PDB 7LWP) refined to 3.0 Å (Fig. 2B, figs. S5 and S6, and table S3). A previously unobserved state (M1; PDB 7LWQ, 3.2 Å) was identified, with two RBDs in the down position and no density visible for the entire S1 subunit of the third protomer (Fig. 2C, figs. S5 and S6, and table S3). The 3-RBD-down states were ~43% of the total population, with the rest of the particles constituting "open" states, including ~47% 1-RBD-up, ~7.5% 2-RBD-up, and ~2.3% of the M1 spike. Thus, we observed a modest decrease in the 3-RBD-down state from ~56% that we had reported for the D614G spike, and the appearance of open states (2-RBD-up and M1) that were not observed for the S-GSAS-D614G dataset (*16*).

Upon closer examination, we noted unusual variability in the S2 subunit of the ΔFV 3-RBD-down structures. We compared these structures either by aligning them using S2 residues 908 to 1035 of the HR1-CH region (fig. S7A) or by calculating difference distance matrices (DDMs) for superposition-free comparisons between pairs of structures (fig. S7, B and C, and supplementary text) (*41*). Both methods revealed considerable variability in S2, which was most pronounced for the 3D-4 structure (Fig. 2A and fig. S7). By contrast, the three 1-RBD-up structures showed little variability in S2, which suggests that cluster 5 mutations largely affect the 3-RBD-down state (fig. S8) (*16*). The variation in the S2 region was unexpected because the S2 subunit had appeared relatively invariant in prior studies (*16*, *42*, *43*).

We next sought to understand the effect of each amino acid substitution on the functional and structural properties of the spike. The ΔFV spike bound ACE2 with improved affinity over the D614G spike by a factor of ~3.5, resulting from a decreased off-rate mediated by the Y453F substitution (Fig. 2D, fig. S9, and table S2). Although neither the I692V substitution nor ΔH69/V70 affected ACE2 binding affinity, ΔH69/V70 contributed to increased affinity for the NTD-directed nAbs DH1050.1 and DH1050.2. The I692V substitution occurs in SD2, where small changes can translate to large movements in the NTD and RBD regions (Fig. 1) (*16*, *19*). In the D614G spike, Ile^{692} contacts Pro^{600}; loss of the methyl due to the I692V substitution increases the distance between Pro^{600} and Val^{692} (fig. S10). We observed disorder in the 3D-4 cryo-EM map, accompanied by the largest separation between Pro^{600} and Val^{692} of all the ΔFV spike 3-RBD-down structures. This local destabilization around the I692V substitution in 3D-4, together with DDM comparisons and superpositions that showed 3D-4 to be the most asymmetric of the 3-RBD-down structures as well as the most variable in the S2 subunit, suggested a role for the I692V substitution in the 3-RBD-down state disorder.

To define and quantify changes in ΔFV spike domain orientations, and to determine how local changes around the SD2 I692V substitution propagate to adjacent domains, we examined its quaternary structure using a vector representation (*19*). This was accomplished by assigning a central coordinate to each domain and calculating angles, dihedrals, and distances between different structural elements (Fig. 2E and supplementary text). Principal components analysis (PCA) of these intraprotomer vector relationships showed that the 3D-4 protomers occupied a distinct cluster (Fig. 2F), consistent with the DDM analysis (fig. S7, B and C). The two RBD-down protomers in M1(A and C) were similar to 3D-1(A), 3D-2(C), 1U-1(A), and 2U(C) protomers along the first principal component (PC1), with M1(A) separating from M1(C) in PC2 into a 3D-1(A)–containing cluster. Both 3D-1(A) and 3D-3(C) occupied extreme positions in the vector set for angles involving the NTD′, subdomains, and the RBD that mimic the 1U-1(A) structure (fig. S11). Because constraints on RBD-down protomers are relaxed in spikes with at least one RBD in the up position, this may represent a particularly stable protomer position. Together, the vector clustering is consistent with structural observations for the 3D-4 structure and indicates that loss of a single S1 protomer in M1 allowed its two other RBD-down protomers to relax to a configuration resembling RBD-down protomers in 1-RBD-up spikes.

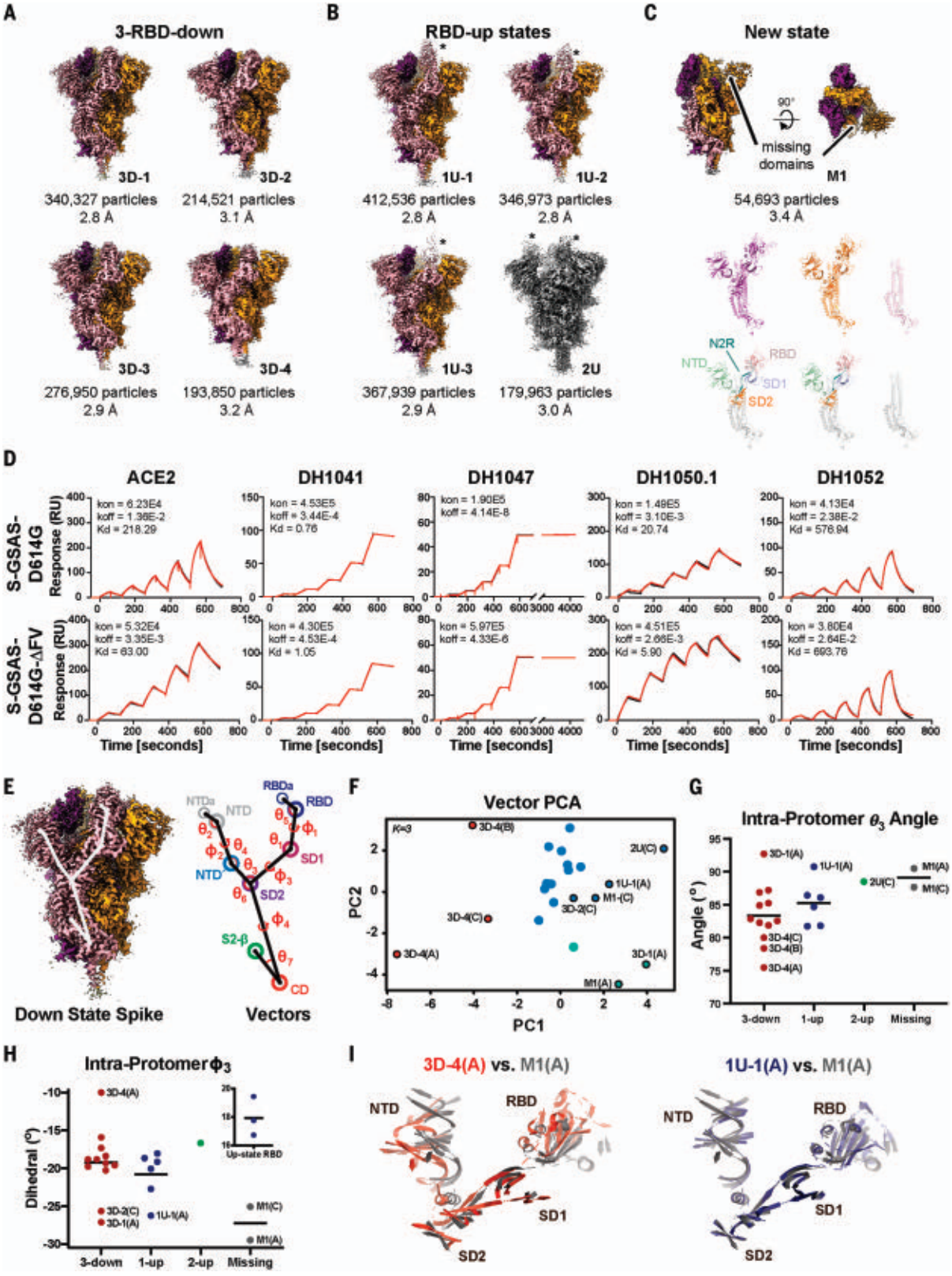


Fig. 2. Structures and antigenicity of the mink-associated ΔFV spike ectodomain. (**A** to **C**) Cryo-EM reconstructions of the ΔFV ectodomain colored by protomer chains. (A) 3-RBD-down states: 3D-1 (EMDB 23549, PDB 7LWL), 3D-3 (EMDB 23548, PDB 7LWK), 3D-2 (EMDB 23546, PDB 7LWI), 3D-4 (EMDB 23547, PDB 7LWJ). (B) RBD-up states, including 3 1-RBD-up states: 1U-1 (EMDB 23550, PDB 7LWM), 1U-2 (EMDB 23551, PDB 7LWN), 1U-3 (EMDB 23552, PDB 7LWO), and a 2-RBD-up state (EMDB 23553, PDB 7LWP). The asterisks are placed next to the RBD in the up position. (C) M1 (EMDB 23554, PDB 7LWQ), a state lacking the S1 subunit and SD2 subdomain of one of the three protomers. Top: Two views of the cryo-EM reconstruction rotated by

90°; middle, the individual protomers colored to match the colors in the top panel; bottom, the protomers with RBDs colored salmon, NTDs green, SD1 blue, SD2 orange, and the S2 subunit gray. (**D**) Binding of ACE2 receptor ectodomain (RBD-directed) and antibodies DH1041 and DH1047 (RBD-directed, neutralizing), DH1050.1 (NTD-directed, neutralizing), and DH1052 (NTD-directed, non-neutralizing) to D614G (top row) and B.1.1.7 (bottom row) spikes, measured by SPR using single-cycle kinetics. The red lines are the binding sensorgrams; the black lines show fits of the data to a 1:1 Langmuir binding model. The on-rate (k_{on}, M^{-1} s^{-1}), off-rate (k_{off}, s^{-1}), and binding affinity (K_D, nM) for each interaction are indicated. RU, response units. The binding of DH1047 to spike was too tight to allow accurate affinity measurement. (**E** to **I**) Vector analysis defining changes in intraprotomer domain dispositions. (E) Left: Map of the 3-RBD-down spike highlighting vector positions. Right: Schematic showing angles and dihedrals between different structural elements in the SARS-CoV-2 S ectodomain. (F) Principal components analysis of the intraprotomer vector magnitudes, angles, and dihedrals. Dot color indicates K-means cluster assignment. (G) Intraprotomer θ_3 angles formed by NTD′, SD2, and SD1. (H) Intraprotomer ϕ_3 dihedral angle describing rotation of the NTD′ relative to the RBD about an SD2-to-SD1 axis. (I) Chain A of the M1 protomer aligned to the chain A of 3D-4 (left) and chain A of 1U-1 (right). The protomers were aligned on SD2; for clarity, only secondary structural elements are shown.

We next examined the angle formed by the NTD′, SD2, and SD1 domain centers, termed θ_3, and a dihedral describing how the NTD′, SD2, SD1, and RBD rotate relative to one another, termed ϕ_3 (Fig. 2, E to H). The 3D-4 protomers occupied a distinct ϕ_3 and θ_3 angle cluster (fig. S12); in particular, the 3D-4(A) protomer ϕ_3 dihedral differed markedly from the primary cluster in the direction of up-state protomers (Fig. 2H, inset). Consistent with the PCA clustering, the θ_3 angles of 3D-1(A), 1U-1(A), and 2U(C) were similar to those of the M1 protomers. The 3D-2(C), 3D-1(A), and 1U-1(A) protomers displayed ϕ_3 dihedrals similar to those of the M1 protomers (Fig. 2, G and H). The similarity of the M1 protomers and the up-state protomers suggests that the M1 state occurs to release strain from 3-RBD-down configurations induced by the cluster 5 mutations. Comparing the 3D-4(A) S1 subunit structure to that of M1(A) demonstrated the marked differences in their RBD positioning, whereas alignment of M1(A) S1 subunit to 1U-1(A) showed their similarity (Fig. 2I).

Comparing the ΔFV spike 3-RBD-down structures to our previously published D614G spike structures (PDB 7KE4, 7KE6, 7KE7, and 7KE8) revealed that the 3D-1 and 3D-2 protomers closely matched 7KE4 and 7KE8, respectively, in their intraprotomer ϕ_3 and θ_3 angles (fig. S11, A and B). Two protomers in the 3D-4(B and C) structure resembled two protomers in the 7KE8(A and B) D614G spike structure in their ϕ_3 dihedrals. Both the 7KE8 and 3D-4 structures displayed marked asymmetry, with the third protomer in each occupying an extreme dihedral angle; in 7KE8(C), the NTD and RBD are rotated toward S2, whereas 3D-4(A) showed a rotation in the opposite direction (fig. S11C). As a result of contact between SD1 and NTD′, this results in global shifts of S1 elements away from S2. These shifts, together with close contact between S2 and these S1 domains, result in changes in S2 structure leading to the variability observed in our structural analysis (fig. S7). The large separation of S1 from S2 in the 3D-4(A) protomer (fig. S11C) suggests that it could be an intermediate that leads to the S1-dislocated M1(B) state. The 3D-3 structure also lacked a close match (fig. S11, A and B). Alignment of 3D-3 to its most similar D614G down-state trimer structure, 7KE7, indicated similar but less extreme differences in domains, which suggests that 3D-3 is yet another intermediate structure leading to the pre-M1 3D-4 state. Thus, by combining cryo-EM classifications and vector analysis, we tracked the origin of the observed instability in the ΔFV spike and found evidence of instability in two 3-RBD-down structures (3D-3 and 3D-4) that leads to dislocation of a S1 protomer in M1.

In summary, our data show that interspecies adaptation involves improved receptor binding affinity of the ΔFV spike mediated primarily by the RBD Y453F substitution. The observed increase in RBD-up states may also contribute to higher levels of ACE2 binding by providing more receptor-accessible sites. We found no evidence in the binding data for immune evasion at the dominant neutralization sites; this is consistent with previous findings that neutralization potency of a panel of RBD antibodies was not notably affected by Y453F or ΔH69/V70 (*38*). Structural analysis revealed destabilization of the 3-RBD-down state and loss of tight regulation of its conformation in the mink-associated ΔFV spike. We can infer from these structures that in the virion-associated spike these changes could have an impact on spike stability, possibly leading to premature S1 shedding.

Structural analysis of the SARS-CoV-2 S protein B.1.1.7 variant

The B.1.1.7 variant emerged in the UK in September 2020 and spread worldwide, with reports of increased transmissibility, virulence, and mortality (*44*). An RBD N501Y substitution results in improved ACE2 affinity (*45*). The N501Y substitution, either on its own or in combination with the NTD ΔH69/V70 deletion or the SD2 P681H mutation, does not substantially affect serum neutralization elicited by current vaccines (*1*, *38*, *46*, *47*). Although susceptible to RBD-directed nAbs such as DH1041, DH1043, and DH1047 (*9*, *38*), B.1.1.7 shows increased resistance to NTD-directed Abs including 4A8 (PDB 7C2L), 5-24, and 4-8 (*10*, *48*). This resistance was attributed to the ΔY144 deletion, which occurs in a NTD loop that forms an antigenic supersite (*49*) also targeted by the DH1050.1 nAb (PDB 7LCN) (*50*).

Our binding data were consistent with published neutralization data (Fig. 1, Fig. 3A, and figs. S2 and S3). B.1.1.7 spike affinity for ACE2 was improved over the D614G spike by a factor of ~5 as a result of the N501Y substitution. We measured nanomolar affinity of the B.1.1.7 spike for NTD-directed nAb DH1050.1, albeit at substantially reduced binding levels relative to the D614G spike (Fig. 3A, figs. S2, S3, and S9, and table S2), consistent with impairment of the NTD antigenic supersite in B.1.1.7 (*49*), while retaining robust binding to most RBD-directed antibodies.

To visualize the impact of the amino acid variations on the spike conformation, we determined cryo-EM structures of the B.1.1.7 spike (Fig. 3, B to D, figs. S13 and S14, and table S3). Multiple populations of the 3-RBD-down and RBD-up states were identified, with a higher proportion of RBD-up particles observed for the B.1.1.7 spike (~1.8:1 RBD-up/RBD-down) relative to the D614G spike (~0.8:1) (*16*) and the mink-associated ΔFV spike (~1.3:1) (Fig. 2, A to C). Three populations of 3-RBD-down spike were refined to 3.2 to 3.6 Å (Fig. 3B, figs. S13 and S14, and table S3), each showing visible asymmetry with weaker density for one of its RBDs (Fig. 3B), suggestive of enhanced mobility. We identified several RBD-up structures, including a typical 1-RBD-up state (Fig. 3C) and two 1-RBD-up populations with the up RBD and its adjacent NTD disordered (Fig. 3D). We identified states with 2- or 3-RBD up (fig. S13G) that were not detected in the D614G spike (*16*). We were unable to obtain high-resolution reconstructions of these populations because of their limited particle numbers and preferred orientations of the particles. Unlike the mink-associated ΔFV spike structures, DDM analysis of the B.1.1.7 structures did not show variability in S2 (fig. S15). The apparent increase in RBD mobility in the B.1.1.7 spike 3-RBD-down structures suggested a reduced barrier for up-state transition due to weakening of down-state contacts. RBDs in their down state contact an adjacent NTD and another RBD via interprotomer protein-protein and protein-glycan contacts (Fig. 3B, inset) (*51*, *52*). Transition from the down to the up state replaces these contacts with differing RBD-to-NTD and RBD-to-RBD contacts (Fig. 3C, inset).

We next sought to understand how variations that are distal from the RBD/NTD region influence the B.1.1.7 spike conformational

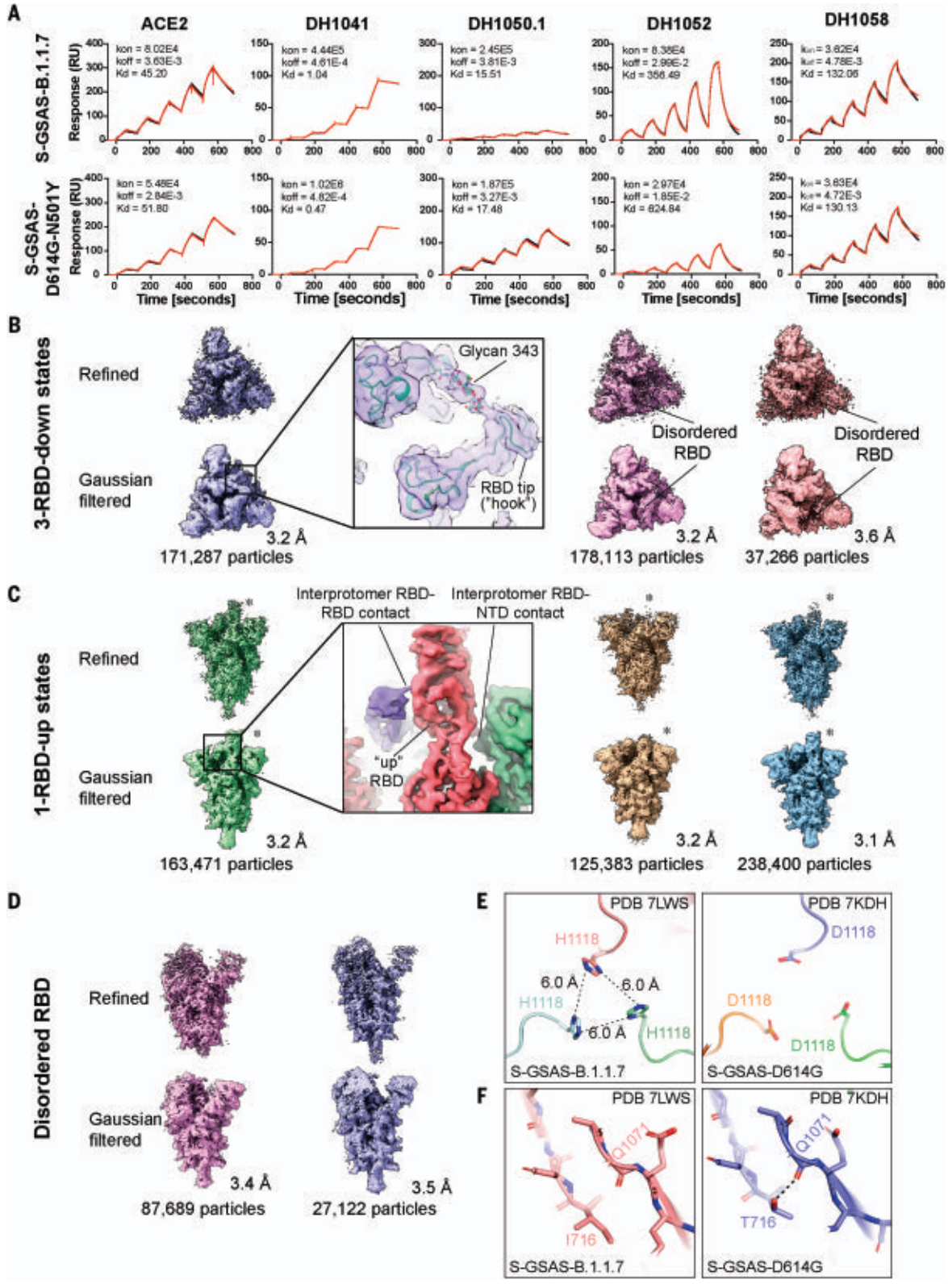


Fig. 3. Antigenicity and structures of the B.1.1.7 spike. (**A**) Binding of ACE2 receptor ectodomain (RBD-directed) and antibodies DH1041 and DH1047 (RBD-directed, neutralizing), DH1050.1 (NTD-directed, neutralizing), and DH1052 (NTD-directed, non-neutralizing) to B.1.1.7 (top) and N501Y (bottom) measured by SPR using single-cycle kinetics. The red lines are the binding sensorgrams; the black lines show fits of the data to a 1:1 Langmuir binding model. The on-rate (k_{on}, M^{-1} s^{-1}), off-rate (k_{off}, s^{-1}), and binding affinity (K_D, nM) for each interaction are indicated. (**B** to **D**) Cryo-EM reconstructions of 3-RBD-down states (B), 1-RBD-up states (C), and 1-RBD-up states with disordered RBD (D). The asterisks are placed next to the RBD in the up position. (**E**) Residue His^{1118} in the B.1.1.7 spike (PDB 7LWS) and Asp^{1118} in the D614G spike (PDB 7DKH). (**F**) Ile^{716} in the B.1.1.7 spike and Thr^{716} in the D614G spike. Dashed line shows H-bond with backbone carbonyl of Gln^{1071}.

distribution. These variations spanned multiple domains including SD1 [Ala^{570} → Asp (A570D)], SD2 [Pro^{681} → His (P681H)], HR1 [Ser^{982} → Ala (S982A)], CD [Asp^{1118} → His (D1118H)], and the linker between SD2 and fusion peptide [Thr^{716} → Ile (T716I)] (Fig. 1A). The P681H substitution located near the furin cleavage site could not be visualized because of the disorder in the cryo-EM map in that region. The D1118H substitution, on the other hand, was well resolved and formed a symmetric histidine triad near the base of the spike (Fig. 3E and fig. S14, B and C). Although the histidines were positioned too far from each other for direct hydrogen bonding, water-mediated interactions are feasible at this separation. Moreover, the cryo-EM reconstructions showed evidence for alternate conformations that could place the histidines into closer proximity (fig. S14B). By contrast, the T716I substitution abrogated an intraprotomer hydrogen bond (H-bond) between the Thr^{716} side-chain and Gln^{1071} main-chain carbonyls (Fig. 3F), suggesting a local destabilizing effect.

The A570D and S982A substitutions (Fig. 4, A to E), in the SD1 and HR1 regions, respectively, appeared to be counterposing. The A570D substitution resulted in an interprotomer H-bond with the Asn^{856} side chain, reinforcing the stacking of the SD1 loop against the HR1 helix of the adjacent protomer (Fig. 4, A and B). The HR1 S982A substitution, on the other hand, resulted in the loss of an interprotomer H-bond between the Ser^{982} and Thr^{547} side chains (Fig. 4, C and D). Comparing the down (PDB 7KDK) and up (PDB 7KDL) protomers in the D614G spike (*16*) showed concerted ~5- to 6-Å shifts in the Ala^{570} and Thr^{547} loop positions, with the Thr^{547} loop in the up protomer shifted farther away from Ser^{982} and no longer within H-bonding distance of it (Fig. 4D). Thus, the S982A mutation appears to disable a latch that modulates the RBD up/down equilibrium, thereby increasing RBD up propensity (Fig. 4E). We had previously engineered a construct, named u1S2q, where modulation of a latch involving the Ala^{570} loop was implicated in shifting its RBD up/down equilibrium (*19*).

To gain insight into how the Ser^{982}-Thr^{547} interprotomer latch affects the spike quaternary structure, we defined a new set of interprotomer vectors (Fig. 4F). Within each S protomer we defined a "unit" comprising the SD1/RBD region and the NTD/NTD′ region of the adjacent protomer with which it interacts. These units are in conformational communication ("Com") through RBD-to-RBD contacts at the apex, as well as through the SD2 subdomain. We examined the relative disposition of the three units and of SD2 by using a vector network spanning the trimer. For each structure, the protomer that contained the disordered RBD (termed Protomer$_B$) showed a marked increase in the intraprotomer angle formed by the NTD′, SD2, and an SD2 anchor (SD2a) point ($\theta_{5'}$) relative to this angle in the other two protomers (Protomer$_A$ and Protomer$_C$). This occurred in conjunction with a shift in the angle between the SD2, SD2a, and SD1 ($\theta_{6'}$; Fig. 4G). These angular changes were accompanied by a rotation of SD1 and SD2a about an axis connecting the NTD′ and SD2 ($\phi_{8'}$) as well as a compensatory rotation of the SD1 to adjacent NTD′ ($\phi_{1'}$). This compensatory shift occurs as a result of differences in the Ala^{570} loop positions. With the SD2 orientation relative to S2 largely similar to that of the other protomers, these movements can be ascribed to the S982A- and A570D-induced movements of SD1. Together, these changes resulted in disengagement of the NTD from the adjacent RBD, explaining the increase in RBD disorder. Thus, the S982A and A570D pairing acts as an allosteric switch through coupled domain movements.

In summary, structural analysis of the B.1.1.7 spike highlights how allosteric effects of variations in distal regions alter RBD disposition. In B.1.1.7, amino acid substitutions that destabilize the 3-RBD-down or closed state to favor RBD-up or open states are balanced by substitutions that stabilize the prefusion spike conformation. Thus, whereas the T716I substitution disrupts an intraprotomer H-bond, the D1118H histidine triad appears to play a stabilizing role. Similarly, whereas the S982A substitution abrogates an H-bond, facilitating RBD-up movement, the A570D substitution adds an H-bond with Asn^{856}, stabilizing interactions between HR1 and SD1. The accumulation of stabilizing contacts in the B.1.1.7 spike, even as it acquires mutations that enable increased presentation of receptor-accessible RBD-up states, may contribute to stabilizing the prefusion spike to prevent premature S1 shedding.

Structural analysis of variants bearing the K417N, E484K, and N501Y RBD mutations

Multiple variants that originated independently in different geographical locations show three amino acid substitutions (K417N, E484K, and N501Y) in the RBD, which suggests convergent evolution and selective advantage of these substitutions. Of these, the E484K mutation is of particular concern because of its location within nAb epitopes, and it has been shown to reduce or eliminate binding to many potent RBD-directed nAbs (*2*). The E484K and K417N-E484K-N501Y ("triple mutant RBD") substitutions abolished binding of the potent class 2 RBD nAbs DH1041 and DH1043 to an RBD construct (Fig. 5A and fig. S16) (*33*). We found, however, that high-affinity binding of DH1041 and DH1043 to S-GSAS-D614G-E484K ("E484K spike") and S-GSAS-D614G-K417N-E484K-N501Y ("triple mutant spike") was retained, albeit at reduced levels (Fig. 5, B and C, and figs. S2, S3, and S9).

To understand why some binding to DH1041 and DH1043 was retained for the E484K variant in the context of a S ectodomain, whereas binding was completely abrogated in the RBD-only construct, we studied the effect of the mutations on RBD conformation using molecular dynamics (MD) simulations to compare the native RBD and the triple mutant RBD (model included residues 327 to 529 in each). We built Markov state models of transitions between conformational states from large ensembles of short MD simulations for both constructs (figs. S17 to S20 and table S4; total simulation time ~260 μs each). The Markov models were characterized by a hook-like folded RBD tip (the "Hook" state), which resembled the conformation observed in x-ray crystal structures (*33*, *53*), and a highly dynamic "Disordered" state in which the RBD tip cycles between a variety of conformations (Fig. 5, D and E, and figs. S18 and S19, C and E). Whereas the native RBD displayed a nearly even proportion of Hook versus Disordered states (Fig. 5D), the triple mutant RBD showed a marked increase in the Disordered state (Fig. 5E). These population differences result from an increased transition rate to the Disordered state from the Hook state combined with a slower transition rate back to the Hook state in the triple mutant RBD relative to the native RBD (figs. S18F and S19F). Monitoring the interactions between the residue 484 side chains in each model indicated that the native Glu^{484} hydrogen bonding with the Phe^{490} backbone in particular acted to stabilize the Hook state (fig. S20). In the Disordered state, the Lys^{484} side chain forms fewer interactions across the RBD relative to Glu^{484} (fig. S20B). Together, these results are consistent with the loss in binding of Abs DH1041 and DH1043 to the RBD E484K variant and indicate that the E484K substitution destabilizes the native conformation of the RBD tip, hindering binding of class 2 RBD–directed SARS-CoV-2–neutralizing Abs.

To visualize the impact of RBD tip conformational variability on the spike, we determined cryo-EM structures of the triple mutant spike (Fig. 6A, figs. S21 and S22, and table S3). We identified 3-RBD-down, 1-RBD-up, and 2-RBD-up states, as well as intermediate states that showed one RBD in the up position and another RBD partially up. 3-RBD-down states accounted for ~12% of the total spike population and showed considerable disorder in their RBDs, with the disorder being most pronounced for one of the three RBDs and its contacting NTD (Fig. 6A).

We next studied spikes that, in addition to the RBD K417N-E484K-N501Y substitutions, also contained multiple residue changes in the NTD, and an Ala^{701} → Val substitution, found in B.1.351 (Fig. 1, Fig. 6, B and C, and figs. S2, S3, S23, and S24). Despite no additional

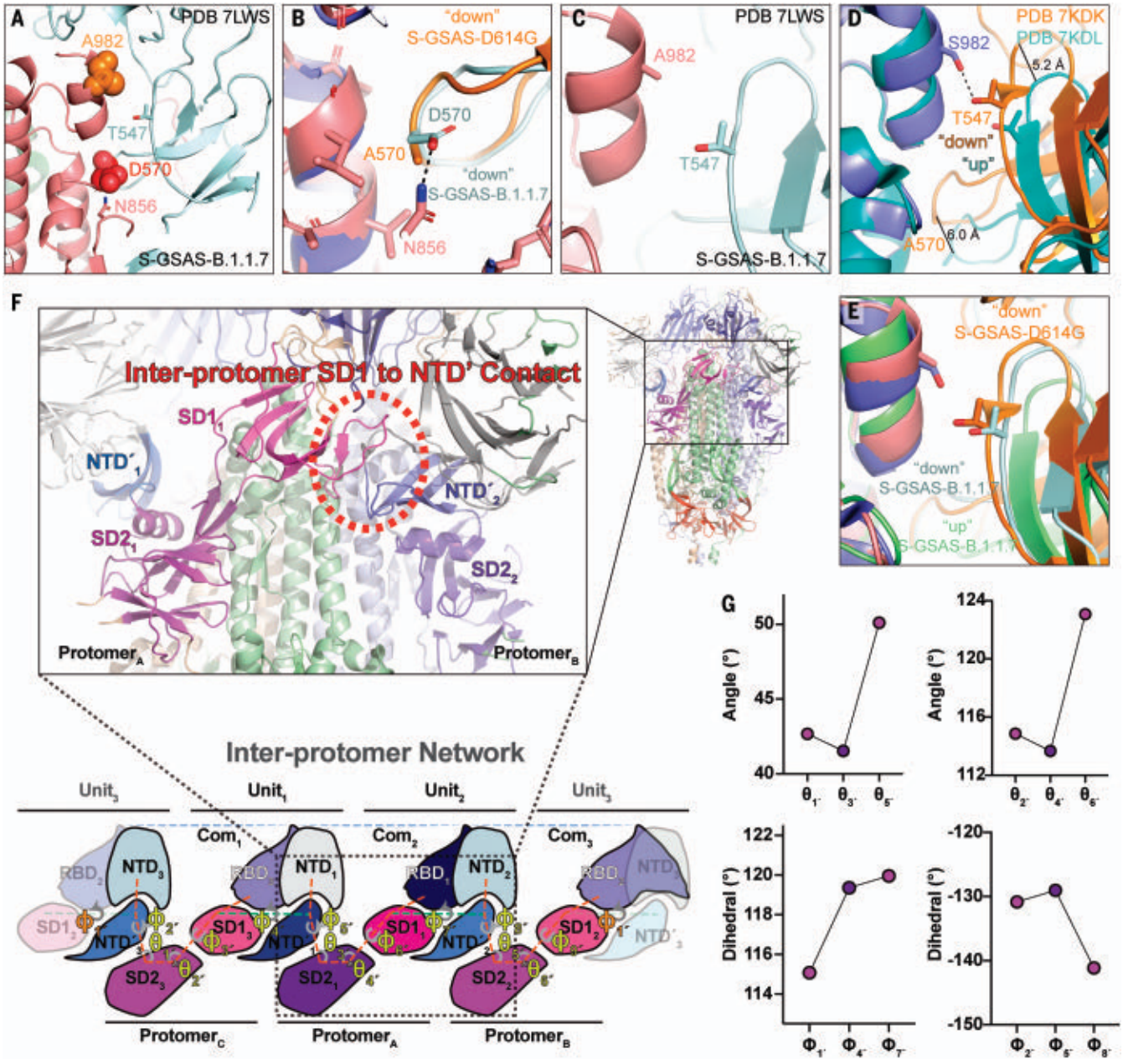


Fig. 4. Details of the B.1.1.7 spike modulation of the Ser^{982}-Ala^{570} latch. (**A**) Zoomed-in view of the region of the A570D (red spheres) and S982A (orange spheres) substitutions in the B.1.1.7 spike; S protomers are colored pale cyan and salmon. (**B**) Overlay of 3-RBD-down structures of the D614G spike (PDB 7KDK; orange and slate blue) and the B.1.1.7 spike (PDB 7LWS; pale cyan and salmon). (**C**) Zoomed-in view of region around the B.1.1.7 spike S982A substitution (PDB 7LWS). Residues Ala^{982} and Thr^{547} are shown in sticks. (**D**) Overlay of 3-RBD-down (PDB 7KDK, orange and slate blue) and 1-RBD-up (PDB 7KDL, teal) structures of S-GSAS-D614G, showing movement of the Thr^{547} and Ala^{570} loops and loss in H-bond between Thr^{547} and Ser^{982} upon transition from the down to the up state. (**E**) Overlay of 3-RBD-down structures of S-GSAS-D614G (PDB 7KDK, orange and slate blue) and S-GSAS-B.1.1.7 (PDB 7LWS, pale cyan and salmon), and 1-RBD-up structure of S-GSAS-B.1.1.7 (PDB 7LWV, green). Relative to the S-GSAS-D614G down state, the Thr^{547} loop in the B.1.1.7 spike down state protomer is shifted toward the loop position in the up protomer. Residues 908 to 1035 were used for the overlays. Hydrogen bonds are shown as dashed lines. (**F**) Top left: Zoomed-in view of the S1 interaction network spanning Protomer$_A$ and Protomer$_B$ highlighting the locations of the NTD's, SD2s, SD1s, and the interprotomer contact point between SD1 and the NTD'. Top right: S ectodomain trimer indicating the zoomed-in location. Bottom: Vector network connecting the protomer NTD', SD2, and SD1 domains. The SD2 anchor point (SD2a) is indicated by the asterisk. Interactive, interprotomer contact units involving SD1/RBD to NTD/NTD' pairs are identified with RBD-to-RBD communication (Com) points highlighted. Dashed box indicates the visible region in the structure at upper left. (**G**) Angular measures for the interprotomer network. Top left: Angle formed by SD2, SD2a, and SD1s. Top right: Angle formed by NTD', SD2, and SD2a. Bottom left: Interprotomer dihedral rotation of SD2a relative to SD2 about an SD1-to-NTD' axis. Bottom right: Interprotomer dihedral rotation between SD1 and SD2 about an NTD'-to-SD2 axis.

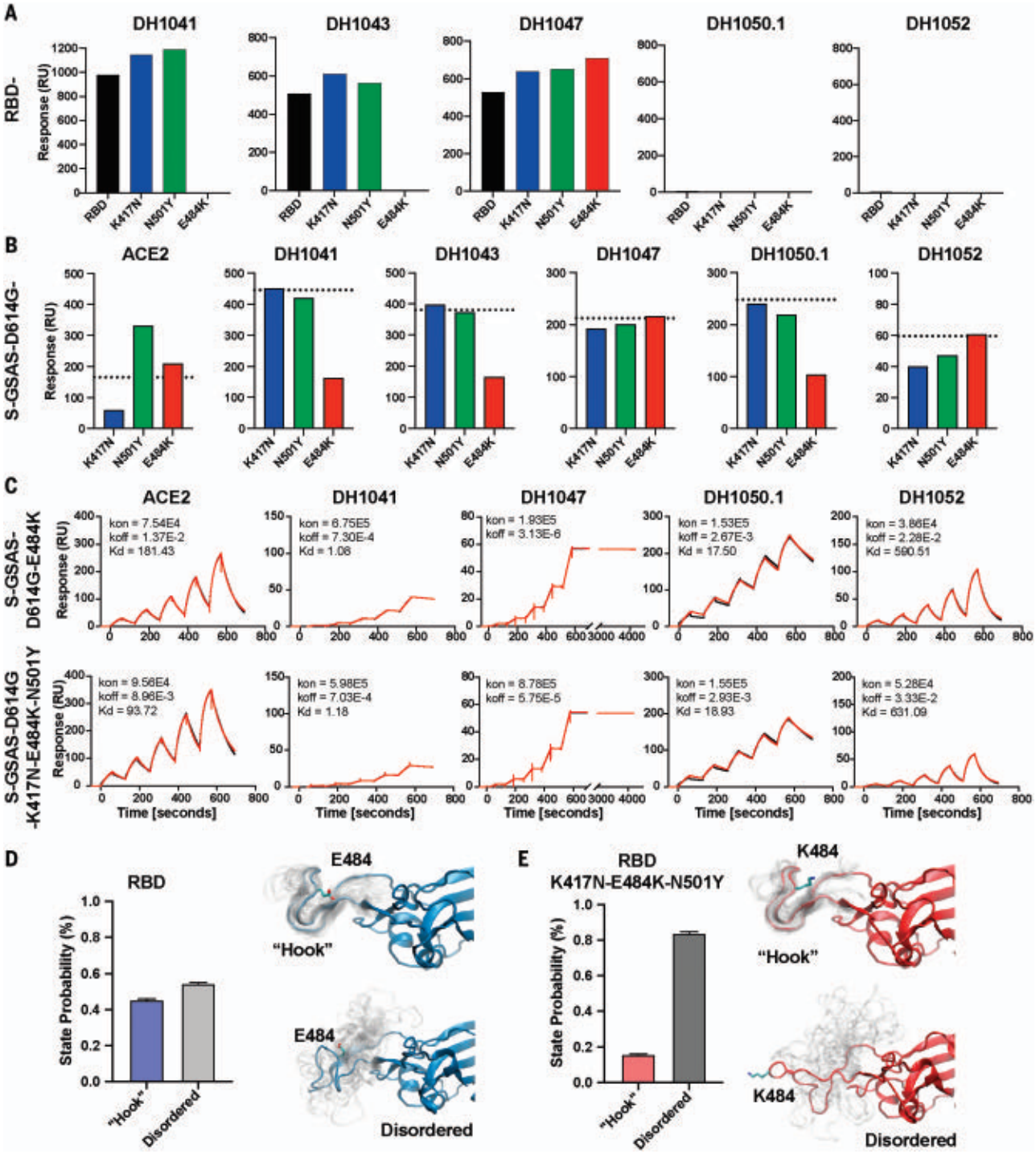


Fig. 5. Antigenic and conformational analysis of the RBD E484K substitution. (**A**) Binding of RBD-directed antibodies DH1041, DH1043, and DH1047 and NTD-directed antibodies DH1050.1 and DH1052 to WT RBD, RBD-K417N, RBD-N501Y, and RBD-E484K, measured by SPR. (**B**) Binding of ACE2; RBD-directed antibodies DH1041, DH1043, and DH1047; and NTD-directed antibodies DH1050.1 and DH1052 to spike variants, measured by SPR. The black dotted lines represent D614G spike binding levels. (**C**) Binding of ACE2 receptor ectodomain (RBD-directed) and antibodies DH1041 and DH1047 (RBD-directed, neutralizing), DH1050.1 (NTD-directed, neutralizing), and DH1052 (NTD-directed, non-neutralizing) to S-GSAS-D614G-E484K (top row) and S-GSAS-D614G-K417N-E484K-N501Y ("triple mutant spike") (bottom row), measured by SPR using single-cycle kinetics. The red lines are the binding sensorgrams; the black lines show fits of the data to a 1:1 Langmuir binding model. The on-rate (k_{on}, M^{-1} s^{-1}), off-rate (k_{off}, s^{-1}), and affinity (K_D, nM) for each interaction are indicated. The binding of DH1047 to spike was too tight to allow accurate affinity measurement. (**D** and **E**) State probabilities from the WT RBD [(D), left] and the K417N-E484K-N501Y variant RBD [(E), left] Markov model stationary distribution. Error bars indicate the 95% confidence interval. The Hook and Disordered states of the WT RBD with 25 configurations are shown in translucent gray [(D), right)]. The K417N-E484K-N501Y variant RBD Hook and Disordered states with 25 configurations are shown in translucent gray [(E), right)]. Residue 484 is depicted in stick representation.

RBD mutations, binding to RBD-directed nAbs was further reduced (Figs. 1D and 6C), showing that amino acid changes outside the RBD have an allosteric effect on the binding of RBD-directed Abs. A cryo-EM dataset of a B.1.351 spike (Fig. 1) revealed a ~6:1 ratio of RBD-up to 3-RBD-down structures (Fig. 6B). A "consensus" 3-RBD-down state with 212,753 particles was refined to 3.7 Å and displayed remarkably weak RBD density in one of the three RBDs that also appeared detached from its interprotomer-contacting NTD (Fig. 6A, PDB 7LYM). Taken together, these data implicate the K417N-E484K-N501Y substitutions in the RBD disorder observed in the 3-RBD-down states and suggest that the E484K-induced conformational disorder in the RBD tip Hook structure may be the source of the increased RBD-up spike populations due to weakened RBD-to-RBD coupling. In the spike, interprotomer interactions made by the RBD in its up state, and secondary contacts that the bound antibody makes with adjacent RBDs, may play a role in stabilizing antibody binding to the E484K mutant (*54*), thereby explaining the retention of high-affinity binding, albeit at lower levels.

We next asked whether the weakened RBD-RBD and RBD-NTD coupling involving the disordered RBD had an impact on spike quaternary structure. Domain interface mutations are limited to the RBD in the triple mutant and B.1.351 spike variants (Fig. 1A). Asymmetry in the S1 subunit was observed when aligning the SD2 subdomain of each protomer (Fig. 6D). Patterns in the interprotomer vector network indicated that the triple mutant and B.1.351 spikes were similar in their protomer-to-protomer relationships (Fig. 6E). The absolute positions, however, displayed marked differences (Fig. 6, D and E, and fig. S25), suggesting that the additional mutations in the B.1.351 spike play a role in further modulating spike conformation. Comparing the interprotomer vector networks of these structures with the 3-RBD-down D614G spike structures indicated that the B.1.351 structure was most similar to the D614G 7KE8 structure, whereas the triple mutant spike lacked similarity to any of the D614G structures (fig. S25). This shift toward a more D614G-like state in B.1.351 may indicate the selection of stabilizing mutations to balance the RBD-destabilizing mutations. Together, these results show that amino acid variations in the RBD alone can have marked impacts on S1 quaternary structure, and accumulation of additional variations outside the RBD may in turn modulate RBD conformational changes.

Comparing SARS-CoV-2 variant S ectodomain quaternary structure

The structural results presented here indicate that the primary consequence of conformational adjustments in the SARS-CoV-2 variants is increased propensity for RBD exposure. Our data implicate destabilization of the 3-RBD-down state and involvement of a disordered RBD in this conformational difference. To compare the different approaches that the variants take toward this destabilization, we examined the interprotomer network of each variant spike (Fig. 4F), together with a new RBD-to-RBD and RBD-to-NTD network (Fig. 7A). It is necessary to define a primary protomer for these comparisons because of the asymmetric nature of the spike. We selected the protomer containing the RBD most distant from its adjacent NTD, often the disordered RBD protomer, for this analysis [this protomer is here designated Protomer$_{A''}$; a double prime (″) designation was used for all vector measures and domain/protomer names to signify this change]. We also included in our analysis an asymmetric 3-RBD-down reconstruction of our engineered u1S2q S ectodomain (*19*), and four of our previously published 3-RBD-down D614G spike reconstructions (*16*). We first examined PCA clustering to identify structurally similar sets (Fig. 7, A and B). The triple mutant and B.1.351 spike structures, as well as ΔFV 3D-1 and 3D-2, clustered with D614G spike structures; the B.1.1.7 and ΔFV 3D-3 structures clustered with u1S2q; and ΔFV 3D-4 differed markedly from all others. The separation of the structures into D614G-like and u1S2q-like is consistent with differing RBD destabilization strategies in the variants that harbor the RBD triple mutants relative to the B.1.1.7 and ΔFV spikes. Examination of the primary vectors reporting on the differences observed in these clusters indicated that the typically disordered RBD protomer, Protomer$_{A''}$, is the driver of differences between the two clusters. Positioning of SD1 relative to SD2, defined by the angle $\theta_{4''}$, in Protomer$_{A''}$, and the S2-to-SD2 and S2-to-NTD′ distances, were each indicators of these differences (Fig. 7, B and C).

The interconnected spike domain network suggests that changes in local quaternary arrangements are likely to induce rearrangements in distant domains (Fig. 7D). We therefore examined correlations in quaternary arrangements of SD2, SD2a, SD1, and NTD′ in the full dataset. The variant discriminating SD2-to-SD1 angle $\theta_{4''}$ (defined in Fig. 4F) displays a considerable number of correlations with quaternary arrangements throughout the network (Fig. 7, E and F, and figs. S26 to S30). This includes correlation with the Protomer$_{B''}$ and Protomer$_{C''}$ SD2-to-SD1 angles $\theta_{2''}$ and $\theta_{6''}$. Correlation was also observed with the interprotomer dihedral angle that defined the rotation of SD2 about axes connecting SD1 and NTD′ from Protomer$_{B''}$ to Protomer$_{C''}$ $\phi_{1''}$ and Protomer$_{C''}$ to Protomer$_{A''}$ ($\phi_{4''}$) (Fig. 7, E and F). These, and correlation with dihedral rotation of Protomer$_{B''}$ SD2 and Protomer $_{C''}$ NTD′ about an axis connecting the SD2 anchor and SD1, $\phi_{9''}$, are mirrored by the Protomer$_{B''}$ SD2-to-SD1 angle, $\theta_{6''}$. The relationships identified show that changes in domain arrangement in one protomer have predictable impacts on the domain arrangements of the other protomers. In the D614G cluster, quaternary arrangements give rise to the marked distance between the disordered RBD and the NTD′ (Fig. 7D). For the triple mutant and B.1.351 spike structures, the RBD tip disorder presumably reduces the stability of its contact with the adjacent RBD, increasing its up-state propensity. Unlike the D614G cluster, in the u1S2q cluster RBDs are all distant from their adjacent NTD (Fig. 7D). Examination of the structures indicated that rearrangements occurred in the orientation of SD1 relative to SD2 and S2. The engineered u1S2q contains mutations only in S2 and in the SD1 Ala^{570} loop that is adjacent to S2. These together increase the up-state population. It is therefore likely that amino acid substitution in SD2 and S2/SD1 in the ΔFV and B.1.1.7 spikes, respectively, are responsible for the increased RBD-up populations in these spikes. Thus, several mechanisms exist by which changes induced in domain interaction strength by spike amino acid substitutions modify RBD positioning.

Discussion

The SARS-CoV-2 spike plays an essential role in virus spread and represents the primary target for neutralizing antibodies. Spike mutations in SARS-CoV-2 variants can have an impact on virus neutralization sensitivity and transmissibility. Although many of the currently circulating variants of interest/concern likely arose from some combination of genetic drift, host adaptation, and immune evasion, the virus will increasingly experience pressure from vaccine-elicited antibody responses. To prepare for the continued evolution of the virus, it is essential to understand how spike variations affect virus transmissibility and neutralization sensitivity. The increased binding to ACE2, mediated both by affinity-enhancing substitutions in the RBD and increased propensity for the receptor-accessible RBD-up states, may contribute to the rapid spread of variants. For the mink-associated variant, increased receptor binding may have helped to establish infection in a new host. Whereas all human-evolved variants studied here showed reduced binding to antibodies at dominant neutralization epitopes, the mink-associated variant retained similar levels of binding to all antibodies tested, underscoring the role of the human immune response in shaping the course of SARS-CoV-2 evolution. For the mink-evolved variant, we uncovered evidence for spike instability, which may be the reason why the variant failed to spread widely when transmitted back to humans. For the human-evolved variants, we found that the S protein

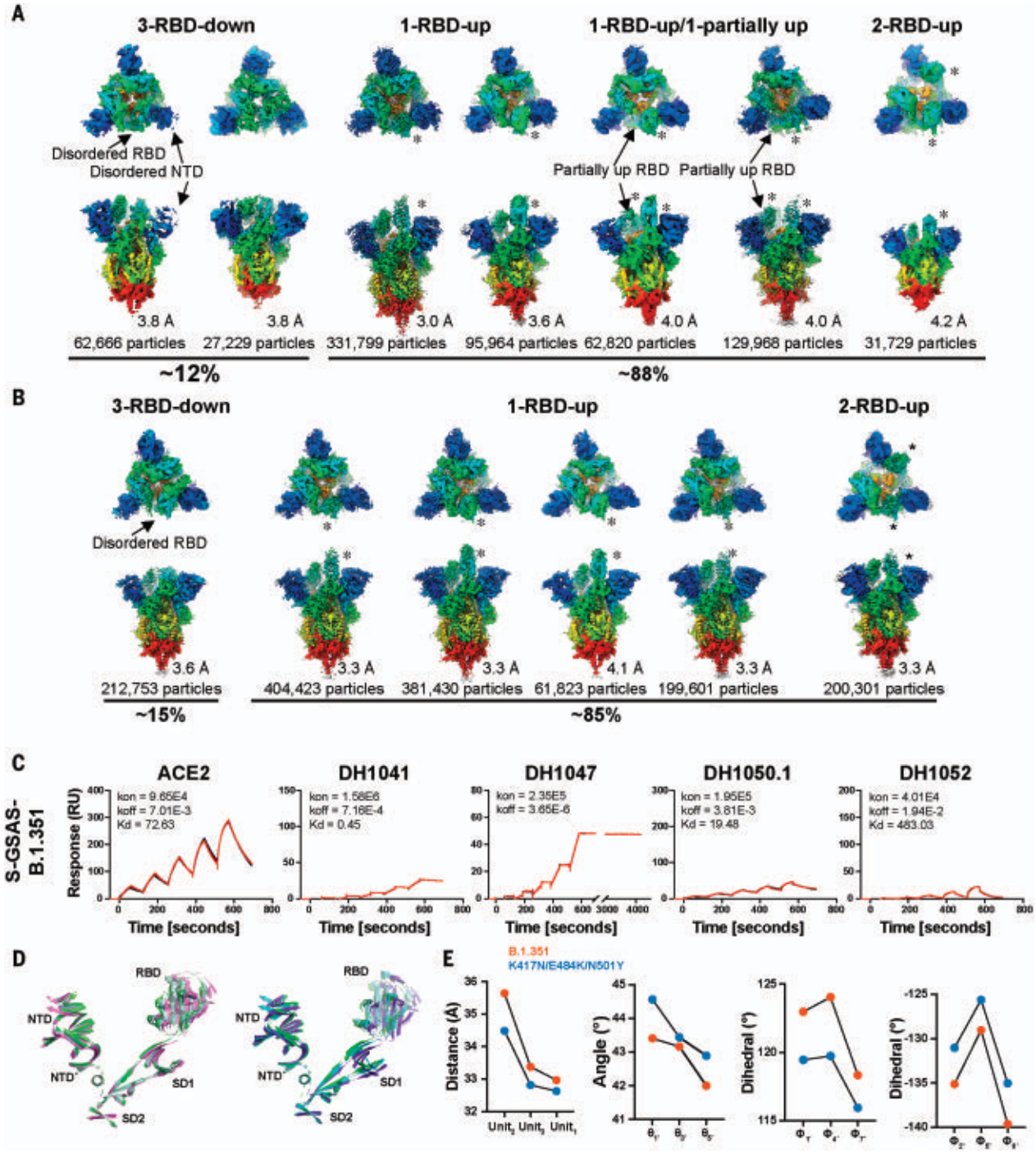


Fig. 6. Analysis of S-GSAS-D614G-K417N-E484K-N501Y ("triple mutant spike") and S-GSAS-B.1.351 (B.1.351 spike). (**A** and **B**) Cryo-EM reconstructions of (A) triple mutant spike and (B) B.1.351 spike, in rainbow colors. (**C**) Binding ACE2 receptor ectodomain (RBD-directed) and antibodies DH1041 and DH1047 (RBD-directed, neutralizing), DH1050.1 (NTD-directed, neutralizing), and DH1052 (NTD-directed, non-neutralizing) to the B.1.351 spike measured by SPR using single-cycle kinetics. The red lines are the binding sensorgrams; the black lines show fits of the data to a 1:1 Langmuir binding model. The on-rate (k_{on}, M^{-1} s^{-1}), off-rate (k_{off}, s^{-1}), and affinity (K_D, nM) for each interaction are indicated. The binding of DH1047 to spike was too tight to allow accurate affinity measurement. (**D**) Cartoon helix and sheet secondary structure elements of the triple mutant spike variant SD2 aligned S1 protomers (left) and B.1.351 variant SD2 aligned S1 protomers (right). (**E**) Angle and dihedral measures for the interprotomer SD2-SD1-NTD′ network. From left to right: RBD to adjacent NTD distance, NTD′-to-SD2 angle, SD1-to-NTD′ dihedral, and NTD′-to-SD2 dihedral.

used different mechanisms for manipulation of its immunodominant regions to converge on a common goal of destabilizing the 3-RBD-down state. In the B.1.1.7 variant, this occurred by modifications in the interaction between SD1 or SD2 and S2, whereas for variants harboring the K417N/E484K/N501Y RBD triple substitutions, RBD destabilization was mediated by RBD-RBD contacts. Together, these results show that these variants have modified the S1 subunit domain interaction network to control the functionally critical disposition of the RBD while acquiring antibody resistance and improved transmissibility. We have provided a structurally detailed view of

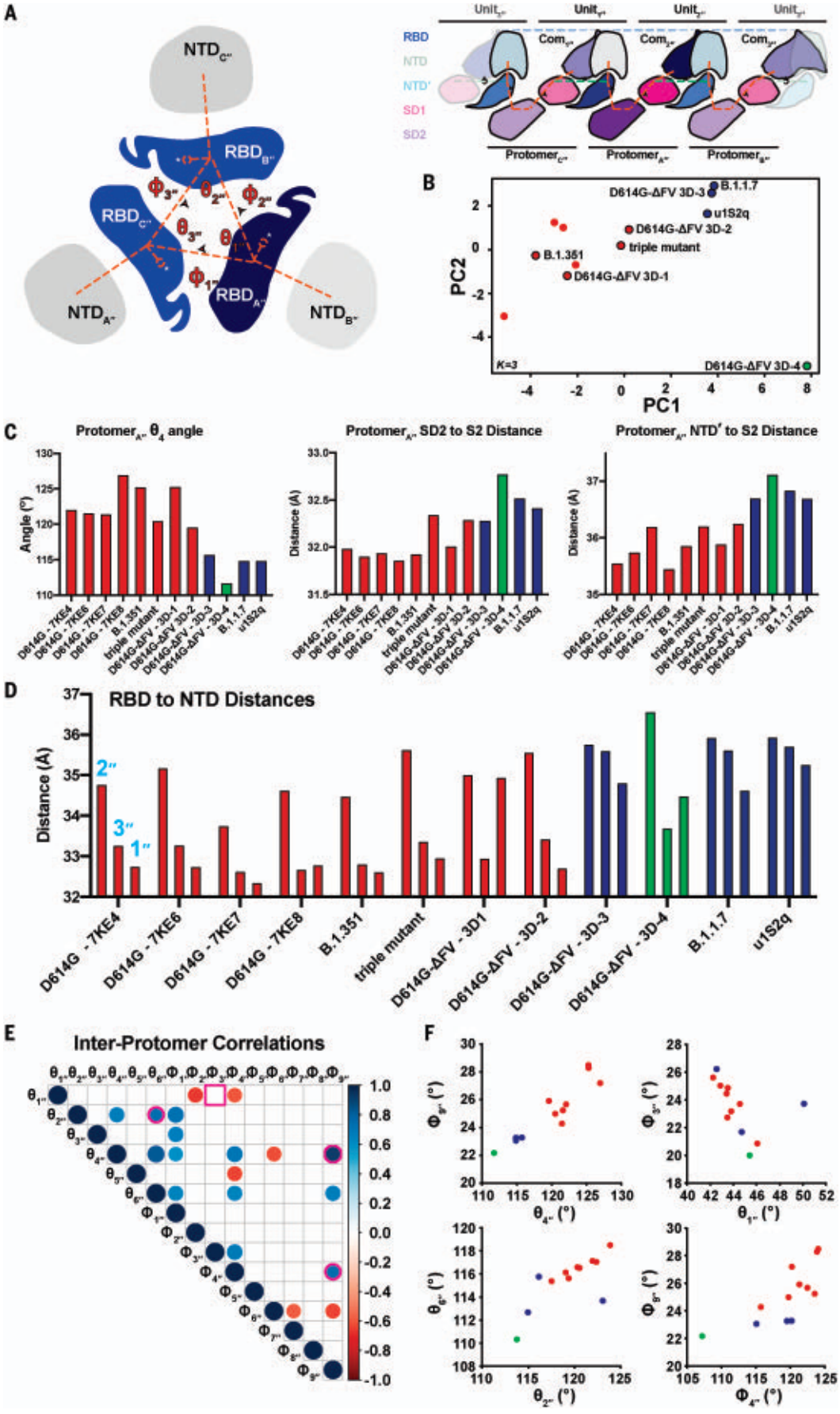

A
RBD
NTD
NTD'
SD1
SD2
B
D614G-ΔFV 3D-3
B.1.1.7
u1S2q
D614G-ΔFV 3D-2
triple mutant
B.1.351
D614G-ΔFV 3D-1
D614G-ΔFV 3D-4
K=3
PC1
PC2
C
Protomer θ4 angle
Protomer SD2 to S2 Distance
Protomer NTD' to S2 Distance
Angle (°)
Distance (Å)
D
RBD to NTD Distances
Distance (Å)
D614G - 7KE4
D614G - 7KE6
D614G - 7KE7
D614G - 7KE8
B.1.351
triple mutant
D614G-ΔFV - 3D1
D614G-ΔFV - 3D-2
D614G-ΔFV - 3D-3
D614G-ΔFV - 3D-4
B.1.1.7
u1S2q
E
Inter-Protomer Correlations
F

Fig. 7. Comparison of interprotomer network and RBD-to-RBD quaternary structure. (**A**) Left: RBD and NTD vectors, angles, and dihedrals. Anchor points are identified with asterisks. Right: Simplified schematic of the SD2, SD2a, SD1, and NTD′ interprotomer contact network. (**B**) Principal components analysis of the interprotomer network and RBD-to-RBD vector measures. Dot color indicates K-means cluster assignment. Clusters correspond to a GSAS-D614G (D614G)–like cluster (red), a u1S2q-like cluster (blue), and outlier ΔFV (ΔFV) 3D-4 (green). (**C**) Top three contributors to PC1 for $Protomer_{A''}$. (**D**) RBD-to-NTD distance for the variants including the previously determined D614G structures and the asymmetric u1S2q structure. (**E**) Significant correlations between the interprotomer angle measures (N = 12, P < 0.05). Pink outlines identify relationships plotted in (F). Square outline identifies nonsignificant correlation in the full structure set that was significant in the D614G cluster–only correlations. (**F**) Selected vector relationship plots. Dot color indicates K-means cluster assignment from the PCA analysis in (B).

these variants and a framework from which to anticipate further changes to the spike as the pathogen evolves.

Materials and methods

Plasmids

Gene syntheses for all plasmids generated by this study were performed and the sequence confirmed by GeneImmune Biotechnology (Rockville, MD). The SARS-CoV-2 spike protein ectodomain constructs comprised the S protein residues 1 to 1208 (GenBank MN908947) with the D614G mutation, the furin cleavage site (RRAR; residues 682 to 685) mutated to GSAS, a C-terminal T4 fibritin trimerization motif, a C-terminal HRV3C protease cleavage site, a TwinStrepTag, and an 8×HisTag. All spike ectodomains were cloned into the mammalian expression vector pαH and have been deposited to Addgene (*42*) (www.addgene.org) under the codes 171743, 171744, 171745, 171746, 171747, 171748, 171749, 171750, 171751, and 171752. For the ACE2 construct, the C terminus was fused a human Fc region (*19*).

Cell culture and protein expression

GIBCO FreeStyle 293-F cells [human embryonic kidney (HEK)] were maintained at 37°C and 9% CO_2 in a 75% humidified atmosphere in FreeStyle 293 Expression Medium (GIBCO). Plasmids were transiently transfected using Turbo293 (SpeedBiosystems) and incubated at 37°C, 9% CO2, 75% humidity with agitation at 120 rpm for 6 days. On the day after transfection, HyClone CDM4HEK293 media (Cytiva) was added to the cells. Antibodies were produced in Expi293F cells (HEK; GIBCO). Cells were maintained in Expi293 Expression Medium (GIBCO) at 37°C, 120 rpm and 8% CO_2 and 75% humidity. Plasmids were transiently transfected using the ExpiFectamine 293 Transfection Kit and protocol (GIBCO) (*9*, *19*, *55*).

Protein purification

On day 6 after transfection, spike ectodomains were harvested from the concentrated supernatant. The spike ectodomains were purified using StrepTactin resin (IBA LifeSciences) and size exclusion chromatography (SEC) using a Superose 6 10/300 GL Increase column (Cytiva) equilibrated in 2 mM Tris, pH 8.0, 200 mM NaCl, 0.02% NaN_3. All steps of the purification were performed at room temperature and in a single day. Protein quality was assessed by SDS-PAGE using NuPage 4 to 12% (Invitrogen). The purified proteins were flash-frozen and stored at −80°C in single-use aliquots. Each aliquot was thawed by a 20-min incubation at 37°C before use. Antibodies were purified by Protein A affinity and digested to their Fab state using LysC. ACE2 with human Fc tag was purified by Protein A affinity chromatography and SEC (*19*). RBD constructs were produced and purified as described (*56*).

SPR

Antibody binding to SARS-CoV-2 spike and RBD constructs was assessed using SPR on a Biacore T-200 (Cytiva) with HBS buffer supplemented with 3 mM EDTA and 0.05% surfactant P-20 (HBS-EP+, Cytiva). All binding assays were performed at 25°C. Spike variants were captured on a Series S streptavidin (SA) chip (Cytiva) by flowing over 200 nM of the spike for 60 s at 10 μl/min flow rate. The Fabs were injected at concentrations ranging from 0.625 nM to 800 nM (twofold serial dilution) using the single-cycle kinetics mode with five concentrations per cycle. For the single-injection assay, the Fabs were injected at a concentration of 200 nM. A contact time of 60 s, dissociation time of 120 s (3600 s for DH1047 for the single-cycle kinetics) at a flow rate of 50 μl/min was used. The surface was regenerated after each dissociation phase with three pulses of a 50 mM NaOH + 1 M NaCl solution for 10 s at 100 μl/min. For the RBDs, the antibodies were captured on a CM5 chip (Cytiva) coated with Human Anti-Fc (using Cytiva Human Antibody Capture Kit and protocol), by flowing over 100 nM antibody solution at a flow rate of 5 μl/min for 120 s. The RBDs were then injected at 100 nM for 120 s at a flow rate of 50 μl/min with a dissociation time of 30 s. The surface was regenerated by three consecutive pulses of 3 M $MgCl_2$ for 10 s at 100 μl/min. Sensorgram data were analyzed using BiaEvaluation software (Cytiva).

Negative-stain electron microscopy

Samples were diluted to 100 μg/ml in 20 mM HEPES pH 7.4, 150 mM NaCl, 5% glycerol, 7.5 mM glutaraldehyde (Electron Microscopy Sciences) and incubated for 5 min before quenching the glutaraldehyde by the addition of 1 M Tris (to a final concentration of 75 mM) and 5 min incubation. A 5-μl drop of sample was applied to a glow-discharged carbon-coated grid (Electron Microscopy Sciences, CF300-Cu) for 10 to 15 s, blotted, stained with 2% uranyl formate (Electron Microscopy Sciences), blotted, and air-dried. Images were obtained using a Philips EM420 electron microscope at 120 kV, 82,000× magnification, and a 4.02 Å pixel size. RELION (*57*) software was used for particle picking and 2D and 3D class averaging.

ELISA assays

Spike ectodomains were tested for antibody- or ACE2-binding in ELISA assays as described (*32*). Assays were run in two formats: antibodies/ACE2-coated or spike-coated. For the first format, the assay was performed on 384-well plates coated at 2 μg/ml overnight at 4°C, washed, blocked, and followed by twofold serially diluted spike protein starting at 25 μg/ml. Binding was detected with polyclonal anti–SARS-CoV-2 spike rabbit serum (developed in our lab), followed by goat anti-rabbit HRP (Abcam, Ab97080) and TMB substrate (Sera Care Life Sciences). Absorbance was read at 450 nm. In the second format, serially diluted spike protein was bound in wells of a 384-well plates, which were previously coated with streptavidin (Thermo Fisher Scientific) at 2 μg/ml and blocked. Proteins were incubated at room temperature for 1 hour, washed, then human mAbs were added at 10 μg/ml. Antibodies were incubated at room temperature for 1 hour, washed, and binding detected with goat anti-human HRP (Jackson ImmunoResearch) and TMB substrate.

Cryo-EM

Purified SARS-CoV-2 spike ectodomains were diluted to a concentration of ~1.5 mg/ml in 2 mM Tris pH 8.0, 200 mM NaCl and 0.02% NaN_3 and 0.5% glycerol was added. A 2.3-μl drop of protein was deposited on a Quantifoil-1.2/1.3 grid (Electron Microscopy Sciences) that had been glow-discharged for 10 s using a PELCO easiGlow Glow Discharge Cleaning System. After a 30-s incubation in >95% humidity, excess protein was blotted away for 2.5 s before being plunge-frozen into liquid ethane using a Leica EM GP2 plunge freezer (Leica Microsystems). Frozen grids were imaged using a Titan Krios (Thermo Fisher) equipped with a K3 detector (Gatan). CryoSPARC (*58*)

software was used for data processing. Phenix (*54*, *59*), Coot (*60*), Pymol (*61*), Chimera (*62*), ChimeraX (*63*), and Isolde (*64*) were used for model building and refinement.

Vector-based structure analysis

Vector analysis of intraprotomer domain positions was performed as described (*19*) using the Visual Molecular Dynamics (VMD) (*65*) software package Tcl interface (*66*). For each protomer of each structure, Cα centroids were determined for the NTD (residues 27 to 69, 80 to 130, 168 to 172, 187 to 209, 216 to 242, and 263 to 271), NTD′ (residues 44 to 53 and 272 to 293), RBD (residues 334 to 378, 389 to 443, and 503 to 521), SD1 (residues 323 to 329 and 529 to 590), SD2 (residues 294 to 322, 591 to 620, 641 to 691, and 692 to 696), CD (residues 711 to 716 1072 to 1121), and a S2 sheet motif (S2s; residues 717 to 727 and 1047 to 1071). Additional centroids for the NTD (NTD_c; residues 116 to 129 and 169 to 172) and RBD (RBD_c; residues 403 to 410) were determined for use as reference points for monitoring the relative NTD and RBD orientations to the NTD′ and SD1, respectively. Vectors were calculated between the following within protomer centroids: NTD to NTD′, NTD′ to SD2, SD2 to SD1, SD2 to CD, SD1 to RBD, CD to S2s, NTD_c to NTD, and RBD to RBD_c. Vector magnitudes, angles, and dihedrals were determined from these vectors and centroids. Interprotomer domain vector calculations for the SD2, SD1, and NTD′ used these centroids in addition to anchor residue Cα positions for each domain including SD2 residue 671 (SD2a), SD1 residue 575 (SD1a), and NTD′ residue 276 (NTD′a). These were selected according to visualization of position variation in all protomers used in this analysis via alignment of all of each domain in PyMol (*61*). Vectors were calculated for the following: NTD′ to NTD'_r, NTD′ to SD2, SD2 to $SD2_r$, SD2 to SD1, SD1 to $SD1_r$, and SD1 to NTD′. Angles and dihedrals were determined from these vectors and centroids. Vectors for the RBD to adjacent RBD and RBD to adjacent NTD were calculated using the above RBD, NTD, and RBD_c centroids. Vectors were calculated for the following: RBD_2 to RBD_1, RBD_3 to RBD_2, and RBD_3 to RBD_1. Angles and dihedrals were determined from these vectors and centroids. PCA, K-means clustering, and Pearson correlation (confidence interval 0.95, $P < 0.05$) analysis of vector sets was performed in R (*67*). Data were centered and scaled for the PCA analyses.

Difference distance matrices (DDMs)

DDMs were generated using the Bio3D package (*68*) implemented in R (*67*).

Adaptive sampling molecular dynamics

The CHARMM CR3022–bound SARS-CoV-2 RBD crystal structure (*69*) (PDB ID 6ZLR) model (*70*, *71*) was used for the adaptive sampling simulations (*66*). The CR3022 antibody, glycan unit, water, and ions were stripped from the model, leaving only the protein portion of the RBD. The final model comprised spike residues 327 to 529. A single Man5 glycan was added at the Asn^{343} position using the CHARMM GUI (*70*) with the P.1/B.1.1.28/B.1.351 RBD mutations K417N, E484K, and N501Y prepared in PyMol. Systems for simulation were built using the AmberTools20 Leap (*72*) program. The unmutated (WT) and P.1/B.1.1.28/B.1.351 (Mut) RBDs were immersed in a truncated octahedral TIP3P water box with a minimum edge distance of 15 Å to the nearest protein atom followed by system neutralization with chlorine atoms resulting in systems sizes of 67,508 and 66,894 atoms for the WT and Mut, respectively. The Amber ff14SB protein (*73*) and Glycam (*74*) forefields were used throughout. All simulations were performed using the Amber20 pmemd CUDA implementation. The systems were first minimized for 10,000 steps with protein atom restraints followed by minimization of the full system without restraints for an additional 10,000 steps. This was followed by heating of the systems from 0 K to 298 K over a period of 20 ps in the NVT ensemble using a 2-fs time step and the particle mesh Ewald method for long-range electrostatics and periodic boundary conditions (*75*). The systems were then equilibrated for 100 ps in the NPT ensemble with the temperature controlled using Langevin dynamics with a frequency of 1.0 ps^{-1} and 1 atm pressure maintained using isotropic position scaling with a relaxation time of 2 ps (*76*). A non-bonded cutoff of 8 Å was used throughout and hydrogen atoms were constrained using the SHAKE algorithm (*77*) with hydrogen mass repartitioning (*78*) used to allow for a 4-fs time step. To generate an ensemble of RBD tip conformations for initiation of the adaptive sampling routine, we performed 100 50-ns simulations in the NVT ensemble with randomized initial velocities for each of the WT and Mut systems. The final frame from each of these simulations was used to initiate the adaptive sampling scheme. Adaptive sampling was performed using the High-Throughput Molecular Dynamics (HTMD v. 1.24.2) package (*79*). Each iteration consisted of 50 to 100 independent simulations of 100 ns. Simulations from each iteration were first projected using a dihedral metric with angles split into their sin and cos components for residues 454 to 491. This was followed by a TICA (*80*) projection using a lag time of 5 ns and retaining five dimensions. Markov state models were then built using a lag time of 50 ns for the selection of new states for the next iteration. A total of 29 adaptive iterations were performed, yielding total simulation times of 274.8 μs and 256.8 μs for the WT and Mut systems, respectively. Simulations were visualized in VMD and PyMol.

Markov state modeling

Markov state models (MSMs) were prepared in HTMD with an appropriate coordinate projection selected using PyEMMA (*81*) (v. 2.5.7). Multiple projections were tested on a 25-μs subset of the Mut simulations that included atomic distance and contact measures between RBD residues as well as backbone torsions of the RBD tip residues using the variational approach to Markov processes score (*82*) (fig. S17 and table S4) (*66*). This led to the selection of a Cα pairwise distance metric between residues 471 to 480 and 484 to 488 for MSM construction. MSMs were prepared in HTMD using a TICA lag time of 5 ns retaining five dimensions followed by K-means clustering using 500 cluster centers. The implied time scales (ITS) plots were used to select a lag time of 30 ns for MSM building. Models were coarse-grained via Perron cluster analysis (PCCA++) using two states and validated using the Chapman-Kolmogorov (CK) test. A bootstrapping routine without replacement was used to calculate measurement errors retaining 80% of the data per iteration for a total of 100 iterations. State statistics were collected for mean first passage times (MFPT), stationary distributions, and root-mean-square deviations (RMSDs) for RBD tip residues 470 to 490. Residue 484 side-chain contacts were calculated from a representative model. A contact was defined as atom pairing within 3.5 Å between either the minimum of either Glu^{484} γ-carboxyl O atoms (for WT) or Lys^{484} ε-amino N atom (for Mut) and backbone or side-chain O or N atoms for residues 348 to 354, 413 to 425, or 446 to 500. The RMSD and contact metric means were model-weighted. Weighted state ensembles containing 250 structures were collected for visualization in VMD.

REFERENCES AND NOTES

1. Z. Wang *et al*., mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* **592**, 616–622 (2021). pmid: 33567448
2. C. K. Wibmer *et al*., SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat. Med.* **27**, 622–625 (2021). pmid: 33654292
3. B. B. Oude Munnink *et al*., Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* **371**, 172–177 (2021). doi: 10.1126/science.abe5901; pmid: 33172935
4. B. Turoňová *et al*., In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **370**, 203–208 (2020). pmid: 32817270
5. Z. Ke *et al*., Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **588**, 498–502 (2020). doi: 10.1038/s41586-020-2665-2; pmid: 32805734
6. M. Hoffmann, H. Kleine-Weber, S. Pöhlmann, A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol. Cell* **78**, 779–784.e5 (2020). doi: 10.1016/j.molcel.2020.04.022; pmid: 32362314
7. G. D. Sempowski, K. O. Saunders, P. Acharya, K. J. Wiehe, B. F. Haynes, Pandemic Preparedness: Developing Vaccines and Therapeutic Antibodies For COVID-19. *Cell* **181**, 1458–1463 (2020). doi: 10.1016/j.cell.2020.05.041; pmid: 32492407
8. K. S. Corbett *et al*., SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature* **586**, 567–571 (2020). doi: 10.1038/s41586-020-2622-0; pmid: 32756549
9. D. Li *et al*., The functions of SARS-CoV-2 neutralizing and infection-enhancing antibodies in vitro and in mice and

nonhuman primates. bioRxiv 424729 [preprint]. 2 January 2021.
10. R. Yan *et al*., Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020). doi: 10.1126/science.abb2762; pmid: 32132184
11. C. O. Barnes *et al*., SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682–687 (2020). doi: 10.1038/s41586-020-2852-1; pmid: 33045718
12. C. O. Barnes *et al*., Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* **182**, 828–842.e16 (2020). doi: 10.1016/j.cell.2020.06.025; pmid: 32645326
13. M. Gui *et al*., Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res.* **27**, 119–129 (2017). doi: 10.1038/cr.2016.152; pmid: 28008928
14. Y. Yuan *et al*., Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* **8**, 15092 (2017). doi: 10.1038/ncomms15092; pmid: 28393837
15. J. Shang *et al*., Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11727–11734 (2020). doi: 10.1073/pnas.2003138117; pmid: 32376634
16. S. M. Gobeil *et al*., D614G Mutation Alters SARS-CoV-2 Spike Conformation and Enhances Protease Cleavage at the S1/S2 Junction. *Cell Rep.* **34**, 108630 (2021). doi: 10.1016/j.celrep.2020.108630; pmid: 33417835
17. L. Yurkovetskiy *et al*., Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* **183**, 739–751.e8 (2020). pmid: 32991842
18. T. Zhou *et al*., Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host Microbe* **28**, 867–879.e5 (2020). doi: 10.1016/j.chom.2020.11.004; pmid: 33271067
19. R. Henderson *et al*., Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nat. Struct. Mol. Biol.* **27**, 925–933 (2020). pmid: 32699321
20. J. Zhang *et al*., Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science* **372**, 525–530 (2021). doi: 10.1126/science.abf2303; pmid: 33727252
21. D. Bestle *et al*., TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Sci. Alliance* **3**, e202000786 (2020). doi: 10.26508/lsa.202000786; pmid: 32703818
22. M. Hoffmann *et al*., SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271–280.e8 (2020). doi: 10.1016/j.cell.2020.02.052; pmid: 32142651
23. S. Matsuyama *et al*., Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 7001–7003 (2020). doi: 10.1073/pnas.2002589117; pmid: 32165541
24. B. Korber *et al*., Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812–827.e19 (2020). doi: 10.1016/j.cell.2020.06.043; pmid: 32697968
25. M. Koopmans, SARS-CoV-2 and the human-animal interface: Outbreaks on mink farms. *Lancet Infect. Dis.* **21**, 18–19 (2021). doi: 10.1016/S1473-3099(20)30912-9; pmid: 33227234
26. H. Tegally *et al*., Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **592**, 438–443 (2021). doi: 10.1038/s41586-021-03402-9; pmid: 33690265
27. M. Mwenda *et al*., Detection of B.1.351 SARS-CoV-2 Variant Strain - Zambia, December 2020. *MMWR Morb. Mortal. Wkly. Rep.* **70**, 280–282 (2021). doi: 10.15585/mmwr.mm7008e2; pmid: 33630820
28. M. H. S. Paiva *et al*., Multiple Introductions Followed by Ongoing Community Spread of SARS-CoV-2 at One of the Largest Metropolitan Areas of Northeast Brazil. *Viruses* **12**, 1414 (2020). doi: 10.3390/v12121414; pmid: 33316947
29. E. Boehm *et al*., Novel SARS-CoV-2 variants: The pandemics within the pandemic. *Clin. Microbiol. Infect.* (2021). doi: 10.1016/j.cmi.2021.05.022; pmid: 34015535
30. S. E. Galloway *et al*., Emergence of SARS-CoV-2 B.1.1.7 Lineage - United States, December 29, 2020-January 12, 2021. *MMWR Morb. Mortal. Wkly. Rep.* **70**, 95–99 (2021). doi: 10.15585/mmwr.mm7003e2; pmid: 33476315
31. K. Leung, M. H. Shum, G. M. Leung, T. T. Lam, J. T. Wu, Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill.* **26**, 2002106 (2021). doi: 10.2807/1560-7917.ES.2020.26.1.2002106; pmid: 33413740
32. R. J. Edwards *et al*., Cold sensitivity of the SARS-CoV-2 spike ectodomain. *Nat. Struct. Mol. Biol.* **28**, 128–131 (2021). doi: 10.1038/s41594-020-00547-5; pmid: 33402708
33. K. O. Saunders *et al*., Neutralizing antibody vaccine for pandemic and pre-emergent coronaviruses. *Nature* 10.1038/s41586-021-03594-0 (2021). doi: 10.1038/s41586-021-03594-0; pmid: 33971664
34. D. R. Martinez *et al*., A broadly neutralizing antibody protects against SARS-CoV, pre-emergent bat CoVs, and SARS-CoV-2 variants in mice. bioRxiv 441655 [preprint]. 28 April 2021.
35. M. McCallum *et al*., N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **184**, 2332–2347.e16 (2021). doi: 10.1016/j.cell.2021.03.028; pmid: 33761326
36. G. Cerutti *et al*., Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* **29**, 819–833.e7 (2021). doi: 10.1016/j.chom.2021.03.005; pmid: 33789084
37. W. B. Williams *et al*., Fab-dimerized glycan-reactive antibodies are a structural category of natural antibodies. *Cell* **184**, 2955–2972.e25 (2021). pmid: 34019795
38. X. Shen *et al*., SARS-CoV-2 variant B.1.1.7 is susceptible to neutralizing antibodies elicited by ancestral Spike vaccines. *Cell Host Microbe* **29**, 529–539.e3 (2021). doi: 10.1016/j.chom.2021.03.002
39. M. Hoffmann *et al*., SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. *Cell* **184**, 2384–2393.e12 (2021). doi: 10.1016/j.cell.2021.03.036; pmid: 33794143
40. P. Wang *et al*., Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. *Cell Host Microbe* **29**, 747–751.e4 (2021). doi: 10.1016/j.chom.2021.04.007; pmid: 33887205
41. F. M. Richards, C. E. Kundrot, Identification of structural motifs from protein coordinate data: Secondary structure and first-level supersecondary structure. *Proteins* **3**, 71–84 (1988). doi: 10.1002/prot.340030202; pmid: 3399495
42. D. Wrapp *et al*., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020). doi: 10.1126/science.abb2507; pmid: 32075877
43. A. C. Walls *et al*., Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281–292.e6 (2020). doi: 10.1016/j.cell.2020.02.058; pmid: 32155444
44. N. G. Davies *et al*., Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* **593**, 270–274 (2021). doi: 10.1038/s41586-021-03426-1
45. T. N. Starr *et al*., Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295–1310.e20 (2020). doi: 10.1016/j.cell.2020.08.012; pmid: 32841599
46. X. Xie *et al*., Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat. Med.* **27**, 620–621 (2021). pmid: 33558724
47. K. Wu *et al*., mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. bioRxiv 427948 [preprint]. 25 January 2021.
48. L. Liu *et al*., Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* **584**, 450–456 (2020). doi: 10.1038/s41586-020-2571-7; pmid: 32698192
49. P. Wang *et al*., Increased Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7 to Antibody Neutralization. bioRxiv 428137 [preprint]. 26 January 2021.
50. A. Krarup *et al*., A highly stable prefusion RSV F vaccine derived from structural analysis of the fusion mechanism. *Nat. Commun.* **6**, 8143 (2015). doi: 10.1038/ncomms9143; pmid: 26333350
51. T. Sztain *et al*., A glycan gate controls opening of the SARS-CoV-2 spike protein. bioRxiv 431212 [preprint]. 16 February 2021.
52. R. Henderson *et al*., Glycans on the SARS-CoV-2 Spike Control the Receptor Binding Domain Conformation. bioRxiv 173765 [preprint]. 26 June 2020.
53. M. Yuan *et al*., Structural basis of a shared antibody response to SARS-CoV-2. *Science* **369**, 1119–1123 (2020). doi: 10.1126/science.abd2321; pmid: 32661058
54. D. Liebschner *et al*., Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019). doi: 10.1107/S2059798319011471; pmid: 31588918
55. P. Acharya *et al*., A glycan cluster on the SARS-CoV-2 spike ectodomain is recognized by Fab-dimerized glycan-reactive antibodies. bioRxiv 178897 [preprint]. 30 June 2020.
56. K. O. Saunders *et al*., SARS-CoV-2 vaccination induces neutralizing antibodies against pandemic and pre-emergent SARS-related coronaviruses in monkeys. bioRxiv 431492 [preprint]. 17 February 2021.
57. S. H. W. Scheres, A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012). doi: 10.1016/j.jmb.2011.11.010; pmid: 22100448
58. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017). doi: 10.1038/nmeth.4169; pmid: 28165473
59. P. V. Afonine *et al*., Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018). doi: 10.1107/S2059798318006551; pmid: 29872004
60. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010). doi: 10.1107/S0907444910007493; pmid: 20383002
61. L. Schrodinger, The PyMOL Molecular Graphics System (2015).
62. E. F. Pettersen *et al*., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
63. T. D. Goddard *et al*., UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018). doi: 10.1002/pro.3235; pmid: 28710774
64. T. I. Croll, ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D* **74**, 519–530 (2018). doi: 10.1107/S2059798318002425; pmid: 29872003
65. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996). doi: 10.1016/0263-7855(96)00018-5; pmid: 8744570
66. S. M. C. Gobeil *et al*., SARS-CoV-2 spike structure vector analysis scripts and molecular dynamics simulation trajectories for the SARS-CoV-2 WT and Mut (K417N+E484K+N501Y) RBDs. Zenodo DOI: 10.5281/zenodo.4926233 (2021).
67. R Core Team, R: A Language and Environment for Statistical Computing (2017).
68. B. J. Grant, L. Skjærven, X. Q. Yao, The Bio1.3D packages for structural bioinformatics. *Protein Sci.* **30**, 20–30 (2021). pmid: 32734663
69. C. Nichols, J. Ng, A. Keshu, F. Fraternali, G. F. De Nicola, A New Crystal Form of the SARS-CoV-2 Receptor Binding Domain: CR3022 Complex-An Ideal Target for In-Crystal Fragment Screening of the ACE2 Binding Site Surface. *Front. Pharmacol.* **11**, 615211 (2020). doi: 10.3389/fphar.2020.615211; pmid: 33381049
70. S. Jo, T. Kim, V. G. Iyer, W. Im, CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008). doi: 10.1002/jcc.20945; pmid: 18351591
71. H. Woo *et al*., Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. *J. Phys. Chem. B* **124**, 7128–7137 (2020). doi: 10.1021/acs.jpcb.0c04553; pmid: 32559081
72. D. A. Case *et al*., Amber 2021 (University of California, San Francisco, 2020).
73. J. A. Maier *et al*., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015). doi: 10.1021/acs.jctc.5b00255; pmid: 26574453
74. K. N. Kirschner *et al*., GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **29**, 622–655 (2008). doi: 10.1002/jcc.20820; pmid: 17849372
75. U. Essmann *et al*., A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995). doi: 10.1063/1.470117
76. R. J. Loncharich, B. R. Brooks, R. W. Pastor, Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers* **32**, 523–535 (1992). doi: 10.1002/bip.360320508; pmid: 1515543
77. J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977). doi: 10.1016/0021-9991(77)90098-5
78. C. W. Hopkins, S. Le Grand, R. C. Walker, A. E. Roitberg, Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **11**, 1864–1874 (2015). doi: 10.1021/ct5010406; pmid: 26574392
79. S. Doerr, M. J. Harvey, F. Noé, G. De Fabritiis, HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J. Chem. Theory Comput.* **12**, 1845–1852 (2016). doi: 10.1021/acs.jctc.6b00049; pmid: 26949976
80. G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, Identification of slow molecular order parameters for Markov

model construction. *J. Chem. Phys.* **139**, 015102 (2013). doi: 10.1063/1.4811489; pmid: 23822324
81. M. K. Scherer *et al*., PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015). doi: 10.1021/acs.jctc.5b00743; pmid: 26574340
82. H. Wu, F. Noé, Variational Approach for Learning Markov Processes from Time Series Data. *J. Nonlinear Sci.* **30**, 23–66 (2020). doi: 10.1007/s00332-019-09567-y

ACKNOWLEDGMENTS

Cryo-EM data were collected at the National Center for Cryo-EM Access and Training (NCCAT) and the Simons Electron Microscopy Center located at the New York Structural Biology Center, supported by the NIH Common Fund Transformative High Resolution Cryo-Electron Microscopy program (U24 GM129539) and by grants from the Simons Foundation (SF349247) and the State of New York. We thank E. Eng, M. Aragon, E. Chua, and J. Mendez for microscope alignments and assistance with cryo-EM data collection. This study used the computational resources offered by Duke Research Computing (http://rc.duke.edu; NIH 1S10OD018164-01) at Duke University. Discovery and initial functional characterization of antibodies DH1041, DH1043, DH1047, DH1058, DH1050.1, DH1050.2, and DH1052 were supported by Defense Advanced Research Projects Agency grant N66001-09-C-2082, and work was performed in the Duke Regional Biocontainment Laboratory, which received partial support for construction from National Institute of Allergy and Infectious Diseases (NIAID) grant UC6-AI058607. **Funding:** Supported by an administrative supplement to NIH grant R01 AI145687 for coronavirus research (P.A. and R.H.); NIAID Division of AID grant AI142596 (B.F.H.); and the State of North Carolina funded by the Coronavirus Aid, Relief, and Economic Security 382 Act (CARES Act) (B.F.H.). **Author contributions:** S.M.-C.G. and P.A. designed and led the study and determined and analyzed cryo-EM structures; S.M.-C.G. designed SARS-CoV-2 ectodomain constructs, expressed and purified proteins, and performed SPR assays; K.J., V.S., and M.F.K. expressed and purified proteins; S.M., K.M., K.W., and R.H. performed structural analysis; K.M. and R.J.E. performed NSEM analysis; R.P. and D.L. performed ELISA assays; K.O.S. provided key reagents; B.K. supervised variant sequences; B.F.H. supervised ELISA assays; S.M.-C.G., P.A., and R.H. wrote the manuscript with help from all authors; R.H. led computational analysis; and P.A. supervised the study and reviewed all data. **Competing interests:** K.O.S., D.L., P.A., and B.F.H. are inventors on a patent application submitted by Duke University that covers the SARS-CoV-2 monoclonal antibodies studied in this paper. R.H., K.O.S., B.F.H., and P.A. are inventors on a patent application submitted by Duke University that covers the construct u1s2q. The other authors declare no competing interests. **Data and materials availability:** Cryo-EM reconstructions and atomic models generated during this study are available at wwPDB and EMBD (www.rcsb.org; http://emsearch.rutgers.edu) under the following accession codes: PDB IDs 7LWI, 7LWJ, 7LWK, 7LWL, 7LWM, 7LWN, 7LWO, 7LWP, 7LWQ, 7LWT, 7LWU, 7LWV, 7LWS, 7LWW, 7LYK, 7LYL, 7LYM, 7LYN, 7LYO, 7LYP, and 7LYQ; EMDB IDs EMDB-23546, EMD-23547, EMD-23548, EMD-23549, EMD-23550, EMD-23551, EMD-23552, EMD-23553, EMD-23554, EMD-23556, EMD-23557, EMD-23558, EMD-23555, EMD-23559, EMD-23593, EMD-23594, EMD-23595, EMD-23596, EMD-23597, EMD-23598, and EMD-23599. Vector analysis, Markov modelling scripts, and molecular dynamics trajectories are available at https://doi.org/10.5281/zenodo.4926233. Plasmids generated in this study have been deposited to Addgene (www.addgene.org) under the codes 171743, 171744, 171745, 171746, 171747, 171748, 171749, 171750, 171751, and 171752. Materials are available from the corresponding authors upon request. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6555/eabi6226/suppl/DC1
Supplementary Text
Figs. S1 to S30
Tables S1 to S7
MDAR Reproducibility Checklist

20 March 2021; accepted 16 June 2021
Published online 24 June 2021
10.1126/science.abi6226

RESEARCH ARTICLES

CORONAVIRUS

Structural basis for enhanced infectivity and immune evasion of SARS-CoV-2 variants

Yongfei Cai[1,2]†, Jun Zhang[1,2]†, Tianshu Xiao[1,2]†, Christy L. Lavine[3], Shaun Rawson[4,5,6], Hanqin Peng[1], Haisun Zhu[7], Krishna Anand[7], Pei Tong[8], Avneesh Gautam[8], Shen Lu[9], Sarah M. Sterling[5,6], Richard M. Walsh Jr.[5,6], Sophia Rits-Volloch[1], Jianming Lu[9,10], Duane R. Wesemann[8], Wei Yang[7], Michael S. Seaman[3], Bing Chen[1,2]*

Several fast-spreading variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have become the dominant circulating strains in the COVID-19 pandemic. We report here cryo–electron microscopy structures of the full-length spike (S) trimers of the B.1.1.7 and B.1.351 variants, as well as their biochemical and antigenic properties. Amino acid substitutions in the B.1.1.7 protein increase both the accessibility of its receptor binding domain and the binding affinity for receptor angiotensin-converting enzyme 2 (ACE2). The enhanced receptor engagement may account for the increased transmissibility. The B.1.351 variant has evolved to reshape antigenic surfaces of the major neutralizing sites on the S protein, making it resistant to some potent neutralizing antibodies. These findings provide structural details on how SARS-CoV-2 has evolved to enhance viral fitness and immune evasion.

The COVID-19 pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (*1*), has led to millions of lives lost and devastating socioeconomic disruptions worldwide. Although the mutation rate of the coronavirus is relatively low because of the proofreading activity of its replication machinery (*2*), several variants of concern have emerged—including the B.1.1.7 lineage first identified in the United Kingdom, the B.1.351 lineage in South Africa, and the B.1.1.28 lineage in Brazil—within a period of several months (*3*–*5*). These variants not only appear to spread more efficiently than the virus from the initial outbreak [i.e., the strain Wuhan-Hu-1; (*1*)] but also may be more resistant to immunity elicited by the Wuhan-Hu-1 strain after either natural infection or vaccination (*6*–*8*). The B.1.1.7 variant is of particular concern because it has been reported to be more deadly (*9*, *10*). Thus, understanding the underlying mechanisms of the increased transmissibility, risk of mortality, and immune resistance of new variants may facilitate development of intervention strategies to control the crisis.

SARS-CoV-2 is an enveloped positive-stranded RNA virus that depends on fusion of viral and target cell membranes to enter a host cell. This first key step of infection is catalyzed by the virus-encoded trimeric spike (S) protein, which is also a major surface antigen and thus an important target for development of diagnostics, vaccines, and therapeutics. The S protein is synthesized as a single-chain precursor and is subsequently cleaved by a furin-like protease into the receptor-binding fragment S1 and the fusion fragment S2 [fig. S1 and (*11*)]. Binding of the viral receptor angiotensin-converting enzyme 2 (ACE2) on the host cell surface to the receptor-binding domain (RBD) of S1, together with a second proteolytic cleavage by another cellular protease in S2 [S2′ site; fig. S1 and (*12*)], induces dissociation of S1 and irreversible refolding of S2 into a postfusion structure, ultimately leading to membrane fusion (*13*, *14*). In the prefusion conformation, S1 folds into four domains—NTD (N-terminal domain), RBD, and two CTDs (C-terminal domains)—that wrap around the prefusion S2 structure. The RBD can adopt two distinct conformations: "up" for a receptor-accessible state and "down" for a receptor-inaccessible state (*15*). Rapid progress in structural biology of the S protein has advanced our knowledge of the SARS-CoV-2 entry process (*15*–*28*). We have previously identified two structural elements, the FPPR (fusion peptide proximal region) and the 630 loop, which appear to modulate the S protein stability as well as the RBD conformation and thus the receptor accessibility (*22*, *28*).

The S protein is the basis of almost all the first-generation COVID-19 vaccines, which were developed using the Wuhan-Hu-1 sequence (*29*, *30*). Several have received emergency use authorization by various regulatory agencies throughout the world because of their impressive protective efficacy and minimal side effects (*31*, *32*). These vaccines appear to have somewhat lower efficacy against the B.1.351 variant than against its parental strain (*6*–*8*, *33*), and this variant became completely resistant to many convalescent serum samples in vitro (*8*). How to address genetic diversity has therefore become a high priority for developing next-generation vaccines. In this study, we have characterized the full-length S proteins from the B.1.1.7 and B.1.351 variants and determined their structures by cryo–electron microscopy (cryo-EM), providing a structural basis for understanding the molecular mechanisms of the enhanced infectivity of B.1.1.7 and the immune evasion of B.1.351.

Biochemical and antigenic properties of the intact S proteins from the new variants

To characterize the full-length S proteins with the sequences derived from natural isolates of the B.1.1.7 (hCoV-19/England/MILK-C504CD/2020) and B.1.351 (hCoV-19/South Africa/KRISP-EC-MDSH925100/2020) variants (fig. S1), we first transfected human embryonic kidney (HEK) 293 cells with the respective expression constructs and compared their membrane fusion activities with those of the full-length S constructs of their parental strains [Wuhan-Hu-1: D614 (Asp at position 614) and its early D614G variant: G614 (Asp-to-Gly mutation at position 614) (*34*)]. All S proteins expressed at comparable levels (fig. S2A), and the cells producing these S proteins fused efficiently with ACE2-expressing cells (fig. S2B). Consistent with our previous findings (*22*, *28*), the G614 and B.1.351 variant S constructs showed slightly higher fusion activity than the D614 and B.1.1.7 variants, but the differences diminished when the transfection level increased.

To produce the full-length S proteins, we added a C-terminal strep-tag to the B.1.1.7 and B.1.351 S (fig. S3A) and expressed and purified these proteins under the conditions established for producing the D614 and G614 S trimers (*22*, *28*). The B.1.1.7 protein eluted in three distinct peaks, representing the prefusion S trimer, postfusion S2 trimer, and dissociated S1 monomer, respectively (*22*), consistent with Coomassie-stained SDS–polyacrylamide gel electrophoresis (SDS-PAGE) analysis (fig. S3B). Nonetheless, the prefusion trimer was the predominant form, accounting for >70% of the total protein, indicating that this trimer is more

[1]Division of Molecular Medicine, Boston Children's Hospital, 3 Blackfan Street, Boston, MA 02115, USA. [2]Department of Pediatrics, Harvard Medical School, 3 Blackfan Street, Boston, MA 02115, USA. [3]Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, MA 02215, USA. [4]SBGrid Consortium, Harvard Medical School, 250 Longwood Avenue, Boston, MA 02115, USA. [5]The Harvard Cryo-EM Center for Structural Biology, Harvard Medical School, 250 Longwood Avenue, Boston, MA 02115, USA. [6]Department of Biological Chemistry and Molecular Pharmacology, Blavatnik Institute, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, USA. [7]Institute for Protein Innovation, Harvard Institutes of Medicine, 4 Blackfan Circle, Boston, MA 02115, USA. [8]Division of Allergy and Immunology and Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA. [9]Codex BioSolutions, Inc., 401 Professional Drive, Gaithersburg, MD 20879, USA. [10]Department of Biochemistry and Molecular and Cellular Biology, Georgetown University School of Medicine, 3900 Reservoir Road, NW, Washington, DC 20057, USA.
*Corresponding author. Email: bchen@crystal.harvard.edu
†These authors contributed equally to this work.

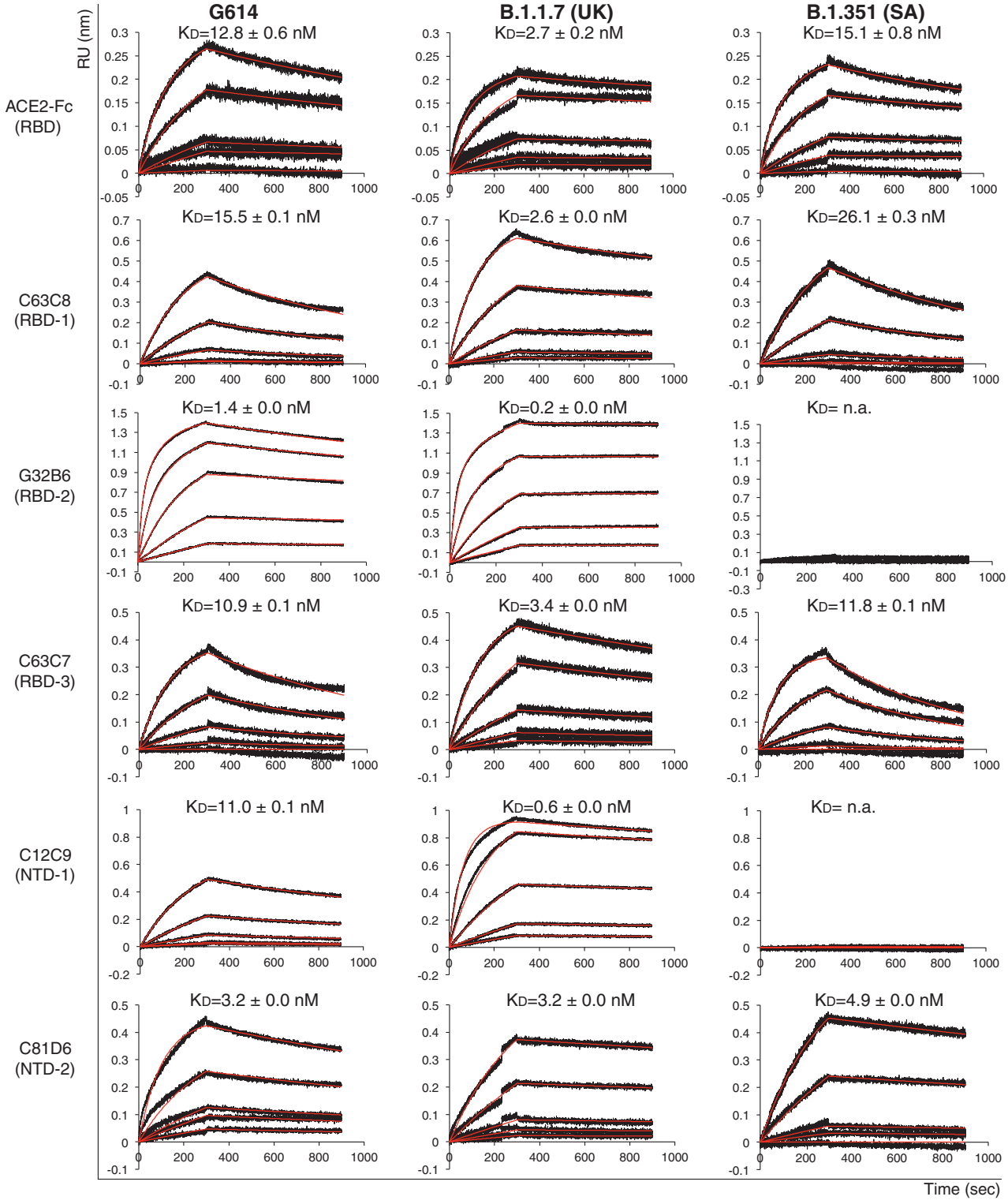


Fig. 1. Antigenic properties of the purified full-length SARS-CoV-2 S proteins. BLI analysis of the association of prefusion S trimers from the G614 "parent" strain and the B.1.1.7 and B.1.351 variants derived from it with soluble ACE2 constructs and with a panel of antibodies representing five epitopic regions on the RBD and NTD [see fig. S4A and (*35*)]. For ACE2 binding, purified S proteins were immobilized to AR2G (Amine Reactive 2nd Generation) biosensors and dipped into the wells containing ACE2 at different concentrations. For antibody binding, various antibodies were immobilized to AHC (Anti-human IgG Fc Capture) biosensors and dipped into the wells containing each purified S protein at different concentrations. Binding kinetics were evaluated using a 1:1 Langmuir model except for dimeric ACE2 and antibody G32B6 targeting the RBD-2, which were analyzed by a bivalent binding model. The sensorgrams are in black and the fits in red. Binding constants are also summarized here and in table S1. All experiments were repeated at least twice with essentially identical results. K_D, dissociation constant; n.a., not available; RU, response unit.

stable than D614, where the prefusion trimer was only <25%. Like the G614 trimer (*28*), B.1.351 protein eluted in a single major peak, corresponding to the prefusion S trimer (fig. S3B), with no obvious peaks for dissociated S1 and S2. SDS-PAGE analysis showed that the prefusion trimer peaks contained primarily the cleaved S1/S2 complex for both the proteins, with the cleavage level moderately higher for B.1.351 than for B.1.1.7. These results indicate that the B.1.351 and G614 S proteins have almost identical biochemical properties, whereas the B.1.1.7 trimer is slightly less stable.

To assess antigenic properties of the prefusion variant S trimers, we measured their binding to soluble ACE2 and S-directed monoclonal antibodies isolated from COVID-19 convalescent individuals by bio-layer interferometry (BLI). These antibodies target various epitopic regions on the S trimer, as defined by clusters of competing antibodies and designated RBD-1, RBD-2, RBD-3, NTD-1, NTD-2, and S2 [fig. S4A; (*35*)]. All but the last two clusters contain neutralizing antibodies. The B.1.1.7 variant bound stronger to the receptor than did its G614 parent, regardless of the ACE2 oligomeric state (Fig. 1, fig. S4B, and table S1). The B.1.351 trimer had higher affinity for monomeric ACE2, but slightly lower affinity for dimeric ACE2, than the G614 trimer. In both cases, affinity for ACE2 of the B.1.351 protein was lower than that of the B.1.1.7 variant. These results suggest that the mutation in the RBD of the B.1.1.7 variant [N501Y (Asn^{501}→Tyr)] enhances receptor recognition, whereas the additional mutations in the B.1.351 variant [K417N (Lys^{417}→Asn) and E484K (Glu^{484}→Lys)] reduce ACE2 affinity to a level close to that of the G614 protein, consistent with the previous data (*36*, *37*). All selected monoclonal antibodies bound G614 S with reasonable affinities, and the B.1.1.7 variant showed a similar pattern but with substantially stronger binding to almost all the antibodies (Fig. 1, fig. S4B, and table S1). By contrast, the B.1.351 variant completely lost binding to the two RBD-2 antibodies, G32B6 and C12A2, as well as to the two NTD-1 antibodies, C12C9 and C83B6, whereas the affinities for the rest of the antibodies were the same as those of the G614 trimer. The BLI data were also consistent with the binding results with the membrane-bound S trimers measured by flow cytometry (fig. S5).

We next assessed the neutralization potency of the antibodies and the trimeric ACE2 construct in blocking infection of these variants in an HIV-based pseudovirus assay. For most antibodies, the neutralization potency correlated with their binding affinity for the membrane-bound or purified S proteins (fig. S6 and table S2). C81D6 and C163E6 recognize two non-neutralizing epitopes, located in the NTD and S2, respectively, and they did not neutralize any of the pseudoviruses. The B.1.1.7 virus is the most sensitive to the trimeric ACE2 and the RBD-up–targeting C63C7, suggesting that the B.1.1.7 trimer may prefer the RBD-up conformation. Thus, the detergent-solubilized S proteins adopt a physiologically relevant conformation, and mutations in B.1.351 have a greater impact on the antibody sensitivity of the virus than those in B.1.1.7.

Structures of the full-length S trimers from the B.1.1.7 and B.1.351 variants

We determined the cryo-EM structures of the full-length S trimers with the unmodified sequences of the B.1.1.7 and B.1.351 variants. Cryo-EM images were acquired on a Titan Krios electron microscope equipped with a Gatan K3 direct electron detector. We used RELION (*38*) for particle picking, two-dimensional (2D) classification, 3D classification, and refinement (figs. S7 to S10), and cryoSPARC (*39*) for validation. 3D classification identified five distinct classes for the B.1.1.7 S trimer—representing one closed prefusion conformation, three one-RBD-up conformations, and one two-RBD-up conformation—and two different classes for the B.1.351 trimer, representing a closed conformation and a one-RBD-up conformation. These structures were refined to 2.9- to 4.3-Å resolution (figs. S7 to S10 and table S3).

The overall architectures of the full-length variant S proteins are very similar to that of the G614 S trimer in the corresponding conformation [figs. S11 and S12; (*28*)]. In the closed, three-RBD-down structure, the four domains of S1—NTD, RBD, CTD1, and CTD2—wrap around the prefusion S2 trimer. In the one-RBD-up conformation, the RBD position has no effect on the central core region of S2, but two NTDs, the immediately adjacent one and the one from the same protomer, shift away from the threefold axis and open up the trimer. The furin cleavage site at the S1/S2 boundary (residues 682 to 685) in these structures remains disordered, and the structures therefore cannot explain the difference in the cleavage level between the B.1.1.7 and B.1.351 trimers; the position of a substitution [P681H (Pro^{681}→His)] in the B.1.1.7 S (fig. S1) close to the cleavage site is likewise not visible. A small class of particles in the two-RBD-up conformation was present only with the B.1.1.7 trimer (fig. S11), possibly because B.1.1.7 S1 is less likely to dissociate.

For the B.1.1.7 S trimer, most particles used for refinement were in the RBD-up conformation (Fig. 2, A to E). We have proposed that the FPPR (residues 828 to 853) and 630 loop (residues 620 to 640) modulate the stability and fusogenic structural rearrangements of the S protein (*22*, *28*). In the closed conformation of the B.1.1.7 trimer, all three FPPR and three 630 loops are disordered (Fig. 2F), which otherwise would help clamp down the RBDs. This explains why the B.1.1.7 trimer is more likely than its parental G614 variant to populate the RBD-up conformation, because the FPPRs and 630 loops are structured in the G614 trimer (*28*). In the one-RBD-up conformation, one 630 loop on the opposite side of the up RBD becomes fully structured, inserting between neighboring NTD and CTDs in the same configuration found in the G614 trimer (*28*). The second 630 loop is partially ordered, whereas the third one remains disordered. A similar pattern is found for three FPPRs, although the structured FPPR adopts a conformation distinct from the one seen in our previous structures of the full-length S proteins (*22*, *28*). Overall, the arrangement of these structural elements appears to stabilize the cleaved S trimer and to prevent the premature S1 dissociation in the one-RBD-up conformation. The three one-RBD-up structures differ only by the degree to which the up RBD and the adjacent NTD of its neighboring protomer shift away from the central threefold axis (fig. S13A). We have suggested that the two-RBD-up conformation might be unstable (*22*, *28*), leading to S1 dissociation and irreversible S2 refolding. If this suggestion is valid, the small class of the two-RBD-up particles probably contains mainly uncleaved S trimers.

The two classes for the B.1.351 S trimer represent the closed prefusion and one-RBD-up states, respectively (Fig. 2, G and H). The configurations of the FPPR and 630 loop closely follow the distribution seen in the G614 trimer: All are structured in the RBD-down conformation, whereas only the FPPR and 630-loop pair is ordered in the one-RBD-up conformation [fig. S12; (*28*)]. These observations are consistent with the similar biochemical stabilities of the B.1.351 and G614 S trimers [fig. S3; (*28*)].

Structural consequences of mutations in the B.1.1.7 variant

We superposed the structures of the B.1.1.7 trimer onto the G614 trimer in the closed conformation, aligning them by the S2 structure (Fig. 3A). An outward rotation of all three S1 subunits in B.1.1.7 leads to a slightly more open conformation. This rotation in B.1.1.7 widens the gap between the NTD and the CTDs of the same protomer (fig. S13B). In the G614 trimer, this gap accommodates the ordered 630 loop that reinforces CTD2 and prevents S1 shedding (*28*). The widened gap in the variant loosens the grip on the 630 loop, accounting for the absence of ordered features in this part of the B.1.1.7 map. There are two mutations that may be responsible for these structural differences. First, Ala^{570} in CTD1 packs against one side of the FPPR in the G614 trimer (Fig. 3B). The A570D (Ala^{570}→Asp) mutation, with a larger side chain, may weaken the packing and destabilize the FPPR. Moreover, in the one-RBD-up conformation of the B.1.1.7 S, in which the FPPR is at least partially structured, Lys^{854}, which in

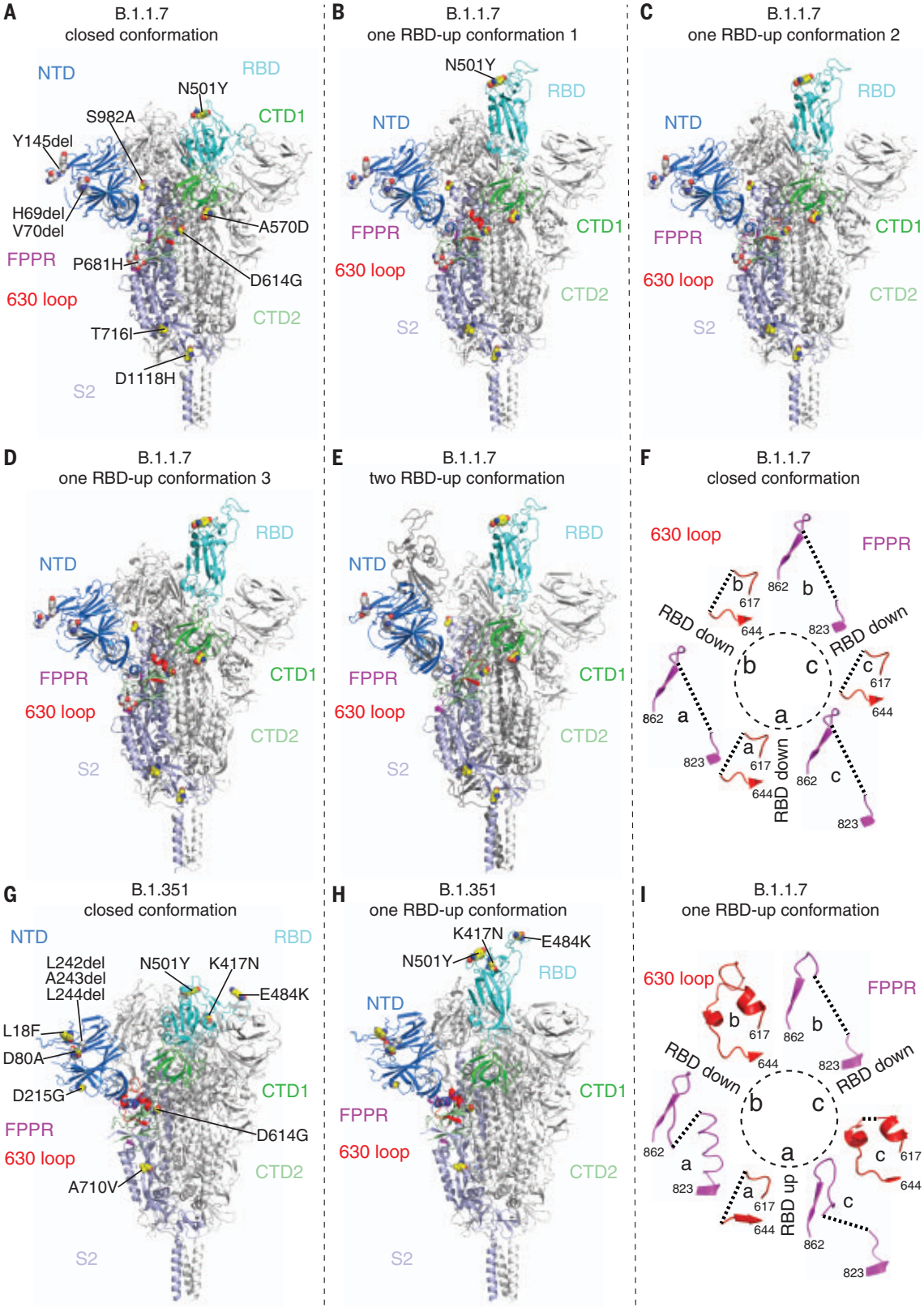


Fig. 2. Cryo-EM structures of the full-length SARS-CoV-2 S proteins from the B.1.1.7 and B.1.351 variants. (**A** to **E**) The structures of the closed prefusion conformation (A), three one-RBD-up conformations [(B) to (D)], and a two-RBD-up conformation (E) of the B.1.1.7 S trimer are shown in ribbon diagram, with one protomer colored as NTD in blue, RBD in cyan, CTD1 in green, CTD2 in light green, S2 in light blue, the 630 loop in red, and the FPPR in magenta. (**G** and **H**) The structures of the closed prefusion conformation (G) and one-RBD-up conformation (H) of the B.1.351 S trimer are shown in ribbon diagram with the same color scheme as in (A) to (E). All mutations in the new variants, as compared with the original virus (D614), are highlighted as sphere models. (**F** and **I**) Structures, in the B.1.1.7 trimer, of segments (residues 617 to 644) containing the 630 loop (red) and segments (residues 823 to 862) containing the FPPR (magenta) from each of the three protomers (a, b, and c). The position of each RBD is indicated. Dashed lines indicate gaps in the chain trace (disordered loops).

the G614 trimer probably forms a hydrogen bond with the main chain carbonyl group of Gly^{614}, flips back in B.1.1.7 to form a salt bridge with the mutant Asp^{570}. Second, S982A (Ser^{982}→Ala) eliminates a hydrogen bond between the central helices of S2 and the carbonyl group of Gly^{545} in CTD1 (Fig. 3C). These two mutations together allow an outward movement of CTD1 by more than 3 Å (fig. S13B), thereby affecting the conformation of the FPPR and 630 loops. In the one-RBD-up conformation, the NTD and CTDs on the

opposite side of the up RBD move closer together, narrowing the gap between them and stabilizing the structured 630 loop.

Other mutations in B.1.1.7 cluster in the NTD, including deletions of His[69], Val[70], and Tyr[145] (Fig. 3D). The first two residues are in a disordered loop in all these S structures, and the structural impact of their deletion is unclear. Tyr[145] is also near a loop (residues 144 to 155), and its deletion apparently causes only some local changes of the loop. The absence of structural changes in the B.1.1.7 NTD is consistent with the absence of effects on its sensitivity to the various NTD-directed antibodies (*35*). Additional mutations [N501Y, T716I (Thr[716]→Ile), and D1118H (Asp[1118]→His)] caused minimal local changes (fig. S14, A to C).

Structural impact of the mutations in the B.1.351 variant

The overall structures of the B.1.351 and G614 trimers were essentially the same for the corresponding states, except for some loop regions in the NTD (Fig. 4A and fig. S15). Three mutations—K417N, E484K, and N501Y—at the ACE2 binding site do not produce any major structural rearrangements (Fig. 4B). The most notable differences are in the NTD, which contains three point mutations [L18F(Leu[18]→Phe), D80A (Asp[80]→Ala), and D215G (Asp[215]→Gly)] and a three-residue deletion (L242del, A243del, and L244del). The L18F and D80A changes lead to reconfiguration of the N-terminal segment despite the disulfide between Cys[15] and Cys[136] that partly anchors the N-terminal peptide (Fig. 4C). D215G appears to have the least structural impact because Asp[215] is a solvent-exposed residue that may compensate for the surface charge from the neighboring, well-exposed Arg[214].

The most consequential changes are probably from the triple-residue deletion, because these nonpolar residues, located on the edge of the NTD core structure formed by four stacking β sheets, are replaced with polar residues His[245]-Arg[246]-Ser[247]. This replacement causes a shift of the nearby loop (residues 144 to 155) and must also reconfigure the adjacent disordered loop (residues 246 to 260), both of which form part of the NTD neutralizing epitopes (*40*). Furthermore, Arg[246] is pointing toward the side chain of Arg[102] near the segment 172 to 188, forcing this loop to rearrange. As shown in Fig. 4D, the 172-to-188 segment wraps around the edge of the NTD core, packing against Leu[242]-Ala[243]-Leu[244] at the edge of the β sheet in the G614 trimer. The triple-residue deletion rearranges the 172-to-188 segment with a movement up to 17 Å (Leu[180]). By substantially altering the conformational preferences of this component of the molecular surface, these mutations likely affect binding of any antibody that has part of its footprint in this region. The additional mutation A701V (Ala[701]→Val) is located in the surface-exposed region of S2 and caused minimal structural changes (fig. S14D).

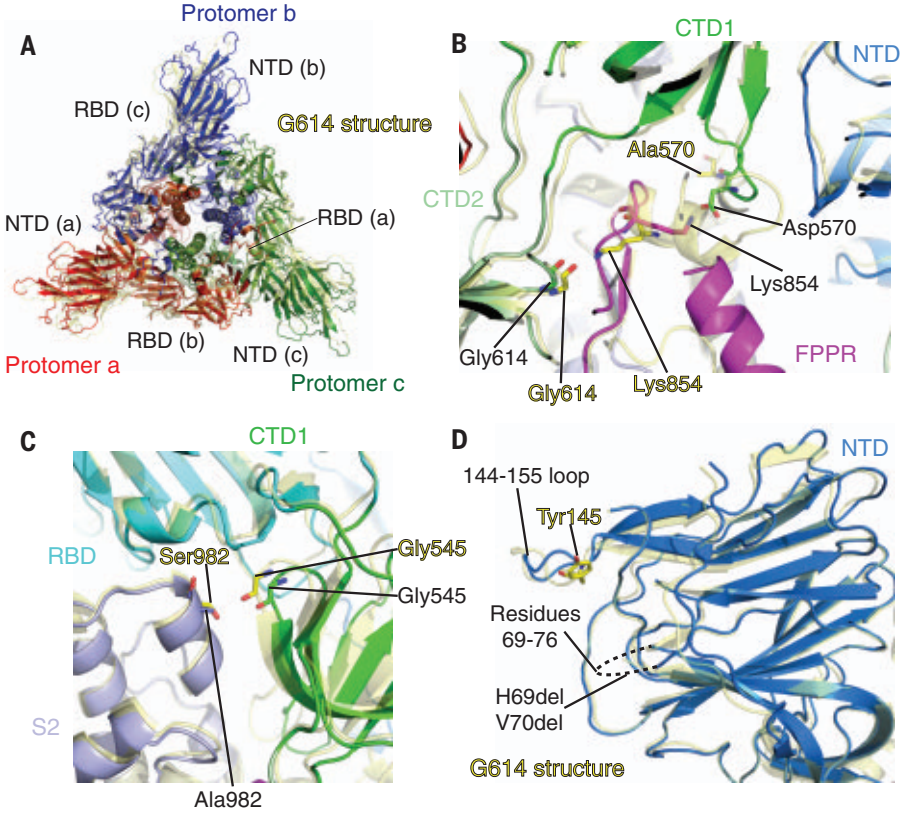


Fig. 3. Structural impact of the mutations in the B.1.1.7 S. (**A**) Top views of superposition of the structure of the B.1.1.7 S trimer in ribbon representation, with the structure of the prefusion trimer of the G614 S [Protein Data Bank (PDB) ID: 7KRQ] shown in yellow. The NTD and RBD of each protomer are indicated. (**B**) A close-up view of the region near the A570D mutation with superposition of the B.1.1.7 trimer structure (one-RBD-up) in green (CTD1) and magenta (FPPR) and the G614 trimer (closed) in yellow. Residues Ala[570], Asp[570], two Gly[614], and two Lys[854] from both structures are shown in stick model. (**C**) A view of the region near the S982A mutation with superposition of the B.1.1.7 trimer structure (closed) in green (CTD1) and magenta (FPPR) and the G614 trimer (closed) in yellow. (**D**) Superposition of the NTD structure of the B.1.1.7 S trimer in blue with the NTD of the G614 S trimer in yellow. Locations of Tyr[145] and the disordered loop containing residues 69 to 76 are indicated.

Discussion

Transmissibility and immune evasion are independent selective forces driving emergence of viral genetic diversity. The changes of most concern in the SARS-CoV-2 S protein would be those that simultaneously enhance transmission, augment disease severity, and evade immune recognition in previously exposed hosts. Our data suggest that the most problematic combination of such mutations is not yet present in the existing variants examined here.

In the B.1.1.7 virus, mutations A570D and S982A lead to an outward shift of the CTD1, thereby relaxing the FPPR and 630 loop, which help retain the RBD in its "down" position in the parental strain. The mutations increase the frequency with which the S trimer samples the RBD-up conformation, allowing B.1.1.7 to better present the receptor binding motif (RBM) to ACE2 on the host cells. Once one RBD flips up, the fully or partially ordered 630 loops of the neighboring protomers stabilize the CTD2, which folds together with the N-terminal segment of S2, and thus prevent the premature S1 dissociation. N501Y in the ACE2 binding site of the RBD also increases the affinity of that domain for the receptor, probably because of the hydrophobic interaction of Tyr[501] with Tyr[41] of ACE2 (*36*) and a possible cation-π interaction with ACE2 Lys[353] (fig. S16). The combination of enhanced RBM presentation and additional local interactions might allow the B.1.1.7 virus to infect cell types with lower ACE2 levels than those of the nasal and bronchial epithelial cells that the virus typically infects; an expanded cell tropism could account for the increased risk of mortality in patients infected with this variant (*9*, *10*). The mutations in B.1.1.7 caused no major structural rearrangements in the RBD and NTD, consistent with

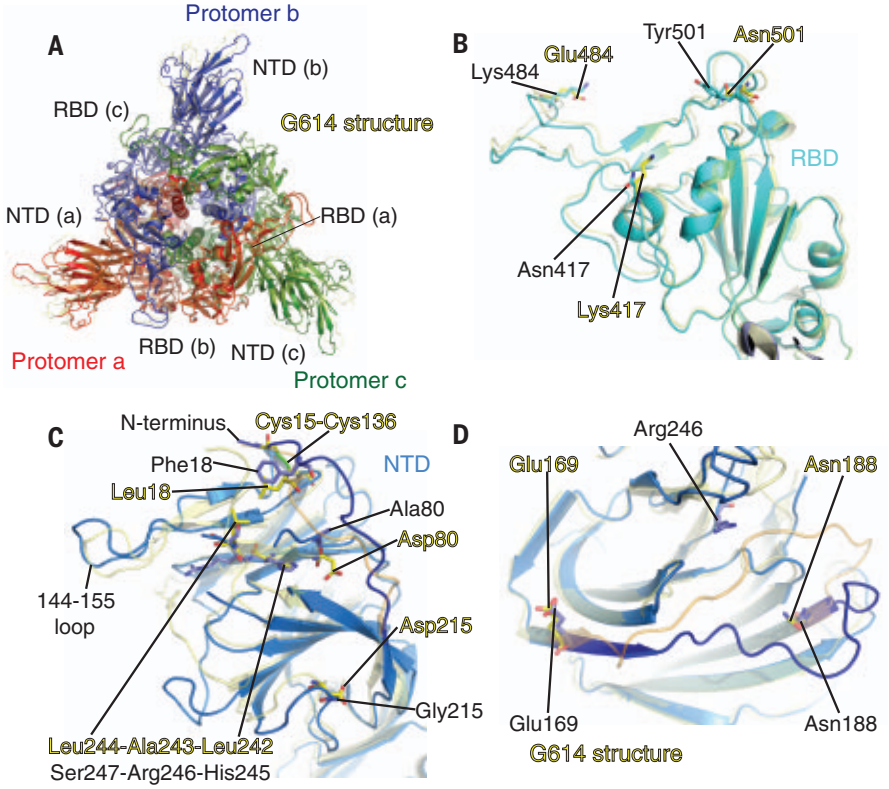


Fig. 4. Structural impact of the mutations in the B.1.351 S. (**A**) Top views of superposition of the structure of the B.1.351 S trimer in ribbon representation, with the structure of the prefusion trimer of the G614 S (PDB ID: 7KRQ) shown in yellow. The NTD and RBD of each protomer are indicated. (**B**) Superposition of the RBD structure of the B.1.351 S trimer in blue with the RBD of the G614 S trimer in yellow. Locations of mutations K417N, E484K, and N501Y are indicated, and these residues are shown in stick model. (**C**) A view of the NTDs from superposition of the structure of the B.1.351 S trimer in blue and the G614 S in yellow. Locations of mutations L18F, D80A, and D215G; the disulfide bond between Cys^{15} and Cys^{136}; and replacement of Leu^{242}-Ala^{243}-Leu^{244} by His^{245}-Arg^{246}-Ser^{247} are indicated, and the residues are shown in stick model. (**D**) Superposition of the NTD structure of the B.1.351 S trimer in blue with the NTD of the G614 S trimer in yellow. Displacement of the segment 169 to 188 and the location of Arg^{246} in the B.1.351 structure are indicated.

minimal changes in the sensitivity of the B.1.1.7 variant to the potently neutralizing antibodies [tables S1 and S2; (*33*)].

In the B.1.351 virus, the S protein largely retains the structure of the G614 trimer with almost identical biochemical stability. N501Y, K417N, and E484K in the RBD have not caused major structural changes, but the loss of salt bridges between Lys^{417} and ACE2 Asp^{30} and Glu^{484} and ACE2 Lys^{31} mitigates the increased receptor affinity imparted by N501Y (fig. S16). K417N and E484K probably lead to loss of binding and neutralization by antibodies that target the RBD-2 epitopes (fig. S4A). The accompanying mutations in the NTD remodel the antigenic surface and greatly reduce the potency of neutralizing antibodies against NTD-1 epitopes. The B.1.351 variant was probably selected under a certain level of immune pressure, because it altered two major neutralizing sites on the S trimer simultaneously with only a slight compromise in its ability to engage a host cell.

The global range of SARS-CoV-2 and the daily vast number of replication events make emergence of new variants inevitable and substantially increases the viral genetic diversity. In many cases, antibody resistance may compromise viral fitness, as in the B.1.351 variant, which resists neutralization by RBD-directed antibodies but also loses the enhanced affinity and transmissibility imparted by N501Y, as a consequence of the immune-escape mutations. It is also possible to combine immune evasion and virulence through continuous viral evolution, such as a B.1.1.7 variant that contains the E484K mutation (B.1.1.7+E484K) (*41*). Such a combination will bring greater challenges for vaccine development compared with the beginning of the pandemic. If SARS-CoV-2 becomes seasonal, innovative strategies already developed against other human pathogens—such HIV-1, hepatitis C virus, and influenza virus—may be applicable to on-going control of the COVID-19 pandemic. The B.1.351 S trimer, which has superior biochemical stability and new epitopes, should be an excellent starting point for developing next-generation vaccines designed to elicit broadly neutralizing antibody responses.

REFERENCES AND NOTES

1. P. Zhou *et al.*, *Nature* **579**, 270–273 (2020).
2. F. Robson *et al.*, *Mol. Cell* **79**, 710–727 (2020).
3. H. Tegally *et al.*, medRxiv 2020.12.21.20248640 [Preprint] (2020). https://doi.org/10.1101/2020.12.21.20248640.
4. F. Grabowski, G. Preibisch, M. Kochańczyk, T. Lipniacki, *Viruses* **13**, 392 (2021).
5. C. M. Voloch *et al.*, *J. Virol.* **95**, e00119–e00121 (2021).
6. K. Wu, A. P. Werner *et al.*, bioRxiv 2021.01.25.427948 [Preprint] (2021). https://doi.org/10.1101/2021.01.25.427948.
7. P. Wang *et al.*, *Nature* **539**, 130–135 (2021).
8. C. K. Wibmer *et al.*, *Nat. Med.* **27**, 622–625 (2021).
9. N. G. Davies *et al.*, *Nature* **593**, 270–274 (2021).
10. R. Challen *et al.*, *BMJ* **372**, n579 (2021).
11. B. J. Bosch, R. van der Zee, C. A. de Haan, P. J. Rottier, *J. Virol.* **77**, 8801–8811 (2003).
12. M. Hoffmann *et al.*, *Cell* **181**, 271–280.e8 (2020).
13. J. K. Millet, G. R. Whittaker, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15214–15219 (2014).
14. M. A. Tortorici, D. Veesler, *Adv. Virus Res.* **105**, 93–116 (2019).
15. D. Wrapp *et al.*, *Science* **367**, 1260–1263 (2020).
16. A. C. Walls *et al.*, *Cell* **181**, 281–292.e6 (2020).
17. J. Lan *et al.*, *Nature* **581**, 215–220 (2020).
18. R. Yan *et al.*, *Science* **367**, 1444–1448 (2020).
19. J. Shang *et al.*, *Nature* **581**, 221–224 (2020).
20. Q. Wang *et al.*, *Cell* **181**, 894–904.e9 (2020).
21. S. Xia *et al.*, *Cell Res.* **30**, 343–355 (2020).
22. Y. Cai *et al.*, *Science* **369**, 1586–1592 (2020).
23. S. Bangaru *et al.*, *Science* **370**, 1089–1094 (2020).
24. B. Turoňová *et al.*, *Science* **370**, 203–208 (2020).
25. Z. Ke *et al.*, *Nature* **588**, 498–502 (2020).
26. H. Yao *et al.*, *Cell* **183**, 730–738.e13 (2020).
27. C. Liu *et al.*, *Structure* **28**, 1218–1224.e4 (2020).
28. J. Zhang *et al.*, *Science* **372**, 525–530 (2021).
29. L. Dai, G. F. Gao, *Nat. Rev. Immunol.* **21**, 73–82 (2021).
30. M. S. Gebre *et al.*, *Cell* **184**, 1589–1603 (2021).
31. L. R. Baden *et al.*, *N. Engl. J. Med.* **384**, 403–416 (2021).
32. F. P. Polack *et al.*, *N. Engl. J. Med.* **383**, 2603–2615 (2020).
33. S. S. Abdool Karim, T. de Oliveira, *N. Engl. J. Med.* **384**, 1866–1868 (2021).
34. B. Korber *et al.*, *Cell* **182**, 812–827.e19 (2020).
35. P. Tong *et al.*, bioRxiv 2021.03.10.434840 [Preprint] (2021). https://doi.org/10.1101/2021.03.10.434840.
36. X. Zhu *et al.*, *PLOS Biol.* **19**, e3001237 (2021).
37. S. M. Gobeil *et al.*, bioRxiv 2021.03.11.435037 [Preprint] (2021). https://doi.org/10.1101/2021.03.11.435037.
38. S. H. Scheres, *J. Struct. Biol.* **180**, 519–530 (2012).
39. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, *Nat. Methods* **14**, 290–296 (2017).
40. X. Chi *et al.*, *Science* **369**, 650–655 (2020).
41. J. Wise, *BMJ* **372**, n359 (2021).

ACKNOWLEDGMENTS

We thank the SBGrid team for technical assistance, K. Arnett for support and advice on the BLI experiments, and S. Harrison and A. Carfi for critical reading of the manuscript. EM data were collected at the Harvard Cryo-EM Center for Structural Biology of Harvard Medical School. We acknowledge support for COVID-19–related structural biology research at Harvard from the Nancy Lurie Marks Family Foundation and the Massachusetts Consortium on Pathogen Readiness (MassCPR). **Funding:** This work was supported by a Fast grant by Emergent Ventures (to B.C.), a COVID-19 Award by Massachusetts Consortium on Pathogen Readiness (MassCPR; to B.C.), and NIH grants AI147884 (to B.C.), AI141002 (to B.C.), and AI127193 (to B.C. and J. Chou). **Author**

contributions: B.C., Y.C., J.Z., and T.X. conceived the project. Y.C. and H.P. expressed and purified the full-length S proteins. T.X. performed BLI and cell-cell fusion experiments. J.Z. and Y.C. prepared cryo grids and performed EM data collection with contributions from S.M.S. and R.M.W. J.Z. and Y.C. processed the cryo-EM data and built and refined the atomic models with help from S.R. C.L.L. and M.S.S. performed the neutralization assays using the HIV-based pseudoviruses. H.Z., K.A., and W.Y. performed the flow cytometry experiments. P.T., A.G., and D.R.W. produced anti-S monoclonal antibodies. S.L. and J.L. created all the expression constructs. S.R.-V. contributed to cell culture and protein production. All authors analyzed the data. B.C., Y.C., J.Z., and T.X. wrote the manuscript with input from all other authors.
Competing interests: W.Y. serves on the scientific advisory boards of Hummingbird Bioscience and GO Therapeutics and is a consultant to GV20 Oncotherapy. All other authors declare no competing interests. **Data and materials availability:** The atomic structure coordinates are deposited in the RCSB Protein Data Bank (PDB) under accession numbers 7N1Q, 7N1T, 7N1U, 7N1V, 7N1W, 7N1X, and 7N1Y; and the electron microscopy maps have been deposited in the Electron Microscopy Data Bank (EMDB) under accession numbers EMD-24121, EMD-24122, EMD-24123, EMD-24124, EMD-24125, EMD-24126, and EMD-24127. All materials generated during the current study are available from the corresponding author under a materials transfer agreement with Boston Children's Hospital. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6555/642/suppl/DC1
Materials and Methods
Figs. S1 to S16
Tables S1 to S3
References (*42*–*51*)
MDAR Reproducibility Checklist

12 April 2021; accepted 15 June 2021
Published online 24 June 2021
10.1126/science.abi9745

CORONAVIRUS

SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern

Matthew McCallum[1]†, Jessica Bassi[2]†, Anna De Marco[2]†, Alex Chen[3]†, Alexandra C. Walls[1]†, Julia Di Iulio[3], M. Alejandra Tortorici[1], Mary-Jane Navarro[1], Chiara Silacci-Fregni[2], Christian Saliba[2], Kaitlin R. Sprouse[1], Maria Agostini[3], Dora Pinto[2], Katja Culap[2], Siro Bianchi[2], Stefano Jaconi[2], Elisabetta Cameroni[2], John E. Bowen[1], Sasha W. Tilles[4], Matteo Samuele Pizzuto[2], Sonja Bernasconi Guastalla[5], Giovanni Bona[6], Alessandra Franzetti Pellanda[6], Christian Garzoni[7], Wesley C. Van Voorhis[4], Laura E. Rosen[3], Gyorgy Snell[3], Amalio Telenti[3], Herbert W. Virgin[3], Luca Piccoli[2]*, Davide Corti[2]*, David Veesler[1]*

A novel variant of concern (VOC) named CAL.20C (B.1.427/B.1.429), which was originally detected in California, carries spike glycoprotein mutations S13I in the signal peptide, W152C in the N-terminal domain (NTD), and L452R in the receptor-binding domain (RBD). Plasma from individuals vaccinated with a Wuhan-1 isolate–based messenger RNA vaccine or from convalescent individuals exhibited neutralizing titers that were reduced 2- to 3.5-fold against the B.1.427/B.1.429 variant relative to wild-type pseudoviruses. The L452R mutation reduced neutralizing activity in 14 of 34 RBD-specific monoclonal antibodies (mAbs). The S13I and W152C mutations resulted in total loss of neutralization for 10 of 10 NTD-specific mAbs because the NTD antigenic supersite was remodeled by a shift of the signal peptide cleavage site and the formation of a new disulfide bond, as revealed by mass spectrometry and structural studies.

COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and is associated with acute respiratory distress syndrome, as well as extrapulmonary complications such as vascular thrombosis, coagulopathy, and a hyperinflammatory syndrome contributing to disease severity and mortality. SARS-CoV-2 infects target cells using the spike glycoprotein (S), which is organized as a homotrimer with each monomer comprising an S_1 and an S_2 subunit (*1*, *2*). The S_1 subunit harbors the receptor-binding domain (RBD) and the N-terminal domain (NTD), as well as two other domains designated here as C and D (*3*, *4*). The RBD interacts with the angiotensin-converting enzyme 2 (ACE2) entry receptor on host cells through a subset of amino acids forming the receptor-binding motif (*1*, *2*, *5*–*7*). The NTD was suggested to bind DC-SIGN, L-SIGN, and AXL, which may act as attachment receptors (*8*, *9*). Both the RBD and the NTD are targeted by neutralizing antibodies (Abs) in infected or vaccinated individuals, and some RBD-specific monoclonal Abs (mAbs) are currently being evaluated in clinical trials or are authorized for use in COVID-19 patients (*10*–*24*). The S_2 subunit is the fusion machinery that merges viral and host membranes to initiate infection and is the target of Abs cross-reacting with multiple coronavirus subgenera because of its higher sequence conservation compared with the S_1 subunit (*25*–*28*).

[1]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. [2]Humabs Biomed SA, a subsidiary of Vir Biotechnology, 6500 Bellinzona, Switzerland. [3]Vir Biotechnology, San Francisco, CA 94158, USA. [4]Center for Emerging and Re-emerging Infectious Diseases, Division of Allergy and Infectious Diseases, Department of Medicine, University of Washington School of Medicine, Seattle, WA 98195, USA. [5]Independent physician, 6828 Balerna, Switzerland. [6]Clinical Research Unit, Clinica Luganese Moncucco, 6900 Lugano, Switzerland. [7]Clinic of Internal Medicine and Infectious Diseases, Clinica Luganese Moncucco, 6900 Lugano, Switzerland.
***Corresponding author. Email: lpiccoli@vir.bio (L.P.); dcorti@vir.bio (D.C.); dveesler@uw.edu (D.V.)**
†These authors contributed equally to this work.

The ongoing global spread of SARS-CoV-2 led to the fixation of the D614G substitution (*29*, *30*), as well as to the emergence of a large number of viral lineages worldwide, including several variants of concern (VOC). Specifically, the B.1.1.7, B.1.351, and P.1 lineages that originated in the United Kingdom, South Africa, and Brazil, respectively, are characterized by the accumulation of mutations in the S gene as well as in other genes (*31*–*33*). Some of these mutations lead to marked reductions in the neutralization potency of several mAbs, convalescent sera, and Pfizer/BioNTech BNT162b2– or Moderna mRNA-1273–elicited Abs (*19*, *34*–*40*). The B.1.1.7 variant has become dominant worldwide because of its higher transmissibility (*33*), underscoring the importance of studying and understanding the consequences of SARS-CoV-2 antigenic drift.

Results

The incidence of the B.1.427/B.1.429 lineages is increasing rapidly

The SARS-CoV-2 B.1.427/B.1.429 variant was reported for the first time at the beginning of 2021 in California, and as of May 2021 had been detected in 34 additional countries (*41*, *42*). The two lineages B.1.427 and B.1.429 (belonging to clade 20C according to Nextstrain designation) share the same S mutations (S13I in the signal peptide, W152C in the NTD, and L452R in the RBD) but harbor different mutations in other SARS-CoV-2 genes (*42*). Molecular clock analysis suggested that the progenitor of both lineages emerged in May 2020, diverging to give rise to the B.1.427 and B.1.429 lineages in June and July 2020 (*42*). The fast rise in the number of cases associated with the B.1.427/B.1.429 lineages led to their classification as a VOC by the US Centers for Disease Control and Prevention (https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html).

As of 30 April 2021, 8441 and 21,072 sequenced genomes had been reported in GISAID (Global Initiative for Sharing all Influenza Data) for the B.1.427 and B.1.429 lineages, respectively. This VOC was detected in California

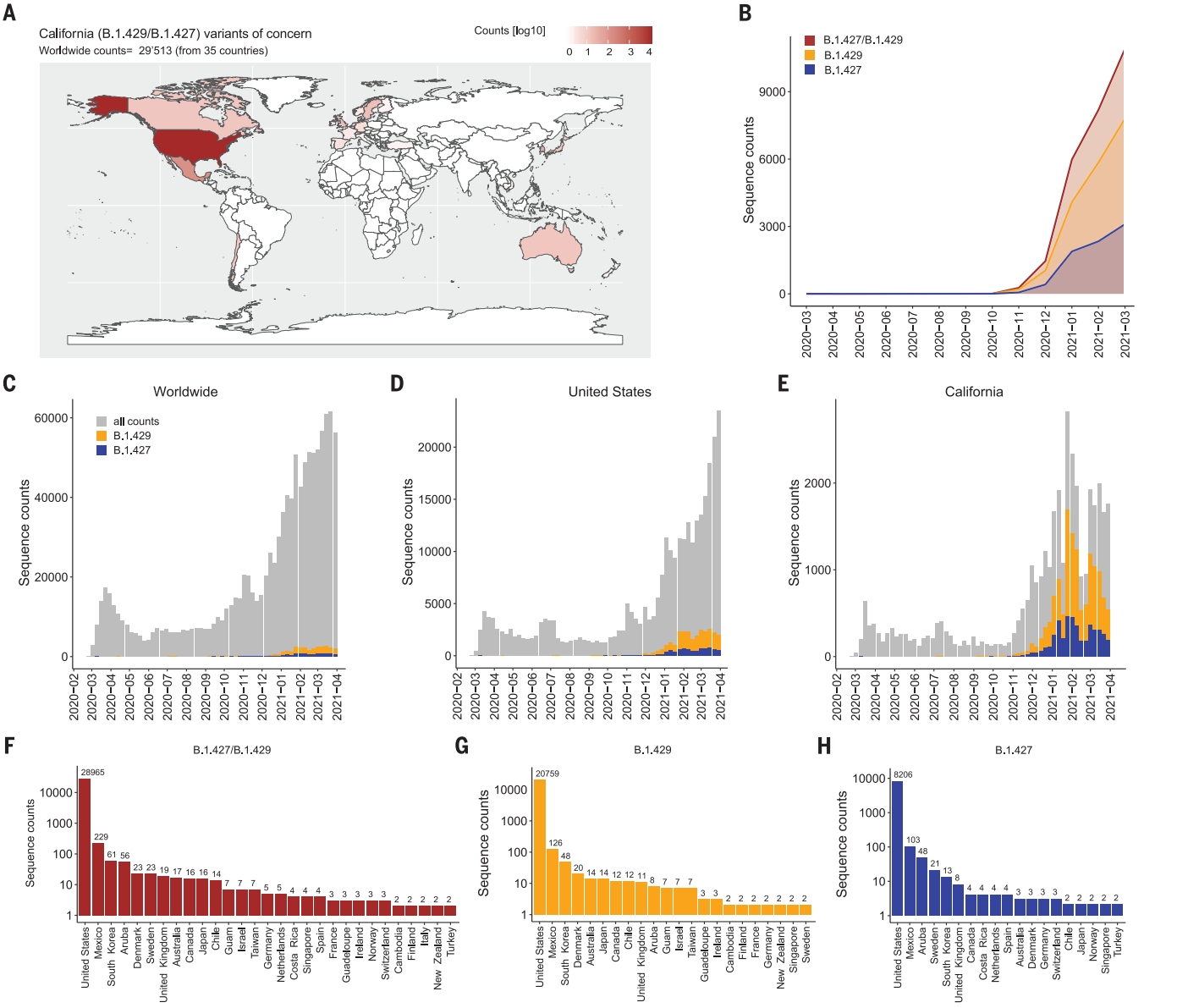


Fig. 1. Geographic distribution and evolution of incidence over time of the SARS-CoV-2 B.1.427/B.1.429 VOC. (**A**) World map showing the geographic distribution and sequence counts of B.1.427/B.1.429 VOC as of 30 April 2021. (**B**) Cumulative and individual B.1.427/B.1.429 VOC sequence counts by month. (**C** to **E**) Total number of SARS-CoV-2 (gray) and B.1.427/B.1.429 VOC (blue/orange) sequences deposited on a monthly basis worldwide (C), in the United States (D), and in California (E). (**F** to **H**) Total number of B.1.427/B.1.429 (F), B.1.429 (G), and B.1.427 (H) sequences deposited by country as of 30 April 2021. Only countries with two or more deposited sequences are shown.

and in other US states, and more recently in 34 additional countries worldwide (Fig. 1, A to H, and table S1). The number of B.1.427/B.1.429 genome sequences deposited increased rapidly after December 2020, with an incidence exceeding 50% in California beginning in February 2021 (Fig. 1, B to E). Collectively, this analysis illustrates the increased incidence of the B.1.427/B.1.429 VOC and its progressive geographical spread from California to other US states and other countries, which is consistent with a recent study suggesting enhanced transmissibility relative to the ancestral isolate (*42*).

B.1.427/ B.1.429 S reduces sensitivity to vaccine-elicited Abs

To assess the impact of the three mutations present in the B.1.427/B.1.429 S glycoprotein on neutralization, we first compared side by side the neutralization potency of mRNA vaccine–elicited Abs against the G614 S and B.1.427/B.1.429 S pseudoviruses. We used plasma from 15 individuals who received two doses of the Moderna mRNA-1273 vaccine and from 15 individuals who received two doses of Pfizer/BioNtech BNT162b2 vaccine collected between 7 and 27 days after booster immunization (table S2). All vaccinees had substantial plasma-neutralizing activity against G614 SARS-CoV-2 S pseudotyped viruses. Using a murine leukemia virus (MLV) pseudotyping system, geometric mean titers (GMTs) showed

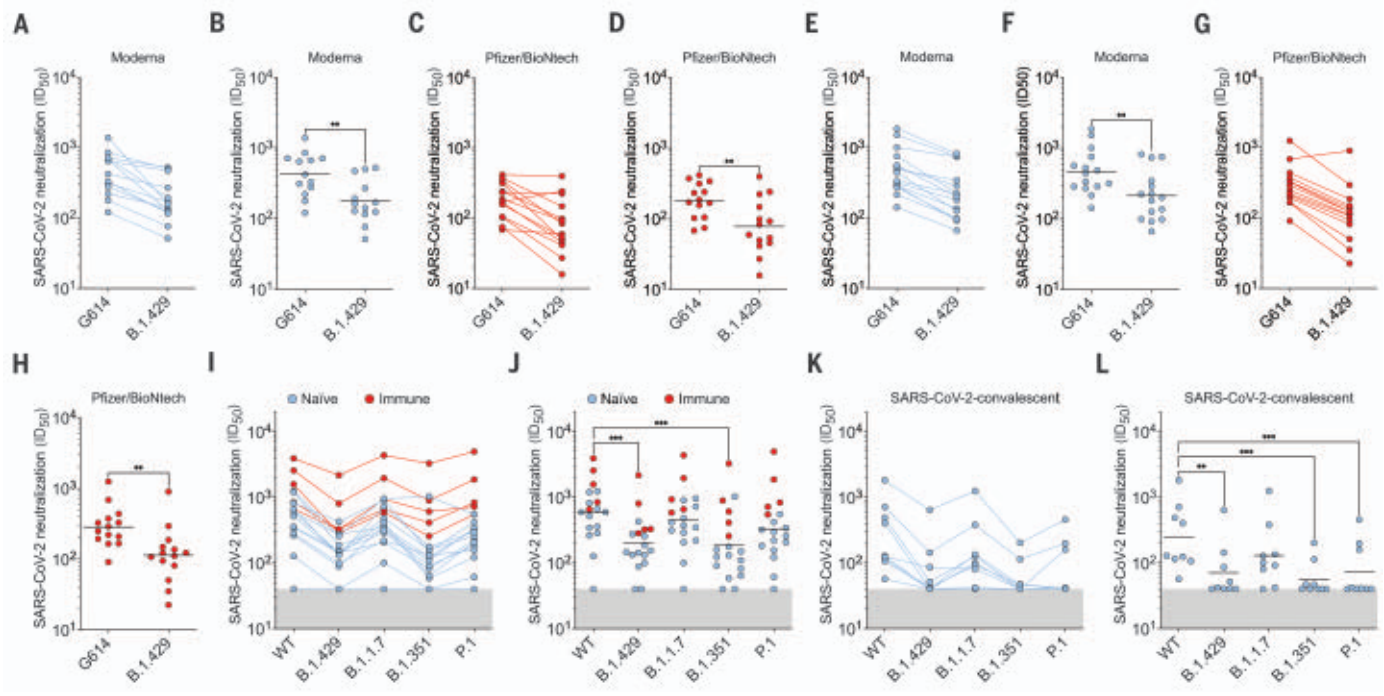


Fig. 2. B.1.427/B.1.429 S pseudotyped virus neutralization by vaccine-elicited and COVID-19 convalescent plasma. (**A**, **B**, **E**, and **F**) Neutralizing Ab titers [mean inhibition dilution (ID_{50})] shown as pairwise connected [(A) and (E)] or GMT [(B) and (F)] against MLV [(A) and (B)] or VSV [(E) and (F)] pseudotyped viruses harboring G614 SARS-CoV-2 S or B.1.427/B.1.429 (B.1.429) S determined using plasma from individuals who received two doses of the Moderna mRNA-1273 vaccine (blue). (**C**, **D**, **G**, and **H**) Neutralizing Ab titers (ID_{50}) shown as pairwise connected [(C) and (G)] or GMT [(D) and (H)] against MLV [(C) and (D)] or VSV [(G) and (H)] pseudotyped viruses harboring G614 SARS-CoV-2 S or B.1.427/B.1.429 (B.1.429) S determined using plasma from individuals who received two doses of the Pfizer/BioNtech BNT162b2 mRNA vaccine (red). (**I** and **J**) Neutralizing Ab ID_{50} (I) and GMT (J) titers against VSV pseudotyped viruses harboring D614 SARS-CoV-2 S, B.1.427/B.1.429 S, B.1.1.7 S, B.1.351 S, or P.1 S determined using plasma from naïve (blue) and previously infected (red) individuals who received two doses of the Pfizer/BioNtech BNT162b2 mRNA vaccine. "Naïve" indicates vaccinated individuals who had not been previously infected with SARS-CoV-2; "immune" refers to vaccinated individuals who had been previously infected with SARS-CoV-2. (**K** and **L**) Neutralizing Ab ID_{50} (K) and GMT (L) titers against VSV pseudotyped viruses harboring D614 SARS-CoV-2 S, B.1.427/B.1.429 S, B.1.1.7 S, B.1.351 S, or P.1 S determined using plasma from convalescent individuals who were infected with WT SARS-CoV-2. Neutralization data shown in (A) to (H) and (I) to (L) were performed using 293T-ACE2 and VeroE6-TMPRSS2, respectively. Data are average of $n = 2$ replicates.

that the average neutralization potency of the Moderna mRNA1273–elicited plasma was reduced 2.4-fold for B.1.427/B.1.429 S (GMT: 178) compared with G614 S (GMT: 424) (Fig. 2, A and B; figs. S1 and S2; and table S3), whereas it was reduced 2.3-fold with Pfizer/BioNtech BNT162b2–elicited plasma (B.1.427/B.1.429 GMT: 78 versus G614 GMT: 182) (Fig. 2, C and D; figs. S1 and S2; and table S3). Using a vesicular stomatitis virus (VSV) pseudotyping system, we observed a 2.2-fold average reduction of Moderna mRNA1273–elicited plasma-neutralizing activity against B.1.427/B.1.429 S (GMT: 213) compared with G614 S (GMT: 464) pseudoviruses (Fig. 2, E and F; figs. S1 and S2; and table S3) and a 2.5-fold average reduction of Pfizer/BioNtech BNT162b2–elicited plasma-neutralizing activity against B.1.427/B.1.429 S (GMT: 113) compared with G614 S (GMT: 285) pseudoviruses (Fig. 2, G and H; figs. S1 and S2; and table S3). We also analyzed plasma from 18 individuals, five of whom were previously infected with wild-type (WT) SARS-CoV-2, who received two doses of the Pfizer/BioNtech BNT162b2 vaccine and whose samples were collected between 14 and 28 days after booster immunization. We compared the neutralization potency of Pfizer/BioNtech BNT162b2 vaccine–elicited Abs against D614 S, B.1.427/B.1.429 S, B.1.1.7 S, B.1.351 S, and P.1 S VSV pseudotyped viruses using Vero E6 expressing TMPRSS2 as target cells. GMT plasma neutralization potency was reduced 2.9-fold for B.1.427/B.1.429 S (GMT: 197) compared with D614 S (GMT: 570), which is a decrease comparable to that observed with B.1.351 (GMT: 180, 3.2-fold reduction) and greater than that observed with B.1.1.7 and P.1 (GMT: 450 and 330, 1.3-fold and 1.7-fold reduction, respectively) pseudotyped viruses (Fig. 2, I and J; figs. S1 and S2; and table S3). These data indicate that the three B.1.427/B.1.429 S residue substitutions led to a modest but significant reduction of neutralization potency from vaccine-elicited Abs.

We also analyzed plasma from nine convalescent donors who experienced symptomatic COVID-19 in early 2020 (and consequently were likely exposed to the Wuhan-1 or a closely related SARS-CoV-2 isolate) collected 15 to 28 days after symptom onset (table S2). The neutralization potency of plasma from the nine convalescent donors was reduced 3.4-fold for B.1.427/B.1.429 S (GMT: 70) compared with G614 S (GMT: 240), similar to what we observed with B.1.351 (4.4-fold, GMT: 55) and P.1 (3.3-fold, GMT: 72) pseudotyped viruses, whereas neutralization of B.1.1.7 was less affected (1.9-fold, GMT: 127) (Fig. 2, K and L; figs. S1 and S2; and table S3). In several cases, the level of neutralizing activity against the VOC was found to be below the limit of detection.

These findings show that the three mutations present in the B1.427/B.1.429 S glycoprotein decrease the neutralizing activity of vaccine-elicited and infection-elicited Abs,

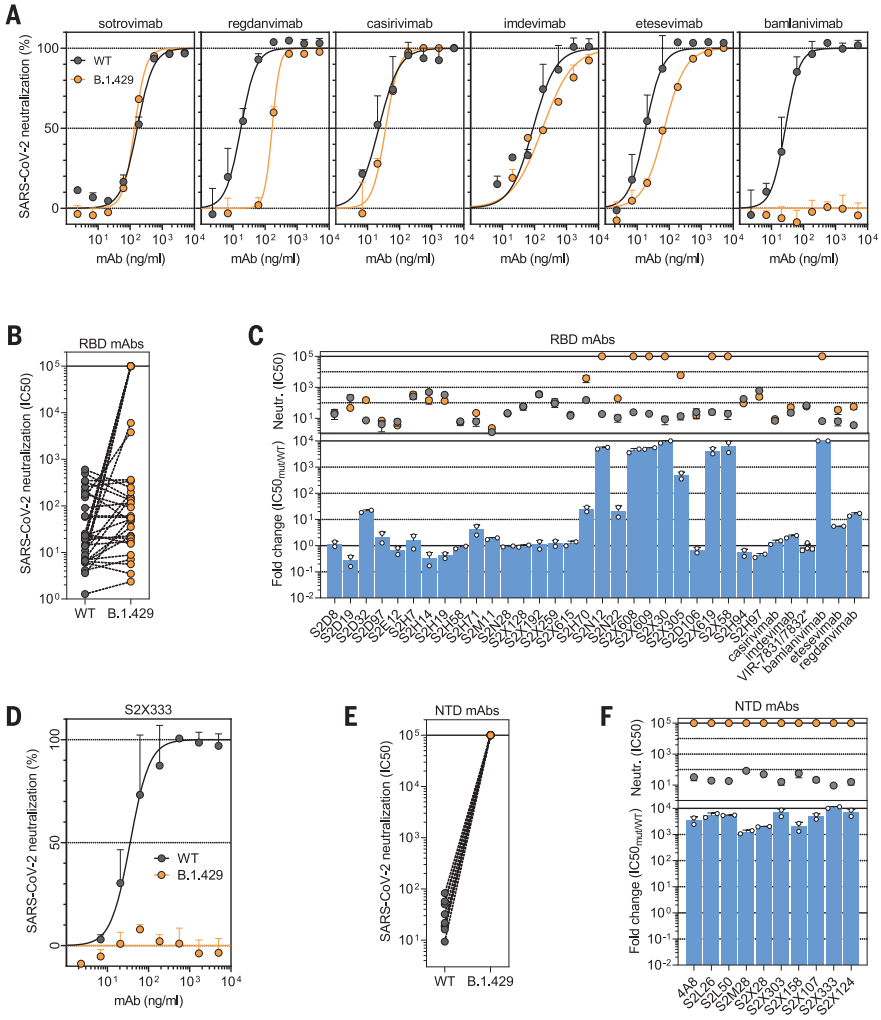


Fig. 3. Neutralization by a panel of RBD- and NTD-specific mAbs against SARS-CoV-2 D614 S and B.1.427/B.1.429 S pseudoviruses. (**A** and **D**) Neutralization of SARS-CoV-2 pseudotyped VSV carrying D614 (gray) or B.1.427/B.1.429 (orange) S protein by clinical stage RBD mAbs (A) and an NTD-targeting mAb (S2X333) (D). Data are representative of $n = 2$ replicates. (**B** and **E**) Neutralization of SARS-CoV-2 S VSV pseudotypes carrying D614 or B.1.427/B.1.429 S by 34 mAbs targeting the RBD (B) and 10 mAbs targeting the NTD (E). Data are the mean of 50% inhibitory concentration (IC_{50}) values (in nanograms per milliliter) of $n = 2$ independent experiments. Non-neutralizing IC_{50} titers were set at 10^5 ng/ml. (**C** and **F**) Neutralization by RBD-specific (C) and NTD-specific (F) mAbs shown as mean IC_{50} values (top) and mean fold change (bottom) for B.1.427/B.1.429 S (orange) relative to D614 S (gray) VSV pseudoviruses. VIR-7831 is a derivative of S309 mAb (sotrovimab). *VIR-7832 (variant of VIR-7831 carrying the LS-GAALIE Fc mutations) shown as squares. Non-neutralizing IC_{50} titers and fold change were set to 10^5 ng/ml and 10^4, respectively.

suggesting that these lineage-defining residue substitutions are associated with immune evasion. However, these data also underscore the higher quality of Ab responses induced by vaccination compared with infection and their enhanced resilience to mutations found in VOC.

B.1.427/B.1.429 S mutations reduce sensitivity to RBD- and NTD-specific Abs

To evaluate the contribution of RBD and NTD substitutions to the reduced neutralization potency of sera from vaccinees and convalescent plasma, we compared the neutralizing activity of 34 RBD and 10 NTD mAbs against the D614 S or B.1.427/B.1.429 S variant using a VSV pseudotyping system (*1*, *43*).

The panel of RBD-specific mAbs (including six clinical mAbs) recognizes distinct antigenic sites, as previously characterized (*10*, *11*, *13*, *20*, *44*, *45*). Briefly, epitopes span the receptor-binding motif (antigenic sites Ia and Ib), a cryptic antigenic site II, the exposed N343 glycan–containing antigenic site IV, and a second cryptic antigenic site V (*10*, *11*). A total of 14 out of 34 mAbs showed a reduced neutralization potency when comparing B.1.427/B.1.429 S and D614 S pseudoviruses (Fig. 3, A to C, and fig. S3). Of the six mAbs in clinical use, regdanvimab (CT-P59), and to a smaller extent etesevimab (LY-CoV016), showed a reduction in neutralization potency, whereas bamlanivimab (LY-CoV555) entirely lost its neutralizing activity. Neutralization mediated by the casirivimab/imdevimab mAb cocktail (REGN10933 and REGN10987) (*14*, *15*) and by VIR-7831 (a derivative of S309, recently renamed sotrovimab) (*10*, *23*, *24*) was unaffected by the B.1.427/B.1.429 S variant. To address the role of the L452R mutation in the neutralization escape from RBD-specific Abs, we tested the binding of the 34 RBD-specific mAbs to WT and L452R mutant RBD using biolayer interferometry (fig. S4). The 10 RBD-specific mAbs experiencing a 10-fold or greater reduction in neutralization potency of the B.1.427/B.1.429 variant relative to D614 S bound poorly to the L452R RBD mutant, demonstrating a direct role of this mutation in immune evasion.

The neutralizing activity of all 10 NTD-specific mAbs tested was abolished as a result of the presence of the S13I and W152C mutations (Fig. 3, D to F). These data indicate that the decreased potency of neutralization of the B.1.427/B.1.429 variant results from evasion of both RBD- and NTD-specific mAb-mediated neutralization.

Structural characterization of the SARS-CoV-2 B.1.427/B.1.429 S trimer

To visualize the changes in SARS-CoV-2 B.1.427/B.1.429 S that contribute to immune evasion, we determined the cryo–electron microscopy (cryoEM) structure of the variant S ectodomain trimer [carrying the HexaPro mutations (*46*)] bound to the RBD-specific mAb S2M11 and the NTD-specific mAb S2L20 at 2.3-Å resolution (Fig. 4A, fig. S5, and table S4). S2M11 was used to lock the RBDs in the closed state (Fig. 4B), and S2L20 was used to stabilize the NTDs (Fig. 4C) (*12*, *13*). Superimposing the regdanvimab (CT-P59)–bound and bamlanivimab (LY-CoV555)–bound SARS-CoV-2 RBD structures to B.1.427/B.1.429 S revealed that the introduced

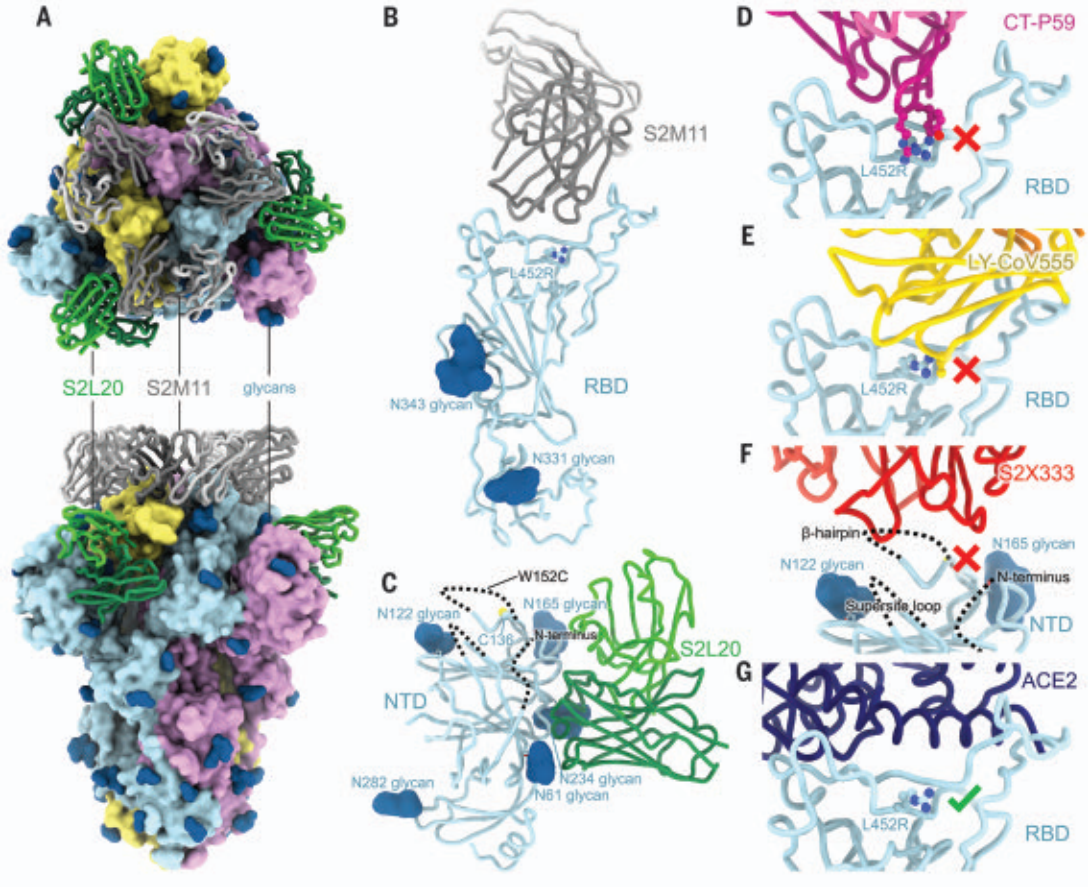


Fig. 4. CryoEM structure of the SARS-CoV-2 B.1.427/B.1.429 S ectodomain trimer. (**A**) Structure of the S trimer (surface rendering) bound to the S2M11 and S2L20 Fabs (ribbons) in two orthogonal orientations. SARS-CoV-2 S protomers are colored pink, cyan, and gold, and the S2L20 Fab heavy and light chains are colored dark and light green, respectively, and the S2M11 Fab heavy and light chains are colored dark and light gray, respectively. Only the Fab-variable domains are resolved in the map. N-linked glycans are rendered as dark blue spheres. (**B**) Magnified view of the S2M11-bound RBD with R452 shown in ball and stick representation. (**C**) Magnified view of the S2L20-bound NTD with disordered N terminus, supersite β-hairpin, and loop regions shown as dashed lines. (**D**) Superimposition of the CT-P59–bound SARS-CoV-2 RBD structure (PDB 7CM4) on the SARS-CoV-2 B.1.427/B.1.429 S cryoEM structure showing that R452 would sterically clash with the mAb. (**E**) Superimposition of the LY-CoV555–bound SARS-CoV-2 RBD structure (PDB 7KMG) on the SARS-CoV-2 B.1.427/B.1.429 S cryoEM structure showing that L452R would sterically clash with the mAb. (**F**) Superimposition of the S2X333-bound SARS-CoV-2 S structure (PDB 7LXW) on the SARS-CoV-2 B.1.427/B.1.429 S cryoEM structure showing that most of the NTD antigenic supersite epitope residues are disordered. (**G**) Superimposition of the ACE2-bound SARS-CoV-2 RBD structure (PDB 7DMU) on the SARS-CoV-2 B.1.427/B.1.429 S cryoEM structure showing that L452R points away from the interface with ACE2.

L452R is sterically incompatible with binding of these mAbs (Fig. 4, D and E), rationalizing the dampening or loss of neutralizing activity.

We subsequently used local refinement to account for the conformational dynamics of the NTD and S2L20 relative to the rest of S and obtained a cryoEM reconstruction of the NTD bound to S2L20 at 3.0-Å resolution (Fig. 4C, fig. S5, and table S4). The structure revealed that the B.1.427/B.1.429 NTD antigenic supersite is severely altered. The N terminus is disordered up to residue 27, as is the supersite β-hairpin (disordered between residues 137 and 158) and the supersite loop (disordered between residues 243 and 264) (Fig. 4F). These structural changes explain the abrogation of binding and neutralization of the panel of NTD-specific mAbs evaluated.

Overlaying an ACE2-bound SARS-CoV-2 RBD structure with the B.1.427/B.1.429 variant S structure showed that the R452 residue points away from and does not contact ACE2, suggesting that this substitution would not affect receptor engagement (Fig. 4G). We next evaluated binding of the monomeric human ACE2 ectodomain to immobilized B.1.427/B.1.429 and WT RBDs using surface plasmon resonance (fig. S6, A and B, and table S5); biolayer interferometry (fig. S6, C to E, and table S5); and binding of B.1.427/B.1.429, B.1.1.7, and WT RBDs to immobilized human ACE2 by enzyme-linked immunosorbent assay (ELISA) (fig. S6F and table SS5). Our results indicate that the B.1.427/B.1.429 and WT RBDs bound to ACE2 with comparable affinities, whereas the B.1.1.7 RBD had a markedly increased affinity for ACE2 (*34*), validating the structural observations.

Disulfide bond rearrangement in the B.1.427/B.1.429 variant NTD antigenic supersite

To investigate further the molecular basis for the loss of NTD-directed mAb-neutralizing activity and structural changes in the NTD, we analyzed the binding of a panel of NTD-specific mAbs to recombinant SARS-CoV-2 NTD variants using ELISA. The S13I signal peptide mutation dampened binding of five mAbs and abrogated binding of five additional mAbs of the 11 neutralizing mAbs evaluated (Fig. 5A and fig. S7). Furthermore, the W152C mutation reduced recognition

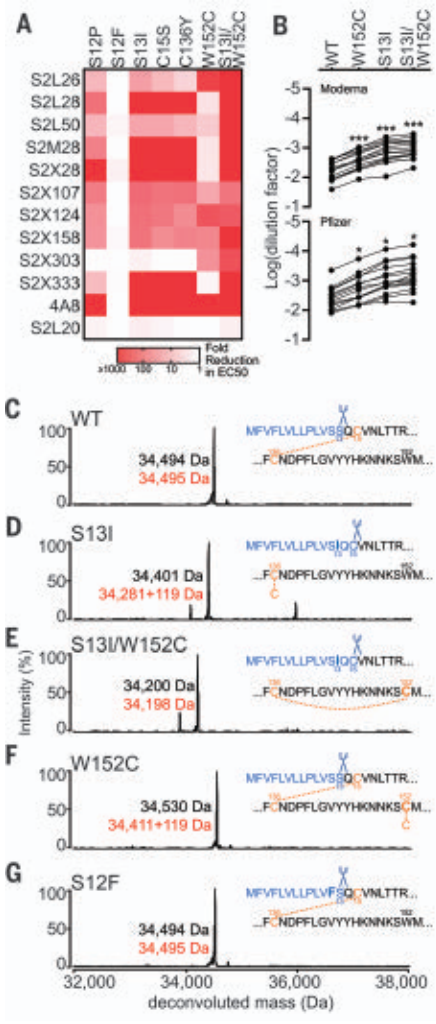


Fig. 5. The B.1.427/B.1.429 S S13I and W152C mutations lead to immune evasion. (**A**) Binding of a panel of 11 neutralizing (antigenic site i) and one non-neutralizing (antigenic site iv) NTD-specific mAbs to recombinant SARS-CoV-2 NTD variants analyzed by ELISA displayed as a heatmap. (**B**) Binding of plasma Abs from vaccinated individuals to recombinant SARS-CoV-2 NTD variants analyzed by ELISA. The mean dilution factor for each mutant was compared by the one-way ANOVA test against WT (*P < 0.05, **P < 0.001). (**C** to **G**) Deconvoluted mass spectra of purified NTD constructs, including the WT NTD with the native signal peptide (C), the S13I NTD (D), the S13I and W152C NTD (E), the W152C NTD (F), and the S12F NTD (G). The empirical mass (black) and theoretical mass (red) are shown beside the corresponding peak. An additional 119 Da were observed for the S13I and W152C NTDs, corresponding to cysteinylation of the free cysteine residue in these constructs (as L-cysteine was present in the expression media). The cleaved signal peptide (blue text) and subsequent residue sequence (black text) are also shown based on the MS results. Mutated residues are shown in bold. Cysteines are highlighted in light orange (unless in the cleaved signal peptide), and disulfide bonds are shown as dotted light orange lines between cysteines. Residues are numbered for reference.

of six NTD-neutralizing mAbs, including a complete loss of binding for two of them, with a pattern complementary to that observed for S13I (Fig. 5A and fig. S7). The B.1.427/B.1.429 S13I/W152C NTD did not bind to any NTD-directed neutralizing mAbs, which are known to target a single antigenic site (antigenic site I) (*12*), whereas binding of the non-neutralizing S2L20 mAb to the NTD antigenic site IV was not affected by any mutants, confirming proper retention of folding, as supported by the structural data (Fig. 5A and fig. S7). Binding of vaccine-elicited plasma to NTD mutants confirmed and extended these observations with polyclonal Abs by showing an increasingly marked reduction in binding titers caused by the W152C, S13I, and S13I/W152C residue substitutions (Fig. 5B and fig. S8).

We previously showed that disruption of the C15/C136 disulfide bond that connects the N terminus to the rest of the NTD by mutation of either residue or alteration of the signal peptide cleavage site abrogates the neutralizing activity of mAbs targeting the NTD antigenic supersite (site I) (*12*). Because the S13I substitution resides in the signal peptide and is predicted to shift the signal peptide cleavage site from S13-Q14 to C15-V16, we hypothesized that this substitution indirectly affects the integrity of NTD antigenic site I, which comprises the N terminus. Mass spectrometry analysis of the S13I and S13I/W152C NTD variants confirmed that signal peptide cleavage occurs immediately after residue C15 (Fig. 5, C to E). As a result, C136, which would otherwise be disulfide linked to C15, is cysteinylated in the S13I NTD because of the presence of free cysteine in the expression medium (Fig. 5D and fig. S9). Likewise, the W152C mutation, which introduces a free cysteine, was also found to be cysteinylated in the W152C NTD (Fig. 5F). It is not clear whether cysteinylation would occur during natural infection with S13I or W152C mutants alone or what contribution cysteinylation plays in immune evasion of S13I or W152C mutants alone. Dampening of NTD-specific neutralizing mAb binding is stronger for the S13I mutant than for the S12P mutant, which we previously showed also shifts the signal peptide cleavage site to C15-V16 (Fig. 5A). Conversely, we did not observe any effect on mAb binding of the S12F substitution, which has also been detected in clinical isolates, in agreement with the fact that this mutation did not affect the native signal peptide cleavage site (i.e., it occurs at the S13-Q14 position), as observed by mass spectrometry (Fig. 5G). In the absence of the C15-C136 disulfide bond, the N terminus is no longer stapled to the NTD, consistent with the structural data showing that the N terminus of the B.1.427/B.1.429 variant becomes disordered relative to the rest of the NTD (Fig. 4C).

Although the S13I and W152C NTD variants were respectively cysteinylated at positions C136 and W152C, the double mutant S13I/W152C was not cysteinylated, suggesting that C136 and W152C had formed a new disulfide bond (Fig. 5, D to F). Tandem mass spectrometry analysis of nonreduced, digested peptides identified linked discontinuous peptides containing C136 and W152C (fig. S9), confirming that a disulfide bond forms between C136 and W152C in the S13I/W152C NTD of the B.1.427/B.1.429 variant. W152C is in the β-hairpin of the antigenic supersite, and the formation of a new disulfide bond with C136 would move residues in the β-hairpin >20 Å. The local structure of the β-hairpin was disordered in the B.1.427/B.1.429 variant (Fig. 4C).

Collectively, these findings demonstrate that the S13I and W152C mutations found in the B.1.427/B.1.429 S variant are jointly responsible for escape from NTD-specific mAbs because of deletion of the two SARS-CoV-2 S N-terminal residues and overall rearrangement of the NTD antigenic supersite. Our data support that the SARS-CoV-2 NTD evolved a compensatory mechanism to form an alternative disulfide bond and that mutations of the S signal peptide occur in vivo in a clinical setting to promote immune evasion. The SARS-CoV-2 B.1.427/B.1.429 S variant therefore relies on an indirect and unusual neutralization escape strategy.

Discussion

Serum- or plasma-neutralizing activity is a correlate of protection against SARS-CoV-2 challenge in nonhuman primates (*47*, *48*), and treatment with several neutralizing mAbs has reduced viral burden and decreased hospitalization and mortality in clinical trials (*10*, *14*, *15*, *22*, *23*, *49*). The observed L452R-mediated immune evasion of B.1.427/B.1.429 S is consistent with previous findings showing that this substitution reduced the binding or neutralizing activity of some mAbs before the discovery of the B.1.427/B.1.429 variant (*50*–*53*). The acquisition of the L452R substitution by multiple lineages across multiple

continents, including the B.1.617.1 and B.1.617.2 lineages emerging in India (*54*), is suggestive of positive selection, which might result from the selective pressure of RBD-specific neutralizing Abs (*55*).

The SARS-CoV-2 NTD undergoes rapid antigenic drift and accumulates a larger number of mutations and deletions relative to other regions of the S glycoprotein (*12*, *56*). For instance, the L18F substitution and the deletion of residue Y144 are found in 8% and 26% of viral genomes sequenced and are present in the B.1.351/P.1 lineages and the B.1.1.7 lineage, respectively. Both of these mutations are associated with reduction or abrogation of mAb binding and neutralization (*12*, *34*). The finding that multiple circulating SARS-CoV-2 variants map to the NTD, several of them in the antigenic supersite (site I), suggests that the NTD is subject to a strong selective pressure from the host humoral immune response. This is further supported by the identification of deletions within the NTD antigenic supersite in immunocompromised hosts with prolonged infections (*57*–*59*) and in the in vitro selection of SARS-CoV-2 S escape variants with NTD mutations that decrease the binding and neutralization potency of COVID-19 convalescent patient sera or mAbs (*12*, *34*, *60*, *61*). The data herein showing immune evasion of all tested NTD-specific mAbs by the B.1.427/B.1.429 variant also support that the NTD antigenic supersite is under host immune pressure.

Similar to how the S13I/W152C mutations facilitate evasion of all tested NTD-specific mAbs, E484K causes broad resistance to many RBD-specific mAbs. The independent acquisition of the E484K mutation in the B.1.351, P.1, and B.1.526 variants and, more recently, in the B.1.1.7 variant (*34*), suggests this could also occur in the B.1.427/B.1.429 lineages. Indeed, four genome sequences with the E484K RBD mutation in the B.1.427 variant have recently been deposited in GISAID. Alternatively, the S13I/W152C mutations could emerge in any of these variants. The S13I mutation was recently detected in the SARS-CoV-2 B.1.526 lineage, which was originally described in New York (*62*, *63*). Understanding the newly found mechanism of immune evasion in emerging SARS-CoV-2 variants, such as the signal peptide modification described herein, is as important as sequence surveillance itself to successfully counter the ongoing pandemic.

REFERENCES AND NOTES

1. A. C. Walls *et al*., *Cell* **181**, 281–292.e6 (2020).
2. D. Wrapp *et al*., *Science* **367**, 1260–1263 (2020).
3. A. C. Walls *et al*., *Nature* **531**, 114–117 (2016).
4. M. A. Tortorici, D. Veesler, *Adv. Virus Res.* **105**, 93–116 (2019).
5. M. Letko, A. Marzi, V. Munster, *Nat. Microbiol.* **5**, 562–569 (2020).
6. P. Zhou *et al*., *Nature* **588**, E6 (2020).
7. M. Hoffmann *et al*., *Cell* **181**, 271–280.e8 (2020).
8. W. T. Soh *et al*., bioRxiv 369264 [Preprint]. 5 November 2020; https://doi.org/10.1101/2020.11.05.369264.
9. S. Wang *et al*., *Cell Res.* **31**, 126–140 (2021).
10. D. Pinto *et al*., *Nature* **583**, 290–295 (2020).
11. L. Piccoli *et al*., *Cell* **183**, 1024–1042.e21 (2020).
12. M. McCallum *et al*., *Cell* **184**, 2332–2347.e16 (2021).
13. M. A. Tortorici *et al*., *Science* **370**, 950–957 (2020).
14. J. Hansen *et al*., *Science* **369**, 1010–1014 (2020).
15. A. Baum *et al*., *Science* **369**, 1014–1018 (2020).
16. N. Suryadevara *et al*., *Cell* **184**, 2316–2331.e15 (2021).
17. G. Cerutti *et al*., *Cell Host Microbe* **29**, 819–833.e7 (2021).
18. X. Chi *et al*., *Science* **369**, 650–655 (2020).
19. Z. Wang *et al*., *Nature* **592**, 616–622 (2021).
20. C. O. Barnes *et al*., *Nature* **588**, 682–687 (2020).
21. D. F. Robbiani *et al*., *Nature* **584**, 437–442 (2020).
22. B. E. Jones *et al*., *Sci. Transl. Med.* **13**, eabf1906 (2020).
23. A. L. Cathcart *et al*., bioRxiv 434607 [Preprint]. 24 June 2021; https://doi.org/10.1101/2021.03.09.434607.
24. D. Corti, L. A. Purcell, G. Snell, D. Veesler, *Cell* **184**, 3086–3108 (2021).
25. C. Wang *et al*., *Nat. Commun.* **12**, 1715 (2021).
26. G. Song *et al*., *Nat. Commun.* **12**, 2938 (2021).
27. D. Pinto *et al*., bioRxiv 442808 [Preprint]. 10 May 2021; https://doi.org/10.1101/2021.05.09.442808.
28. M. M. Sauer *et al*., *Nat. Struct. Mol. Biol.* **28**, 478–486 (2021).
29. B. Korber *et al*., *Cell* **182**, 812–827.e19 (2020).
30. J. A. Plante *et al*., *Nature* **592**, 116–121 (2021).
31. H. Tegally *et al*., *Nature* **592**, 438–443 (2021).
32. N. R. Faria *et al*., on behalf of CADDE Genomic Network, "Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings" (virological.org, 2021); https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586.
33. N. G. Davies *et al*., *Science* **372**, eabg3055 (2021).
34. D. A. Collier *et al*., *Nature* **593**, 136–141 (2021).
35. P. Wang *et al*., *Nature* **593**, 130–135 (2021).
36. S. A. Madhi *et al*., *N. Engl. J. Med.* **384**, 1885–1898 (2021).
37. S. Cele *et al*., *Nature* **593**, 142–146 (2021).
38. D. Planas *et al*., *Nat. Med.* **27**, 917–924 (2021).
39. C. K. Wibmer *et al*., *Nat. Med.* **27**, 622–625 (2021).
40. X. Xie *et al*., *Nat. Med.* **27**, 620–621 (2021).
41. A. Rambaut *et al*., *Nat. Microbiol.* **5**, 1403–1407 (2020).
42. X. Deng *et al*., *Cell* **184**, 3426–3437.e8 (2021).
43. J. K. Millet, G. R. Whittaker, *Bio Protoc.* **6**, e2035 (2016).
44. T. N. Starr *et al*., bioRxiv 438709 [Preprint]. 8 April 2021.
45. M. A. Tortorici *et al*., bioRxiv 438818 [Preprint]. 8 April 2021; https://doi.org/10.1101/2021.04.07.438818.
46. C. L. Hsieh *et al*., *Science* **369**, 1501–1505 (2020).
47. K. McMahan *et al*., *Nature* **590**, 630–634 (2021).
48. P. S. Arunachalam *et al*., *Nature* **594**, 253–258 (2021).
49. A. Baum *et al*., *Science* **370**, 1110–1115 (2020).
50. Z. Liu *et al*., *Cell Host Microbe* **29**, 477–488.e4 (2021).
51. Q. Li *et al*., *Cell* **182**, 1284–1294.e9 (2020).
52. T. N. Starr, A. J. Greaney, A. S. Dingens, J. D. Bloom, *Cell Rep. Med.* **2**, 100255 (2021).
53. T. N. Starr *et al*., *Cell* **182**, 1295–1310.e20 (2020).
54. V.-V. Edara *et al*., bioRxiv 443299 [Preprint]. 10 May 2021; https://doi.org/10.1101/2021.05.09.443299.
55. V. Tchesnokova *et al*., bioRxiv 432189 [Preprint]. 11 March 2021 https://doi.org/10.1101/2021.02.22.432189.
56. E. C. Thomson *et al*., *Cell* **184**, 1171–1187.e20 (2021).
57. V. A. Avanzato *et al*., *Cell* **183**, 1901–1912.e9 (2020).
58. B. Choi *et al*., *N. Engl. J. Med.* **383**, 2291–2293 (2020).
59. K. R. McCarthy *et al*., *Science* **371**, 1139–1142 (2021).
60. E. Andreano *et al*., bioRxiv 424451 [Preprint]. 28 December 2020; https://doi.org/10.1101/2020.12.28.424451.
61. Y. Weisblum *et al*., *eLife* **9**, e61312 (2020); https://doi.org/10.1101/2020.12.28.424451.
62. M. K. Annavajhala *et al*., medRxiv 2021.02.23.21252259 [Preprint]. 12 June 2021; https://doi.org/10.1101/2021.02.23.21252259.
63. A. P. West Jr., C. O. Barnes, Z. Yang, P. J. Bjorkman, SARS-CoV-2 lineage B.1.526 emerging in the New York region detected by software utility created to query the spike mutational landscape. bioRxiv 431043 [Preprint]. 22 April 2021; https://doi.org/10.1101/2021.02.14.431043.

ACKNOWLEDGMENTS

We thank H. Tani (University of Toyama) for providing the reagents necessary for preparing VSV pseudotyped viruses. **Funding:** This study was supported by the National Institute of Allergy and Infectious Diseases (grant DP1AI158186 and HHSN272201700059C to D.V. and grant U01 AI151698-01 to W.C.V.V.), a Pew Biomedical Scholars Award (D.V.), Investigators in the Pathogenesis of Infectious Disease Awards from the Burroughs Wellcome Fund (D.V.), Fast Grants (D.V.), the Natural Sciences and Engineering Research Council of Canada (M.M.), and the Pasteur Institute (M.A.T.). **Author contributions:** L.P., D.C., and D.V. conceived the study. M.M., J.B., A.D.M., A.C., A.C.W., J.D.I., and M.A.T. designed the study and the experiments. M.M., E.C., and K.C. performed mutagenesis for mutant expression plasmids. M.M., J.E.B., E.C., and S.J. performed mutant expression. S.B.G., G.B., A.F.P., C.G., S.T., and W.V. contributed to donor recruitment and plasma sample collection. A.C.W., M.A.T., M.J.N., J.B., A.D.M., D.P., C.S., and C.S.-F. produced pseudoviruses and performed pseudovirus neutralization assays. J.d.I. and A.T. performed bioinformatic analysis. M.M., J.B., A.D.M., A.C.W., L.E.R., G.S., L.P., D.C., and D.V. analyzed the data and prepared the manuscript with input from all authors. M.S.P., L.P., G.S., H.W.V., D.C., and D.V. supervised the study. **Competing interests:** A.D.M., J.B., A.C., J.D.I., C.S.-F., C.S., M.A., D.P., K.C., S.B., S.J., E.C., M.S.P., L.E.R., G.S., A.T., H.W.V., L.P., and D.C. are employees of and may hold shares in Vir Biotechnology, Inc. D.C. is currently listed as an inventor on multiple patent applications that disclose the subject matter described in this manuscript. H.W.V. is a founder of PierianDx and Casma Therapeutics. Neither company provided funding for this work or is performing related work. D.V. is a consultant for Vir Biotechnology, Inc. The Veesler laboratory has received a sponsored research agreement from Vir Biotechnology, Inc. The remaining authors declare no conflicts of interest. **Data and materials availability:** The cryoEM map and coordinates have been deposited to the Electron Microscopy Databank and Protein Data Bank with accession numbers EMD-24236, EMD-24237, PDB 7N8H, and PDB 7N8I (see table S4 for details). Materials generated in this study are available from the corresponding authors upon request but may require a completed Materials Transfer Agreement signed with Vir Biotechnology, Inc. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6555/648/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S6
References (*64*–*90*)
MDAR Reproducibility Checklist

31 March 2021; accepted 25 June 2021
Published online 1 July 2021
10.1126/science.abi7994

PLANT GENOMICS

De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes

Matthew B. Hufford[1], Arun S. Seetharam[1,2], Margaret R. Woodhouse[3], Kapeel M. Chougule[4], Shujun Ou[1], Jianing Liu[5], William A. Ricci[6], Tingting Guo[7], Andrew Olson[4], Yinjie Qiu[8], Rafael Della Coletta[8], Silas Tittes[9,10], Asher I. Hudson[9,10], Alexandre P. Marand[5], Sharon Wei[4], Zhenyuan Lu[4], Bo Wang[4], Marcela K. Tello-Ruiz[4], Rebecca D. Piri[11], Na Wang[6], Dong won Kim[6], Yibing Zeng[5], Christine H. O'Connor[8,12], Xianran Li[7], Amanda M. Gilbert[8], Erin Baggs[13], Ksenia V. Krasileva[13], John L. Portwood II[3], Ethalinda K. S. Cannon[3], Carson M. Andorf[3], Nancy Manchanda[1], Samantha J. Snodgrass[1], David E. Hufnagel[1,14], Qiuhan Jiang[1], Sarah Pedersen[1], Michael L. Syring[1], David A. Kudrna[15], Victor Llaca[16], Kevin Fengler[16], Robert J. Schmitz[5], Jeffrey Ross-Ibarra[9,10,17], Jianming Yu[7], Jonathan I. Gent[6], Candice N. Hirsch[8], Doreen Ware[18,4], R. Kelly Dawe[5,6,11]*

We report de novo genome assemblies, transcriptomes, annotations, and methylomes for the 26 inbreds that serve as the founders for the maize nested association mapping population. The number of pan-genes in these diverse genomes exceeds 103,000, with approximately a third found across all genotypes. The results demonstrate that the ancient tetraploid character of maize continues to degrade by fractionation to the present day. Excellent contiguity over repeat arrays and complete annotation of centromeres revealed additional variation in major cytological landmarks. We show that combining structural variation with single-nucleotide polymorphisms can improve the power of quantitative mapping studies. We also document variation at the level of DNA methylation and demonstrate that unmethylated regions are enriched for cis-regulatory elements that contribute to phenotypic variation.

Maize is the most widely planted crop in the world and an important model system for the study of gene function. The species is known for its extreme genetic diversity, which has allowed for broad adaptation throughout the tropics and intensive use in temperate regions. Nevertheless, most current genomic resources are referenced to a single inbred, B73, which contains only 63 to 74% of the genes and/or low-copy sequences in the full maize pan-genome (*1*–*4*). Moreover, there is extensive structural polymorphism in noncoding and regulatory genomic regions that has been shown to contribute to variation in numerous traits (*5*). In recent years, additional maize genomes have been assembled, which has allowed limited characterization of the species' pan-genome (*2*, *6*–*10*). However, comparisons across genome projects are often confounded by differences in assembly and annotation methods.

The maize nested association mapping (NAM) population was developed to study the genetic architecture of quantitative traits (*11*). Twenty-five founder inbred lines were strategically selected from a larger association panel (*12*) to represent the breadth of maize diversity, including lines from the non–stiff-stalk temperate heterotic group; lines from tropical and subtropical regions of Africa, Asia, and the Americas; and both sweet corn and popcorn germplasm (*13*). Each NAM parental inbred was crossed to B73 and selfed to generate 25 populations of 200 recombinant inbred lines that combine the advantages of linkage and association mapping for important agronomic traits (*14*). Biological infrastructure continues to be developed around these lines [e.g., (*15*, *16*)], but comprehensive genomic resources are needed to fully realize the power of the NAM population.

Consistency and quality of genome assemblies

Here, we describe assembled and annotated genomes for the 25 NAM founder inbreds and an improved reference assembly of B73 (table S1). The 26 genomes were sequenced to high depth (63–85×) by PacBio long-read technology, assembled into contigs by a hybrid approach (*17*), scaffolded by Bionano optical maps, and ordered into pseudomolecules by using linkage data from the NAM recombinant inbred lines and maize pan-genome anchor markers (*4*). Assembly and annotation statistics improve upon nearly all available maize assemblies, including the previous B73 reference genome (*18*), with the total length of placed scaffolds (2.102 to 2.162 billion base pairs) at the estimated genome size of maize, a mean scaffold N50 of 119.2 Mb [contig N50 of 25.7 million base pairs (Mbp)], complete gene space [mean of 96% complete benchmarking universal single-copy orthologs (BUSCO)] (*19*), and, on the basis of the LTR Assembly Index (mean of 28) (*20*), full assembly of the transposable element (TE)–laden portions of the genome (Table 1 and table S2). Improvements in contiguity and completeness can be attributed to recent advances in sequence and optical map data, as well as more-effective assembly algorithms (*21*).

Gene identification and diversity in gene content

We sequenced mRNA from 10 tissues for each inbred. These data were used for evidence-based gene annotation of each line, which was then improved by using B73 full-length cDNA and expressed sequence tags. The evidence set was augmented with ab initio gene models and the gene structures refined for all accessions through phylogeny-based methods. This pipeline revealed an average of 40,621 (SE = 117) protein-coding and 4998 (SE = 100) noncoding gene models per genome. Most genes share orthologs with the grass (Poaceae) family and species in the Andropogoneae tribe of grasses, which includes maize and sorghum (Fig. 1A). The accuracy of the annotations, which was measured by the congruence between annotations and supporting evidence (annotation edit distance) (*22*), is higher than that of previous reference maize annotations (fig. S1) (*2*, *6*, *10*, *18*, *23*).

We next assessed the gene catalog of the pan-genome. Genes with high sequence similarity, located within blocks of homologous sequence in pairwise comparisons, were grouped together as one pan-gene. In many instances, a gene was not annotated by our computational pipeline, yet at least 90% of the gene was present in the correct homologous location; when this occurred, the pan-gene was considered present (fig. S2A) (*17*), even though in some cases, the absence of

[1]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA. [2]Genome Informatics Facility, Iowa State University, Ames, IA 50011, USA. [3]USDA-ARS Corn Insects and Crop Genetics Research Unit, Iowa State University, Ames, IA 50011, USA. [4]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. [5]Department of Genetics, University of Georgia, Athens, GA 30602, USA. [6]Department of Plant Biology, University of Georgia, Athens, GA 30602, USA. [7]Department of Agronomy, Iowa State University, Ames, IA 50011, USA. [8]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA. [9]Center for Population Biology, University of California, Davis, CA 95616, USA. [10]Department of Evolution and Ecology, University of California, Davis, CA 95616, USA. [11]Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA. [12]Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN 55108, USA. [13]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA. [14]Virus and Prion Research Unit, National Animal Disease Center, USDA-ARS, Ames, IA, 50010, USA. [15]Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA. [16]Corteva Agriscience, Johnston, IA 50131, USA. [17]Genome Center, University of California, Davis, CA 95616, USA. [18]USDA-ARS NAA Robert W. Holley Center for Agriculture and Health, Agricultural Research Service, Ithaca, NY 14853, USA.
*Corresponding author. Email: kdawe@uga.edu

Table 1. Quality metrics for genome assemblies and gene model annotations. Darker shading indicates higher quality. The NAM lines are shaded on the basis of their primary grouping (gold, stiff-stalk heterotic group; blue, non–stiff-stalk heterotic group; gray, mixed tropical-temperate ancestry; purple, popcorn; orange, sweet corn; green, tropical). Hp301 and P39 have the lowest amounts of TR-1 and subtelomere repeats, respectively. Our methods can overestimate assembly when repeats are in low abundance (*17*).

	B73_v4	B73_v5	B97	Ky21	M162W	Ms71	OH43	OH7B	M37W	Mo18W	Tx303	Hp301	P39	Il14H	CML52	CML69	CML103	CML228	CML247	CML277	CML322	CML333	KI3	KI11	NC350	NC358	Tx8
Assembly Size (Mb)	2134	2182	2193	2172	2184	2214	2177	2165	2192	2223	2216	2141	2139	2125	2308	2225	2162	2301	2215	2191	2219	2231	2216	2274	2291	2227	2271
Contig N50 (Mb)	1.2	52.36	49.77	19.07	27.81	34.1	28.63	13.62	39.62	24.98	27.97	35.6	35.78	19.64	11.2	21.34	11.34	9.553	11.43	6.255	30.48	28.82	16.18	31.4	49	25.84	11.61
Scaffold N50 (Mb)	10.69	160.85	137.68	115.38	111.38	96.45	105.60	140.13	105.37	111.10	99.16	135.87	147.88	135.80	92.05	107.57	129.92	108.07	101.10	96.85	102.20	99.84	107.93	110.07	100.66	96.95	100.58
Pseudomolecule % N	1.43	0.175	0.156	0.306	0.175	0.23	0.121	0.407	0.087	0.338	0.314	0.198	0.117	0.158	0.936	0.296	0.241	1.207	0.459	0.426	0.144	0.148	0.392	0.121	0.072	0.175	0.567
BUSCO (% complete)	95.70	95.76	95.69	96.04	96.04	95.97	95.76	95.76	95.97	96.60	95.83	95.63	95.76	95.63	95.76	95.90	95.56	96.25	96.32	96.18	95.42	96.32	96.87	95.83	96.18	96.18	96.11
LTR Assembly Index (LAI)	26.68	27.84	28.06	28.06	28.09	27.91	27.89	28.04	28.09	27.81	27.71	28.05	27.61	27.83	27.92	28.34	28.3	27.93	28.44	27.95	28.44	28.33	28.27	27.64	27.96	28.22	27.9
CentC (% assembled)	17.52	87.94	38.89	56.78	44.54	75.47	66.73	47.55	79.84	75.75	76.42	65.64	69.96	55.31	45.25	56.92	54.96	55.32	47.43	29.33	69.96	69.24	43.82	74.21	64.18	52.69	62.81
Knob180 (% assembled)	5.651	18.63	8.24	8.96	7.89	7.79	4.34	8.35	6.91	10.73	7.49	22	22.24	55.54	4.89	2.21	10.57	3.97	3.65	3.2	4.73	3.83	3.44	12.71	13.9	5.61	2.7
TR-1 (% assembled)	23.01	89.43	66.41	15.67	38.98	25.13	8.26	38.76	79.14	13.93	34.96	110.8	42.22	86.81	6.93	3.17	8.45	3.3	5.54	9.55	4.08	11.42	10.76	20.3	12.6	5.29	2.48
rDNA arrays (% assembled)	0.352	9.41	7.16	10.71	8.76	6.05	6.5	9.74	7.5	16.34	6.38	13.54	6.53	6.89	13.5	9.77	8.33	5.64	8.48	8.02	11.56	11.64	7.33	3.97	10.9	9.44	7.88
Subtelomere (% assembled)	1.963	90.31	69	52.2	73.43	60.75	60.06	48.66	70.63	70.47	68.71	65.19	116.5	97.04	19.65	85.34	34.06	85.15	52.76	32.95	94.6	70.27	86.82	91.63	88.51	64.07	83.64
Gene Length (average)	4163	4477	4403	4371	4403	4349	4408	4332	4377	4327	4278	4405	4232	4337	4477	4446	4436	4318	4564	4348	4280	4432	4439	4442	4445	4406	4382
Genic Space Annotated (%)	8.03	8.17	8.12	8.23	8.38	8.12	8.25	7.97	8.17	8.05	7.97	8.2	8.22	8.24	7.93	8.07	8.23	7.9	8.38	8.04	7.94	8.04	8.26	7.8	7.86	7.88	8.07

annotation may reflect fractionation and/or pseudogenization.

Across the 26 genomes, a total of 103,033 pan-genes were identified. Previous analysis reported ~63,000 pan-genes on the basis of transcriptome assemblies of seedling RNA sequencing (RNA-seq) reads from 500 individuals (*1*). The superior contiguity of our assemblies and the application of both ab initio and evidence-based annotation using RNA-seq from a diverse set of 10 tissues likely account for the increased sensitivity. More than 80% of pan-genes were identified within just 10 inbred lines on the basis of a bootstrap resampling of genomes (Fig. 1B). When considered separately, temperate and tropical lines have differentiated sets of pan-genes but show a comparable rate of pan-gene increase as lines are added, suggesting they have similar gene-content diversity (Fig. 1B).

Pan-genes, excluding tandem duplicates (*17*), were classified as core (present in all 26 lines), near-core (present in 24 to 25 lines), dispensable (present in 2 to 23 lines), and private (present in only 1 line) (Fig. 1C). The portion of genes classified into each of these groups was consistent across genotypes, with an average of 58.41% (SE = 0.07%) belonging to the core genome, 8.23% (SE = 0.05%) to the near-core genome, 31.75% (SE = 0.09%) to the dispensable genome, and 1.60% (SE = 0.08%) private genes (Fig. 1C; fig. S2, B and C; and table S3). In total, 32,052 genes are in the core or near-core portion of the pan-genome, and 70,981 are genes in the dispensable or private portion. The core genes (and gene families enriched for core genes) (table S4) are generally from higher phylostrata levels (i.e., Viridiplantae and Poaceae), whereas those in the near-core and dispensable sets either share orthologs only with closely related species or are maize specific (fig. S2E). Some private genes may be spurious annotations that result from imperfect masking of repeat sequences, as most core and near-core genes are syntenic to sorghum (57.78%), whereas this is rarely the case for dispensable and private genes (1.83% syntenic). Core genes were expressed in more tissues (Fig. 1D) and had higher transcript abundance (fig. S2F) when compared with genes present in fewer individuals. However, across the relatively small number of tissues (eight or more per line) profiled for this analysis, 18% of dispensable and 32% of private genes were expressed in at least one tissue. A total of 16,751 pan-genes were tandemly duplicated in at least one genome, of which 7040 were duplicated in a single genome. On a per-gene basis in genomes with at least one tandem duplicate, the average copy number is 2.20 (SE = 0.01) (fig. S2D).

Partial tetraploidy and tempo of fractionation

The maize ancestor underwent a whole-genome duplication (WGD) allopolyploidy event 5 to 20 million years ago (Fig. 2A) (*24*, *25*). Evidence for WGD is found in the existence of two separate genomes that are broken and rearranged yet still show clear synteny to sorghum (*24*, *26*). Many duplicated genes have since undergone loss, or fractionation, reducing maize to its current diploid state (*26*, *27*). Furthermore, fractionation is biased toward one homoeologous genome (M2, more fractionated) over the other (M1, less fractionated) (*26*). The M1 and M2 subgenomes are composed almost exclusively of core (87.25%) and near-core (6.19%) pan-genes (Figs. 1C and 2A). The broad architecture of syntenic regions relative to sorghum is consistent across the NAM genomes (fig. S3).

Given the ancient time frame of the WGD in maize and the rapid tempo of fractionation observed in other species (*28*, *29*), little variation in the retention of specific homoeologs is expected at the species level. In fact, prior work in temperate maize suggested that most fractionation occurred before domestication (*6*, *30*). However, our diverse set of genomes allows for a more complete characterization of fractionation within the species. Because fractionation can occur at the level of small deletions (*27*, *31*), we evaluated both partial and complete homoeolog loss beginning with a conservative set of 16,195 maize pan-orthologs. We determined that 7043 were single-copy orthologs, in which the homoeologous gene was likely deleted before maize speciation (Fig. 2A). In addition, we identified 4576 homoeologous pairs (Fig. 2A), of which 2155 had the same exon structure of the sorghum ortholog in both homoeologs. In 1281 pairs, at least one copy of the gene differed from its sorghum ortholog but did not vary among NAM lines, likely representing fractionation that pre-dated *Zea mays*. Another 1140 pairs varied across the genomes in their pattern of exon retention, segregating for deletions or structural differences in at least one copy of the gene. This segregating set was manually curated (data S1) to remove loci where exons or flanking sequence could not be confidently identified (Fig. 2A), resulting in a curated set of 494 homoeolog pairs segregating for fractionation, which represents >10% of pairs present in the pan-genome. Of these, 281 M2 homoeologs had exon loss compared with 236 M1 homoeologs, a 19% difference ($P < 0.05$, χ^2 test), which suggests ongoing biased fractionation. Analysis of gene ontology terms revealed putative functional differences between fully fractionated and segregating fractionated loci (fig. S4 and data S1).

Population genetic theory predicts that mutations segregating within a species, such as the segregating fractionation deletions we have identified, arose within the past $4N_e$ generations, where N_e represents the effective population size of the species. Using the N_e of the maize progenitor teosinte as an upward bound for maize [$N_e = 150,000$; (*32*)], we can infer that most segregating fractionation arose within the past 600,000 generations. Therefore, most segregating fractionation substantially postdates the WGD. Theory also predicts that

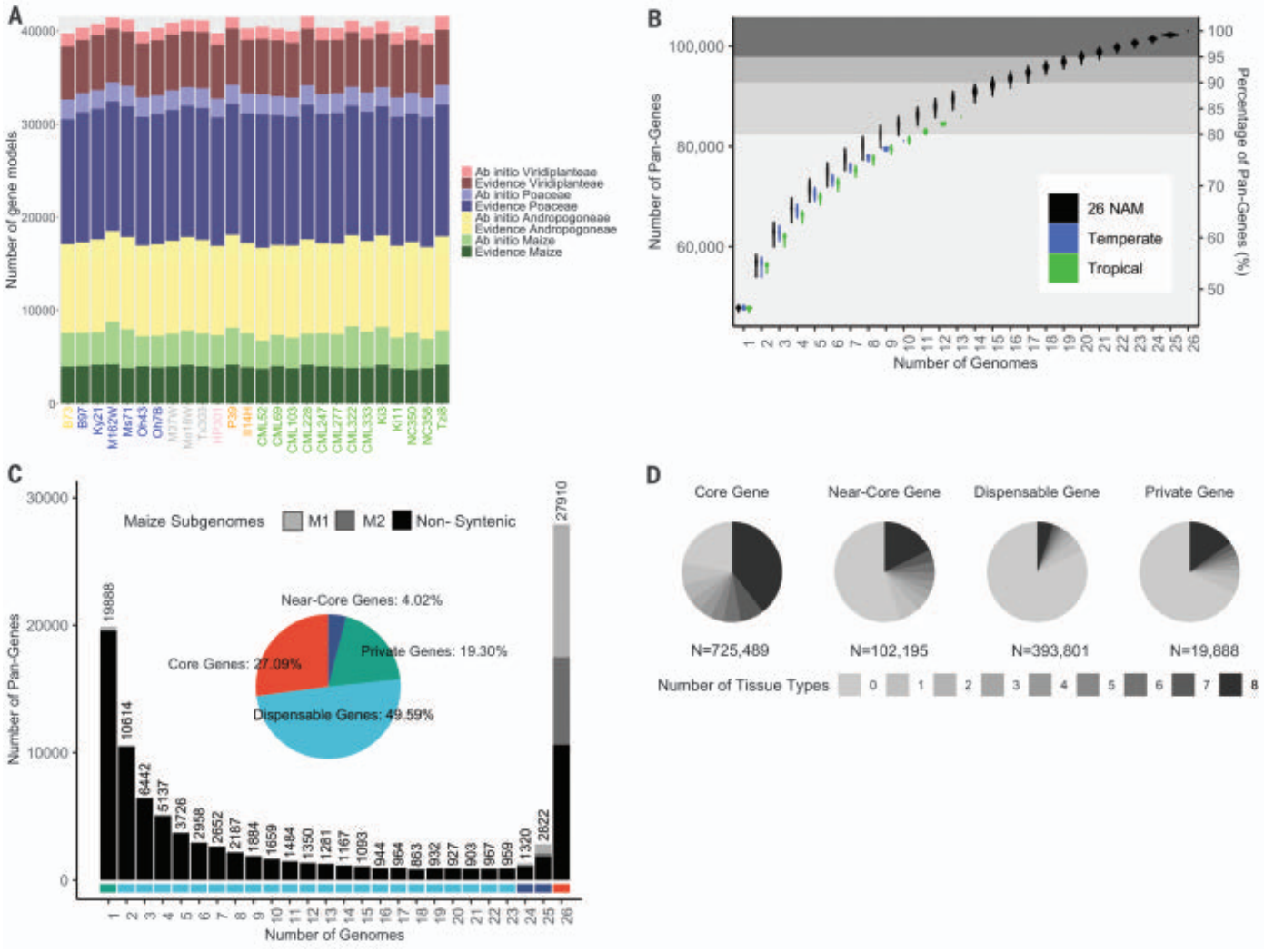


Fig. 1. Pan-genome analysis of the gene space. (**A**) Pan-genes categorized by annotation method and phylostrata. Genes annotated with evidence have mRNA support, whereas ab initio genes are predicted on the basis of DNA sequence alone. Genes within progressing phylostrata [species *Z. mays* (maize), tribe Andropogoneae, family Poaceae, kingdom Viridiplantae] are more conserved. (**B**) Number of pan-genes added with each additional genome assembly. Order of genomes being added into the pan-genome was bootstrapped 1000 times. Tropical lines include CML52, CML69, CML103, CML228, CML247, CML277, CML322, CML333, Ki3, Ki11, NC350, NC358, and Tzi8; temperate lines include B73, B97, Ky21, M162W, Ms71, Oh43, Oh7B, HP301, P39, and Il14H. (**C**) Proportion of pan-genes in the core, near-core, dispensable, and private fractions of the pan-genome. For (B) and (C), tandem duplicates were considered as a single pan-gene and coordinates were filled in when a gene was not annotated, but an alignment with >90% coverage and 90% identity was present within the correct homologous block. (**D**) Number of tissues with expression (reads per kilobase per million reads > 1) for each gene in each genome on the basis of their pan-genome classification. Tissues in this analysis include root, shoot, V11 base, V11 middle, V11 tip, anther, tassel, and ear.

rare deletions should be younger than those segregating at intermediate frequency. We constructed the unfolded site frequency spectrum (SFS) of segregating fractionation deletions and compared this with the unfolded SFS of noncoding single-nucleotide polymorphisms (SNPs) using sorghum to define the ancestral state (Fig. 2B). The data reveal a similar frequency distribution in deletions and SNPs, with a preponderance of rare variants in both, suggesting that a subset of fractionation may be quite young, with diploidization potentially continuing in modern maize. We also evaluated patterns of co-exon retention in non–stiff-stalk temperate, tropical, and flint-derived maize, observing population-specific fractionation (Fig. 2C). This variation in homoeolog retention at the population level confirms previous suppositions about the tempo of fractionation (*33*) and may reflect relaxed constraint on retained homoeologs after the domestication and migration of maize to temperate climates.

The repetitive fraction of the pan-genome

TEs were annotated in each assembly by using structural features and sequence homology (*34*). Individual TE libraries from each inbred were then combined to form a pan-genome library, which was used to identify TE sequences missed by individual libraries. The annotations reveal that DNA transposons and LTR retrotransposons constitute 8.5 and 74.4% of the genome, respectively (table S5 and fig. S5). A total of 27,228 TE families were included in the pan-genome TE library, of which 59.7% were present in all 26 NAM founders, and 2.5% were specific to one genome (fig. S6). The average percentages of intact and fragmented TEs were 30.5 and 69.5% (SE = 0.06%), respectively. As reported previously, *Gypsy* LTR retrotransposon families are more abundant in pericentromeric regions, whereas *Copia*

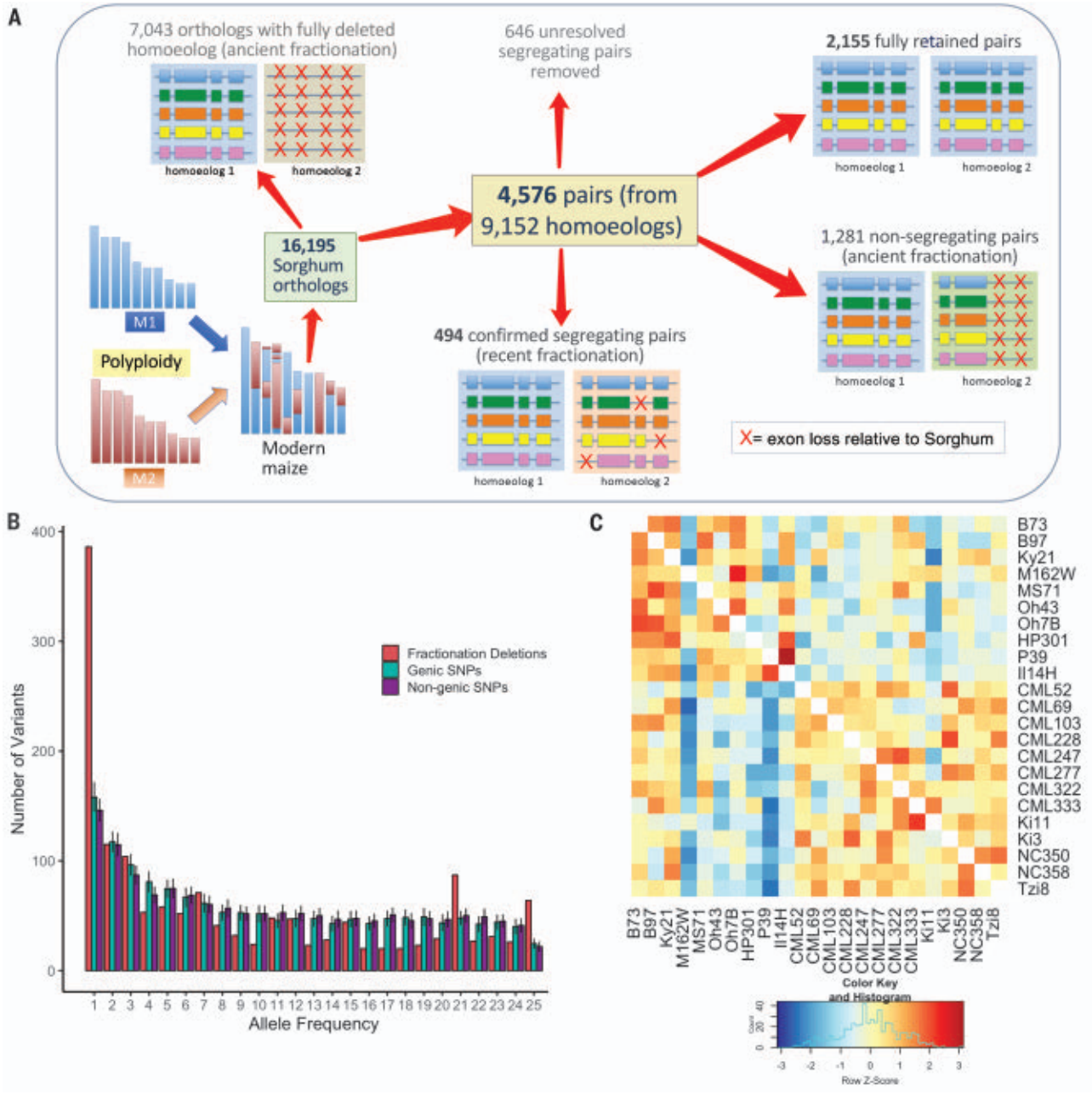


Fig. 2. The tempo of fractionation in maize. (**A**) Schematic showing how genes were categorized. A total of 16,195 conservatively chosen orthologs were subdivided into classes, representing retained pairs, ancient fractionation, and recent fractionation. (**B**) Unfolded SFS of segregating exon loss and noncoding SNPs (genic and nongenic) by using sorghum to define the ancestral state. (**C**) Heatmap of the number of co-retained exons between any two NAM lines. Lines with mixed ancestry (M37W, Mo18W, Tx303) are excluded. Colors indicate the *z*-score (the difference measured in standard deviations between a single pairwise comparison and all others in the row).

LTR retrotransposons are enriched in the gene-dense chromosome arms (fig. S7) (*35*). Tropical lines have significantly more *Gypsy* elements than temperate lines (P = 0.002, t test), with mean *Gypsy* content of 1018 and 988 Mbp, respectively (table S5 and fig. S5). This may reflect increasing constraint on *Gypsy* proliferation in temperate lines that have, on average, smaller genomes (Table 1).

In some maize lines, >15% of the genome is composed of tandem repeat arrays, including the centromere repeat CentC, the two knob repeats knob180 and TR-1, subtelomere, and telomere repeats (*36*, *37*). Repeats of this type remain a major impediment to assembly. A mean of 60% of CentC, 70% of the 4-12-1 subtelomeric sequence (*38*), 28.9% of TR-1, 1% of knob180, and 0.09% of ribosomal DNA repeat units were incorporated in the final assemblies (Table 1).

A total of 110 (of 260) functional centromeres identified by CENH3 chromatin immunoprecipitation sequencing (*39*, *40*) were fully assembled, and of these, 88 are gapless (fig. S8A) (*40*). Chromosomes with very long CentC arrays (such as chromosomes 1, 6, and 7) often have assembly gaps, and the precise location of the centromere could not be determined. In other cases, the centromeres include fully assembled small CentC arrays or the functional centromeres are located to one side of the CentC tracts in regions dominated by retrotransposons (Fig. 3A). By projecting all

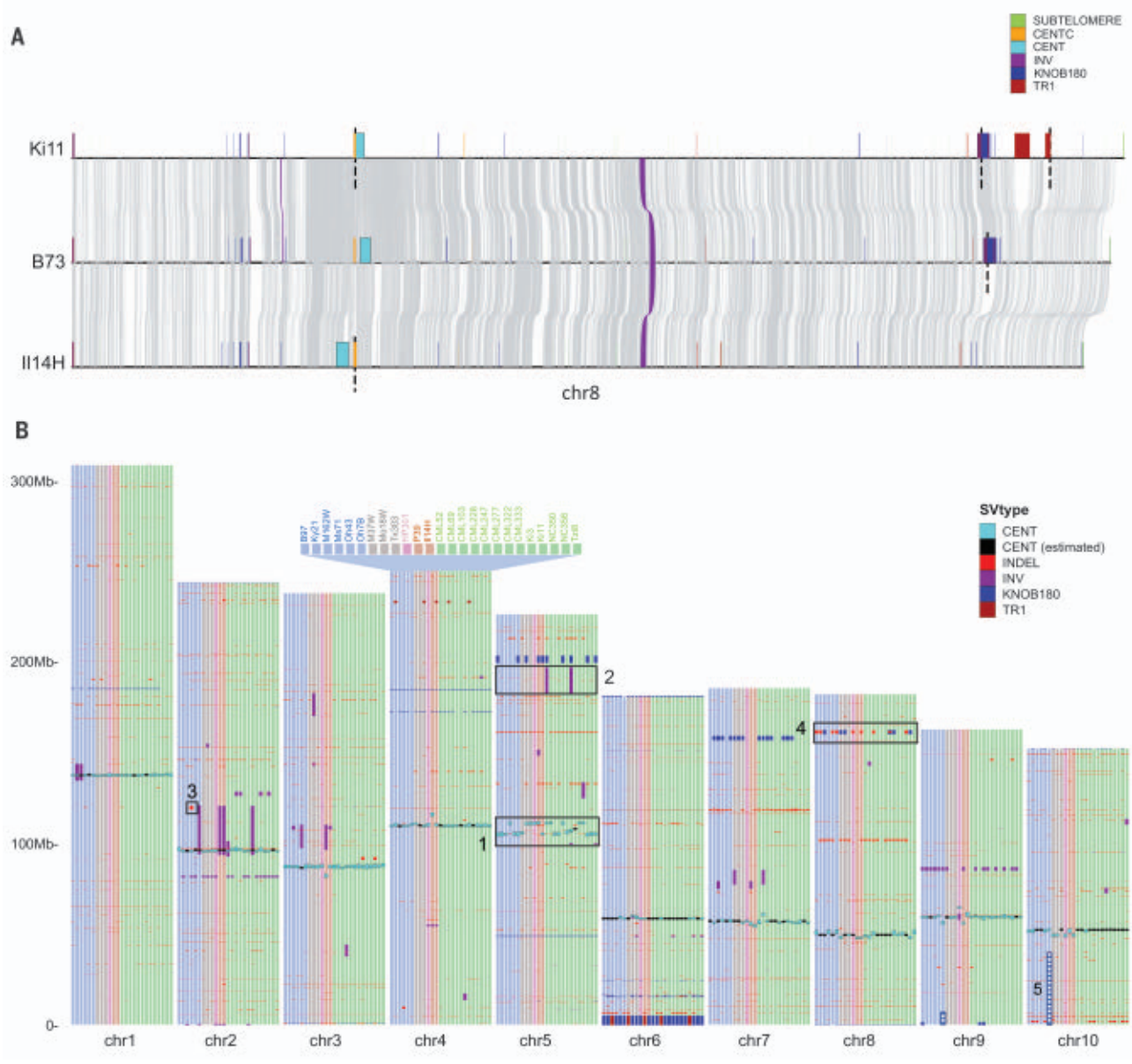


Fig. 3. Structural variation in the NAM founders. (**A**) Pairwise alignments between Ki11, B73, and Il14H on chromosome 8. Gray links represent syntenic aligned regions; gaps of unknown size (scaffold gaps) are marked by dashed lines. INV, inversion. (**B**) Large (>100 kbp) SVs, centromeres, and knobs across the NAM lines versus the B73 reference. The subset of SVs larger than 1 Mbp were manually curated, and only those containing genes are represented. Features 1 to 5 highlight major SVs: (1) multiple centromere movement events; (2) a major inversion previously hypothesized on the basis of suppressed recombination; (3) a large deletion in the Ms71 inbred; (4) knob polymorphism; (5) reciprocal translocation between chromosome 9 and 10 in the Oh7B inbred (both segments placed in their standard positions for display).

centromere locations onto B73, we were able to identify 12 centromere movement events (3 on chr5 and chr9 and 2 on chr3, chr8, and chr10), which clarifies and extends prior evidence for centromere shifting (Fig. 3B and fig. S8B) (*39*). The variation in CentC abundance and positional polymorphism made it possible to gaplessly assemble at least two variants of all 10 centromeres (fig. S8A).

Both knob180 and TR-1 arrays are subject to meiotic drive and accumulate when a chromosome variant known as Abnormal chromosome 10 (Ab10) is present (*37*, *41*). Although Ab10 is absent from modern inbreds, its legacy remains in the form of many large knobs. Most knob180 and TR-1 repeat arrays were identified in midarm positions (81.9%), where meiotic drive is most effective. Long knob180 and TR-1 repeat arrays can occur separately but are more frequently intermingled in fragmented arrays along with transposons (Fig. 3A and fig. S9) (*42*). Analysis of classical (cytologically visible) knobs on chromosomes 1S, 2S, 2L, 3L, 4L, 5L, 6L, 7L, 8L, and 9S revealed that their locations are syntenic and that several are composed of a series of disjointed smaller knobs (Fig. 3A and fig. S10). In some lines, knobs are not visible cytologically but can still be detected as smaller arrays at the sequence level; however, many show strict presence-absence variation among the NAM founder inbreds.

Tandem repeat arrays are also commonly found at the ends of chromosome arms (table S6). Among the 520 chromosome ends, 57.9% contained knob180 repeats, and 30.5% contained

subtelomere repeats. At least 65.6% of chromosome ends were fully assembled as indicated by the presence of telomere sequences.

Structural variation and impact on phenotype

Comparative analyses among the NAM genotypes to B73 revealed a cumulative total of 791,101 structural variants (SVs) >100 bp in size. Tropical lines, which are the most divergent from B73, include a substantially higher number of SVs than temperate lines (mean = 32,976 versus 29,742; $P = 0.00013$) (tables S7 and S8). SVs are more common on chromosome arms where recombination is highest (fig. S11), similar to SNPs and other forms of genetic variation (*43*). Almost half (49.6%) of SVs were <5 kbp in size, with 25.7% being <500 bp. Across all size classes, SVs are skewed toward rare variants (fig. S12). Several large SVs were found segregating within the 26 NAM genomes (Fig. 3B), including 35 distinct inversion polymorphisms and 5 insertion-deletion polymorphisms >1 Mbp. For example, a 14.6-Mbp inversion on chromosome 5 in the CML52 and CML322 lines, which was previously hypothesized on the basis of suppressed recombination in the NAM RILs (*11*), is confirmed in this study through assembly. Additionally, there is a 1.9-Mbp deletion with seven genes on chromosome 2 in the MS71 inbred and a 1.8-Mbp deletion with two genes on chromosome 8 found in eight lines. Our data also capture a very large reciprocal translocation (involving >47 Mbp of DNA) between the short arms of chromosomes 9 and 10 in Oh7B that had been previously detected in cytological studies (Fig. 3B) (*38*).

The high proportion of rare SVs in maize suggests that these may be a particularly deleterious class of variants, as observed in other species (*44*, *45*). Indels and inversions occur in regions that have 49.8% fewer genic base pairs than the genomic background. Furthermore, SVs are 17% less likely to be found in conserved regions than SNPs (odds ratios of 0.27 and 0.58 for SVs and SNPs, respectively; Fisher's exact test; $P < 0.001$). Approximate Bayesian computation modeling revealed that selection against SVs is at least as strong as that against nonsynonymous substitutions (fig. S13) (*17*). These results suggest that, when they occur, SVs are particularly consequential and relevant to fitness.

To estimate the phenotypic impact of SVs, we assessed the genetic basis of 36 complex traits (*14*) using 71,196 filtered SVs in 4027 recombinant inbred lines derived from the NAM founder inbreds (fig. S14A) (*11*). The analysis revealed that SVs explain a high percentage of phenotypic variance for disease traits (60.10 ~ 61.75%) and less for agronomic or morphological (20.04 ~ 61.04%) and metabolic traits (4.79 ~ 26.78%). Much of the phenotypic variation was also explained by SNPs, which were much more numerous (288-fold more) relative to our conservative set of SVs (fig. S14A). When the SNP and SV data were integrated into one linear mixed model, the combined markers only slightly surpassed values from SNPs, consistent with the fact that most SVs are in high linkage disequilibrium with SNPs (fig. S14A).

We also carried out genome-wide association analyses (GWASs) to identify specific SVs contributing to phenotypic variation for the same suite of traits (fig. S14, B to G). Among the detected GWAS signals, 93.05% overlapped with those identified with SNPs, and 6.95% were specific to SVs (no significant SNP detected within 5 Mbp of significant SVs). There

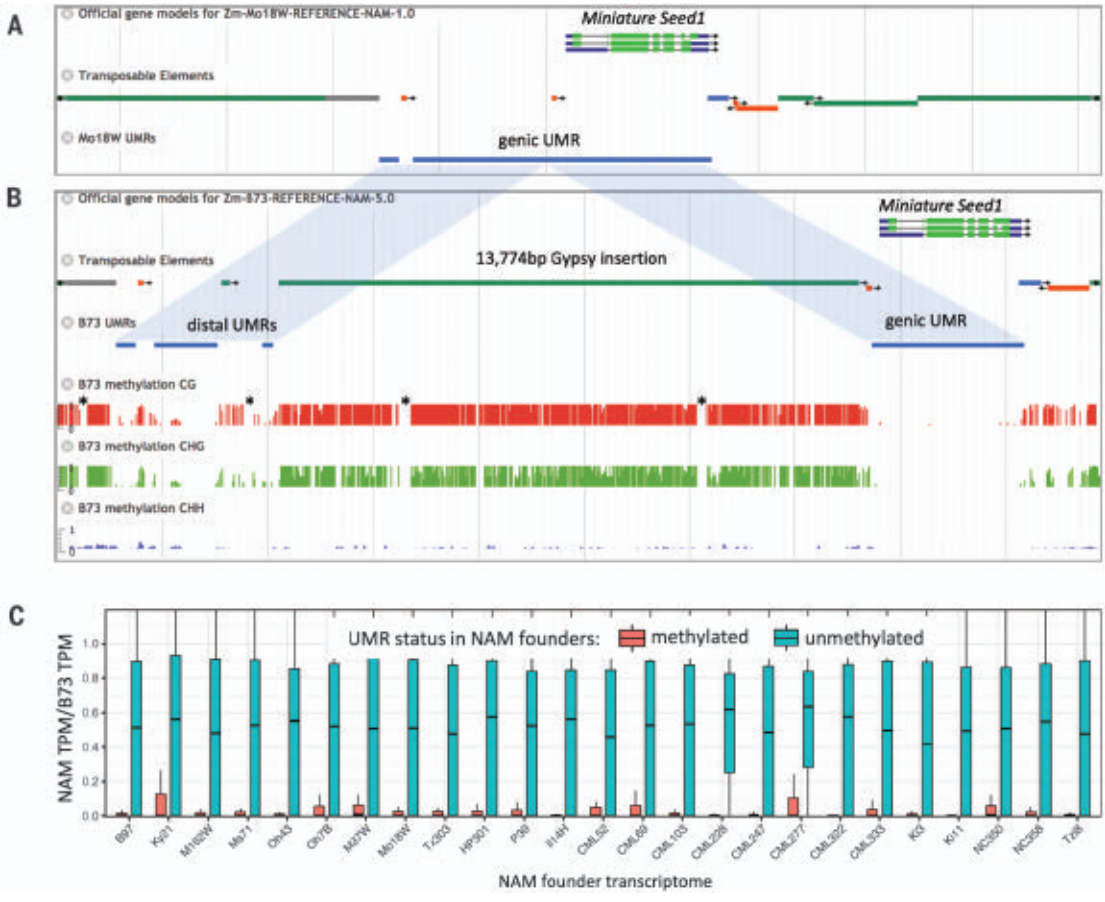


Fig. 4. UMR variation across the NAM founders. (**A**) Annotation of the *Miniature seed1* gene in the Mo18W inbred. An image from the MaizeGDB browser shows gene, TE, and UMR tracks. TE tracks are color-coded by superfamily: green-gray, long terminal repeats; red, terminal inverted repeats; and blue, long interspersed nuclear elements. The gray vertical lines show 2.5-kbp intervals. (**B**) Annotation and underlying methylation data for *Miniature seed1* in the B73 inbred. The insertion of a *Gypsy* element moved part of the proximal UMR to a position 14 kbp upstream from the TSS. Methylation tracks indicate base pair–level methylation values from 0 to 100%. Asterisks indicate gaps in coverage, which are visible in separate tracks (fig. S28). (**C**) Relationship between methylation and gene expression. UMRs were mapped to B73 to identify UMRs that overlap with TSS. The *y* axis indicates the ratio of transcripts per million (TPM; compared with B73) when the region is methylated (red) or unmethylated (teal).

was a significant enrichment of SVs associated with phenotypes in genic regions (z = 8.022, P < 1.04e-15) (fig. S15). The most significant association between an SV and a trait not identified with SNP markers was a quantitative trait locus for northern leaf blight on chromosome 10 (fig. S14F). This SV is within a gene encoding a thylakoid lumenal protein; such proteins could be linked to plant immunity through the regulation of cell death during viral infection (*46*). We anticipate that the effects of SVs may be even more pronounced in larger association panels, where extensive historical recombination may help disentangle their effects from nearby SNPs.

Disease resistance in plants is frequently associated with SV in the form of tandem arrays of resistance genes. Complex arrays of resistance genes are retained, potentially through birth-death dynamics in an evolutionary arms race with pathogens or through balancing selection for the maintenance of diverse plant defenses (*47*). Nucleotide-binding, leucine-rich-repeat (NLR) proteins provide a common type of resistance. Our data reveal that there are fewer NLR genes in maize than in other Poaceae (fig. S16) and that most NAM lines have lost the same clades of NLRs as sorghum (fig. S17). Only one line (CML277) retains the MIC1 NLR clade, which is particularly fast-evolving in Poaceae (*48*). Nevertheless, there is clear NLR variation among the NAM lines (fig. S18), and tropical genomes contain a significantly higher number of NLR genes than temperate genomes (t test, P = 0.006), suggesting ongoing coevolution with pathogens, particularly where disease pressure is high.

The annotated NLR genes were significantly enriched for overlap with SVs (bootstrap permutation test, P < 0.001). An extreme example is found at the *rp1* (resistance to *Puccinia sorghi1*) locus on the short arm of chromosome 10, which is known to be highly variable (*49*). We observed exceptional diversity in the NAM lines with as few as 4 *rp1* copies in P39 and as many as 30 in M37W (table S9). However, because of its repetitive nature, only 18 NAM lines have gapless assemblies of the *rp1* locus.

SVs linked to transposons have been shown, through the modulation of gene expression, to underlie flowering-time adaptation in maize during tropical-to-temperate migration (*50*, *51*). Our SV and TE-annotation pipelines identified the adaptive *CACTA*-like insertion that was previously reported upstream of the flowering-time locus *ZmCCT10* (*51*). We also surveyed 173 genes linked to flowering-time (*52*, *53*) and discovered three genes (*GL15*, *ZCN10*, and *Dof21*) with TE-derived SVs <5 kbp upstream of their transcription start sites (TSSs). These SVs distinguish temperate from tropical lines (t < −2.346, P < 0.0358) (fig. S19) and show significant correlation (F > 8.658, P < 0.001) with expression levels.

Discovery of candidate cis-regulatory elements through DNA methylation

On the basis of sequence alone, it can be difficult to identify functional sequences in the intergenic spaces. One approach is to score for unmethylated DNA, which provides both a tissue-independent indicator of gene regulatory elements and evidence that annotated genes are active (*5*, *54*, *55*). We sequenced enzymatic methyl sequencing libraries from each NAM line and identified methylated bases in three sequence contexts, CG, CHG, and CHH (where H is A, T, or C). Results are consistent across genes and transposons, demonstrating the quality of the libraries (figs. S20 and S21). There is minor variation in total methylation across inbreds, with CML247 being noteworthy for uniformly lower CG methylation in several tissues (fig. S22). Such natural variation in methylation is also observed in *Arabidopsis* ecotypes (*56*).

Each of the three methylation (m) contexts reveals information on the locations of repeats, genes, and regulatory elements. mCHH levels are generally low except at heterochromatin borders, whereas mCHG and mCG are abundant in repetitive regions. Both mCHG and mCG are depleted from regulatory elements, and mCHG is depleted from exons (*57*). However, mCG is often present in exons (Fig. 4) (*58*). Thus, to identify unmethylated regions (UMRs) that correspond to regulatory elements and gene bodies, we defined UMRs using a method that takes into account mCHG and mCG but does not exclude high mCG–only regions (the term UMR is used for simplicity; some regions contain CG methylation). Comparison of the 26 methylomes revealed uniformity in number and length of UMRs, averaging ~180 Mbp in total length in each genome (figs. S23 and S24). To confirm the accuracy of the UMR data, we also identified accessible chromatin regions using ATAC sequencing for each inbred. We expect chromatin to be accessible mainly in the subset of genes expressed in the tissue sampled (primarily leaves) and to show concordance with UMRs. The data reveal that a mean of 99% of genic and 96% of nongenic (distal) accessible chromatin regions overlap with UMRs in each genome (figs. S25 and S26).

To assess methylation diversity, we mapped UMRs from all inbreds to the B73 genome. Approximately 95% of genic UMRs overlap across genomes in pairwise comparisons (fig. S27). UMR polymorphism is higher in the intergenic space, particularly among UMRs >5 kbp from genes, where typically ~75% of UMRs overlap (fig. S27). Even when the UMR sequence is conserved, its position relative to the closest gene may vary substantially among inbreds. This is exemplified by the *Miniature Seed1* gene, in which a UMR proximal to the promoter in Mo18W is displaced nearly 14 kbp upstream in B73 by a single *Huck* element (*Gypsy* LTR superfamily) (Fig. 4). The *Huck* insertion is present in 23 of 26 genomes, and in 2 of these (Oh43 and CML322), additional nested TE insertions increased the distance between the gene and the UMR to 27 kbp. Although UMR polymorphism correlates with genetic distance across NAM lines (fig. S29), UMRs from Tzi8 were not substantially shared with other tropical genomes.

Adaptive variation in DNA methylation has been observed in maize (*59*), most likely through effects on gene expression. To estimate how well UMRs predict transcription, we identified a conservative subset of UMR overlapping genes that were unmethylated in B73 but methylated in at least one other methylome. These differentially methylated regions were strongly correlated with differences in gene expression (Fig. 4 and fig. S30). We further evaluated the enrichment of significant GWAS SNPs across 36 traits in UMRs. From genome-wide estimates, UMRs show 2.50- to 3.26-fold enrichment across traits for significant associations. Roughly 18% of SNPs identified by GWAS lie outside of genic regions but within UMRs (table S10), which is consistent with the view that UMRs can be used to identify functional, noncoding regions (*5*, *54*, *55*).

Summary

Our analysis of 26 genomes uncovered variation in both the genic and repetitive fractions of the pan-genome. Tropical, temperate, and flint-derived popcorn and sweet corn germplasm are differentiated in distinctive ways, including their pan-gene complement, homoeolog retention after polyploidy, abundance of TEs, NLR disease-resistance gene copy number, and methylation profiles. The available data will have broad utility for genetic and genomic studies and facilitate rapid associations to phenotyping information. For example, the genic presence-absence variation that was identified in this study may be imputed across additional mapping populations to clarify its contribution to heterosis through complementation (*60*). More generally, these resources should motivate a shift away from the single-reference mindset to a multireference view in which any one of 26 inbreds, each with different experimental and agronomic advantages, can be deployed for the purposes of basic discovery and crop improvement.

REFERENCES AND NOTES

1. C. N. Hirsch *et al.*, *Plant Cell* **26**, 121–135 (2014).
2. C. N. Hirsch *et al.*, *Plant Cell* **28**, 2700–2714 (2016).
3. M. Jin *et al.*, *Sci. Rep.* **6**, 18936 (2016).
4. F. Lu *et al.*, *Nat. Commun.* **6**, 6914 (2015).
5. W. A. Ricci *et al.*, *Nat. Plants* **5**, 1237–1249 (2019).
6. S. Sun *et al.*, *Nat. Genet.* **50**, 1289–1295 (2018).

7. G. Haberer *et al.*, *Nat. Genet.* **52**, 950–957 (2020).
8. N. Yang *et al.*, *Nat. Genet.* **51**, 1052–1059 (2019).
9. G. Lin *et al.*, *Genome Biol.* **22**, 175 (2021).
10. N. M. Springer *et al.*, *Nat. Genet.* **50**, 1282–1288 (2018).
11. M. D. McMullen *et al.*, *Science* **325**, 737–740 (2009).
12. S. A. Flint-Garcia *et al.*, *Plant J.* **44**, 1054–1064 (2005).
13. J. Yu, J. B. Holland, M. D. McMullen, E. S. Buckler, *Genetics* **178**, 539–551 (2008).
14. J. G. Wallace *et al.*, *PLOS Genet.* **10**, e1004845 (2014).
15. S. R. Eichten *et al.*, *Plant Cell* **25**, 2783–2797 (2013).
16. R. J. Schaefer *et al.*, *Plant Cell* **30**, 2922–2942 (2018).
17. See supplementary materials.
18. Y. Jiao *et al.*, *Nature* **546**, 524–527 (2017).
19. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, *Bioinformatics* **31**, 3210–3212 (2015).
20. S. Ou, J. Chen, N. Jiang, *Nucleic Acids Res.* **46**, e126 (2018).
21. S. Ou *et al.*, *Nat. Commun.* **11**, 2288 (2020).
22. K. Eilbeck, B. Moore, C. Holt, M. Yandell, *BMC Bioinformatics* **10**, 67 (2009).
23. M. Law *et al.*, *Plant Physiol.* **167**, 25–39 (2015).
24. Z. Swigonová *et al.*, *Genome Res.* **14** (10a), 1916–1923 (2004).
25. X. Wang *et al.*, *Mol. Plant* **8**, 885–898 (2015).
26. J. C. Schnable, N. M. Springer, M. Freeling, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4069–4074 (2011).
27. M. R. Woodhouse *et al.*, *PLOS Biol.* **8**, e1000409 (2010).
28. J. C. Schnable, M. Freeling, E. Lyons, *Genome Biol. Evol.* **4**, 265–277 (2012).
29. T. Mandáková, S. Joly, M. Krzywinski, K. Mummenhoff, M. A. Lysak, *Plant Cell* **22**, 2277–2290 (2010).
30. A. B. Brohammer, T. J. Y. Kono, N. M. Springer, S. E. McGaugh, C. N. Hirsch, *Plant J.* **93**, 131–141 (2018).
31. H. Tang *et al.*, *Genetics* **190**, 1563–1574 (2012).
32. T. M. Beissinger *et al.*, *Nat. Plants* **2**, 16084 (2016).
33. F. Cheng *et al.*, *Nat. Plants* **4**, 258–268 (2018).
34. S. Ou *et al.*, *Genome Biol.* **20**, 275 (2019).
35. R. S. Baucom *et al.*, *PLOS Genet.* **5**, e1000732 (2009).
36. P. Bilinski *et al.*, *PLOS Genet.* **14**, e1007162 (2018).
37. R. K. Dawe *et al.*, *Cell* **173**, 839–850.e18 (2018).
38. P. S. Albert, Z. Gao, T. V. Danilova, J. A. Birchler, *Cytogenet. Genome Res.* **129**, 6–16 (2010).
39. K. L. Schneider, Z. Xie, T. K. Wolfgruber, G. G. Presting, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E987–E996 (2016).
40. N. Wang, J. Liu, W. A. Ricci, J. I. Gent, R. K. Dawe, *Genetics* **217**, iyab020 (2021).
41. K. W. Swentowsky *et al.*, *Genes Dev.* **34**, 1239–1251 (2020).
42. J. Liu *et al.*, *Genome Biol.* **21**, 121 (2020).
43. J.-M. Chia *et al.*, *Nat. Genet.* **44**, 803–807 (2012).
44. H. J. Abel *et al.*, *Nature* **583**, 83–89 (2020).
45. E. V. Leushkin, G. A. Bazykin, A. S. Kondrashov, *Genome Biol. Evol.* **5**, 514–524 (2013).
46. S. Seo *et al.*, *Plant Cell* **12**, 917–932 (2000).
47. H. Mizuno *et al.*, *Sci. Rep.* **10**, 872 (2020).
48. P. C. Bailey *et al.*, *Genome Biol.* **19**, 23 (2018).
49. S. H. Hulbert, J. L. Bennetzen, *Mol. Gen. Genet.* **226**, 377–382 (1991).
50. C. Huang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E334–E341 (2018).
51. Q. Yang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16969–16974 (2013).
52. Z. Dong *et al.*, *PLOS ONE* **7**, e43450 (2012).
53. Y.-X. Li *et al.*, *Plant J.* **86**, 391–402 (2016).
54. R. Oka *et al.*, *Genome Biol.* **18**, 137 (2017).
55. P. A. Crisp *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23991–24000 (2020).
56. T. Kawakatsu *et al.*, *Cell* **166**, 492–505 (2016).
57. J. I. Gent *et al.*, *Genome Res.* **23**, 628–637 (2013).
58. A. J. Bewick, R. J. Schmitz, *Curr. Opin. Plant Biol.* **36**, 103–110 (2017).
59. G. Xu *et al.*, *Nat. Commun.* **11**, 5539 (2020).
60. N. M. Springer *et al.*, *PLOS Genet.* **5**, e1000734 (2009).
61. M. B. Hufford *et al.*, HuffordLab/NAM-genomes: publication. prerelease, Zenodo (2021); https://zenodo.org/record/4781590).

ACKNOWLEDGMENTS

We appreciate the sequencing services provided by the University of Arizona, Oregon State University, Brigham Young University, and the University of Georgia, as well as coordination among sequencing centers provided by Pacific Biosciences. We appreciate the early contributions to maize genome assembly in the cloud from DNAnexus. The authors further acknowledge the High Performance Computing facility at Iowa State University [partially funded by the National Science Foundation (NSF) 1726447], Minnesota Supercomputing Institute, the Georgia Advanced Computing Resource Center, Cold Spring Harbor Laboratory high-performance computing center (NIH S10 OD0286321-01), and the participants of the Virtual Maize Annotation Jamboree who evaluated the initial gene predictions for benchmarking and improvements in the final gene annotations. **Funding:** Primary support for this work came from a generous grant from the NSF (IOS-1744001). Additional support came from NSF IOS-1546727 to C.N.H., USDA 2018-67013-27571 to C.N.H., USDA-ARS 8062-21000-041-00D, NSF IOS-1127112 and NIH-OD S10 OD028632 to D.W., NSF IOS-1546719 to M.B.H., NSF IOS-1822330 to J.R.-I. and M.B.H., USDA Hatch project (CA-D-PLS-2066-H to J.R.-I.), NSF IOS-1856627 to R.J.S., an NSF Postdoctoral Fellowship in Biology (DBI-1905869 to A.P.M.), NSF Graduate Research Fellowships (1650042 to A.I.H. and 1744592 to S.J.S.), NSF Research Traineeship (DGE-1545463) to Iowa State University (trainee S.J.S.), USDA-ARS 58-5030-8-064 to M.B.H. and C.M.A., USDA-ARS project 5030-21000-068-00D to C.M.A. and M.R.W., and NSF IOS-1546657 to J.Y. **Author contributions:** Conceptualization—R.K.D., D.W., M.B.H., C.N.H., and J.I.G.; Data curation—M.R.W., A.S.S., K.M.C., S.O., J.L., S.W., A.P.M., Z.L., B.W., M.K.T.-R., J.L.P., E.K.S.C., and C.M.A.; Formal analysis—A.S.S., M.R.W., K.M.C., S.O., J.L., W.A.R., T.G., A.O., Y.Q., R.D.C., S.T., A.P.M., A.I.H., S.W., Z.L., B.W., M.K.T.-R., R.D.P., Y.Z., C.H.O., X.L., A.M.G., E.B., J.L.P., N.M., S.J.S., Q.J., S.P., M.L.S., K.F., and J.I.G.; Funding acquisition—R.K.D., D.W., M.B.H., C.N.H., J.I.G., and R.J.S.; Investigation—D.w.K., D.A.K., N.W., D.E.H., V.L., K.F., and J.I.G.; Methodology—M.B.H., D.W., C.N.H., A.S.S., M.R.W., K.F., W.A.R., J.L., R.J.S., J.I.G., J.R.-I., J.Y., and R.K.D.; Project administration—R.K.D., M.B.H., D.W., C.N.H., and J.I.G.; Software—A.S.S., D.E.H., N.M., and S.O.; Supervision—M.B.H., D.W., R.K.D., C.N.H., J.I.G., K.V.K., R.J.S., J.R.-I., and J.Y.; Visualization—M.R.W., A.S.S., J.L., W.A.R., Y.Q., K.M.C., S.O., R.D.P., S.J.S., C.N.H., and J.I.G.; and Writing—M.B.H., R.K.D., C.N.H., and J.I.G. **Competing interests:** R.J.S. is a cofounder of REquest Genomics, LLC, a company that provides epigenomic services. All other authors declare no competing interests. **Data and materials availability:** Genome assemblies and annotations can be accessed at https://maizegdb.org/NAM_project and http://maize-pangenome.gramene.org. Raw data used for the assemblies including PacBio, Illumina, and Bionano data are available through ENA BioProject IDs PRJEB31061 and PRJEB32225. RNA-seq data are available at ENA ArrayExpress E-MTAB-8633 and E-MTAB-8628. Enzymatic methyl sequencing reads are available at ENA ArrayExpress E-MTAB-10088. ATAC-sequencing reads are available under NCBI GEO accession GSE165787. Other files, tables and supplemental data can be found in CyVerse https://datacommons.cyverse.org/browse/iplant/home/shared/NAM/NAM_genome_and_annotation_Jan2021_release. Links to the NLR trees can be found at https://itol.embl.de/shared/xCJbI9ndshEK. Scripts used to generate and analyze data are available as a Zenodo repository (*61*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6555/655/suppl/DC1
Materials and Methods
Figs. S1 to S30
Tables S1 to S10
References (*62*–*176*)
Data S1

14 January 2021; accepted 24 June 2021
10.1126/science.abg5289

PHENYLKETONURIA

A noncoding RNA modulator potentiates phenylalanine metabolism in mice

Yajuan Li[1]†, Zhi Tan[2]†‡, Yaohua Zhang[1]†, Zhao Zhang[3], Qingsong Hu[1]§, Ke Liang[1], Yao Jun[1], Youqiong Ye[3], Yi-Chuan Li[3], Chunlai Li[1]¶, Lan Liao[4], Jianming Xu[5], Zhen Xing[1]#, Yinghong Pan[6], Sujash S. Chatterjee[6], Tina K. Nguyen[1], Heidi Hsiao[1], Sergey D. Egranov[1], Nagireddy Putluri[5], Cristian Coarfa[5], David H. Hawke[7], Preethi H. Gunaratne[6], Kuang-Lei Tsai[3], Leng Han[8], Mien-Chie Hung[9,10], George A. Calin[11], Fares Namour[12,13], Jean-Louis Guéant[12,13], Ania C. Muntau[14], Nenad Blau[15], V. Reid Sutton[16], Manuel Schiff[17,18], François Feillet[13,19]*, Shuxing Zhang[2,20]*, Chunru Lin[1,20]*, Liuqing Yang[1,20,21]***

The functional role of long noncoding RNAs (lncRNAs) in inherited metabolic disorders, including phenylketonuria (PKU), is unknown. Here, we demonstrate that the mouse lncRNA *Pair* and human *HULC* associate with phenylalanine hydroxylase (PAH). *Pair*-knockout mice exhibited excessive blood phenylalanine (Phe), musty odor, hypopigmentation, growth retardation, and progressive neurological symptoms including seizures, which faithfully models human PKU. *HULC* depletion led to reduced PAH enzymatic activities in human induced pluripotent stem cell–differentiated hepatocytes. Mechanistically, *HULC* modulated the enzymatic activities of PAH by facilitating PAH-substrate and PAH-cofactor interactions. To develop a therapeutic strategy for restoring liver lncRNAs, we designed GalNAc-tagged lncRNA mimics that exhibit liver enrichment. Treatment with GalNAc-*HULC* mimics reduced excessive Phe in *Pair*$^{-/-}$ and *Pah*$^{R408W/R408W}$ mice and improved the Phe tolerance of these mice.

The predominant view of human genetic diseases revolves around the identification of mutations and dysfunctions of coding genes. However, ~98% of human genomic mutations occur within noncoding regions, and an understanding of the functional roles of noncoding RNAs in human genetic diseases is lacking. Phenylketonuria (PKU, OMIM 261600) and its milder variant, hyperphenylalaninemia (HPA), are genetic disorders caused by a deficiency in the hydrolysis of L-phenylalanine (Phe) to L-tyrosine (Tyr) (*1*). More than 1000 PAH variants [see the Phenylalanine Hydroxylase Gene Locus-Specific Database (PAHvdb) database: http://www.biopku.org/home/pah.asp] have been identified

and associated with PAH deficiency. Most of the variants are missense, usually resulting in protein misfolding and/or impairment of catalytic functions (*2*). Approximately one in every 10,000 infants is affected by this disease (*3*). On the basis of blood Phe concentration during diagnosis or screening in the neonatal period, PKU can be categorized as PKU (Phe >900 μmol/L), mild PKU (Phe <900 but >360 μmol/L), or mild HPA (Phe above the normal limit but <360 μmol/L) (*4*, *5*). Excessive Phe concentrations in patients with untreated PKU cause brain damage and associated mental retardation (*6*). Although PKU is largely considered a monogenic disorder, clinical evidence has indicated that PAH variants are incompletely correlated with the metabolic phenotypes of patients with PKU (*7*–*9*). In some patients diagnosed with PKU, no mutations could be identified in the *PAH* gene, suggesting that unknown factors may contribute to this disorder (*10*). Consistent with this notion, the recent discovery of biallelic mutations in the *DNAJC12* gene in patients exhibiting high blood Phe concentrations without mutations in *PAH* or genes involved in BH_4 (tetrahydrobiopterin, a PAH cofactor) metabolism (*11*) suggested the possibility that non-*PAH* genes may affect PAH function and subsequently increase blood Phe concentrations.

Newborns diagnosed with PKU are typically treated with a Phe-restricted diet and/or BH_4 supplementation. Patients with untreated PKU exhibit intellectual disability, behavioral issues, seizures, and psychiatric disorders (*12*). For patients with certain genotypes of mild PKU, supplementation with BH_4 has been suggested as an enzyme enhancement therapy (*13*). However, patients with a severe form of PKU (Phe >900 μmol/L and usually >1200 μmol/L) require additional treatment considerations. These patients also frequently respond poorly to BH_4 treatment (*14*). Given the potential immunogenic nature of Phe ammonia lyase (PAL) enzyme supplementation or substitution therapy (*15*, *16*) and the uncertainty of gene therapy (*17*), additional therapeutic strategies are urgently needed.

RESULTS

Depletion of lncRNAs drives PKU

Long intergenic noncoding RNAs (lincRNAs) and long noncoding RNAs (lncRNAs) are transcripts with low coding potential. Aiming to investigate the biological importance of lncRNAs, we determined the lncRNA profiles of livers from mouse embryonic day 18.5 (E18.5) embryos and from 2-month-old adult mice (Fig. 1A and fig. S1, A to C), finding that 2210408F21Rik (NR_040259) (renamed *Pair:* PAH-activating lincRNA) is one of the most up-regulated lncRNAs in the livers of adults compared with those from E18.5 embryos and exhibits low coding potential [Coding-Non-Coding Identifying Tool (CNIT) score −0.3544 (*18*); fig. S1, D to H, and table S1]. To deplete this mouse lncRNA, we introduced a two-nucleotide mutation using CRISPR/Cas9 to interrupt the splicing site after the first exon (Fig. 1B). Northern blotting indicated that *Pair* has two major isoforms with molecular weights of 730 bp and 1.5 kb, which were both depleted upon the introduction of the splicing-site mutation (Fig. 1C). Compared with wild-type (WT) and heterozygous littermates, $Pair^{-/-}$ mice exhibited similar expression of neighboring genes with no detectable alterations in major organ development (fig. S2, A to D).

Both male and female $Pair^{-/-}$ mice exhibited hypopigmentation (Fig. 1D), growth retardation, and elevated serum Phe concentrations (Fig. 1, E to G) reminiscent of human PKU. $Pair^{-/-}$ livers showed no detectable changes in the expression of BH_4 biogenesis genes or PAH protein abundance (fig. S2, E to G). $Pair^{-/-}$ livers exhibited enzymatic deficiency in converting Phe to Tyr, and $Pair^{+/-}$ livers showed impaired PAH enzymatic activity (fig. S2, H and I). $Pair^{+/-}$ mice exhibited blood Phe concentrations within the normal range; however, these animals showed elevated blood Phe concentrations upon Phe challenge compared with $Pair^{+/+}$ mice (fig. S2J). These results suggested that $Pair^{+/-}$ livers exhibit partial PAH deficiency. Sanger sequencing of the *Pah* gene suggested that Pah^{enu2} mice harbor a T788C mutation, whereas $Pair^{-/-}$ mice harbor a WT *Pah* gene (table S2 and fig. S2, K and L).

The median life span of $Pair^{-/-}$ mice is 15.2 months (Fig. 1H), and >70% of the mice exhibited seizures starting at the age of 8 to 10 months (Fig. 1, I to J, and movies S1 to S3). After cardiopulmonary resuscitation, *Pair*-knockout mice experiencing a seizure could be rescued (movie S4). $Pair^{-/-}$ brains were smaller and exhibited reduced tyrosine hydroxylase (TH)–positive neurons compared with WT and heterozygous littermates (Fig. 1, K and L). $Pair^{-/-}$ livers and serum had diminished tyrosine concentrations. Tyrosine is catalytically produced by PAH in both serum and liver tissue (fig. S3, A to E), confirming the presence of PAH deficiency. The role of *Pair* in seizures, in addition to regulating PAH enzymatic activities, could not be ruled out. Our data suggest that $Pair^{-/-}$ mice model human PKU.

Pair and *HULC* associate with PAH

To understand the molecular mechanism of *Pair*, we performed *Pair* pulldown using biotinylated sense or antisense *Pair* in mouse livers (Fig. 2A and table S3). Sense, but not antisense, *Pair* associated with mouse PAH (Fig. 2A). Streptavidin beads-only and polyadenosine RNA were used as negative controls, and the association of ELAV1 with the AR_3′ untranslated region served as a positive control (*19*) (Fig. 2A). We then performed a cross-linking immunoprecipitation (CLIP) assay using mouse $Pair^{+/+}$ and $Pair^{-/-}$ livers or human liver tissues from two healthy donors (fig. S4, A and B, and table S4). The presence of PAH-RNA complexes was diminished upon *Pair* knockout (fig. S4A). PAH-lncRNA(s) complexes were detected in the two human liver donors (fig. S4B). The PAH-RNA complexes (fig. S4, A and B, blue boxes 1 to 3) were subjected to reverse transcription and Sanger sequencing (table S5). As expected, mouse PAH associated with mouse *Pair* (nucleotides 460 to 496) in $Pair^{+/+}$ livers (Fig. 2B, bottom row). Human PAH associated with one human lncRNA gene, *HULC* (nucleotides 183 to 216) (Fig. 2B, top row). Although *HULC* has been suggested to be up-regulated

[1]Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [2]Intelligent Molecular Discovery Laboratory, Department of Experimental Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, TX 77054, USA. [3]Department of Biochemistry and Molecular Biology, The University of Texas Health Science Center at Houston McGovern Medical School, Houston, TX 77030, USA. [4]Genetically Engineered Mouse Core, Advanced Technology Cores, Baylor College of Medicine, Houston, TX 77030, USA. [5]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA. [6]Department of Biochemistry and Biology, University of Houston, Houston, TX 77030, USA. [7]Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [8]Center for Epigenetics and Disease Prevention, Institute of Biosciences and Technology, Texas A&M University, Houston, TX 77030, USA. [9]Graduate Institute of Biomedical Sciences, Research Center for Cancer Biology, and Center for Molecular Medicine, China Medical University, Taichung 404, Taiwan. [10]Department of Biotechnology, Asia University, Taichung 413, Taiwan. [11]Department of Translational Molecular Pathology, Division of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [12]Department of Molecular Medicine and Reference Center for Inborn Errors of Metabolism, University Hospital of Nancy, Nancy F-54000, France. [13]INSERM, U1256, NGERE - Nutrition, Genetics, and Environmental Risk Exposure, University of Lorraine, Nancy F-54000, France. [14]University Children's Hospital, University Medical Center Hamburg Eppendorf, Hamburg 20246, Germany. [15]Division of Metabolism, University Children's Hospital Zurich, CH-8032 Zurich, Switzerland. [16]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. [17]Necker Hospital, APHP, Reference Center for Inborn Error of Metabolism and Filière G2M, Pediatrics Department, University of Paris, Paris 75007, France. [18]Inserm UMR_S1163, Institut Imagine, Paris 75015, France. [19]Pediatric Department Reference Center for Inborn Errors of Metabolism Children University Hospital Nancy, Nancy F-54000, France. [20]The Graduate School of Biomedical Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [21]Center for RNA Interference and Non-Coding RNAs, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

***Corresponding author. Email: f.feillet@chru-nancy.fr (F.F.); shuzhang@mdanderson.org (S.Z.); clin2@mdanderson.org (C.L.); lyang7@mdanderson.org (L.Y.)**

†These authors contributed equally to this work. ‡Present address: Center for Drug Discovery, Department of Pathology & Immunology, Baylor College of Medicine, Houston, TX 77030, USA. §Present address: The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui 230001, PR China. ¶Present address: Department of Liver Surgery, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, PR China. #Present address: Sanofi U.S., Cambridge, MA 02139, USA. **Present address: UPMC Genome Center, Pittsburgh, PA 15232, USA.

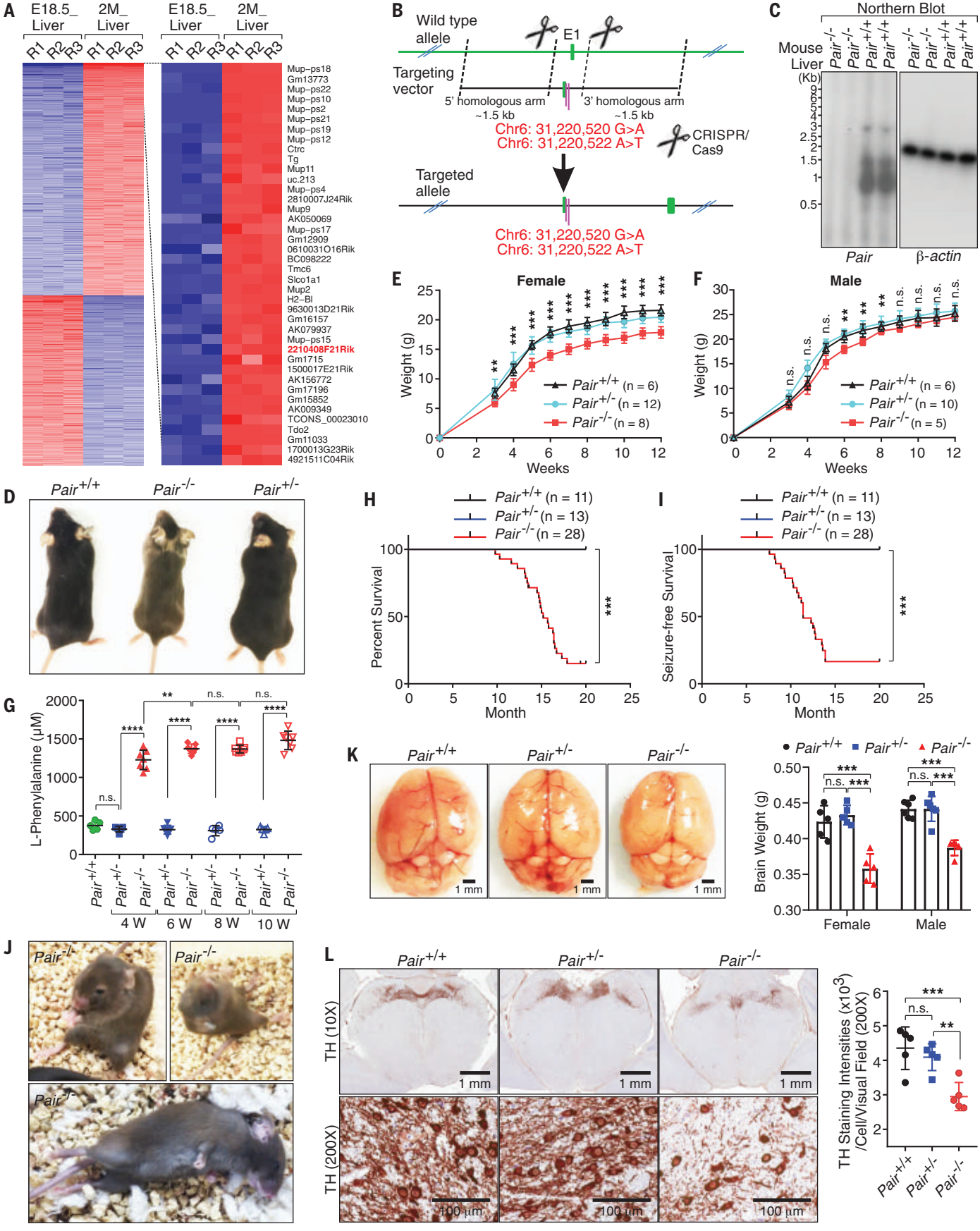

A
E18.5_ Liver
2M_ Liver
E18.5_ Liver
2M_ Liver
R1 R2 R3 R1 R2 R3 R1 R2 R3 R1 R2 R3
Mup−ps18
Gm13773
Mup−ps22
Mup−ps10
Mup−ps2
Mup−ps21
Mup−ps19
Mup−ps12
Ctrc
Tg
Mup11
uc.213
Mup−ps4
2810007J24Rik
Mup9
AK050069
Mup−ps17
Gm12909
0610031O16Rik
BC098222
Tmc6
Slco1a1
Mup2
H2−Bl
9630013D21Rik
Gm16157
AK079937
Mup−ps15
2210408F21Rik
Gm1715
1500017E21Rik
AK156772
Gm17196
Gm15852
AK009349
TCONS_00023010
Tdo2
Gm11033
1700013G23Rik
4921511C04Rik
B
Wild type allele
E1
Targeting vector
5' homologous arm
~1.5 kb
3' homologous arm
~1.5 kb
Chr6: 31,220,520 G>A
Chr6: 31,220,522 A>T
CRISPR/ Cas9
Targeted allele
Chr6: 31,220,520 G>A
Chr6: 31,220,522 A>T
C
Northern Blot
Mouse Liver
Pair-/- Pair-/- Pair+/+ Pair+/+ Pair-/- Pair-/- Pair+/+ Pair+/+
(Kb)
9
6
5
4
3
2.5
2
1.5
1
0.5
Pair
β-actin
D
Pair+/+
Pair-/-
Pair+/-
E
Female
Weight (g)
Weeks
Pair+/+ (n = 6)
Pair+/- (n = 12)
Pair-/- (n = 8)
F
Male
Weight (g)
Weeks
Pair+/+ (n = 6)
Pair+/- (n = 10)
Pair-/- (n = 5)
G
L-Phenylalanine (μM)
4 W
6 W
8 W
10 W
H
Pair+/+ (n = 11)
Pair+/- (n = 13)
Pair-/- (n = 28)
Percent Survival
Month
I
Pair+/+ (n = 11)
Pair+/- (n = 13)
Pair-/- (n = 28)
Seizure-free Survival
Month
J
Pair-/-
Pair-/-
Pair-/-
K
Pair+/+
Pair+/-
Pair-/-
1 mm
Brain Weight (g)
Female
Male
L
Pair+/+
Pair+/-
Pair-/-
TH (10X)
TH (200X)
1 mm
100 μm
TH Staining Intensities (x10^3) /Cell/Visual Field (200X)

Fig. 1. *Pair*$^{-/-}$ mouse mimics human PKU disease. (**A**) lncRNA profiling of livers from E18.5 embryos and from 2-month-old adult mice (n = 3). 2210408F21Rik is highlighted in red. (**B**) Schematic of using CRISPR-Cas9 to generate a *Pair*$^{-/-}$ mouse model. (**C**) Northern blot detection of the expression of *Pair* in the indicated mouse livers. β-actin was used as a loading control. (**D**) Representative images of 12-month-old *Pair* mice. (**E** and **F**) Comparison of body weights of the indicated female (E) and male (F) mice at the age of 3 to 12 weeks. Data are shown as mean ± SD, one-way ANOVA. (**G**) Blood Phe concentrations in the indicated mice was tested every 2 weeks starting from 4 weeks of age (n = 5 *Pair*$^{+/+}$, n = 5 *Pair*$^{-/-}$, n = 7 *Pair*$^{+/-}$ animals). Data are shown as mean ± SD, one-way ANOVA. (**H**) Cumulative survival curve of cohorts of indicated littermates (log-rank test). (**I**) Cumulative seizure-free survival curve of cohorts of indicated littermates (log-rank test). (**J**) Representative images of *Pair*$^{-/-}$ mice experiencing seizures. (**K**) Left: representative images of brains from indicated mice at the age of 12 months (n = 5). Right: scatter plots represent brain weight quantification for both female and male indicated littermates at the age of 12 months (n = 5). Mean ± SD, one-way ANOVA. (**L**) Coronal sections of adult mouse brains subjected to immunohistochemical staining and quantitative data for TH$^+$ neurons in the substantia nigra compact/ventral tegmental area of the indicated mice at the age of 12 months (n = 5). Mean ± SD, one-way ANOVA. Top row, 10× magnification; bottom row, 200× magnification. n.s., not significant at $P > 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

in liver cancer (*20*), Northern blotting indicated that *HULC*, similarly to *Pair*, is specifically expressed in normal liver tissues (fig. S4, C to E), suggesting the biological relevance of *HULC* in liver homeostasis and function.

RNA immunoprecipitation (RIP) assay confirmed that *Pair* and *HULC* associate with PAH protein in mouse and human livers, respectively (fig. S5, A to C). We demonstrated that the regulatory (amino acids 1 to 142) and catalytic (amino acids 142 to 411) domains of PAH are required for PAH-*Pair* interactions (fig. S5, D to G). Next, using primary cultured mouse or human hepatocytes, we demonstrated that PAH protein faithfully colocalizes with *Pair* or *HULC* in the cytosol but not with another cytosolic lncRNA, *Tug1* (Fig. 2C and fig. S5, H to I).

We substituted each nucleotide between *Pair* (nucleotides 460 to 496) and *HULC* (nucleotides 183-216) using the most common transition and transversion types (*21*) and expressed WT or mutant *Pair* or *HULC* in *Pair*$^{-/-}$ or *HULC*-deficient hepatocytes (fig. S6, A and B). RIP assay suggested that *Pair* (nucleotides 470 to 488) and *HULC* (nucleotides 183 to 200) are required for PAH-*Pair* and PAH-*HULC* interactions (fig. S6, A and B). Furthermore, *Pair* 479A>G and *HULC* 191A>G abolished PAH-*Pair* and PAH-*HULC* interactions, respectively (fig. S6, A and B).

SHAPE assays (*22*, *23*) indicated that *HULC* U190 and A191 (capillary electrophoresis sizes 290 and 289) exhibited chemical labeling, suggesting the presence of a loop structure flanked by low chemical probing (Fig. 2D and fig. S6C). Furthermore, *HULC* harbors three additional stem-loop structures, as revealed by SHAPE assays (fig. S7, A to D). These three stem-loop structures exhibited distinct structures compared with the stem-loop at nucleotides 184 to 216 (fig. S7, E to H). The chemical labeling of *HULC*^{A191} was verified by in vivo SHAPE (fig. S8, A and B). We further looked at the secondary structure of *Pair*, finding that *Pair* nucleotides 467 to 484 also exhibited a stem-loop structure, with A479 showing robust chemical probing (fig. S9, A to C). *Pair* nucleotides 467 to 484 exhibited a similar three-dimensional structure as *HULC* nucleotides 184 to 216 but not the above three stem-loops of *HULC* (fig. S9D).

WT *Pair* and WT *HULC*, but not the *Pair* 479A>G or *HULC* 191A>G mutants (referred to as the *Pair* mutant or the *HULC* mutant, respectively), associated with recombinant PAH, as revealed by RNA electrophoretic mobility shift assay (EMSA) assay (Fig. 2, E and F). Expression of MS2-tagged WT *Pair/HULC*, but not the mutants, rescued the association with PAH protein in *Pair*$^{-/-}$ hepatocytes (Fig. 2G and fig. S10A). We then performed a rescue CLIP assay by expressing exogenous WT *Pair/HULC* or mutants in mouse *Pair*$^{-/-}$ hepatocytes (Fig. 2H and fig. S10B). Exogenous *Pair* or *HULC* WT, but not the mutants, associated with PAH similarly to endogenous *Pair* or *HULC* (Fig. 2H). Similar amounts of PAH protein were immunoprecipitated using anti-PAH antibody (Fig. 2H, bottom). These findings suggested that the human lncRNA *HULC* and mouse *Pair* both associate with PAH.

HULC/Pair modulate the enzymatic activity of PAH

To demonstrate the underlying molecular mechanisms of *HULC/Pair* in the enzymatic activity of PAH, we first determined that there are ~800 *HULC* RNA, ~700 *Pair* RNA, and ~4000 PAH protein molecules per human or mouse hepatocyte (fig. S10, C to E). In *Pair*$^{-/-}$ or *HULC*-deficient hepatocytes, exogenous expression of *Pair/HULC* in a dose-dependent manner led to the reduction of cellular Phe concentrations (fig. S10, F and G). The concentrations of amino acids other than Phe and Tyr were minimally affected upon *Pair* depletion (fig. S10H).

The enzymatic activity of PAH requires the presence of a cofactor, BH_4 (*24*). Mutations of PAH affecting PAH-Phe or PAH-BH_4 interactions impair the catalytic activity of PAH (*25*). Previous research indicated that the Phe molecule associates with the regulatory domain of PAH as an allosteric activator (*26*–*28*). Using the crystal structure information of PAH (PDB 6HYC), protein structural bioinformatics analysis suggested that *HULC* nucleotides 184 to 216 associate with the regulatory domain of PAH and allosteric Phe (Fig. 3A). *HULC* A^{191} forms hydrogen bonds with both Thr^{63} and His^{64} (Fig. 3A). The mutation of A^{191} to G^{191} causes a change at position 6 from an amino group (6-NH_2) to a carbonyl group (6-CO). This change leads to the loss of the hydrogen bond between A^{191} and His^{64} because carbonyl groups are hydrogen bond acceptors, whereas amino groups are often hydrogen bond donors. The amino group (2-NH_2) of G^{191} may also cause steric hindrance with Thr^{63}. Phe forms hydrogen bonds with PAH through Asn^{61} and Leu^{62} and it also forms stacking interactions with A^{191} (Fig. 3B). These interactions stabilize the whole structural complex in a conformation that makes the active site fully accessible to Phe as a substrate and BH_4 as a cofactor. Our findings suggested that *HULC* serves as an important factor that stabilizes the interaction between allosteric Phe and PAH, as previous hypothesized (*26*). Aside from A^{191}, *HULC* interacts with PAH through several other residues to achieve the binding specificity of the PAH-*HULC* interaction: A^{195} forms a hydrogen bond with Tyr^{166}; A^{214} forms a hydrogen bond with Arg^{157}; and G^{202} also interacts with Tyr^{154} through a hydrogen bond (Fig. 3C). *Pair* adapts a binding mode to PAH that is similar to *HULC*: Both *Pair* and *HULC* have two nucleotides that stick out to form a "T" shape and exhibit stacking interactions with His^{64} of PAH, and these two nucleotides (A and U) stabilize each other through stacking interactions (Fig. 3D and fig. S10I). The only difference between *Pair* and *HULC* with regard to this behavior is that the order of the sequences of these two nucleotides in *Pair* is UA; in *HULC*, the order is AU (Fig. 3D).

We reasoned that the *HULC*-PAH interaction may facilitate the binding of PAH to Phe or BH_4. To address this hypothesis, we synthesized biotinylated Phe and BH_4 (referred to as Bio-Phe and Bio-BH_4, respectively) (fig. S11A). Compared with *Pair*$^{+/+}$ livers, *Pair*$^{-/-}$ livers exhibited a similar abundance of PAH protein; however, the PAH-Phe and PAH-BH_4 interactions were impaired after *Pair* depletion (Fig. 3E). We then determined that digoxin-tagged *HULC* showed detectable interactions

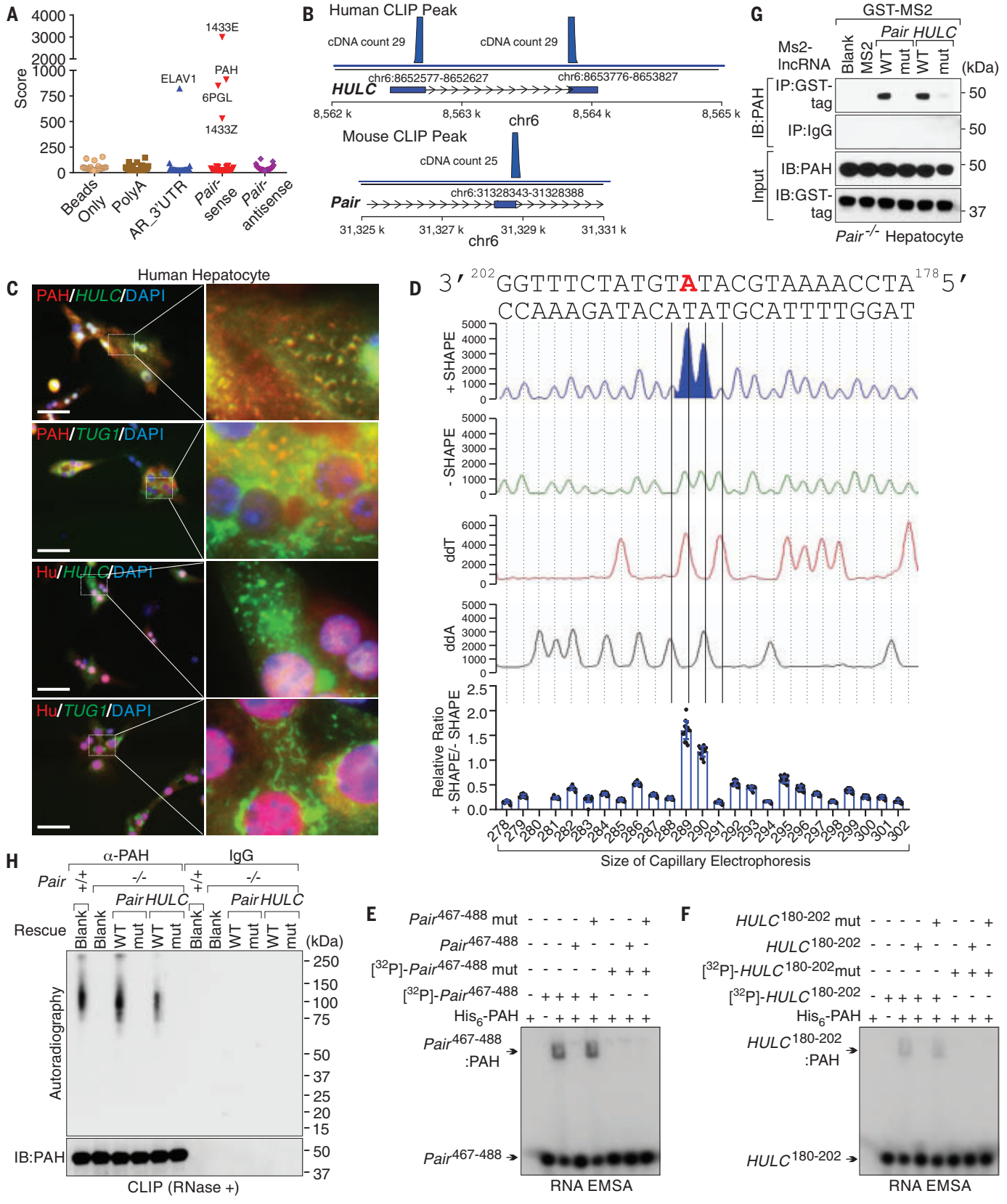


Fig. 2. *Pair* and *HULC* associate with PAH. (**A**) Protein candidates interacting with *Pair* were revealed by LC-MS. The *x*-axis indicates different experimental groups. (**B**) PAH-binding sites along *HULC* and *Pair* identified by CLIP assay. The figure represents the read coverage along *HULC* and *Pair* transcripts obtained by mapping Sanger sequencing reads to the representative transcript. cDNA counts are shown. (**C**) Representative images of immunolabeling with fluorescent

in situ hybridization to simultaneously detect colocalization of the indicated RNA and protein molecules in human primary hepatocytes. *TUG1* and Hu were used as RNA and protein controls, respectively. Scale bars, 50 μm. (**D**) Resolution of SHAPE reactivity by capillary electrophoresis. Four-color electropherograms for the products of one multicolor run by 6-FAM-labeled NMIA modified RNA (+SHAPE), VIC-labeled control RNA (-SHAPE) of *HULC* nucleotides 178 to 202, NED-labeled ddA, and a PET-labeled ddT are shown. Bottom panel: normalized SHAPE reactivity of *HULC* nucleotides 178 to 202. *x*-axis indicates the sizes of capillary electrophoresis. Data are shown as mean ± SD of $n = 15$ independent experiments. (**E** and **F**) EMSA using recombinant His-tagged PAH and [γ-32P]–labeled *Pair* nucleotides 467 to 488 (E) or *HULC* nucleotides 180 to 202 (F) WT or mutant oligonucleotides. Unlabeled *Pair* or *HULC* WT or mutant RNA oligonucleotides were included as competitors. (**G**) MS2-TRAP assay using *Pair*$^{-/-}$ hepatocytes expressing indicated plasmids performed by immunoblotting using the indicated antibodies. (**H**) Immunoblotting (bottom) or autoradiography (top) of CLIP assay using the indicated hepatocytes expressing the indicated plasmids.

with bacterially expressed PAH proteins but undetectable associations with Bio-Phe or Bio-BH_4 (fig. S11B). WT PAH was included as a positive control (fig. S11B). PAH G46S, F55L, and P281L have been suggested to impair interactions between Bio-Phe and/or Bio-BH_4 (*29*, *30*), which we confirmed (fig. S11B). lncRNA *LINK-A* and the interaction between *LINK-A* and PIP_3 (*31*) were included as negative controls (fig. S11B).

The N-terminal regulatory domain of PAH can undergo a conformational change to switch between "open" and "closed" states (*30*). We hypothesized that the binding of *HULC* may stabilize the PAH-Phe-*HULC* complex and stabilize the PAH protein in the open state (fig. S11C). We applied limited proteolysis (LiP) followed by liquid chromatography–mass spectrometry (LC-MS) analysis (LiP-MS) (*32*) to address this hypothesis (fig. S11C). Open-loop regions (gray) and Lip-resistant regions (dark blue) were determined (Fig. 3F and table S6). Region amino acids 57 to 66 showed resistance to LiP in the presence of *HULC* but not the *HULC* mutant, suggesting that this region associates with *HULC* (Fig. 3F, top panel, magenta). Amino acids 57 to 66 of PAH exhibited recovery in the presence of Phe (Fig. 3F, bottom panel, magenta), which was consistent with the structural modeling, demonstrating that Thr^{63} and His^{64} associate with *HULC* and Phe as a complex. Furthermore, LiP-MS suggested a few regions that were dynamically regulated upon *HULC* binding (Fig. 3F, green).

Our data thus suggested that *HULC* associates with PAH in vitro and facilitates the potential conformational change of PAH. Consistent with this notion, expression of WT *Pair* or *HULC* rescued PAH-Phe and PAH-BH_4 interactions and reversed cellular Phe accumulation, whereas the *Pair* and *HULC* mutants failed to do so (Fig. 3, G and H, and fig. S11, D and E).

To determine the functional role of the *HULC*-PAH interaction in vivo, *HULC*-deficient human induced pluripotent stem cells (hiPSCs) were further differentiated into hepatocytes expressing WT *HULC* or the A191G mutant (fig. S12A). Expression of WT *HULC* or the A191G mutant showed no detectable effects on the abundance of PAH protein (fig. S12B). Upon depletion of *HULC*, the conversion of ^{14}C-Phe to ^{14}C-Tyr was reduced with concurrent elevation of cellular Phe concentrations (fig. S12, C and D). Expression of WT *HULC*, but not the A191G mutant, restored the enzymatic activity of PAH and reversed the cellular accumulation of Phe (fig. S12, C and D). Similarly, WT PAH, but not the TH63-64PN mutant, expressed in *PAH*-deficient hiPSC-hepatocytes rescued the enzymatic deficiency of PAH (fig. S12, E to H). Taken together, our findings suggest that *Pair*-PAH and *HULC*-PAH interactions facilitate the PAH-driven catalysis of Phe to Tyr.

HULC mimics restore PAH enzymatic activity

We reasoned that supplementing lncRNAs might improve the catalytic activities of the PAH mutants, leading to reduced serum Phe concentrations and improved symptoms in patients with PKU. We designed Scramble (Scr) and *HULC* mimics representing WT *HULC* nucleotides 181 to 201 and *HULC* A191G-mutated sequences for the following studies. We first selected 17 PAH mutants that were identified from patients with PKU and are known to affect the enzymatic activity of PAH (*33*, *34*) and collected the bacterially expressed WT and mutant PAH (Fig. 4A and fig. S13A). EMSA assay indicated that 13 of the 17 PAH mutants, as well as WT PAH, associated with the *HULC* mimics; conversely, PAH TH63-64PN, R157N, N207S, and S349L failed to associate with the *HULC* mimics (fig. S13B). The denatured WT PAH proteins were included as a negative control (fig. S13B). We further quantified the interactions between PAH proteins (WT/mutants) and *HULC* mimics (WT), finding that the PAH protein interacted with the *HULC* mimics with a K_d value of 131.6 nM (fig. S13C). PAH TH63-64PN, R157N, N207S, and S349L, but not the other PAH mutants, exhibited decreased binding affinities (fig. S13C).

We measured the binding affinities between PAH (WT/mutants) and Bio-Phe or Bio-BH_4 in the presence of *HULC* mimics (figs. S14 and S15, A to O) and summarized the change in binding affinities compared with PAH WT in Fig. 4, A and B. lncRNA mimics representing *LINK-A* 1100 to 1117 (*31*) were included as a negative control (Fig. 4, A and B). For WT PAH protein, the presence of *HULC* mimics enhanced PAH-Phe (Fig. 4A) and PAH-BH_4 interactions (Fig. 4B), whereas *HULC* mutant mimics failed to do so (Fig. 4, A and B, and figs. S14 and S15, A to O). The PAH mutants we tested all exhibited impaired binding affinities toward Phe and/or BH_4 compared with WT PAH protein (Fig. 4, A and B, and figs. S14 and S15, A to O). The presence of *HULC* mimics, but not *HULC* mutant, enhanced the affinity of Phe and/or BH_4 to the PAH mutants. A PAH catalytic pocket deletion mutant (Δ245-379) was included as a negative control (figs. S14O and S15O).

We further demonstrated that WT PAH effectively catalyzed Phe to Tyr, which was abolished in all 17 PAH mutants we tested (Fig. 4C, left section, bar #2). In the presence of *HULC* mimics, but not *LINK-A* mimics or *HULC* mutant, WT PAH showed enhanced enzymatic activity in converting Phe to Tyr (Fig. 4C, middle section, bar #2). Furthermore, 11 of the 17 PAH mutants also showed improved enzymatic activities in converting Phe to Tyr (Fig. 4C, middle section, red). The enzymatic activities of the rest of the PAH mutants were not affected by *HULC* mimics (Fig. 4C, middle section, light blue and brown).

We then determined the k_{cat} (also called the turnover number) of bacterially expressed human WT and mutant PAH in the presence of Scr, WT *HULC*, or mutant *HULC* mimics (Fig. 4D and fig. S16, A to S). Recombinant PAH proteins with deleted catalytic domains were included as a negative control (fig. S16A). In the presence of the WT *HULC* mimics, the k_{cat} of WT PAH was increased (Fig. 4D and fig. S16B). All PAH mutants we tested exhibited impaired k_{cat} under identical conditions (Fig. 4D and fig. S16, C to S). The presence of WT *HULC* mimics, but not mutant *HULC* mimics, enhanced the k_{cat} of PAH F39L, A47V, F55L, I65S, P275L, P281L, I283N, F299C, A300S, I318T, and R408W, but not the enzymatic activities of PAH TH63-64PN, R157N, N207S, and S349L (Fig. 4D and fig. S16, C to S). Therefore, our data suggested that the presence of *HULC* facilitates the enzymatic activity of WT PAH and a cohort of mutants observed in patients with PKU.

HULC/Pair enhances the enzymatic activity of the PAH R408W mutant

About 19.2 to 73% of patients with PKU harbor a R408W mutant (*35*, *36*), and patients with this mutation respond poorly to currently available treatment options. We hypothesized that overexpression of *HULC/Pair* might improve the enzymatic activity of PAH R408W. To

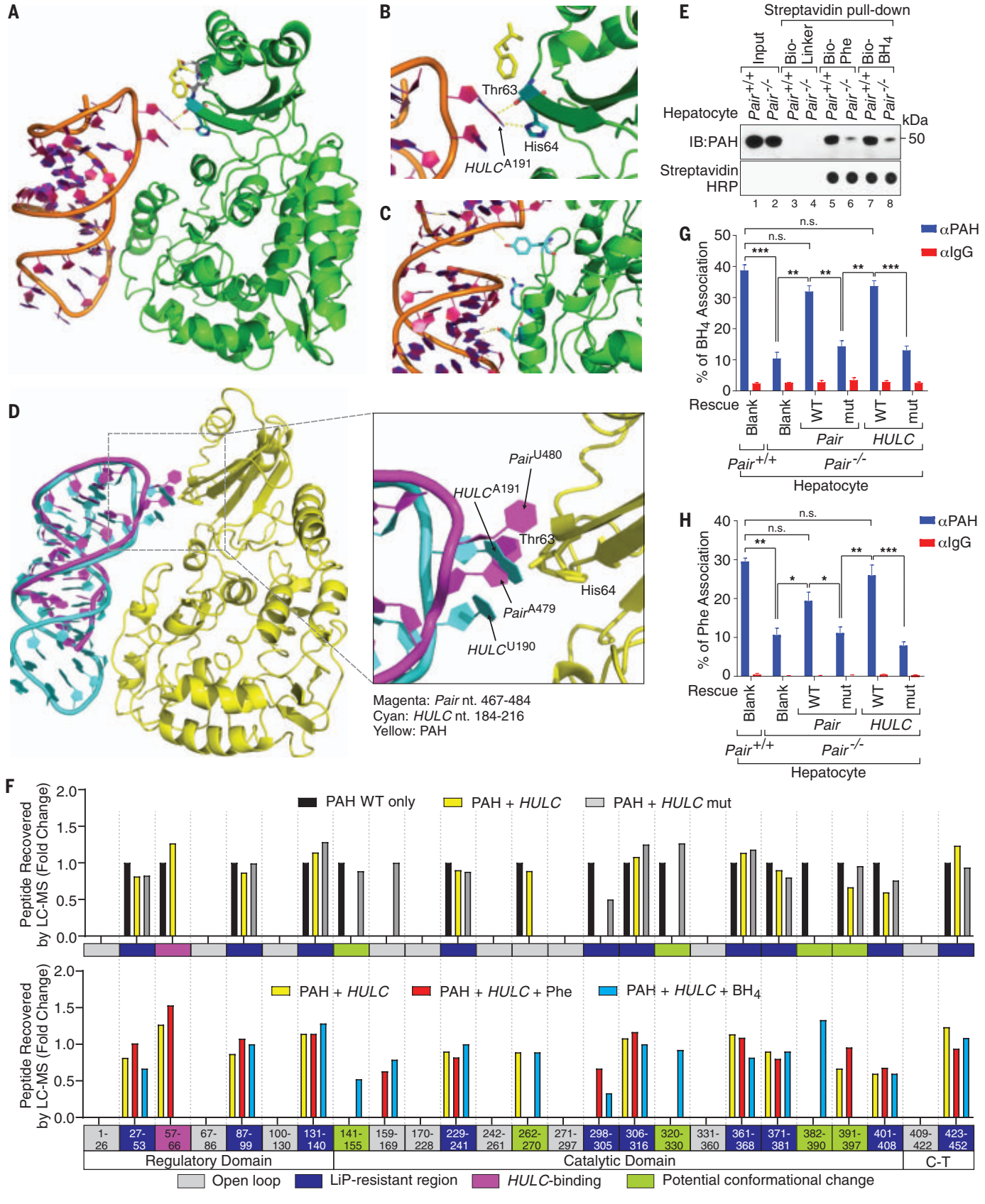


Fig. 3. *HULC* and *Pair* modulate the enzymatic activities of PAH. (**A**) *HULC* binds to PAH to stabilize the allosteric Phe-induced open conformation of PAH. (**B**) Magnified view of *HULC*'s interaction with the regulatory domain of PAH. *HULC* A^{191} forms hydrogen bonds with His^{64} and polar interaction with Thr^{63}. Phe

forms stacking interaction with *HULC* A^{191}. Color representation in both (A) and (B): green indicates PAH; orange and magenta indicate *HULC*; yellow indicates allosteric Phe; and cyan indicates Thr^{63} and His^{64}. (**C**) *HULC* A^{195} forms a hydrogen bond with Tyr^{166}; *HULC* A^{214} forms a hydrogen bond with Arg^{157}; *HULC* G^{202} forms a hydrogen bond with Tyr^{154}. Green indicates PAH; orange and magenta indicate *HULC*; and Cyan indicates Tyr^{154}, Arg^{157}, and Tyr^{166}. (**D**) Superimposition of the three-dimensional structures of *HULC* nucleotides 184 to 216 (cyan) and *Pair* nucleotides 467 to 484 (magenta) after docking to PAH (yellow). *Pair* A^{479}, U^{480} and *HULC* U^{190}, A^{191} are indicated by arrows. (**E**) Streptavidin pull-down using Bio-Phe/Bio-BH_4 followed by immunoblotting detection using anti-PAH antibody in the indicated hepatocytes. (**F**) Fold change of peptides recovered from LiP-MS. Top panel: comparison of PAH WT only, PAH incubated with *HULC*, and PAH incubated with a *HULC* mutant sequence. Bottom panel: comparison of PAH incubated with *HULC*, PAH incubated with *HULC* and Phe, and PAH incubated with *HULC* and BH_4. The *x*-axis indicates the amino acid position of full-length PAH; the *y*-axis indicates the fold change of peptide recovery number. (**G** and **H**) ELISA measurement of the percentage of PAH-associated BH_4 (G) or PAH-associated Phe (H) in the indicated hepatocytes expressing the indicated mimics. Data are shown as mean ± SEM of $n = 5$ independent experiments, one-way ANOVA. n.s., not significant at $P > 0.05$; $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

address this, we first generated a $Pah^{R408W/R408W}$ mouse strain using CRISPR/Cas9 (fig. S17, A and B). Similar to $Pair^{-/-}$ mice, $Pah^{R408W/R408W}$ mice exhibited hyperpigmentation, growth retardation, seizures, and elevated blood Phe concentrations, with no detectable alterations in the protein stability of PAH or the expression of *Pair* (fig. S17, C to H). These observations provided genetic evidence confirming that *Pair/HULC* and PAH act in a linear pathway. In primary cultured mouse hepatocytes isolated from $Pair^{-/-}$ or $Pah^{R408W/R408W}$ mice, we determined cellular Phe, Tyr, and tryptophan (Trp) concentrations, finding that expression of WT *Pair*/*HULC*, but not the mutant (fig. S17I), restored PAH-Phe and PAH-BH_4 interactions and cellular Phe and Tyr concentrations, with no effect on Trp status (fig. S17, I to L).

We collected skin fibroblasts from healthy donors harboring WT *PAH* genes and from a patient with PKU (ID: GM02406) who harbored F299C and R408W mutations. We reprogrammed these fibroblasts into hiPSCs. These hiPSCs were further induced into hepatocytes (fig. S18, A and B). Two hiPSC clones were used, referred to as PKU #7 and PKU #17 (fig. S18, A to C). These hiPSC-derived hepatocytes exhibited similar expression of *HULC*, *PAH,* and hepatic markers (fig. S18, B and C). The hiPSC-derived hepatocytes derived from both PKU #7 and PKU #17 clones exhibited increased Phe concentrations compared with the control (healthy donor), confirming PAH enzymatic activity deficiency (fig. S18D). Expression of full-length *HULC* in these hiPSC-derived hepatocytes reduced Phe concentrations in PKU #7 and PKU #17 hepatocytes (fig. S18, D and E). This suggested that a supply of *HULC* might enhance the enzymatic activity of PAH in F299C and R408W mutants.

GalNAc-HULC mimics improve Phe metabolism in mice

To design lncRNA mimics that could provide therapeutic value in vivo, we synthesized Scr, WT, and mutant *HULC* mimics using 2′-fluoro–modified RNA monomers to provide nuclease resistance in vivo. The potential secondary structure of WT or mutant *HULC* mimics remained identical (fig. S19A). Both full-length *HULC* and the mimics exhibited similar binding affinities for PAH (fig. S19B). Furthermore, the *HULC* A191G mutant and *HULC* mutant mimics both abolished these interactions (fig. S19B).

To facilitate the liver enrichment of *HULC* mimics, we applied three types of *HULC* mimics: *HULC* mimics alone, peptides representing ApoE1 (Apo)–tagged *HULC* mimics, and N-acetylgalactosamine (GalNAc)–tagged *HULC* mimics through intravenous or subcutaneous injection. GalNac-conjugated oligonucleotides have recently been suggested to assist with liver-targeted small interfering RNA delivery (*37*). The 3′-triantennary GalNAc was conjugated to the *HULC* mimic.

We considered that a 3-day treatment trial (Fig. 5A) would serve as the most convenient method to determine the efficacy of these mimics. Because PAH R408W–harboring patients respond to current BH_4 supplementation treatments poorly, we treated female and male $Pah^{R408W/R408W}$ mice with the indicated mimics (fig. S19, C to D). GalNac-*HULC* administered by intravenous injection exhibited the highest efficacy in reducing excessive Phe concentrations in both female and male $Pah^{R408W/R408W}$ mice (fig. S19, C to D).

We then determined the concentrations of biotin-labeled *HULC* or GalNAc-*HULC* mimics in the major organs, finding that a substantial amount of biotin-labeled GalNAc-tagged *HULC* mimics was detected in the liver between 3 and 72 hours after dosing, whereas the lungs and spleen exhibited no detectable accumulation of GalNAc-*HULC* (fig. S19, E to G). Biotin was undetectable in the liver 12 hours after injection (fig. S19H).

As a proof of concept, we applied GalNAc-tagged Scr or *HULC* mimics using short-term (3-day) and medium-term (12-day) treatment regimens (Fig. 5A). To rule out the indirect effects of the GalNAc tag, GalNAc-*HULC* mutant mimics were included (Fig. 5B). Both male and female $Pair^{-/-}$ mice exhibited reduced serum Phe concentrations 24 hours after injection with GalNAc-*HULC* mimics, but not after injection with the Scr or *HULC* mutant mimics (Fig. 5B). During the medium-term treatment, administration of GalNAc-*HULC* mimics reduced blood Phe concentrations in $Pair^{-/-}$ mice throughout the treatment term (Fig. 5C). In the $Pah^{R408W/R408W}$ mice, administration of GalNAc-*HULC* mimics similarly reduced blood Phe concentrations during the short-term and medium-term treatments (Fig. 5, D to E). $Pah^{R408W/R408W}$ mice subjected to GalNAc-*HULC* mimics showed increased blood Tyr concentrations compared with animals given scramble mimics (Fig. 5F). To evaluate Phe clearance capacity, $Pah^{R408W/R408W}$ mice were subjected to pretreatment with Scr or *HULC* mimics followed by a Phe challenge (Fig. 5G). The area under the curve showed a reduction in Phe concentrations after GalNAc-*HULC* mimic treatment compared with GalNAc-Scr (Fig. 5H).

To determine whether the administration of GalNAc-*HULC* facilitates higher tolerance to dietary Phe intake in PKU animals, $Pah^{R408W/R408W}$ were fed a Phe-free diet for 3 days, followed by water containing increasing doses of Phe. Administration of GalNAc-*HULC* mimics allowed $Pah^{R408W/R408W}$ animals to maintain relatively low blood Phe concentrations (<600 μM) upon Phe challenge up to 3.0 mg/ml (Fig. 5I), suggesting that GalNAc-*HULC* mimics may also be able to improve tolerance to dietary Phe.

Sapropterin, a synthetic formulation of BH_4, has been used as an enzymatic enhancer for patients with HPA (*38*). Patients with PKU harboring R408W are resistant to sapropterin treatment (*14*). We tested the efficacy of sapropterin in $Pair^{-/-}$ and $Pah^{R408W/R408W}$ mice, finding that supplementation of sapropterin had no effect on the blood Phe of $Pair^{-/-}$ and $Pah^{R408W/R408W}$ animals (fig. S20, A and B). A combinational treatment of GalNAc-*HULC* and sapropterin had a cooperative effect in reducing the blood Phe concentrations in $Pah^{R408W/R408W}$ animals (Fig. 5J). These findings suggest that a supply of GalNAc-*HULC* mimics may enhance the association between PAH and BH_4, thereby improving the therapeutic effect of sapropterin. The *HULC* mimics showed no detectable effects on the body weight, liver function, or kidney function of $Pair^{-/-}$ and $Pah^{R408W/R408W}$ mice (fig. S20, C to G). Our data suggest that application of *HULC* mimics could enhance the enzymatic activities of certain mutated PAH proteins,

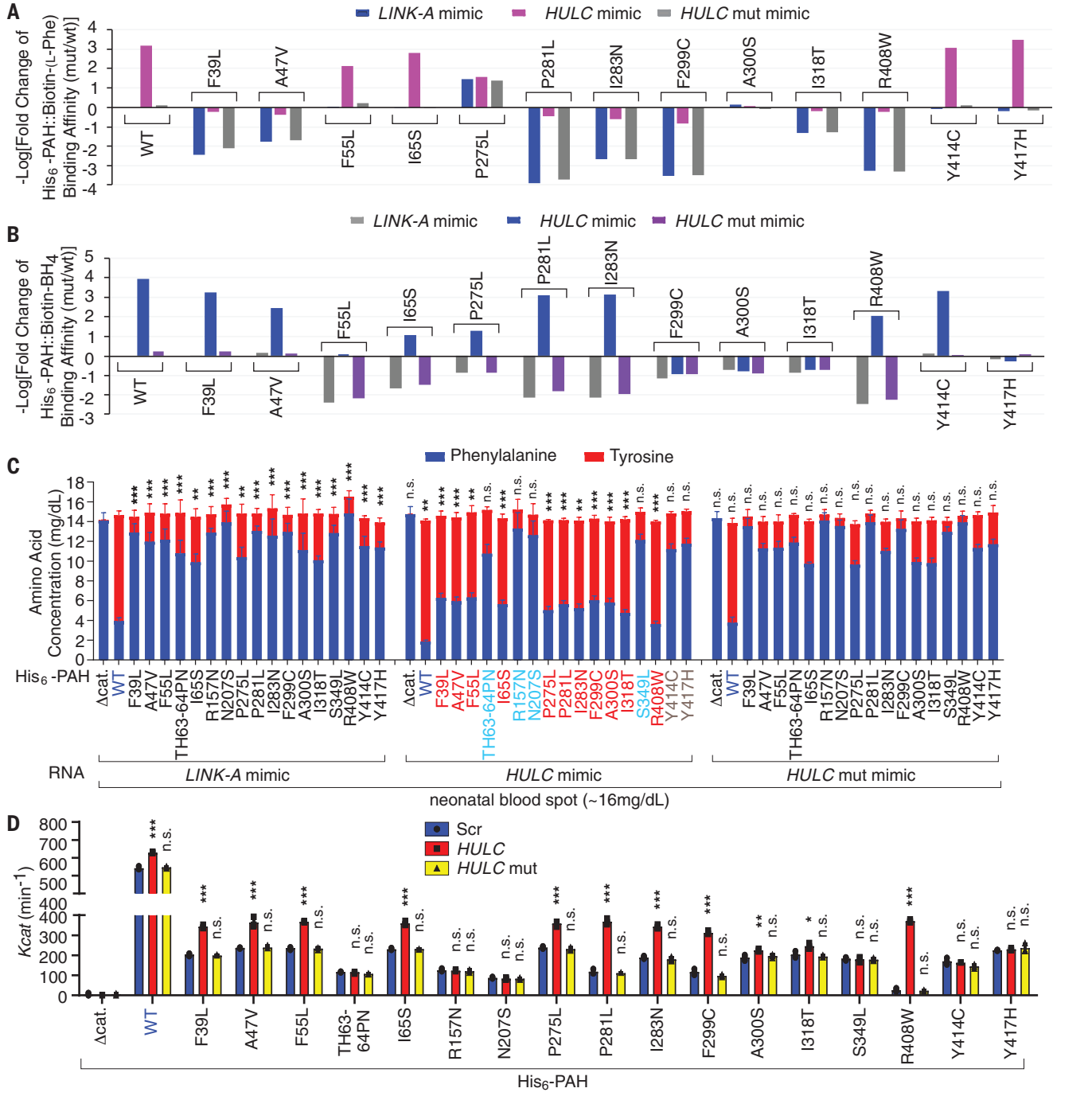


Fig. 4. *HULC* mimics facilitate PAH-Phe and PAH-BH_4 interactions. (**A** and **B**) Log_2 of relative fold change of His-PAH WT/mutants and Biotin-Phe (A) or His-PAH WT/mutants and biotin-BH_4 (B) binding affinity in the presence of the indicated lncRNA mimics. The fold change was normalized using His-PAH WT in the presence of the *LINK-A* mimic. (**C**) Phe and Tyr concentrations in neonatal blood spots in the presence of His-tagged PAH WT or indicated mutants and the indicated lncRNA mimics. Data are shown as mean ± SD of $n = 3$ independent experiments, Student's t test. The lncRNA mimic representing *LINK-A* nucleotides 1100 to 1117 was included as a negative control. (**D**) Determination of k_{cat} of recombinant PAH WT or indicated mutant proteins in the presence of the indicated mimics. Data are shown as the mean ± SD of $n = 3$ independent experiments, one-way ANOVA. n.s., not significant at $P > 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

offering a potential intervention for patients with PKU.

DISCUSSION

PKU has long been considered a single-gene disease with an autosomal-recessive inheritance pattern, in which mutations at the PAH locus lead to impaired enzymatic function and contribute to a hyperphenylalaninemia metabolic phenotype, resulting in cognitive phenotypes such as mental retardation. Recent genotype-phenotype correlation analyses of PAH mutations suggested a substantial discrepancy between genotypes and their predicted metabolic or cognitive phenotypes (*8*, *9*, *39*, *40*). Variations in genomic, epigenomic, transcriptomic, proteomic, and metabolic systems could all contribute to the PAH

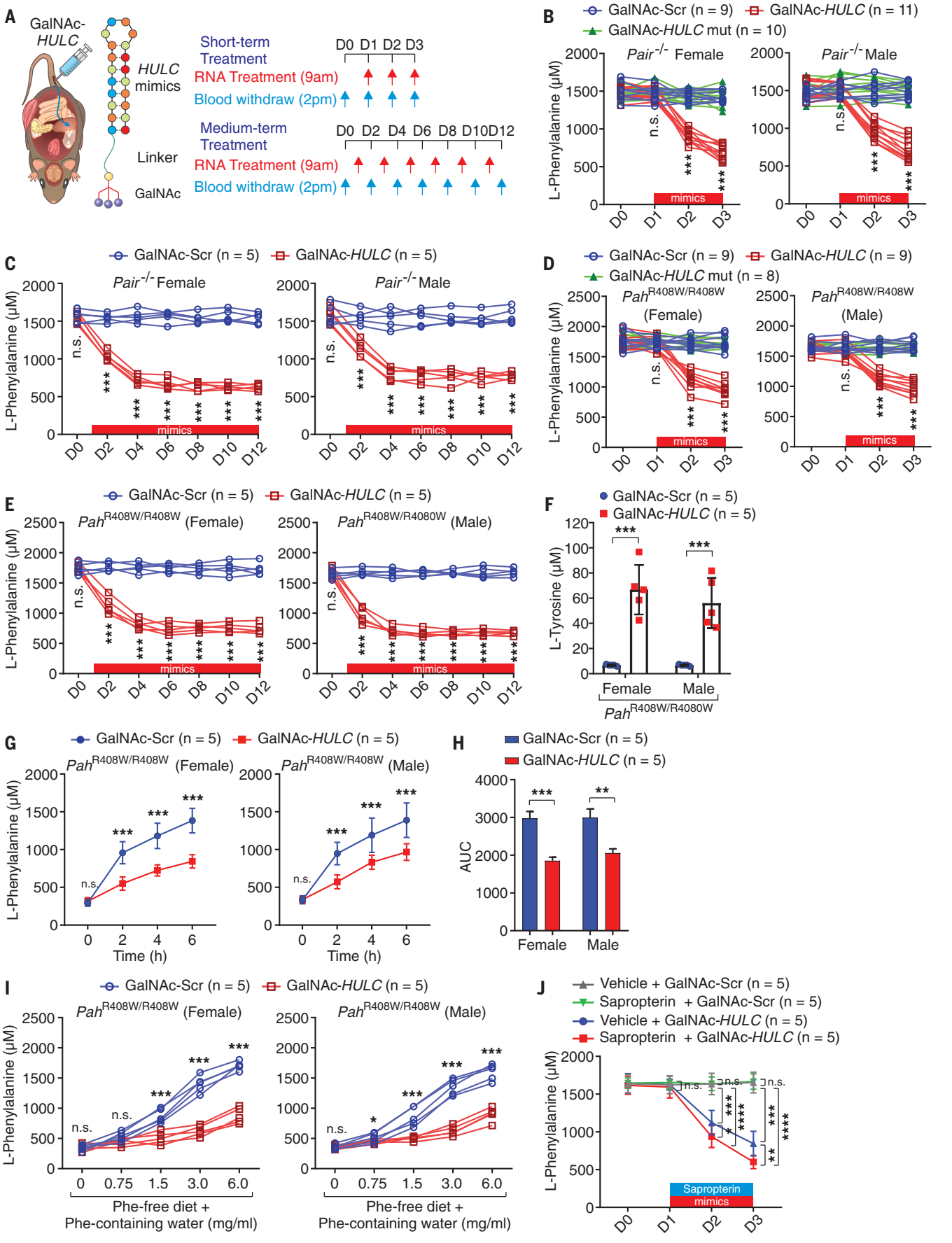

A
GalNAc-HULC
HULC mimics
Linker
GalNAc
Short-term Treatment
D0 D1 D2 D3
RNA Treatment (9am)
Blood withdraw (2pm)
Medium-term Treatment
D0 D2 D4 D6 D8 D10D12
B
GalNAc-Scr (n = 9)
GalNAc-HULC (n = 11)
GalNAc-HULC mut (n = 10)
Pair-/- Female
Pair-/- Male
L-Phenylalanine (μM)
mimics
C
GalNAc-Scr (n = 5)
GalNAc-HULC (n = 5)
D
GalNAc-Scr (n = 9)
GalNAc-HULC (n = 9)
GalNAc-HULC mut (n = 8)
PahR408W/R408W (Female)
PahR408W/R408W (Male)
E
GalNAc-Scr (n = 5)
GalNAc-HULC (n = 5)
PahR408W/R408W (Female)
PahR408W/R4080W (Male)
F
GalNAc-Scr (n = 5)
GalNAc-HULC (n = 5)
L-Tyrosine (μM)
Female
Male
PahR408W/R4080W
G
GalNAc-Scr (n = 5)
GalNAc-HULC (n = 5)
PahR408W/R408W (Female)
PahR408W/R408W (Male)
L-Phenylalanine (μM)
Time (h)
H
GalNAc-Scr (n = 5)
GalNAc-HULC (n = 5)
AUC
Female
Male
I
GalNAc-Scr (n = 5)
GalNAc-HULC (n = 5)
PahR408W/R408W (Female)
PahR408W/R408W (Male)
L-Phenylalanine (μM)
Phe-free diet + Phe-containing water (mg/ml)
J
Vehicle + GalNAc-Scr (n = 5)
Sapropterin + GalNAc-Scr (n = 5)
Vehicle + GalNAc-HULC (n = 5)
Sapropterin + GalNAc-HULC (n = 5)
L-Phenylalanine (μM)
Sapropterin
mimics

Fig. 5. GalNAc-*HULC* lncRNA mimics alleviate PKU symptoms. (**A**) Graphic illustration of GalNAc-tagged *HULC* mimics (left) and the treatment schedules (right). (**B**) Blood Phe concentrations were monitored every day for short-term treatment in female and male *Pair*$^{-/-}$ mice with the indicated mimics. Data were analyzed with Student's *t* test. (**C**) Blood Phe concentrations were monitored every other day for medium-term treatment in female and male *Pair*$^{-/-}$ mice treated with the indicated mimics. Data were analyzed with Student's *t* test. (**D**) Blood Phe concentrations were monitored every day for short-term treatment in female or male *Pah*$^{R408W/R408W}$ mice treated with the indicated mimics. Data were analyzed with Student's *t* test. (**E**) Blood Phe concentrations were monitored every other day for medium-term treatment in female or male *Pah*$^{R408W/R408W}$ mice treated with the indicated mimics. Data were analyzed with Student's *t* test. (**F**) Measurement of blood tyrosine concentrations for medium-term treatment in female and male *Pah*$^{R408W/R408W}$ mice treated with the indicated mimics, Data were analyzed with Student's *t* test. (**G** and **H**) Phe clearance test (G) or area under the curve (AUC) (H) of female or male *Pah*$^{R408W/R408W}$ mice subjected to a Phe-free diet and pretreatment with the indicated mimics. Data are shown as mean ± SD, Student's *t* test. (**I**) Phe tolerance test of female or male *Pah*$^{R408W/R408W}$ mice subjected to a Phe-free diet and pretreatment with indicated mimics for 3 days, followed by water containing 0, 0.75, 1.5, 3.0, or 6.0 mg/ml of Phe (with the dose increasing every 2 days). Data were analyzed with Student's *t* test. (**J**) Blood Phe concentrations were monitored every day for short-term treatment in *Pah*$^{R408W/R408W}$ mice subjected to sapropterin alone or in combination with the indicated mimics. Data are shown as mean ± SD, one-way ANOVA. n.s., not significant at $P > 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

deficiency in patients with PKU (*41*). It has been shown that patients, even siblings sharing identical mutant PAH genotypes, can have greatly differing cognitive and metabolic phenotypes (*10*). This strongly suggests that a PAH genotype cannot consistently and reliably predict the monogenic phenotype (*10*). Consistent with this notion, the recent discovery of biallelic mutations in the *DNAJC12* gene in patients with high Phe concentrations suggested the possibility that non-PAH genes contribute to this disease (*42*). Therefore, factors modulating PAH expression, PAH protein stability, BH_4 biogenesis, and PAH enzymatic activity may play important roles in maintaining PAH enzymatic proficiency. More than 95% of genetic mutations occur at the noncoding regions of the human genome, yet the functional importance of lncRNAs in human genetic diseases remains elusive. We have demonstrated that depletion of the lncRNA *Pair* leads to phenotypes that model human PKU, including hypopigmentation, growth retardation, seizures, and neuronal loss. Whether the presence of the *HULC* genomic variants contributes to the severity of PKU requires further investigation given the heterogeneity of *PAH* genotypes and the lack of knowledge of *HULC* genomic variant status in current data collections.

The nature of genetic diseases has resulted in the development of compensatory lifestyle considerations that have marked, potentially negative impacts on the overall quality of life of patients. BH_4 and its synthetic analog, sapropterin, have been used as enzyme cofactors for a subset of patients with PKU, particularly those with mild PKU (*43*–*45*). However, patients with PKU and Phe >1200 μmol/L are less likely to respond to BH_4 supplementation (*46*). PKU could potentially be genetically corrected using adenoviral or related vectors (*47*), but their specificity, stability, and the adenoviral genes responsible for immune response require further research studies (*48*). Enzyme replacement or substitution therapy, in which PAH or PAL fusion proteins (Palynziq) are intravenously injected, leads to decreased plasma Phe (*15*, *16*, *49*, *50*) but also to potentially concerning immune response–related symptoms experienced by some patients with PKU (*51*, *52*). Thus, the development of additional therapeutic strategies is needed, for example, an approach using liver-enriched lncRNA mimics that specifically enhance PAH-substrate and PAH-cofactor binding affinities and enhance enzymatic activity in vitro and in vivo.

The *HULC* mimics that we have developed target lncRNAs as a therapeutic strategy against a human inherited metabolic disorder. Considering lncRNA mimics as a potential therapeutic option has the following advantages: (i) flexibility in synthesizing lncRNA mimics according to the target sequence and tagging with organ-targeting peptides for tissue-specific distribution [the target sequence could be designed to improve the enzymatic activity of key enzyme(s) in other metabolic disorders]; (ii) profound stability in vivo because of the modified covalent bonds between nucleotides, which are resistant to DNase and RNase; (iii) liver-enriched tissue distribution assisted by the GalNAc-tag; and (iv) low potential for organ toxicity, as supported by our observation that administration of lncRNA mimics resulted in no detectable effects on liver or kidney function. It is likely that combining *HULC* mimics with current dietary restrictions, sapropterin supplements, and enzyme replacement or substitution therapies would further improve patient outcomes, particularly for patients with severe PKU.

Low conservation between mouse and human lncRNAs has greatly hindered the discovery of disease-driving lncRNAs and their roles in human diseases; however, it is possible that RNA "domains" might be conserved between human and mouse lncRNAs. *HULC* and *Pair* associate with the PAH protein at the N-terminal regulatory domain. The presence of *HULC* could rescue PAH enzymatic activity in *Pair*-deficient cells and vice versa, shedding light on the therapeutic potential of the vast, currently unconsidered, and poorly conserved lncRNAs.

PAH is self-regulated by the N-terminal regulatory domain. The crystal structure of PAH suggests that the C-terminal helix is required for the tetramerization of PAH. *HULC/Pair* is unlikely to affect the oligomerization of PAH, as supported by the finding that *HULC* mimics showed no notable effect on the enzymatic activities of PAH Y414C and Y417H. The N-terminal regulatory domain can undergo a conformational change to switch between open and closed states (*30*). We reason that the binding of *HULC* to the N-terminal regulatory domain of PAH may stabilize the PAH-Phe-*HULC* complex, causing the PAH protein to operate in the open state and allowing Phe and BH_4 to access the enzyme. Our finding is consistent with the previous hypothesis that an "unknown" allosteric factor stabilizes the interaction between PAH and Phe, serving as a therapeutic chaperone (*26*). Overall, our findings suggest the functional importance of lncRNAs in Phe metabolism.

REFERENCES AND NOTES

1. I. Følling, *Acta Paediatr. Suppl.* **407**, 4–10 (1994).
2. N. Blau, *Hum. Mutat.* **37**, 508–515 (2016).
3. Medical Research Council Steering Committee for the MRC/DHSS Phenylketonuria Register, *Br. Med. J. (Clin. Res. Ed.)* **282**, 1680–1684 (1981).
4. F. J. van Spronsen *et al.*, *Lancet Diabetes Endocrinol.* **5**, 743–756 (2017).
5. K. M. Camp *et al.*, *Mol. Genet. Metab.* **112**, 87–122 (2014).
6. A. E. Martynyuk, F. J. van Spronsen, E. A. Van der Zee, *Mol. Genet. Metab.* **99**, S100–S105 (2010).
7. A. L. Pey, L. R. Desviat, A. Gámez, M. Ugarte, B. Pérez, *Hum. Mutat.* **21**, 370–378 (2003).
8. E. Kayaalp *et al.*, *Am. J. Hum. Genet.* **61**, 1309–1317 (1997).
9. H. Murad, A. Dabboul, F. Moassas, D. Alasmar, W. Al-Achkar, *Gene* **528**, 241–247 (2013).
10. C. R. Scriver, P. J. Waters, *Trends Genet.* **15**, 267–272 (1999).
11. Y. Anikster *et al.*, *Am. J. Hum. Genet.* **100**, 257–266 (2017).
12. L. Penrose, J. H. Quastel, *Biochem. J.* **31**, 266–274 (1937).
13. S. Kure *et al.*, *J. Pediatr.* **135**, 375–378 (1999).
14. I. Karačić *et al.*, *Mol. Genet. Metab.* **97**, 165–171 (2009).
15. R. Eavri, H. Lorberboum-Galski, *J. Biol. Chem.* **282**, 23402–23409 (2007).
16. J. A. Hoskins *et al.*, *Lancet* **1**, 392–394 (1980).
17. A. Rebuffat, C. O. Harding, Z. Ding, B. Thöny, *Hum. Gene Ther.* **21**, 463–477 (2010).
18. J. C. Guo *et al.*, *Nucleic Acids Res.* **47** (W1), W516–W522 (2019).
19. J. Wang *et al.*, *Int. J. Mol. Sci.* **14**, 10015–10041 (2013).

20. K. Panzitt *et al*., *Gastroenterology* **132**, 330–342 (2007).
21. D. W. Collins, T. H. Jukes, *Genomics* **20**, 386–396 (1994).
22. K. E. Watters, A. M. Yu, E. J. Strobel, A. H. Settle, J. B. Lucks, *Methods* **103**, 34–48 (2016).
23. M. Kubota, C. Tran, R. C. Spitale, *Nat. Chem. Biol.* **11**, 933–941 (2015).
24. B. Thöny, Z. Ding, A. Martínez, *FEBS Lett.* **577**, 507–511 (2004).
25. H. Erlandsen, R. C. Stevens, *J. Inherit. Metab. Dis.* **24**, 213–230 (2001).
26. E. K. Jaffe, *Mol. Genet. Metab.* **121**, 289–296 (2017).
27. S. Zhang, P. F. Fitzpatrick, *J. Biol. Chem.* **291**, 7418–7425 (2016).
28. C. B. Frederick, K. L. Dooley, R. L. Kodell, W. G. Sheldon, F. F. Kadlubar, *Fundam. Appl. Toxicol.* **12**, 346–357 (1989).
29. T. Gjetting, M. Petersen, P. Guldberg, F. Güttler, *Am. J. Hum. Genet.* **68**, 1353–1360 (2001).
30. C. Carluccio, F. Fraternali, F. Salvatore, A. Fornili, A. Zagari, *PLOS ONE* **8**, e79482 (2013).
31. A. Lin *et al*., *Nat. Cell Biol.* **19**, 238–251 (2017).
32. Y. Feng *et al*., *Nat. Biotechnol.* **32**, 1036–1044 (2014).
33. P. J. Waters, *Hum. Mutat.* **21**, 357–369 (2003).
34. A. Gámez, B. Pérez, M. Ugarte, L. R. Desviat, *J. Biol. Chem.* **275**, 29737–29742 (2000).
35. J. Zschocke, C. A. Graham, D. J. Carson, N. C. Nevin, *Am. J. Hum. Genet.* **57**, 1311–1317 (1995).
36. C. A. O'Neill *et al*., *Acta Paediatr. Suppl.* **407**, 43–44 (1994).
37. Y. Huang, *Mol. Ther. Nucleic Acids* **6**, 116–132 (2017).
38. M. Sanford, G. M. Keating, *Drugs* **69**, 461–476 (2009).
39. L. L. Santos *et al*., *Genet. Mol. Res.* **9**, 1–8 (2010).
40. D. Bercovich *et al*., *J. Hum. Genet.* **53**, 407–418 (2008).
41. S. F. Garbade *et al*., *Genet. Med.* **21**, 580–590 (2019).
42. N. Blau, A. Martinez, G. F. Hoffmann, B. Thöny, *Mol. Genet. Metab.* **123**, 1–5 (2018).
43. J. Schaub *et al*., *Arch. Dis. Child.* **53**, 674–676 (1978).
44. N. Blau, H. Erlandsen, *Mol. Genet. Metab.* **82**, 101–111 (2004).
45. F. K. Trefz, D. Scheible, G. Frauendienst-Egger, H. Korall, N. Blau, *Mol. Genet. Metab.* **86**, S75–S80 (2005).
46. G. Gramer, P. Burgard, S. F. Garbade, M. Lindner, *J. Inherit. Metab. Dis.* **30**, 556–562 (2007).
47. R. C. Eisensmith, S. L. Woo, *Acta Paediatr. Suppl.* **407**, 124–129 (1994).
48. Z. Ding, C. O. Harding, B. Thöny, *Mol. Genet. Metab.* **81**, 3–8 (2004).
49. C. N. Sarkissian *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2339–2344 (1999).
50. J. A. Hoskins, J. Gray, *Res. Commun. Chem. Pathol. Pharmacol.* **35**, 275–282 (1982).
51. N. Longo *et al*., *Orphanet J. Rare Dis.* **13**, 108 (2018).
52. K. C. Mahan, M. A. Gandhi, S. Anand, *Curr. Med. Res. Opin.* **35**, 647–651 (2019).

ACKNOWLEDGMENTS

Human iPSCs were generated at the Human Stem Cell Core (HSCC) at Baylor College of Medicine by J. J. Kim. **Funding:** This work was partially supported by the HSCC. The Proteomics Facility was supported in part by NIH grant no. 1S10OD012304-01 to D.H. This project was partially supported by the University of Houston Seq-N-Edit Core with funding from the UH Division of Research, the UH College of NSM, and the Department of Biology & Biochemistry; NRUF MINOR CORE 17 Grant to P.H.G.; UH Small Core Equipment Program Grant to P.H.G.; the core facilities at BCM, including the Metabolomics Core (supported by NIH P30CA125123), the CPRIT Proteomics and Metabolomics Core Facility (N.P.) (supported by RP170005), and the Dan L. Duncan Cancer Center. L.H. was supported by Cancer Prevention & Research Institute of Texas (CPRIT) grants RP150085 and RP190570. Z.T. and S.Z. were partially supported by CPRIT grant RP170333, NIH grant R01CA225955, and NSF grant CHE-1411859. C.R.L. and L.Q.Y. were supported by an MD Anderson Cancer Center Institutional Research Grant. **Author contributions:** L.Q.Y. and C.R.L. conceived the project and designed the experiments. Y.J.L. executed the primary studies. Y.J.L. developed genetic mouse models and related experiments with assistance from Q.S.H. and Y.H.Z. T.K.N., X.Z., L.C.L., H.H., L.L, J.M.X. Z.T., and S.Z. performed the proteomics and modeling studies. J.Y., Y.Q.Y., Z.Z., and L.H. performed bioinformatics analysis. Y.J.L. performed the histological staining and corresponding analysis with assistance from K.L. D.H. executed mass spectrometry analysis. L.Q.Y., S.S.C., Y.H.P., and P.H.G. performed the CLIP assay. N.P. performed and analyzed and C.C. analyzed metabolic profiling. F.N., J.L.G., and F.F. processed clinical specimens with assistance of from M.S. and L.Q.Y. Y.J.L. performed protein purification with assistance from Y.C.L. and K.L.T. S.Z., M.C.H., G.A.C., M.S., A.C.M., N.B., V.R.S., and F.F. contributed to experimental design and data interpretation. S.D.E. assisted with manuscript drafting. L.Q.Y. and C.R.L. wrote the manuscript. **Competing interests:** The RNA mimics, including HULC, biotinylated HULC, and GalNAc-tagged HULC mimics, are in the process of receiving a patent for PKU/HPA patient treatment (no. MDA19-013; inventors L.Y., C.L., Y.L., and S.D.E.). **Data availability:** The raw lncRNA array data for this manuscript are available at GEO under the accession number GSE120207. All other data are available in the main text or the supplementary materials. Because of the COVID-19 pandemic, our genetically modified mouse colonies died out, and we are in the process of reestablishing our homozygous mouse populations. The skin fibroblasts and hiPSCs were obtained from the Coriell Institute with MTA and are subjected to restrictions on redistribution and sharing. Requests for these cells should be obtained from the Coriell Institute. All other research materials are available from the corresponding author upon reasonable request.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6555/662/suppl/DC1
Materials and Methods
Figs. S1 to S20
Tables S1 to S8
References (*53*–*72*)
Movies S1 to S4
MDAR Reproducibility Checklist

9 December 2019; resubmitted 31 August 2020
Accepted 25 June 2021
10.1126/science.aba4991

QUANTUM SENSING

Quantum-enhanced sensing of displacements and electric fields with two-dimensional trapped-ion crystals

Kevin A. Gilmore[1,2]*†, Matthew Affolter[1], Robert J. Lewis-Swan[3,4], Diego Barberena[5,6], Elena Jordan[1]‡, Ana Maria Rey[5,6]*, John J. Bollinger[1]*

Fully controllable ultracold atomic systems are creating opportunities for quantum sensing, yet demonstrating a quantum advantage in useful applications by harnessing entanglement remains a challenging task. Here, we realize a many-body quantum-enhanced sensor to detect displacements and electric fields using a crystal of ~150 trapped ions. The center-of-mass vibrational mode of the crystal serves as a high-*Q* mechanical oscillator, and the collective electronic spin serves as the measurement device. By entangling the oscillator and collective spin and controlling the coherent dynamics via a many-body echo, a displacement is mapped into a spin rotation while avoiding quantum back-action and thermal noise. We achieve a sensitivity to displacements of 8.8 ± 0.4 decibels below the standard quantum limit and a sensitivity for measuring electric fields of 240 ± 10 nanovolts per meter in 1 second. Feasible improvements should enable the use of trapped ions in searches for dark matter.

The development of protocols and quantum platforms that harness entanglement or correlations to sense or measure a physical quantity with an advantage relative to classical alternatives is of both practical and fundamental interest (*1*, *2*). Quantum-enhanced sensors have the potential to enable the measurement of previously undetectable signals that could improve our understanding of the universe. Two prominent examples are gravitational wave detection, where nonclassical states of light are now being adopted to achieve unparalleled sensitivity (*3*, *4*), and searches for dark matter, where a quantum advantage could enable sensing of the weak, nongravitational interaction of dark matter with normal matter (*5*–*8*).

In parallel, mechanical oscillators have become established as exquisite quantum tools to measure small displacements due to weak forces and electric fields (*9*–*17*), such as those generated by axion-like dark matter and hidden photons (*18*). At the simplest level, a weak, resonant force interacting with the mechanical oscillator generates a small coherent displacement β [normalized to twice the size of the oscillator's zero-point motion

[1]National Institute of Standards and Technology, Boulder, CO 80305, USA. [2]Department of Physics, University of Colorado, Boulder, CO 80309, USA. [3]Homer L. Dodge Department of Physics and Astronomy, University of Oklahoma, Norman, OK 73019, USA. [4]Center for Quantum Research and Technology, University of Oklahoma, Norman, OK 73019, USA. [5]JILA, NIST, and Department of Physics, University of Colorado, Boulder, CO 80309, USA. [6]Center for Theory of Quantum Matter, University of Colorado, Boulder, CO 80309, USA.
***Corresponding author. Email: kevin.gilmore@colorado.edu (K.A.G.); arey@jila.colorado.edu (A.M.R.); john.bollinger@nist.gov (J.J.B.)**
†Present address: Honeywell Quantum Solutions, Broomfield, CO 80021, USA. ‡Present address: Physikalisch-Technische Bundesanstalt, 38116 Braunschweig, Germany.

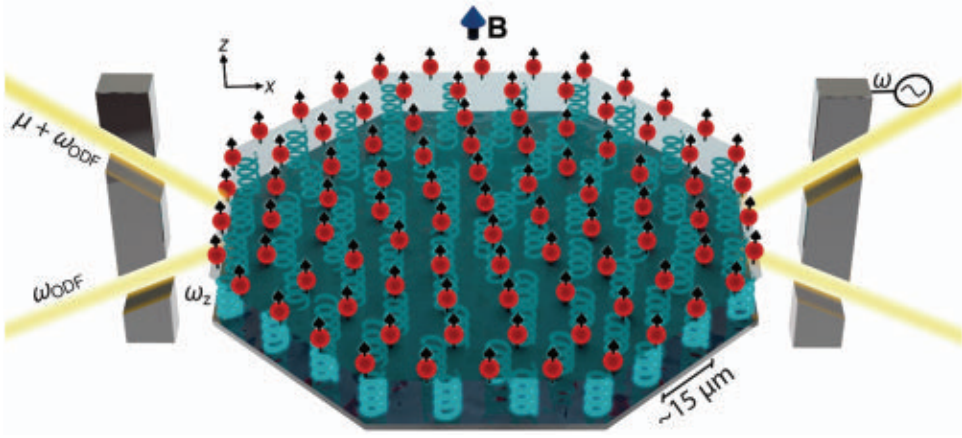


Fig. 1. Trapped-ion crystal quantum sensor. An ensemble of beryllium ions (red dots) confined within a Penning trap self-arrange into a 2D triangular lattice. The Penning trap is characterized by axial magnetic field B = 4.45 T and axial trap frequency $\omega_z = 2\pi \times 1.59$ MHz. Cylindrical electrodes (cross section shown, in gray) generate a harmonic confining potential along the $\hat{Z}$ axis, whereas radial confinement is provided by the Lorentz force from **E** × **B**-induced rotation in the axial magnetic field. The valence electron spin (black arrows) of the ions realizes a collective spin that serves as the measurement device. The center-of-mass (COM) motional mode of the ion crystal realizes a high-Q mechanical oscillator (blue springs) with frequency ω_z. Crossed optical beams (yellow lines) with a beat-note frequency $\mu \approx \omega_z$ generate a spin-dependent optical-dipole force that couples the spins and the COM oscillator. An AC voltage source is applied to the trap endcap electrode to generate a coherent axial oscillation with calibrated amplitude Z_c and frequency $\omega = \omega_z$.

(*19*)] of the oscillator amplitude over time. Typically, the displacement is then inferred by making a measurement on a complementary internal or coupled degree of freedom of the system.

The precision $\Delta\beta$ to which this displacement β can be determined using classical resources that use uncorrelated states, such as a vacuum or a coherent state, is fundamentally bounded by the so-called standard quantum limit (SQL) that limits the attainable sensitivity to $\Delta\beta \geq 1/2$. However, by introducing entanglement between the oscillator and the measurement degree of freedom before the oscillator is excited, one can attain sub-SQL precision (*20*–*23*). This requires subtle control over both the oscillator and the measurement system to minimize undesirable classical noise and to evade quantum back-action (*24*).

We demonstrate a quantum advantage for both displacement and electric field sensing in a two-dimensional (2D) trapped-ion crystal of ~150 ions, 200 μm in diameter. The center-of-mass (COM) normal mode of the crystal realizes the mechanical oscillator, and the internal electronic spin of the ions serves as the accessible measurement probe. A spin-dependent optical dipole force (ODF) that resonantly couples the spins and the COM mode is used to generate metrologically useful entanglement, which enables precise sensing of displacements of the oscillator (*21*). By implementing a many-body echo protocol (*23*) to subsequently disentangle the spins and the oscillator, we are able to estimate the displacement—ideally free from thermal noise—through simple global measurements of the spins. With this technique, we achieve a sensitivity of 8.8 ± 0.4 dB below the SQL for displacements and 4.0 ± 0.5 dB below the SQL for electric fields. In practical terms, we demonstrate an ultimate electric field measurement sensitivity of $240 \pm 10\,\mathrm{nV\,m^{-1}}/\sqrt{\mathrm{Hz}}$, an improvement by a factor of 300 over prior off-resonant classical protocols in trapped ions (*14*, *16*) and by more than an order of magnitude over state-of-the-art electrometers based on Rydberg atoms (*25*). Relative to similar protocols previously performed using a single Rydberg atom (*22*) or a single trapped ion (*20*, *26*) to encode a spin-½ degree of freedom, our protocol demonstrates an enhanced sensitivity resulting from quantum entanglement in a mesoscopic ion crystal, while also benefiting from the intrinsic reduction of zero-point motion in the collective mechanical oscillator as a consequence of the large ion number.

Entanglement-enhanced sensor

Our quantum sensor consists of a single-plane Coulomb crystal of $N \sim 150$ $^9Be^+$ ions confined in a Penning trap (*14*, *16*, *27*–*29*) (Fig. 1). The $^2S_{1/2}$ ground-state valence electron spin $|\uparrow\rangle$ ($|\downarrow\rangle$) ≡ $|m_J = +1/2\rangle$ ($|m_J = -1/2\rangle$) encodes a spin-½ degree of freedom in each ion, which can be coherently controlled by external microwaves resonant with the 124-GHz frequency splitting of the electronic spin states in the presence of a B = 4.5 T magnetic field. The motion of the ion crystal can be decomposed into $2N$ in-plane modes and N axial modes, with the latter coupled to the spin degree of freedom by a spin-dependent ODF produced by a pair of off-resonant laser beams detuned from the nearest optical transitions by ~20 GHz. The ODF beams generate a 1D traveling-wave potential at a frequency μ, which, in contrast to prior settings for sensing in our experiment (*14*, *16*), is now chosen to be near-resonant with the COM mode at a frequency $\omega_z/(2\pi) = 1.59$ MHz, such that the COM is the dominant motional contribution to the observed dynamics. In this limit, and assuming that the ions have an axial extent that is small relative to the wavelength of the traveling-wave optical potential, the system can be well approximated by the Hamiltonian

$$\hat{H}_{\mathrm{ODF}} = \frac{\hbar g}{\sqrt{N}}\left(\hat{a} + \hat{a}^\dagger\right)\hat{J}_z - \hbar\delta\hat{a}^\dagger\hat{a} \quad (1)$$

(*19*, *30*). Here, $\hbar$ is the Planck constant divided by 2π; $\hat{a}^\dagger$ and $\hat{a}$ are the COM phonon creation and annihilation operators that couple uniformly to all spins with strength g; $\delta = \mu - \omega_z$ is the detuning from the COM mode, ideally tuned to be on resonance in our protocol, $\delta = 0$; and $\hat{J}_\alpha = 1/2\sum_{j=1}^{N}\hat{\sigma}_\alpha^{(j)}$ are the collective spin operators, where $\hat{\sigma}_\alpha^{(j)}$ are Pauli operators for the jth spin.

In the experimental sequence to sense small displacements of the COM oscillator (Fig. 2A), the ions are prepared in the state $|\uparrow\rangle_N$ by optical pumping before a microwave $\pi/2$ pulse is applied to rotate the spins to align along the $\hat{x}$ axis of the Bloch sphere. To entangle the spins and the phonons, we apply the ODF Hamiltonian ($\hat{H}_{\mathrm{ODF}}$) for time τ. Then, a calibrated AC voltage is applied to an endcap electrode for a time $t_{\mathrm{drive}} \ll \tau$ to drive up a small coherent displacement β of the COM oscillator. Recent work (*16*) has established a stable relative phase between the displacement β and the ODF, which is experimentally optimized to ensure the largest signal. To optimally detect the displacement, we implement a many-body echo to undo the entangling dynamics by applying a second ODF drive with opposite sign $-\hat{H}_{\mathrm{ODF}}$ for an identical time τ. Experimentally this is done by applying a microwave π pulse about the $\hat{x}$ axis to map $|\uparrow\rangle \rightarrow |\downarrow\rangle$, which is equivalent to setting $g \rightarrow -g$. Finally, the displacement is estimated via a projective measurement on the spin degree of freedom. Specifically, to make an effective measurement of $\hat{J}_y$, we rotate the

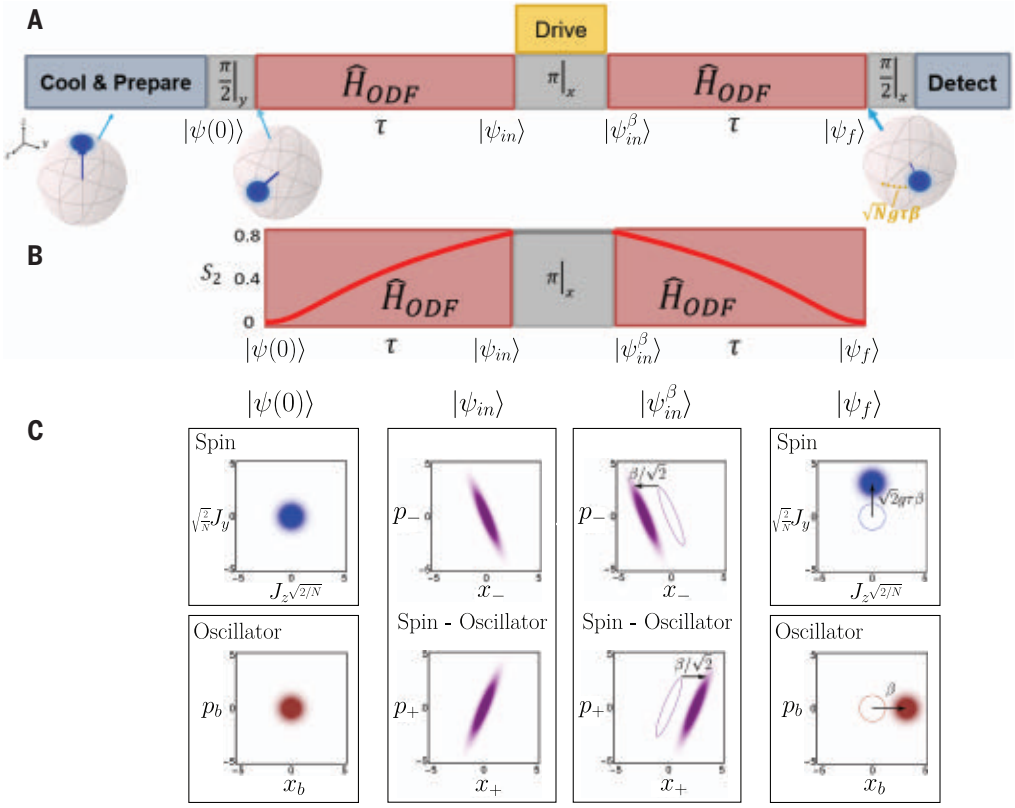


Fig. 2. Displacement-sensing protocol. (**A**) The ions are Doppler-cooled and optically pumped into the state $|\uparrow\rangle$, and a microwave $\pi/2$ pulse then rotates the spin ensemble to align along the $\hat{x}$ axis to prepare the initial state $|\psi(0)\rangle$. A resonant ODF drive, described by $\hat{H}_{ODF}$ (Eq. 1), is then applied for a duration τ to yield the state $|\psi_{in}\rangle$ in which the spin and the COM oscillator are entangled. To generate a small spin-independent displacement β of the oscillator, we apply a weak drive to an endcap electrode resonant with the COM frequency, yielding the perturbed state $|\psi_{in}^{\beta}\rangle$. Simultaneously, a π pulse is applied to flip $|\uparrow\rangle \rightarrow |\downarrow\rangle$ and thus effectively reverse the sign of $\hat{H}_{ODF}$ when a subsequent resonant ODF drive is applied. A final $\pi/2$ pulse transforms the accumulated spin precession into a measurable change in the fraction of spins in $|\uparrow\rangle$ that is read out via a projective measurement. (**B**) The produced entanglement is illustrated by the growth of the Rényi entropy S_2 (*19*), which reaches a maximum after the first ODF drive. The second application of the ODF realizes a many-body echo of the initial dynamics and disentangles the spin and the oscillator—demonstrated by the vanishing Rényi entropy after the second ODF drive—while also acting to map the small spin-independent displacement β into a spin rotation of $\phi = 2g\tau\beta/\sqrt{N}$ about $\hat{z}$. (**C**) A phase-space illustration of the sequence. The initial state (leftmost panels) features the characteristic Gaussian and isotropic quantum noise distribution of a coherent state in both the spin and the oscillator degrees of freedom. The center panels illustrate that the entanglement can be interpreted as squeezing of the composite spin-oscillator quantum noise in a coupled basis. A phase-space illustration of the final decoupled state $|\psi_f\rangle$ is shown at the right. In (B) and (C) we use typical values of $g\tau = 2$ for illustration.

spins by a microwave $\pi/2$ pulse about the $\hat{x}$ axis and then measure $\hat{J}_z$ using state-dependent fluorescence imaging (*19*).

To simplify our analysis of the experimental protocol, we assume that the ions are initialized in the motional ground state, such that the state of the spins and COM mode after the first microwave pulse is $|\psi(0)\rangle = |0\rangle_b \otimes |(N/2)_x\rangle$. Here, we describe the spins using Dicke states $\hat{J}_\alpha|m_\alpha\rangle = m_\alpha|m_\alpha\rangle$ with $\alpha = x, y, z$ and $|0\rangle_b$ is the bosonic vacuum. The first ODF generates a spin-dependent coherent displacement of the oscillator,

$$|\psi_{in}\rangle = \exp\left[-i\frac{\hbar g\tau}{\sqrt{N}}\left(\hat{a}+\hat{a}^\dagger\right)\hat{J}_z\right]|\psi(0)\rangle = \sum_{m_z=-N/2}^{N/2} c_{m_z}|\alpha_{m_z}\rangle_b \otimes |m_z\rangle \quad (2)$$

resulting in strong spin-motion entanglement. This can be directly quantified by the purity of the reduced density matrix of the oscillator degree of freedom after tracing over the spin [e.g., Rényi entanglement entropy S_2 (*19*)] during the first ODF sequence, as shown in Fig. 2B. In Eq. 2 we introduce $|\alpha_{m_z}\rangle_b$ as coherent states with complex amplitude $\alpha_{m_z} = -im_z g\tau/\sqrt{N}$, and c_{m_z} are defined as the expansion coefficients $|(N/2)_x\rangle \equiv \sum_{m_z} c_{m_z}|m_z\rangle$.

Because $N \gg 1$ in our 2D trapped-ion crystal, we are able to formulate a particularly simple understanding of the metrological utility of the spin-boson entanglement created in the state $|\psi_{in}\rangle$. The ODF Hamiltonian can be approximately recast as a squeezing Hamiltonian, $\hat{H}_{ODF} \equiv g(\hat{x}_+^2 - \hat{x}_-^2)/2$, where $\hat{x}_\pm = \pm(\hat{a} + \hat{a}^\dagger \pm \sqrt{4/N}\hat{J}_z)/2$ and $\hat{p}_\pm = \pm i(\hat{a}^\dagger - \hat{a} \pm i\sqrt{4/N}\hat{J}_y)/2$. Here, $\hat{x}_\pm$ and $\hat{p}_\pm$ are quadrature operators of a pair of effective oscillators resulting from a hybridization of the spin and boson degrees of freedom. Note that the normalization of the spin operators in the definition of $\hat{x}_\pm$ and $\hat{p}_\pm$ means that the $\sim\sqrt{N}$ projection noise of the original coherent spin state contributes $\sim\mathcal{O}(1)$ noise to the hybrid quadratures, similar to the vacuum noise of the mechanical oscillator. In this language, the action of $\hat{H}_{ODF}$ is to generate squeezing along specific quadratures in the independent $(\hat{x}_+, \hat{p}_+)$ and $(\hat{x}_-, \hat{p}_-)$ phase spaces (Fig. 2C). In the original basis of the spin and boson degrees of freedom, this squeezing manifests as strong correlations and entanglement between the spins and bosons.

The subsequently applied displacement of the COM oscillator, assumed to be orthogonal to the spin-dependent displacements, is described by the unitary transformation $\hat{U}_\beta = \exp[\beta(\hat{a} - \hat{a}^\dagger)] = \exp(i\sqrt{2}\beta\hat{p}_b)$, where $\sqrt{2}i\hat{p}_b = \hat{a} - \hat{a}^\dagger$ ($\sqrt{2}\hat{x}_b = \hat{a} + \hat{a}^\dagger$) are the oscillator quadratures. This leads to the displaced state $|\psi_{in}^{\beta}\rangle = \hat{U}_\beta|\psi_{in}\rangle$, which can be equivalently framed as a shift in the spin-boson quadrature $\hat{x}_+ \rightarrow \hat{x}_+ + \beta/\sqrt{2}$ $(\hat{x}_- \rightarrow \hat{x}_- - \beta/\sqrt{2})$. The squeezing generated by the initial ODF drive means that, in principle, a small displacement quickly leads to a state that is distinguishable from $|\psi_{in}\rangle$. This displacement, parametrized by β, can be estimated with a sensitivity (see Fig. 2) limited by the quantum Cramer-Rao bound for displacements (*31*), $(\Delta\beta)^2 = 1/(4 + 4g^2\tau^2)$ (*19*). This bound can be recast in the more familiar form of the Heisenberg limit for displacements, $(\Delta\beta)^2_{HL} \approx 1/n$ (*21*), where $n \sim g^2\tau^2$ is the occupancy of the effective oscillator before the displacement. This is to be contrasted

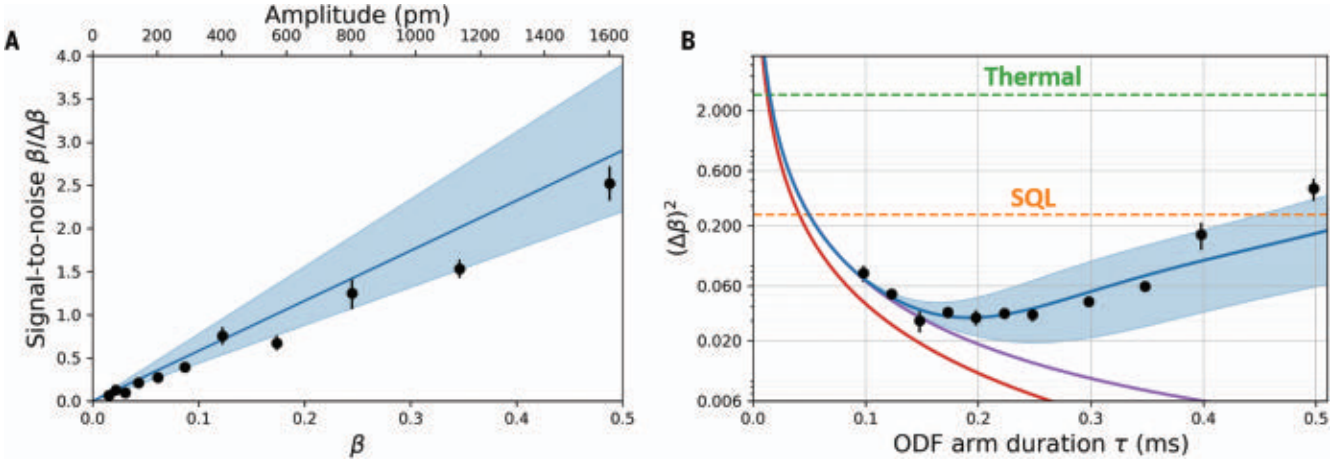


Fig. 3. Performance of mechanical displacement sensor. (**A**) Signal-to-noise ratio $\beta/(\Delta\beta) \equiv Z_c/(\Delta Z_c)$ as a function of displacement β (amplitude $Z_c \equiv 2z_0\beta/\sqrt{N}$, where $z_0 = \sqrt{\hbar/(2m\omega_z)}$ is the spatial extent of the ground-state wave function). Each experimental data point (black markers; error bars indicate statistical and systematic uncertainty) corresponds to the SNR of a single measurement with fixed τ = 200 μs. Good agreement is found with the theoretical model (blue line) using independently calibrated values of $g/(2\pi)$ = 3.91 kHz and COM frequency fluctuations of $\sigma/(2\pi)$ = 40 Hz. The shaded area indicates the confidence region for $\sigma/(2\pi) \in [20, 60]$ Hz, as well as spin depolarization characterized by Γ = 500 s^{-1}. The theoretical calculation also includes an observed 18% increase in background noise above the expected projection noise (*19*). (**B**) Sensitivity $(\Delta\beta)^2$ to the dimensionless displacement β as a function of ODF duration τ. Experimental data (black markers; error bars indicate statistical and systematic uncertainty) is obtained from a small fixed physical displacement Z_c = 775 ± 28 pm with a single-measurement SNR of ~1. We compare the observations to the theoretical predictions of an idealized model (red line), a model that includes depolarization due to spin decoherence (purple line, Γ = 610 s^{-1}), and a model that also includes COM fluctuations of σ = 40 Hz (solid blue line); the shaded area indicates the confidence region for $\sigma/(2\pi) \in [20, 60]$ Hz. The SQL $(\Delta\beta)^2_{SQL} = ¼$, corresponding the extent of the ground-state wave function, is indicated by a dashed orange line. For completeness, we also indicate the classically attainable sensitivity when accounting for excess thermal fluctuations, $\bar{n} = 5$ at the Doppler cooling limit $(\Delta\beta)^2_{th} = (2\bar{n}+1)(\Delta\beta)^2_{SQL}$ (dashed green line). The optimal experimental sensitivity of 8.8 ± 0.4 dB below the SQL (~19 dB below the thermal noise limit) corresponds to the average of the five points centered around τ = 200 μs (*19*).

with the SQL for displacements $(\Delta\beta)^2_{SQL} = ¼$, which is the sensitivity attainable with a coherent state.

Nonetheless, fully attaining the sensitivity promised by the state $|\psi_{in}\rangle$ is not straightforward. In particular, the high degree of spin-boson entanglement in $|\psi_{in}\rangle$ means that a simple measurement of either the spins or phonons independently is insufficient to precisely infer β (*19*). Instead, a more sophisticated measurement that accounts for correlations between the degrees of freedom is required. We address this challenge and remove undesirable quantum back-action effects by the application of a time-reversal step that perfectly disentangles the spin and the oscillator (see Fig. 2) and maps the displacement to accessible spin observables. Specifically, the final state, obtained after applying the many-body echo to $\left|\psi^{\beta}_{in}\right\rangle$ (Eq. 2), is equivalent to

$$|\psi_f\rangle = \exp\left[i\frac{\hbar g\tau}{\sqrt{N}}\left(\hat{a}+\hat{a}^{\dagger}\right)\hat{J}_z\right]\left|\psi^{\beta}_{in}\right\rangle \equiv \exp\left(\frac{2ig\tau\beta}{\sqrt{N}}\hat{J}_z\right)\hat{U}_{\beta}|\psi(0)\rangle \quad (3)$$

(*19*). Writing the final state in the latter form demonstrates the power of the many-body echo: The displacement of the COM mode is encoded into a collective spin rotation of angle $\varphi = 2g\tau\beta/\sqrt{N}$, which can be read out by simple collective spin measurements (*23*, *32*–*34*). Accounting for quantum projection noise of the collective spin, which limits angular resolution of small rotations to $\Delta\varphi \geq 1/\sqrt{N}$, the displacement can thus be estimated with a sensitivity $(\Delta\beta)^2 = 1/(4g^2\tau^2)$ that approaches the Cramer-Rao bound for $g\tau \gg 1$. Moreover, we note that the form of the collective spin rotation is independent of the initial state of the phonons and hence is insensitive to any thermal occupation of the COM mode.

Displacement sensing

Assessing the performance of the implemented displacement-sensing protocol (Fig. 3), we see that the achievable sensitivity of the current experiment is primarily, but not fundamentally, limited by small shot-to-shot fluctuations of the COM mode frequency away from resonance as a consequence of impurity ions in the crystal and thermal occupation of the in-plane modes (*35*). These fluctuations limit our ability to perfectly reverse the entangling dynamics. Including COM frequency fluctuations with a root-mean-square spread σ, as well as single-particle decoherence of the spins at rate Γ due to light scattering generated by the applied ODF beams and an initial thermal occupation $\bar{n} \approx 5$ of the COM mode, the signal-to-noise ratio (SNR) of a single measurement of β is reduced to

$$\frac{\beta}{\Delta\beta} = \frac{2g\tau\beta\exp(-\Gamma\tau)}{\sqrt{1+\exp(-2\Gamma\tau)\left[(2\bar{n}+1)g^2\sigma^2\tau^4+\frac{4}{9}g^4\sigma^2\tau^6\right]}} \quad (4)$$

(*19*). Physically, the prefactor $\exp(-\Gamma\tau)$ comes from depolarization of the collective spin due to decoherence, whereas the τ^4 and τ^6 terms of the denominator describe residual spin-phonon entanglement and excess projection noise introduced by the imperfect time reversal. In Fig. 3A, we find good agreement between the experimentally determined SNR and the theoretical prediction (Eq. 4) with independently calibrated values of g, σ, and Γ (*19*), justifying our understanding of the various noise processes. In particular, we observe $\beta/\Delta\beta \to 0$ as $\beta \to 0$, which indicates that there are no additional systematic errors from effects beyond the applied displacement.

To assess the performance of the displacement-sensing protocol, we determine the variance $(\Delta\beta)^2$ of a single measurement as a function of the ODF interaction time τ for a small fixed displacement β = 0.24. Again, we find good agreement with the theoretical estimates of $(\Delta\beta)^2$ presented in Fig. 3B. Optimizing over τ, we demonstrate a sensitivity of $(\Delta\beta)^2$ = 3.3 ×

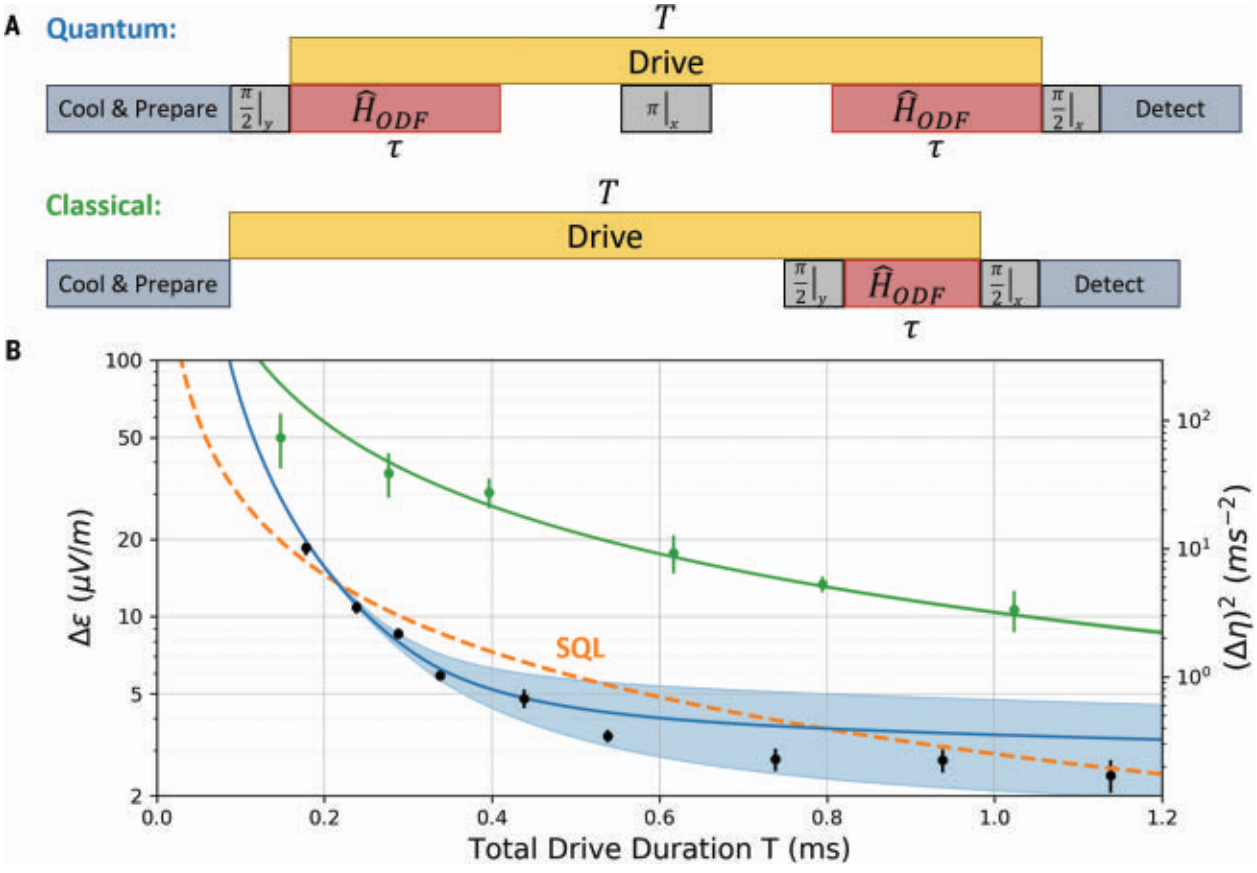


Fig. 4. Performance of the electric field sensor. (**A**) Schematic timeline of the quantum and classical protocols used to sense weak electric fields. The quantum protocol involves sequentially applying a pair of ODF drives with equal duration τ to first entangle and then disentangle the spin and motional degrees of freedom for readout via the spins, while a weak spin-independent drive is applied concurrently for a duration $T \geq 2\tau$. The classical protocol identically applies a spin-independent drive for duration T, but instead the ODF is only turned on for $\tau \leq T$ at the end of the sequence to enable spin readout of the overall displacement. (**B**) Electric field sensitivity $\Delta\varepsilon$ [and $(\Delta\eta)^2$, right y axis] as a function of the total spin-independent drive duration T. We plot experimental data from both classical and quantum protocols (green and black markers, respectively; error bars indicate statistical and systematic uncertainty) with τ chosen to ideally optimize $(\Delta\eta)^2$ (*19*). Good agreement is found with theoretical models of both protocols (green and blue solid lines) using independently calibrated values of $g/(2\pi) = 3.88$ kHz and COM mode frequency fluctuations of $\sigma/(2\pi) \approx 40$ Hz, as well as $\Gamma = 520\ s^{-1}$. The shaded area indicates the confidence region of $\sigma/(2\pi) \in [20, 60]$ Hz, which is not visible on this scale for the classical protocol. As reference, we contrast with the SQL $(\Delta\eta)^2_{SQL} = (2T)^{-2}$ (orange dashed line). Experimental data indicate that the quantum protocol attains a sensitivity that is optimally 4 dB below the SQL (*19*), or up to 14 dB better than the sensitivity attainable with the classical protocol. Taking the average of the three points with longest drive duration T results in an electric field sensitivity of $\Delta E = 240 \pm 10\ nV\ m^{-1}/\sqrt{Hz}$.

10^{-2}, or 8.8 ± 0.4 dB below the SQL. We emphasize that no technical noise has been subtracted in obtaining this result. Our result can be recast in terms of the absolute physical displacement Z_c of the COM mode using the relation $Z_c = 2z_0\beta/\sqrt{N}$, where $z_0 = \sqrt{\hbar/(2m\omega_z)}$ is the size of the ground-state motional wave function of a single ion and the enhancement proportional to $\sqrt{N}$ arises from the increased mass of the COM mode when more ions are present. Accounting for the 8-ms duration (e.g., accounting for preparation and readout) of a single measurement, we equivalently determine that the optimal sensitivity is $\Delta Z_c = 36 \pm 1.5\ pm/\sqrt{Hz}$. Note that in the absence of our time-reversal protocol, excess projection noise introduced by the thermal occupation of the phonon mode, $\bar{n} \approx 5$, would have limited the achievable sensitivity to $(\Delta\beta)^2_{th} = (2\bar{n}+1)(\Delta\beta)^2_{SQL}$ (*19*), or ~19 dB worse than our result.

Electric field sensing

Our ability to resonantly drive a small displacement of the COM mode over a long period of time, while maintaining a stable phase lock to a resonant ODF drive that generates spin-phonon entanglement, places our experiment in an excellent position for quantum-enhanced measurements of weak AC electric fields at the frequency of the COM mode.

In the protocol for sensing an electric field (Fig. 4), a fixed AC voltage is applied to an endcap electrode to drive a continuous displacement of the COM mode for a total time T. Simultaneously, an ODF drive is initially turned on for a time $\tau \leq T/2$ to generate spin-boson entanglement. At some later time, a second ODF pulse with inverted coupling $g \rightarrow -g$ (realized again by a π pulse about $\hat{x}$ of the spins after the first ODF drive) is applied for an identical duration τ and timed to finish with the conclusion of the COM displacement. Collectively, the dynamics of the system is described by the modified Hamiltonian

$$\hat{H}_{sens} = \hat{H}_{ODF}(t) + i\eta\left(\hat{a} - \hat{a}^{\dagger}\right) \qquad (5)$$

(*19*), where η is a parameter related to the electric field to be measured. Note that the time dependence of $\hat{H}_{ODF}(t)$ indicates the modulation $g \rightarrow g(t)$.

For comparison purposes and to emphasize the quantum advantage, we compare the above protocol with a purely "classical" scheme where the continuous displacement of the COM mode

is similarly applied for a total time T but we use only a single ODF drive for readout. Such a protocol ideally gives the sensitivity $(\Delta\eta)_C^2 = (1+g^2\tau^2)/\left[g^2\tau^2(2T-\tau)^2\right]$, which is always worse than the SQL defined as $(\Delta\eta)_{SQL}^2 = 1/(4T^2)$. Note that the total displacement of the oscillator is ideally $\beta \equiv \eta T$. In contrast, our quantum-enhanced sensing scheme leads to $(\Delta\eta)^2 = 1/\left[4g^2\tau^2(T-\tau)^2\right]$, which yields sub-SQL scaling $(\Delta\eta)^2 \to 4/(g^2T^4)$ for $g\tau = gT/2 \gg 1$.

Figure 4 shows the experimentally obtained single-measurement electric field sensitivity of our protocol, in comparison to the SQL and the classical scheme. The ODF duration τ is chosen by numerically optimizing the theoretically predicted sensitivity including technical noise due to COM fluctuations σ and the thermal phonon occupation $\bar{n}$ (*19*). Under current experimental conditions, for intermediate driving time $T = 538\ \mu s$, the quantum protocol allows us to attain a sensitivity of ~4.0 ± 0.5 dB below the SQL. Moreover, we find that in contrast to the quantum protocol, which is ideally insensitive to the initial phonon state, the classical scheme is strongly limited by thermal noise. Quantitatively, the quantum protocol provides a ~14 dB improvement in sensitivity relative to the comparable classical scheme, demonstrating that entanglement can provide not only a fundamental but also a practical advantage when all relevant noise sources are taken into account.

For long drive times ($T \gtrsim 1$ ms), we find that the sensitivity is ultimately bounded by COM frequency noise, $(\Delta\eta)^2 \gtrsim \sigma^2(2\bar{n}+1)/4$ (*19*). Experimentally we determine a single-measurement sensitivity of $\Delta\varepsilon = 2.3 \pm 0.1\ \mu V\ m^{-1}$ at $T = 1.14$ ms, in good agreement with theory (Fig. 4). Accounting for the total duration of the experimental trial, $T_{shot} = 8.73$ ms, we thus obtain a best electric field sensitivity of $\Delta E = \Delta\varepsilon\sqrt{T_{shot}} = 220 \pm 10\ nV\ m^{-1}/\sqrt{Hz}$. Averaging the three experimental results with the longest drive time T gives a sensitivity of $\Delta E = 240 \pm 10\ nV\ m^{-1}/\sqrt{Hz}$ at ~1.6 MHz. This compares favorably with electric field sensors using Rydberg atoms (*36*) that can reach sensitivities of $5.5\ \mu V\ m^{-1}/\sqrt{Hz}$ (*25*) at frequencies on the order of 10 GHz, and also approaches the sensitivity of meter-length classical antennas (*37*) but in a micrometer-scale device.

Conclusions and outlook

We have demonstrated a quantum-enhanced sensor of mechanical displacements and weak electric fields in a crystal composed of ~150 trapped ions. Currently, the primary limitations to the sensor's performance are COM frequency fluctuations and a thermal phonon population. However, these issues are not fundamental in nature and could be resolved in next-generation experiments. Indeed, feasible improvements to the stability of the COM mode frequency to the level of $\sigma/(2\pi) \approx 1$ Hz combined with a reduction of the phonon temperature via electromagnetically induced transparency (EIT) cooling, which has already been demonstrated in the same platform (*38*), should allow us to reach an electric field sensitivity of ~10 $nV\ m^{-1}/\sqrt{Hz}$.

This degree of sensitivity could enable trapped-ion crystals as detectors of hidden photons and axions with a frequency (effective mass) compatible with the currently accessible range of COM mode frequencies (10 kHz to 10 MHz). Although comparisons to other state-of-the-art platforms for sensing dark matter should be made carefully and consistently, one prominent advantage of the trapped-ion system is that the strong magnetic field required to couple axions to photons (*18*) is intrinsically present in the experimental platform. A further benefit is the capability to tune the COM frequency over the aforementioned range, which should be contrasted with dark matter searches using superconducting qubits in cavities (*39*) that must be operated near zero magnetic field and can be cumbersome to tune.

Our largest limitation is a finite-volume effect (*40*) that suppresses the axion-generated electric field because the characteristic size of the region over which the magnetic field is present is small relative to the long Compton wavelength of axions at typical COM frequencies. The sensitivity of our experiment, which detects an electric field, is suppressed quadratically in this small ratio, whereas experiments that detect a dark matter–generated magnetic field are only suppressed to linear order (*19*). Nonetheless, even with this extra suppression, we estimate that our current platform with an electric field sensitivity of 10 $nV\ m^{-1}/\sqrt{Hz}$ at 1.6 MHz could attain a sensitivity to axions with an axion-photon coupling as small as $g_{\alpha\gamma\gamma} \sim 1.6 \times 10^{-11}\ GeV^{-1}$ after 1 day of averaging (*19*). This is competitive with current experiments that search for axions at similar frequency scales, which have reported limits of $g_{\alpha\gamma\gamma} \lesssim 1 \times 10^{-11}\ GeV^{-1}$ (*41*, *42*). Additional optimization of the trapped-ion platform, such as raising the ion number to $N \gtrsim 10^6$ by the use of 3D crystals (*43*, *44*) or increasing the size and strength of the superconducting magnetic field, could also allow us to further enhance our sensing capability by as much as three orders of magnitude.

REFERENCES AND NOTES

1. J. P. Dowling, G. J. Milburn, *Philos. Trans. R. Soc. A* **361**, 1655–1674 (2003).
2. C. L. Degen, F. Reinhard, P. Cappellaro, *Rev. Mod. Phys.* **89**, 035002 (2017).
3. M. Tse *et al.*, *Phys. Rev. Lett.* **123**, 231107 (2019).
4. Virgo Collaboration, *Phys. Rev. Lett.* **123**, 231108 (2019).
5. ADMX Collaboration, *Phys. Rev. Lett.* **120**, 151301 (2018).
6. L. Zhong *et al.*, *Phys. Rev. D* **97**, 092001 (2018).
7. M. Malnou *et al.*, *Phys. Rev. X* **9**, 021023 (2019).
8. K. M. Backes *et al.*, *Nature* **590**, 238–242 (2021).
9. K. C. McCormick *et al.*, *Nature* **572**, 86–90 (2019).
10. F. Wolf *et al.*, *Nat. Commun.* **10**, 2929 (2019).
11. S. Schreppler *et al.*, *Science* **344**, 1486–1489 (2014).
12. S. Kolkowitz *et al.*, *Science* **335**, 1603–1606 (2012).
13. R. D. Delaney, A. P. Reed, R. W. Andrews, K. W. Lehnert, *Phys. Rev. Lett.* **123**, 183603 (2019).
14. K. A. Gilmore, J. G. Bohnet, B. C. Sawyer, J. W. Britton, J. J. Bollinger, *Phys. Rev. Lett.* **118**, 263602 (2017).
15. W. Wang *et al.*, *Nat. Commun.* **10**, 4382 (2019).
16. M. Affolter, K. A. Gilmore, J. E. Jordan, J. J. Bollinger, *Phys. Rev. A* **102**, 052609 (2020).
17. R. A. Thomas *et al.*, *Nat. Phys.* **17**, 228–233 (2021).
18. M. S. Turner, *Phys. Rep.* **197**, 67–97 (1990).
19. See supplementary materials.
20. C. Hempel *et al.*, *Nat. Photonics* **7**, 630–633 (2013).
21. F. Toscano, D. A. R. Dalvit, L. Davidovich, W. H. Zurek, *Phys. Rev. A* **73**, 023803 (2006).
22. M. Penasa *et al.*, *Phys. Rev. A* **94**, 022313 (2016).
23. R. J. Lewis-Swan *et al.*, *Phys. Rev. Lett.* **124**, 193602 (2020).
24. N. S. Kampel *et al.*, *Phys. Rev. X* **7**, 021008 (2017).
25. M. Jing *et al.*, *Nat. Phys.* **16**, 911–915 (2020).
26. S. C. Burd *et al.*, *Science* **364**, 1163–1165 (2019).
27. J. J. Bollinger, J. W. Britton, B. C. Sawyer, in *Non-Neutral Plasma Physics VIII, AIP Conf. Proc. 1521* (2013), pp. 200–209.
28. B. C. Sawyer, J. W. Britton, J. J. Bollinger, *Phys. Rev. A* **89**, 033408 (2014).
29. J. G. Bohnet *et al.*, *Science* **352**, 1297–1301 (2016).
30. A. Safavi-Naini *et al.*, *Phys. Rev. Lett.* **121**, 040503 (2018).
31. S. L. Braunstein, C. M. Caves, *Phys. Rev. Lett.* **72**, 3439–3443 (1994).
32. O. Hosten, R. Krishnakumar, N. J. Engelsen, M. A. Kasevich, *Science* **352**, 1552–1555 (2016).
33. E. Davis, G. Bentsen, M. Schleier-Smith, *Phys. Rev. Lett.* **116**, 053601 (2016).
34. S. P. Nolan, S. S. Szigeti, S. A. Haine, *Phys. Rev. Lett.* **119**, 193601 (2017).
35. A. Shankar *et al.*, *Phys. Rev. A* **102**, 053106 (2020).
36. A. Facon *et al.*, *Nature* **535**, 262–265 (2016).
37. D. H. Meyer, Z. A. Castillo, K. C. Cox, P. D. Kunz, *J. Phys. At. Mol. Opt. Phys.* **53**, 034001 (2020).
38. E. Jordan *et al.*, *Phys. Rev. Lett.* **122**, 053603 (2019).
39. A. V. Dixit *et al.*, *Phys. Rev. Lett.* **126**, 141302 (2021).
40. S. Chaudhuri *et al.*, *Phys. Rev. D* **92**, 075012 (2015).
41. J. A. Devlin *et al.*, *Phys. Rev. Lett.* **126**, 041301 (2021).
42. C. P. Salemi *et al.*, arXiv 2102.06722 [hep-ex] (12 February 2021).
43. W. M. Itano *et al.*, *Science* **279**, 686–689 (1998).
44. A. Mortensen, E. Nielsen, T. Matthey, M. Drewsen, *Phys. Rev. Lett.* **96**, 103001 (2006).

ACKNOWLEDGMENTS

We thank J. Thompson and A. Shankar for helpful discussions, and J. Teufel and K. Cox for careful review of our manuscript. **Funding:** Supported by US Department of Energy (DOE), Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator (QSA), a DOE Office of Science HEP QuantISED award, AFOSR grants FA9550-18-1-0319 and FA9550-20-1-0019, by DARPA and ARO grant W911NF-16-1-0576, the DARPA ONISQ program, ARO single investigator award W911NF-19-1-0210, NSF grants PHY1820885, JILA-PFC PHY-1734006, and QLCI-2016244, and NIST. **Author contributions:** K.A.G., M.A., E.J., and J.J.B. collected and analyzed the experimental data. R.J.L.-S., D.B., and A.M.R. developed the theoretical model. All authors discussed the results and contributed to the preparation of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions of this study are provided in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6555/673/suppl/DC1
Supplementary Text
Figs. S1 to S11
References (*45*–*55*)

15 March 2021; accepted 25 June 2021
10.1126/science.abi5226

CHEMICAL KINETICS

Watching a hydroperoxyalkyl radical (•QOOH) dissociate

Anne S. Hansen[1], Trisha Bhagde[1], Kevin B. Moore III[2], Daniel R. Moberg[2], Ahren W. Jasper[2], Yuri Georgievskii[2], Michael F. Vansco[1]†, Stephen J. Klippenstein[2]*, Marsha I. Lester[1]*

A prototypical hydroperoxyalkyl radical (•QOOH) intermediate, transiently formed in the oxidation of volatile organic compounds, was directly observed through its infrared fingerprint and energy-dependent unimolecular decay to hydroxyl radical and cyclic ether products. Direct time-domain measurements of •QOOH unimolecular dissociation rates over a wide range of energies were found to be in accord with those predicted theoretically using state-of-the-art electronic structure characterizations of the transition state barrier region. Unimolecular decay was enhanced by substantial heavy-atom tunneling involving O-O elongation and C-C-O angle contraction along the reaction pathway. Master equation modeling yielded a fully a priori prediction of the pressure-dependent thermal unimolecular dissociation rates for the •QOOH intermediate—again increased by heavy-atom tunneling—which are required for global models of atmospheric and combustion chemistry.

The atmospheric oxidation of volatile organic compounds (VOCs) and low-temperature (≤1000 K) combustion of hydrocarbon fuels proceed via a well-established sequence of radical reactions (*1*). These radical chain reactions are known to drive the formation and growth of secondary organic aerosols (SOAs) in the atmosphere (*2*) and provide the key chain-branching pathways in low-temperature combustion and autoignition (*3*, *4*). Although most aspects of the reaction mechanism are understood, a critical class of carbon-centered hydroperoxyalkyl radical intermediates, commonly called •QOOH, have eluded direct experimental observation of their structure, stability, and dissociation dynamics. Here, we report the infrared (IR) spectroscopic signature of a carbon-centered •QOOH radical, namely 2-hydroperoxy-2-methylprop-1-yl [•$CH_2(CH_3)_2COOH$], and its energy-resolved and time-dependent evolution to hydroxyl (OH) radical products. This •QOOH radical is transiently formed in the oxidation of isobutane or other branched hydrocarbons.

The oxidation of gaseous organic compounds is initiated by reaction with radical species, primarily OH radicals, through either H-atom abstraction from an alkyl group or OH addition to an unsaturated site, forming carbon-center radicals (R•) that rapidly react with O_2 to generate organic peroxy radicals (ROO•) (*5*) (Fig. 1). ROO• then isomerizes through intramolecular H-atom transfer to produce the carbon-centered hydroperoxyalkyl radical •QOOH, composed of a hydroperoxide group (-OOH) and a new carbon radical center (•Q). The •QOOH intermediates can be transiently stabilized and/or dissociate to release OH radical and stable cyclic ether products, thereby facilitating OH radical chain propagation. Unimolecular decomposition of •QOOH typically competes against its bimolecular reaction with O_2 that forms oxygen-centered peroxy radicals •OOQOOH. These peroxy radicals can undergo further isomerization and decomposition to yield multiple OH radicals and thereby give rise to chain branching. Alternatively, repeated cycles of hydrogen transfer (ROO• to •QOOH) and O_2 addition (to form •OOQOOH) occur during the oxidation of abundant VOCs, such as isoprene and α/β-pinene (*6*–*8*). In this study, we directly characterized the vibrational structure of the •QOOH intermediate in isobutane oxidation and its stability relative to the transition state (TS) barrier (Fig. 1), which controls the rate of its decomposition to OH radical products over a wide range of temperature and pressure conditions.

Despite the ubiquity of •QOOH radicals in atmospheric and combustion chemistry, the few previous experimental studies have focused primarily on product characterization and indirect kinetic measurements associated with •QOOH chemistry (*9*–*11*). Prior experimental studies aimed at investigating the •QOOH formed in isobutane oxidation observed only its products, specifically OH radicals, 2,2-dimethyl epoxide (DMEPOX), and •OOQOOH peroxy radicals, the latter upon addition of O_2, and provided temperature- and pressure-dependent kinetic measurements for the appearance of OH radical products, as elaborated below (*9*, *10*). To the best of our knowledge, Savee *et al.* reported the only prior observation of a •QOOH radical using photoionization to detect a resonantly stabilized •QOOH formed in the oxidation of 1,3-cycloheptadiene (*11*). The resonance stabilization in this system increased the stability of the •QOOH radical and substantially raised the barrier for its unimolecular decomposition to OH products (~30 kcal mol^{-1}) compared with the approximately threefold lower barriers predicted for a wide range of •QOOH radicals that form OH and cyclic ether products (*12*), including the prototypical •QOOH radical reported here (*13*).

Experimental results

In the present study, •QOOH was generated within a quartz capillary reactor tube using photolytically generated Cl atoms to abstract

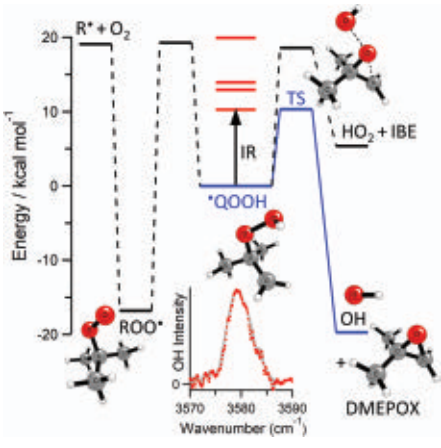


Fig. 1. Reaction pathways for the production and unimolecular decay of a carbon-centered hydroperoxyalkyl radical intermediate (•QOOH) in isobutane oxidation. IR excitation yields a spectroscopic fingerprint of the stabilized •QOOH radicals and initiates their time-dependent unimolecular decay. IR activation of •QOOH provides sufficient energy to surmount or tunnel through the reaction barrier [transition state (TS)], leading to OH radical and cyclic ether [2,2-dimethyl epoxide (DMEPOX)] products; structures of key stationary points are shown. The inset shows the IR action spectrum for the fundamental OH stretch (ν_{OH}) of jet-cooled •QOOH (10 K) at 3579.5 cm^{-1} (10.23 kcal mol^{-1}) acquired with LIF detection of OH products. The reaction pathway from •QOOH to OH and DMEPOX products (blue) is based on the current ANL0F results. Other regions (black), adapted from (*13*), include *tert*-butyl radicals (R•) that react with O_2 to generate peroxy radical (ROO•; structure shown), which can isomerize over a substantial barrier to produce •QOOH. An additional high barrier leads from •QOOH to HO_2 and isobutene (IBE) products.

[1]Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104, USA. [2]Chemical Sciences and Engineering Division, Argonne National Laboratory, Lemont, IL 60439, USA.
*Corresponding author. Email: sjk@anl.gov (S.J.K.); milester@sas.upenn.edu (M.I.L.)
†Present address: Chemical Sciences and Engineering Division, Argonne National Laboratory, Lemont, IL 60439, USA.

a H atom from a methyl group of *tert*-butyl hydroperoxide (TBHP), building on experimental methods developed previously (*9*, *10*). Distinctive to this study, the newly formed •QOOH was collisionally stabilized and rapidly cooled in a pulsed supersonic expansion. •QOOH was then characterized using frequency- and time-resolved IR pump–ultraviolet (UV) probe methods carried out under collision-free conditions. Details of the experimental methods are given in the supplementary materials.

Our search for an IR fingerprint of •QOOH radical intermediates was conducted using IR action spectroscopy (*14*), in which IR activation of jet-cooled and isolated •QOOH was followed by UV laser-induced fluorescence (LIF) detection of OH products on the A-X (1,0) $Q_1(3)$ transition at suitable IR-UV time delays ranging from 10 to 400 ns. The spectroscopic search was guided by theoretical prediction (B2PLYP-D3/cc-pVTZ level of theory) of the anharmonic vibrational frequencies and IR intensities for •QOOH transitions using second order vibrational perturbation theory (VPT2) (table S1). On the basis of prior theoretical predictions of the TS barrier (*9*, *13*, *15*, *16*), the search for •QOOH was initiated in the first OH stretch overtone region ($2\nu_{OH}$), where a strong IR transition was theoretically predicted at 6984 cm^{-1}. IR spectral scans revealed a strong feature for the $2\nu_{OH}$ transition of •QOOH at 6971.5 cm^{-1}, which was readily distinguished from the $2\nu_{OH}$ transition of the stable TBHP precursor at 7017.8 cm^{-1} (*17*). Addition of O_2 to the carrier gas eliminated the IR-induced •QOOH ($2\nu_{OH}$) signal, as expected because of the rapid reaction of •QOOH with O_2 in the reactor tube (*9*, *10*).

The search was extended to successively lower energies to identify other vibrational transitions arising from •QOOH that would also provide sufficient energy to release OH products required for detection. A weaker feature was observed at 4879.0 cm^{-1} and ascribed to a combination band involving OH stretch and OOH bend that was predicted theoretically at 4880 cm^{-1}. Continuing to lower energy, another weak feature was detected at 4534.0 cm^{-1}, which agreed with a theoretically predicted combination band at 4562 cm^{-1}. This vibration is a distinctive signature of •QOOH involving the CH_2 stretch and CH_2 bend at the carbon radical center.

Unexpectedly, IR action spectroscopy also revealed a strong feature at 3579.5 cm^{-1}, shown in Fig. 1 (inset), which appeared at an energy substantially below the most definitive prior theoretical prediction of the TS barrier (12.0 kcal mol^{-1} = 4200 cm^{-1}) (*13*). This feature was readily assigned to the strong fundamental OH stretch (ν_{OH}) of •QOOH predicted at 3580 cm^{-1}. The spectral breadth of the feature arose from its rotational band contour (fig. S1) at a rotational temperature of 10 K with homogeneous broadening originating from rapid intramolecular vibrational energy distribution. The observation of this •QOOH feature by IR action spectroscopy demonstrates that the TS barrier is substantially lower than predicted and/or quantum mechanical tunneling plays a role in the unimolecular decay of •QOOH to OH products.

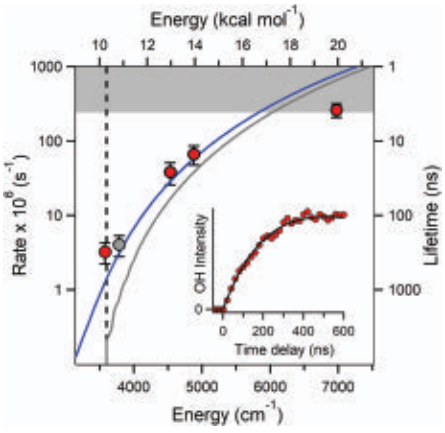


Fig. 2. Energy-dependent unimolecular decay rates for hydroperoxyalkyl radical intermediates (•QOOH). Experimental rates and corresponding lifetimes (with error bars representing ±1σ uncertainty from repeated measurements) are obtained through direct time-resolved appearance of OH products after IR activation of stabilized •QOOH originating from the lowest-energy conformer (red) and a higher-energy conformer (gray). The gray shaded region indicates rates or lifetimes limited by the experimental time resolution (4 ns). Excellent agreement is obtained with statistical unimolecular decay rates based on a high-level calculation of the TS barrier of 10.3 kcal mol^{-1} (vertical dashed line) with quantum mechanical tunneling (blue); computed rates without tunneling (gray) are shown for comparison. Heavy-atom tunneling (763*i* cm^{-1}) involving the O-O stretch and C-C-O bend substantially enhances the computed •QOOH unimolecular decay rate. The inset shows a representative time trace for the appearance of OH products after IR activation of the fundamental OH stretch (ν_{OH}) of jet-cooled •QOOH at 3579.5 cm^{-1}.

Critical additional insight on the unimolecular decay mechanism was obtained from direct time-resolved measurements for the appearance of OH products through UV LIF detection after IR activation of •QOOH (Fig. 2 and fig. S3). Excitation of •QOOH at 6971.5 cm^{-1} resulted in prompt appearance of the OH products within the temporal profile of the combined IR-UV laser pulses, which indicates a laser-limited rise time of <3.8 ± 0.8 ns. The unimolecular decay of •QOOH and corresponding appearance of OH products slowed at lower excitation energies to $6.7 \pm 1.9 \times 10^7$ s^{-1} (associated lifetime of 15.0 ± 4.3 ns) at 4879.0 cm^{-1} and $3.8 \pm 1.3 \times 10^7$ s^{-1} (26.0 ± 8.6 ns) at 4534.0 cm^{-1}. Excitation of •QOOH at 3579.5 cm^{-1} resulted in a rate that was an order of magnitude slower for the appearance of OH products at $3.2 \pm 1.0 \times 10^6$ s^{-1} (309 ± 96 ns), with corresponding time trace shown as the inset in Fig. 2. The experimental unimolecular decay rates $k(E)$ for •QOOH and corresponding lifetimes are shown in Fig. 2; uncertainties are ±1σ based on repeated measurements. In each case, the OH product temporal profile decayed on a microsecond time scale (fig. S2 and table S3), which originated from IR-activated •QOOH and/or its OH products moving out of the spatial volume probed by the UV laser.

An additional weak spectral feature was observed at 3597.8 cm^{-1}, shifted <20 cm^{-1} from the •QOOH ν_{OH} transition. This feature was attributed to the ν_{OH} transition for one or both of the low-lying conformers of •QOOH (table S1), with predicted energies (table S6) only ~0.5 kcal mol^{-1} higher and estimated population <15% of the lowest energy conformer. The low torsional barriers separating the conformers (figs. S7 and S9) enable their rapid interconversion before unimolecular dissociation of •QOOH. In this case, the rate for appearance of OH products was $4.1 \pm 1.3 \times 10^6$ s^{-1} (247 ± 77 ns).

Theoretical analysis

The energy-resolved dissociation rates $k(E)$ measured experimentally provide a stringent test of both the properties of the potential energy surface and the theoretical methods used to predict rates for dissociation of the •QOOH radical to OH radical and cyclic ether products. The most definitive prior exploration of the energetics of this •QOOH radical, involving a high-level focal point analysis, indicated a zero-point corrected dissociation barrier of 12.0 kcal mol^{-1} derived from an estimated CCSD(T) complete basis set (CBS) limit (to cc-pV6Z) with additional core-valence, relativistic, diagonal Born-Oppenheimer, and zero-point corrections (*13*). Earlier electronic structure studies predicted barriers ranging from 10.5 to 15.6 kcal mol^{-1} (table S7) based on more limited approximations to the coupled cluster result at the CBS limit (*9*, *13*, *15*, *16*). Here, we used substantially higher-level electronic structure methods, specifically the inclusion of a CCSDT(Q) correction for multireference effects and a CCSD(T)–F12/cc-pVDZ–F12–based vibrational analysis, to obtain more reliable estimates of the barrier height and the rovibrational properties of the •QOOH reactant, the TS, and the products. For simplicity, we label these calculations as ANL0F because they can be viewed as a revision of the ANL0 methodology (*18*) to incorporate a CCSD(T)–F12–based vibrational analysis. The major theoretical advances in

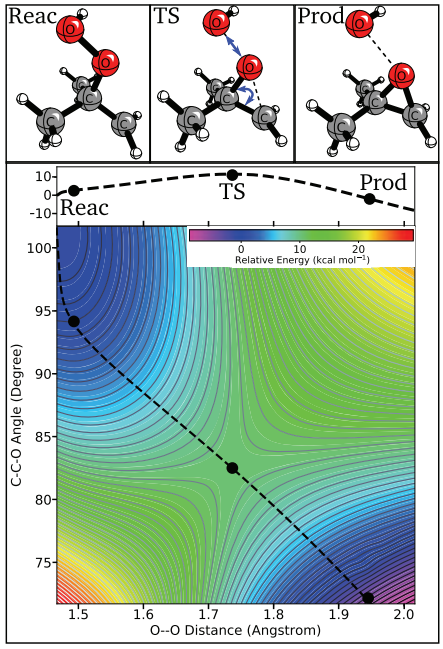


Fig. 3. Reaction coordinate for unimolecular dissociation of hydroperoxyalkyl radical intermediates (•QOOH). Snapshots (top) showing the O-O bond elongation and C-C-O angle contraction (blue arrows) from the •QOOH reactant (Reac) through the TS barrier at 10.3 kcal mol^{-1} to the OH radical and cyclic ether (DMEPOX) products (Prod) and the associated energy profile along the O-O coordinate (middle; kilocalories per mole). A two-dimensional reaction coordinate (bottom) derived from a constrained scan of the O-O distance and C-C-O angle demonstrates that the minimum energy path (black dashed line) progresses directly along these two internal coordinates from the reactants through the saddle point to the products, perpendicularly bisecting the uphill contours.

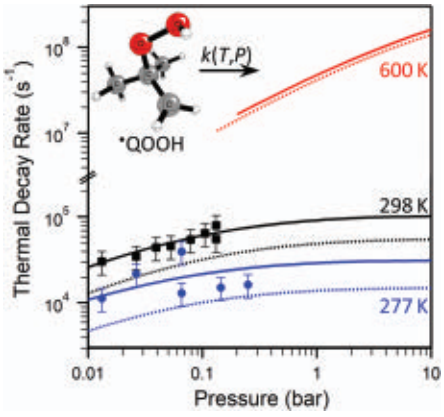


Fig. 4. Pressure-dependent thermal rates for unimolecular dissociation of hydroperoxyalkyl radical intermediates (•QOOH). Master equation modeling of the pressure-dependent thermal rate constant $k(T,P)$ for unimolecular dissociation of the •QOOH intermediate as a function of N_2 pressure at various temperatures [277 K (blue), 298 K (black), and 600 K (red)] based on the experimentally validated TS barrier (10.3 kcal mol^{-1}) with heavy-atom tunneling (solid) and without tunneling (dashed). The fully a priori predictions of the falloff curves are in good agreement with prior experimental observations of thermal rate constants and associated error bars from 250 to 300 K with N_2 bath gas from 0.01 to 0.25 bar reported in (*10*). The master equation simulations indicate that •QOOH is not well-described by a simple phenomenological rate coefficient at pressures <0.2 bar at 600 K because the thermalization and reaction time scales overlap. Analytic representations of the tunneling-corrected $k(T,P)$ are provided in table S10.

this work addressed the multireference character at the TS, quantum mechanical tunneling through the barrier, and multidimensional torsional effects for the •QOOH reactants and OO fission saddle point, which are detailed in the supplementary materials.

In brief, our ANL0F calculations predicted a TS barrier of 10.3 kcal mol^{-1}, 1.7 kcal mol^{-1} lower than the prior high-level evaluation, and a lower imaginary frequency of 763*i* cm^{-1} (versus 876*i* and 907*i* cm^{-1} for the prior UHF- and ROHF-based CCSD(T)/ANO0 calculations, respectively) along the reaction coordinate. The reaction coordinate of the minimum energy pathway from the •QOOH radical to the OH radical and DMEPOX products (Fig. 3) was composed of nearly equal contributions from two simple internal coordinates: O-O bond elongation that leads to OH radical products and C-C-O angle contraction that enables ring closure to form the cyclic ether product. Furthermore, our analysis showed that quantum mechanical tunneling occurred exclusively through heavy-atom motion along these coordinates.

These high-level electronic structure calculations, along with heavy-atom tunneling, provided an accurate starting point for statistical Rice-Ramsperger-Kassel-Marcus (RRKM) calculations of the energy-dependent unimolecular decay rates $k(E)$ for •QOOH (see supplementary materials). The computed microcanonical dissociation rates $k(E)$ were in nearly quantitative agreement with the experimental rates for the appearance of OH products over the energy range from ~10 to 14 kcal mol^{-1} (Fig. 2), where $k(E)$ was predicted to increase nearly 100-fold. Comparison of the computed dissociation rates for •QOOH with and without tunneling showed that tunneling substantially enhanced the rate at energies both above and below the apparent threshold. This enhancement is particularly notable given that the reaction coordinate motions involved heavy C and O atoms exclusively. Nevertheless, the imaginary tunneling frequency was large enough to increase $k(E)$ by a factor of 2 or more for experimentally observed energies that exceeded the reaction threshold by up to 2 kcal mol^{-1}.

Extension to thermal conditions

Two prior experimental studies have probed the appearance of OH radical products after the reaction of Cl atoms with TBHP under thermal conditions (*9*, *10*). The thermal rate constants $k(T,P)$ reported for the appearance of OH products at 298 K in N_2 and He bath gases were on the order of 10^4 s^{-1} and pressure dependent (up to 0.3 bar) (Fig. 4 and fig. S15). The observed rates indicated that the transient •QOOH intermediate formed in this reaction had a lifetime of only ~10 μs (~10^5 s^{-1} decay rate) under thermal conditions in the high-pressure limit. These studies are in stark contrast with the present study, where the •QOOH radical intermediate was stabilized and characterized through its IR spectroscopic fingerprint and energy-dependent unimolecular decay rates $k(E)$ obtained under collision-free conditions.

The microcanonical rates $k(E)$ measured experimentally and validated theoretically by first principles calculations in this work could be directly related to the thermal unimolecular rate constants $k(T,P)$ that are commonly used in global chemical models to predict oxidation in atmospheric and combustion environments (*19*, *20*). The high-pressure limit of the thermal dissociation rate constant simply corresponds to a Boltzmann average of the microcanonical rates. Thus, the present experimental observation of the microcanonical rates over a wide range of energy effectively determines the high-pressure limit of the thermal rate constant, although additional averaging over rotational states of the •QOOH and TS is required. Notably, however, most chemical processes, including prior reports of $k(T,P)$ for •QOOH (*9*, *10*), occur at pressures where a competition between collisional excitation and unimolecular dissociation results in a substantial falloff in the rate constant from its high-pressure limit. As a result, we used the master equation approach, which explicitly evaluates the effect of this competition through simulations of time-dependent variations in the energy- and angular momentum–dependent state populations, to predict the $k(T,P)$ values (*21*). The rates for collision-induced transitions

in energy and angular momentum, which play a key role in such master equation analyses of the falloff behavior, are generally estimated through simple ad hoc or empirically fitted approaches. Here, we instead evaluated them through first principles trajectory simulations (*22*), as described in more detail in the supplementary materials. This first principles approach provides far more rigor in extrapolations to pressures and temperatures outside the experimentally observed ranges.

Our predicted thermal rate constants $k(T,P)$ for unimolecular decomposition of •QOOH as a function of temperature (277, 298, and 600 K) and N_2 pressure across the falloff region to the high-pressure limit are shown in Fig. 4 (and in fig. S15 for He). These first principles theoretical predictions were in notably good agreement with prior experimental observations $k(T,P)$ for N_2 (Fig. 4) and He (fig. S15) bath gases, which span from 250 to 300 K and 0.01 to 0.25 bar (*9*, *10*). This accordance lends confidence to our theoretical predictions of $k(T,P)$ for the higher pressures and temperatures that are of more importance to atmospheric and combustion modeling efforts. Predictions for the pressure-dependent rates at 600 K are included in Fig. 4 because this temperature is representative of optimal chain-branching conditions with peak loss of fuel in low-temperature combustion (*23*). Notably, the present predictions showed substantial deviations from those based on the prior theory-experiment comparisons, as discussed in the supplementary materials.

Moreover, our study showed that quantum mechanical tunneling involving the heavy-atom O-O stretch and C-C-O bend motions also plays an important role in the thermal unimolecular decay of •QOOH, effectively doubling the pressure-dependent decay rates at 277 and 298 K (Fig. 4). Including tunneling, the unimolecular decomposition rate of •QOOH at 298 K in the high-pressure limit was predicted to be $1.0 \times 10^5\ s^{-1}$, which is twofold larger than prior estimates (see supplementary materials for other temperatures). Tunneling even increased the thermal rates at higher temperatures relevant to combustion, although only by ~15% at 600 K.

It is interesting to note that the microcanonical rates at high energies, as observed experimentally, play a direct role in a more complete treatment of the oxidation process. In particular, the •QOOH are typically produced from an ROO• species through passage over a fairly high-energy barrier (Fig. 1). At these high incipient energies, a substantial fraction of the •QOOH dissociates before it ever has a chance to be thermally stabilized through collisions. A master equation treatment of the full R• + O_2 kinetic process automatically incorporates the effect of such higher-energy dissociations in yielding predicted rate constants for "well-skipping" processes (*3*, *4*), such as the R• + O_2 → OH + cyclic-ether and the ROO• → OH + cyclic-ether processes. Such treatment depends heavily on the rate descriptions at much higher energies than are accessed in a typical thermal dissociation experiment. Thus, the determination of the microcanonical rates over a broad energy range, as in the present experimental effort, helps ensure the correct prediction of such well-skipping rate behavior.

A prototypical hydroperoxyalkyl (•QOOH) radical intermediate, which is transiently formed in isobutane oxidation, has been stabilized and identified under isolated conditions by its IR spectral signature. The present IR spectroscopic observations, in particular the strong and distinctive fundamental OH stretch (ν_{OH}) transition, demonstrated a general means for the detection of •QOOH radical intermediates in both controlled laboratory conditions and standard oxidation environments. IR activation of •QOOH also initiated its unimolecular decay to OH radical and cyclic ether products, which was used in this work to obtain energy-resolved and time-dependent measurement of its microcanonical dissociation rates $k(E)$. These measurements provided a critical test of the reaction pathway for its decomposition, which state-of-the-art electronic-structure methods showed has a TS barrier of 10.3 kcal mol^{-1} and an imaginary frequency of $763i$ cm^{-1} that are both ~15% lower than earlier predictions (*13*). Heavy-atom tunneling involving O-O elongation and C-C-O angle contraction effectively doubled the rate of unimolecular decay. Master equation modeling based on first principles trajectory simulations yields pressure-dependent thermal unimolecular rate constants $k(T,P)$, again substantially enhanced by tunneling. The theoretical analyses presented here could be readily generalized to provide similar-quality rate predictions for other •QOOH radical intermediates, which are needed for a wide range of real-world applications relevant to atmospheric and combustion chemistry.

REFERENCES AND NOTES

1. D. R. Glowacki, M. J. Pilling, *ChemPhysChem* **11**, 3836–3843 (2010).
2. F. Bianchi *et al*., *Chem. Rev.* **119**, 3472–3509 (2019).
3. D. L. Osborn, *Annu. Rev. Phys. Chem.* **68**, 233–260 (2017).
4. J. Zádor, C. A. Taatjes, R. X. Fernandes, *Pror. Energy Combust. Sci.* **37**, 371–421 (2011).
5. J. J. Orlando, G. S. Tyndall, *Chem. Soc. Rev.* **41**, 6294–6317 (2012).
6. F. Paulot *et al*., *Science* **325**, 730–733 (2009).
7. J. Peeters, J.-F. Müller, T. Stavrakou, V. S. Nguyen, *J. Phys. Chem. A* **118**, 8625–8643 (2014).
8. M. Ehn *et al*., *Nature* **506**, 476–479 (2014).
9. J. Zádor *et al*., *Phys. Chem. Chem. Phys.* **15**, 10753–10760 (2013).
10. C. A. Whelan *et al*., *J. Phys. Chem. A* **123**, 10254–10262 (2019).
11. J. D. Savee *et al*., *Science* **347**, 643–646 (2015).
12. J. Bugler, J. Power, H. J. Curran, *Proc. Combust. Inst.* **36**, 161–167 (2017).
13. K. B. Moore III, J. M. Turney, H. F. Schaefer III, *J. Chem. Phys.* **146**, 194304 (2017).
14. F. Liu, J. M. Beames, A. S. Petit, A. B. McCoy, M. I. Lester, *Science* **345**, 1596–1598 (2014).
15. J. D. DeSain, C. A. Taatjes, J. A. Miller, S. J. Klippenstein, D. K. Hahn, *Faraday Discuss.* **119**, 101–120 (2001).
16. C.-J. Chen, J. W. Bozzelli, *J. Phys. Chem. A* **103**, 9731–9769 (1999).
17. A. S. Hansen *et al*., *J. Chem. Phys.* **154**, 164306 (2021).
18. S. J. Klippenstein, L. B. Harding, B. Ruscic, *J. Phys. Chem. A* **121**, 6580–6602 (2017).
19. H. J. Curran, *Proc. Combust. Inst.* **37**, 57–81 (2019).
20. S. M. Saunders, M. E. Jenkin, R. G. Derwent, M. J. Pilling, *Atmos. Chem. Phys.* **3**, 161–180 (2003).
21. J. A. Miller, S. J. Klippenstein, *J. Phys. Chem. A* **110**, 10528–10544 (2006).
22. A. W. Jasper *et al*., *Science* **346**, 1212–1215 (2014).
23. L.-S. Tran *et al*., *Proc. Combust. Inst.* **37**, 511–519 (2019).
24. A. S. Hansen *et al*., Zenodo (2021); http://doi.org/10.5281/zenodo.4895832.

ACKNOWLEDGMENTS

The authors thank Z. Liu and S. Pandit for early contributions to the experimental research. **Funding:** The research at the University of Pennsylvania was primarily supported through the US Department of Energy–Basic Energy Sciences under grant DE-FG02-87ER13792. Partial equipment support was provided by the National Science Foundation under grant CHE-1955068. A.S.H. gratefully acknowledges support from the Carlsberg Foundation (CF18–0614) and the Independent Research Fund Denmark (9036–00016B). VPT2 calculations used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation grant no. ACI-1548562 through the allocation TG-CHE190088. K.B.M. was supported by the Molecular Structure and Dynamics program of the US Army Research Office under award no. W911NF1710531. Other ANL contributions were supported by the US Department of Energy (USDOE), Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences, under DOE contract no. DE-AC02-06CH11357. The effort of S.J.K. was supported by the core Gas Phase Chemical Physics program, and A.W.J. and D.R.M. were supported through the Argonne-Sandia Consortium on High-Pressure Combustion Chemistry, FWP no. 2009 ANL 59044. **Author contributions:** M.I.L. conceived the idea and designed the experimental research. A.S.H., T.B., and M.F.V. performed the experiments and analyzed the experimental data. A.S.H. carried out the spectroscopic calculations. S.J.K. and A.W.J. conceptualized and developed the theoretical research. Software development was performed by K.B.M., Y.G., and D.R.M. Calculations and analysis were carried out by S.J.K., K.B.M., and A.W.J. and include visualizations by K.B.M. A.S.H., T.B., K.B.M., A.W.J., S.J.K., and M.I.L. discussed and interpreted the experimental and theoretical results. S.J.K., K.B.M., and A.W.J. each participated in writing the theoretical results; A.S.H. and M.I.L. contributed to writing the experimental results. M.I.L. and S.J.K. supervised and guided the overall research. M.I.L. and S.J.K. wrote the final paper with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text, supplementary materials, and Zenodo (*24*), the latter of which contains the underlying data for figures and input for the Master Equation System Solver. Codes are available either on GitHub or commercially, as detailed in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6555/679/suppl/DC1
Materials and Methods
Supplementary Text
Optimized Geometries and Frequencies
Figs. S1 to S18
Tables S1 to S10
References (*25*–*60*)

18 April 2021; accepted 22 June 2021
10.1126/science.abj0412

REPORTS

METALLURGY

Suppressing atomic diffusion with the Schwarz crystal structure in supersaturated Al–Mg alloys

W. Xu, B. Zhang, X. Y. Li*, K. Lu*

High atomic diffusivity in metals enables substantial tuneability of their structure and properties by tailoring the diffusional processes, but this causes their customized properties to be unstable at elevated temperatures. Eliminating diffusive interfaces by fabricating single crystals or heavily alloying helps to address this issue but does not inhibit atomic diffusion at high homologous temperatures. We discovered that the Schwarz crystal structure was effective at suppressing atomic diffusion in a supersaturated aluminum–magnesium alloy with extremely fine grains. By forming these stable structures, diffusion-controlled intermetallic precipitation from the nanosized grains and their coarsening were inhibited up to the equilibrium melting temperature, around which the apparent across-boundary diffusivity was reduced by about seven orders of magnitude. Developing advanced engineering alloys using the Schwarz crystal structure may lead to useful properties for high-temperature applications.

Because of the nature of interatomic bonding, atomic diffusivity is notably higher in metals relative to that in ceramics and compounds with covalent or ionic bonds (*1*, *2*). This feature enables substantial tuneability of structures at various length scales by tailoring the diffusion-controlled processes during the synthesis and subsequent treatments, resulting in a broad spectrum of properties and performance in metallic materials. For instance, Al alloys can be hardened with precipitation of intermetallic phases by aging around room temperature (*3*, *4*). Strength and ductility of steels can be widely adjusted by manipulating diffusional phase transformations in thermomechanical treatments (*5*, *6*). However, high atomic diffusivity makes structure and the customized properties of metals unstable when they are exposed to elevated temperatures or mechanical loading (*7*). Such instability becomes a major bottleneck in the development of metallic materials, greatly limiting their technological applications at high temperatures.

Resisting diffusion of atoms in metals is challenging, especially at high temperatures. Heavily alloying a metal with foreign elements is limited by curbing lattice diffusion, as illustrated by the minor changes of diffusivity in the high-entropy alloys, in which several different metal elements are mixed together within a crystalline lattice (*8*, *9*). Interfaces or grain boundaries (GBs) associated with more open structures are believed to be fast diffusion channels for atoms relative to the crystalline lattice (*1*, *10*). Diffusion along GBs may be slowed down by optimizing the GB segregation of other elements (*11*–*13*). However, GB alloying is restricted by increasing tendencies of forming second phases with a higher degree of alloying (*14*, *15*).

Eliminating diffusive interfaces by forming single crystals is a standard strategy for reducing diffusivity, for example, as practiced in the fabrication of single-crystal blades of superalloys for high-temperature applications in turbo engines (*16*). Nonetheless, even in single-crystal metals, the high diffusivity cannot be restrained at high homologous temperatures. Substitutional diffusion and self-diffusion are known to be governed by the vacancy diffusion mechanism (*1*). The equilibrium vacancy concentration in crystalline lattice rises markedly at higher homologous temperatures, unavoidably elevating atomic diffusivity.

We recently discovered a metastable structure in pure Cu with extremely fine grains: a Schwarz crystal structure with minimal interfaces constrained by twin boundaries (*17*). Although it contains an extremely high density of interfaces, this structure exhibits a very high thermal stability against grain coarsening at high temperatures close to the melting point. Therefore, exploring whether this stable Schwarz crystal structure is capable of inhibiting diffusion of atoms in alloys at high temperatures is of great interest.

Al is a high-diffusivity metal, and Mg is one of its most diffusive alloying elements (*18*). We observed diffusion behaviors in a supersaturated

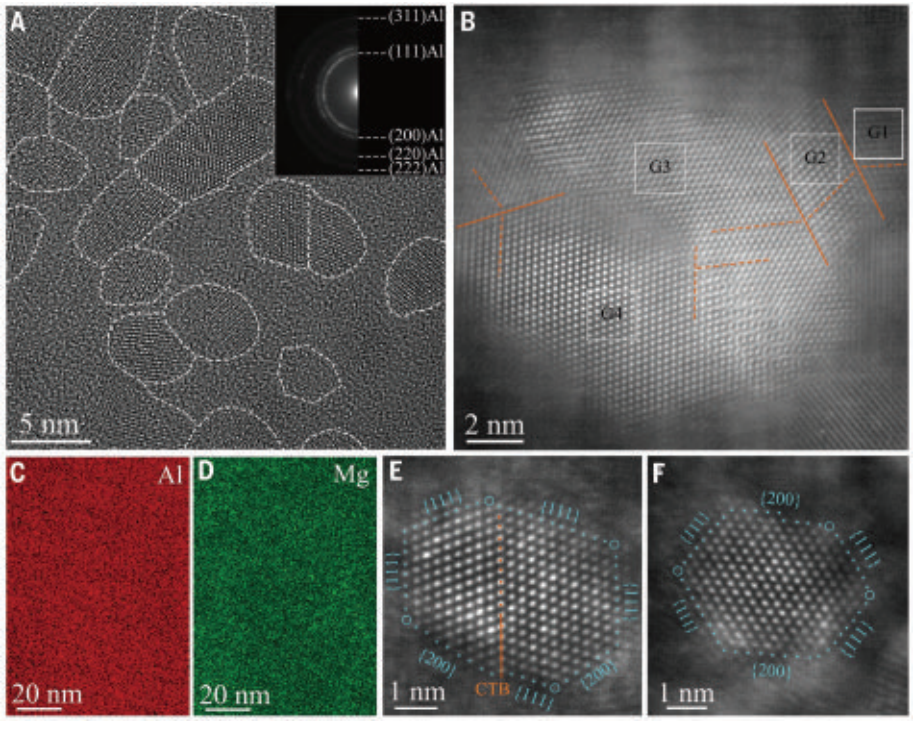


Fig. 1. Structure characterization of the as-prepared SC-8 sample. (**A**) Typical high-resolution TEM image showing nanosized grains outlined by white dashed lines. Inset is a corresponding selected area electron diffraction (SAED) pattern. (**B**) High-resolution high-angle annular dark-field (HAADF)–scanning transmission electron microscope (STEM) image of an aggregate in the manifold structure (fig. S3). Orange solid lines represent coherent twin boundaries (CTBs) and orange dashed lines are {111} planes. (**C** and **D**) Elemental mappings of Al and Mg. (**E** and **F**) High-resolution TEM images of a grain with a CTB (E) and a twin-free grain (F) with a beam direction along the [110] zone axis. Missing atoms at several corners are circled.

Shenyang National Laboratory for Materials Science, Institute of Metal Research, Chinese Academy of Sciences, Shenyang 110016, China.
*Corresponding author. Email: xyli@imr.ac.cn (X.Y.L.); lu@imr.ac.cn (K.L.)

Al–Mg alloy with a Schwarz crystal structure. Several diffusional processes, including precipitation of intermetallic phase, grain coarsening, and melting, were examined at different temperatures. We found that the Schwarz crystal structure was effective at suppressing atomic diffusion before structural melting.

We deformed a thin foil of coarse-grained Al–15 weight percent (wt %) Mg alloy (Al-15% Mg) with a single phase of supersaturated Al (Mg) solid solution using a high-pressure torsion apparatus at 77 K under a hydrostatic pressure of 10 GPa. With the applied strain exceeding ~20, the sample was structurally refined into the nanometer scale. Approximately equiaxed nanosized grains with random orientations were formed in the sample. The grain size distribution was uniform, with an average size of 8 nm (sample SC-8; Fig. 1A). Electron diffraction and X-ray diffraction (XRD) analysis verified that the sample was a fully crystalline, face-centered cubic Al solid solution (α-Al) without any other phase. Composition analysis using energy-dispersive X-ray spectroscopy (EDS) displayed homogeneous distributions of Al and Mg atoms throughout the sample (Fig. 1, C and D). The measured lattice constant of α-Al grains from quantitative XRD was 0.4119 ± 0.0001 nm, corresponding to a Mg concentration of ~15.9 wt % in terms of Vegard's law (*19*), which is close to the nominal composition. These results suggest that supersaturated Mg atoms were homogeneously distributed in the nanograined structure rather than clustering or segregating on GBs, as was documented in Al–Mg alloys deformed at room temperature (*20*).

For comparison, we prepared another sample with larger grains from the same alloy using a smaller plastic strain (~7.0). The sample consisted of equiaxed α-Al grains with random orientations (Fig. 2A and fig. S1). Grain sizes varied from 10 nm to ~100 nm, with an average of 50 nm (sample NG-50). Distributions of Al and Mg atoms were homogeneous in the supersaturated grains (fig. S2) and identical to that in SC-8. The Mg content estimated from the lattice constant was ~15.7 wt %, consistent with that in the SC-8 sample.

In addition to the obvious difference in grain size, the two samples had distinct structural characteristics. In the SC-8, we observed typical manifold structures of continuous networks consisted of irregularly shaped aggregates (fig. S3), akin to that in the Schwarz crystal Cu samples (*17*). The aggregates with smooth boundaries were made up of individual grains a few nanometers in size, among which twin relationships were often identified (Fig. 1B). Such manifold structures were not detected in the NG-50 sample.

Individual grains of the SC-8 sample exhibited diverse geometries with many faceted boundary planes (Fig. 1, E and F). Statistical measurements showed that ~57% of faceted boundaries were {111} atomic planes and ~26% were {100} planes. These frequencies are consistent with that in the Cu Schwarz crystal (*17*) and that of the truncated octahedron projected along the [110] axis (*21*), suggesting that the grain shapes resembled that of the truncated octahedron. Analogous to the Cu Schwarz crystal, atomic images at several corners of the grains were missing, as circled in Fig. 1E. Few faceted boundaries were seen in NG-50. GBs have wavy or curved high-angle boundaries (fig. S1), with morphologies similar to the so-called "high-energy nonequilibrium GBs" in the ultrafine-grained Al alloys (*22*).

In the SC-8 sample, we frequently pictured twins in some tiny grains, whereas some grains were defect free (Fig. 1, E and F). Along the [110] axis, we found that ~30% of grains contained

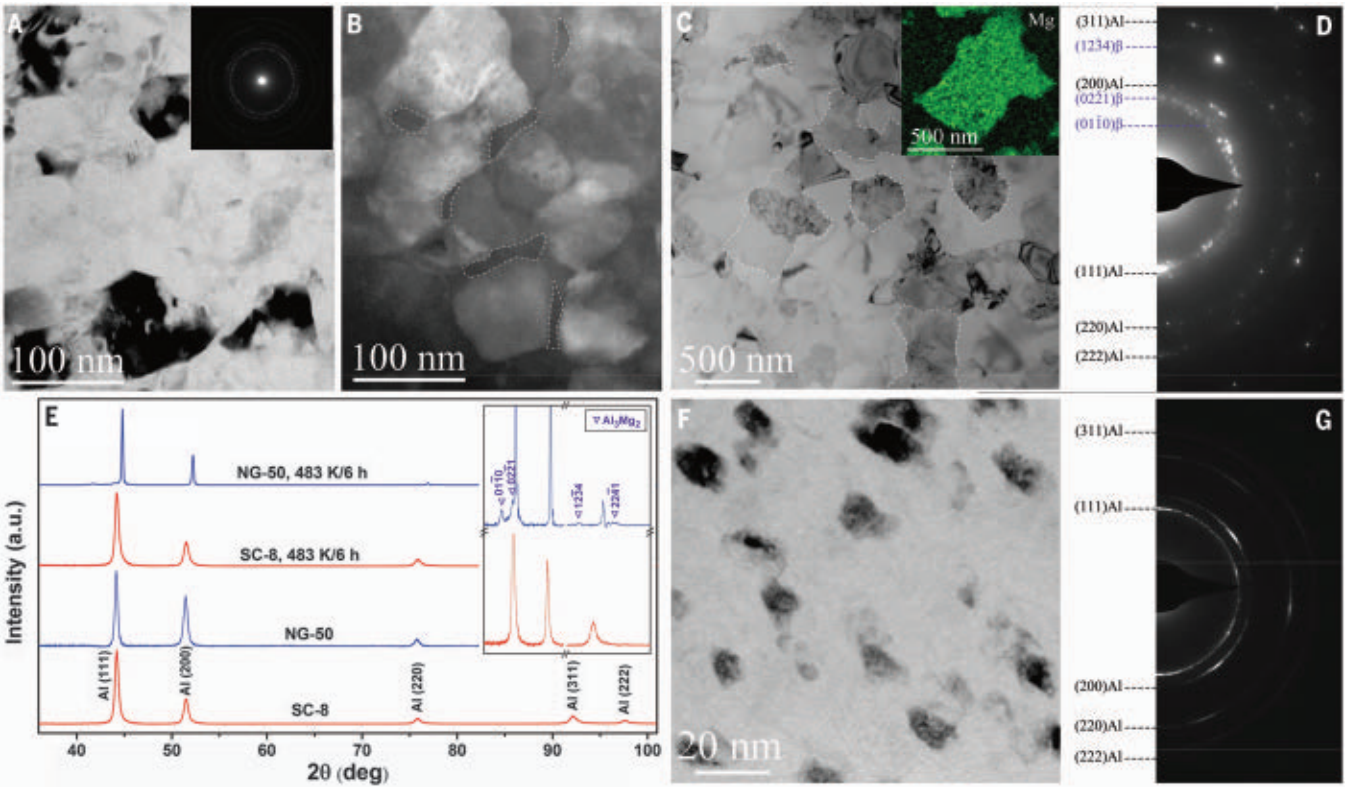


Fig. 2. Structure evolution upon annealing. (**A**) Typical bright-field TEM image of the as-prepared NG-50 sample with a SAED pattern (inset). (**B**) HAADF-STEM image of the NG-50 sample annealed at 423 K for 1 hour showing nanosized precipitates at GBs outlined by white dashed lines. (**C**) Bright-field TEM image of precipitates in the NG-50 annealed at 483 K for 6 hours. Inset is an elemental mapping of Mg of a precipitate. (**D**) Corresponding SAED pattern from (C) indexed with Al_3Mg_2 (β) and α-Al. (**E**) XRD patterns of the as-prepared and the annealed samples as indicated. The inset depicts magnified patterns of the two annealed samples. (**F** and **G**) Typical bright-field TEM image with a SAED pattern of the SC-8 sample annealed at 483 K for 6 hours.

one or more through-grain twins with fully coherent {111} twin boundaries in the as-prepared state. In the NG-50 sample, we did not see twins, as usually detected in ultrafine-grained Al alloys after plastic deformation. Because stacking fault energies for Al and its alloys are high (*23*), twinning might be possible only for extremely small grains when the shear stress is high enough. The critical grain size for the transition from full dislocation to partials can be calculated following classical dislocation theory (*24*) or in terms of the dislocation splitting distance. The estimated and observed sizes reported earlier were ~10 to 20 nm (*25*). This explains why twinning was predominant in the SC-8 sample, which may induce relaxation of GBs (*17*, *26*), whereas twinning was inhibited in the NG-50 sample.

The as-prepared SC-8 Al–Mg sample was structurally characterized by Schwarz crystal structures similar to those in the Cu sample with comparable grain sizes. However, in sharp contrast to the Schwarz crystal structures, the nanostructures in the NG-50 sample were not different from those in conventional nanograined metals and alloys prepared from severe plastic deformation. With these two different nanostructures, we examined comparatively the diffusion-controlled processes in the supersaturated Al–Mg alloy to clarify the effects of the Schwarz crystal structures on atomic diffusion.

For Al-15%Mg alloys, the equilibrium phases are α-Al and Al_3Mg_2 <700 K, a single α-Al at 700 to 736 K, and α-Al mixed with liquid at 736 to 855 K (fig. S4) (*27*). Several diffusional processes were expected in the supersaturated Al–Mg samples as annealed at different temperatures. In the NG-50 sample, precipitation of the Al_3Mg_2 phase along GBs of nanograins began as annealed >423 K (Fig. 2B) because of the much higher atomic diffusivity along GBs than that in the lattice (*1*). By raising annealing temperatures to 483 to 523 K, the amount and sizes of precipitates increased with enlarged diffusivities both along GBs and in the lattice (Fig. 2, C and D). Coarsening of α-Al grains was observable simultaneously (Figs. 2C and 3). Increasing the annealing duration also promoted the precipitation reaction. More precipitates with larger sizes were formed by increasing the annealing duration from 1 to 6 hours at 483 K. Further, Mg content in α-Al grains was lowered by annealing to ~6.0% at 523 K, consistent with the measured lattice constant of α-Al (Fig. 3), suggesting that the supersaturated Mg atoms in Al grains diffuse out to supply formation of the precipitates. The precipitation behaviors and the onset temperatures in the NG-50 agree well with those reported in Al–Mg alloys of similar compositions [373 to 423 K (*28*, *29*)].

The lattice constant and Mg concentration of α-Al grains in the NG-50 increased with annealing temperatures >523 K (Fig. 3) because of the dissolution of Al_3Mg_2 precipitates and diffusion of the dissolved Mg atoms back into α-Al grains. As temperature exceeded 700 K, precipitates were completely dissolved, forming a single α-Al phase with the same Mg concentration and lattice constant as that in the as-prepared state. Grains of α-Al continued to coarsen with dissolution of Al_3Mg_2 (Fig. 3). Grain sizes increased to the micrometer scale in the single-phase region with much enhanced atomic diffusivities of Mg and Al at high homologous temperatures. In terms of the measured increment in average grain sizes after annealing, we estimated the apparent across-boundary diffusivity at 723 K as being ~7.7×10^{-13} m^2 s^{-1}, which is consistent with the documented lattice diffusivity of Al (1.1×10^{-14} m^2 s^{-1}) (*30*).

After annealing the SC-8 sample in the same temperature ranges, however, we observed a different scenario. After annealing at 373 to 723 K, no precipitation was detected at all in the SC-8 using transmission electron microscope (TEM) observations and XRD analysis (Fig. 2E). Microhardness of the sample remained unchanged after annealing at 483 K for 48 hours (fig. S5). The distributions of Mg and Al were as homogeneous as that in the as-prepared state (Fig. 4 and fig. S6). Only the α-Al phase with an unchanged lattice constant was identified in the annealed samples (Fig. 3), suggesting that Mg atoms stay in the supersaturated α-Al nanograins without segregation or formation of intermetallic phases in the entire temperature range studied.

The sizes of α-Al grains in the annealed sample remained stable in the SC-8 sample. Statistical measurements showed that the average grain size of α-Al increased from 8 to 10 nm after annealing at 673 K for 1 hour (Fig. 3), corresponding to a minor drop in microhardness from 4.8 to 4.1 GPa (fig. S8). The average grain size increased to ~12 nm as annealed at 723 K for 1 hour (Fig. 3 and fig. S7), whereas the geometry, lattice constant, and solute distribution of α-Al grains remained unchanged. High-resolution TEM observations denoted truncated octahedron–shaped grains in the annealed sample (fig. S9) that were not obviously different from that in the as-prepared state. Using the grain size change in the SC-8 sample annealed at 723 K, we estimated the apparent across-boundary diffusivity following the same equation as in the NG-50 sample: ~1.5×10^{-20} m^2 s^{-1}, or about seven orders of magnitude lower than that in the NG-50 sample.

The annealing-induced Al_3Mg_2 precipitation from α-Al nanograins and their coarsening were inhibited in the SC-8 sample in the temperature range studied, differing sharply from that in the NG-50 and the ultrafine-grained Al–Mg alloys (*20*). Both the precipitation reaction and the GB migration process that govern grain coarsening were kinetically controlled by diffusion of both elements. The density of GBs in the SC-8 sample was several times higher than that in NG-50, and GBs are often believed to be fast diffusion channels for atoms. Therefore, these diffusional processes may be expected to be kinetically promoted in the SC-8 sample rather than being inhibited. However, the experimental observations were the opposite of this expectation.

The huge difference in the obtained diffusivities between the two samples implies that

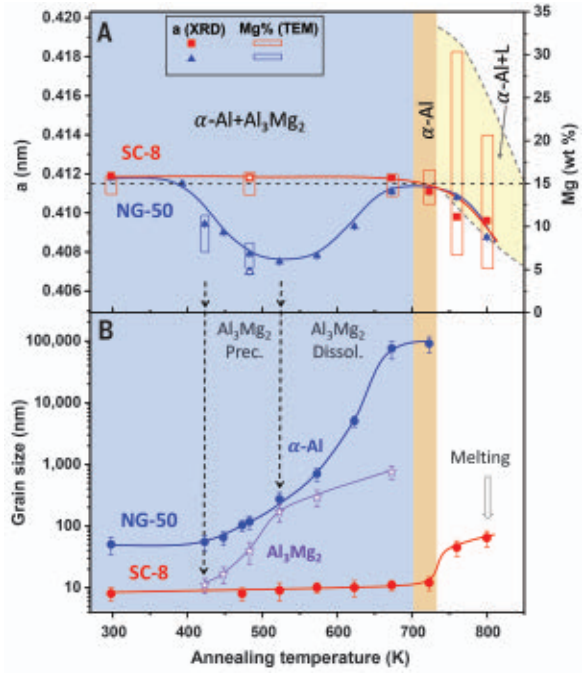


Fig. 3. Stability of lattice constant and grain size. (**A**) Variation of measured lattice constant (*a*) from XRD analysis and Mg concentration from EDS analysis in the two samples as a function of annealing temperature (for a duration of 1 hour). The open square and open triangle represent the SC-8 and NG-50 sample annealed at 483 K for 6 hours, respectively. (**B**) Variation of the average size of α-Al grains in the two samples with annealing temperature. Al_3Mg_2 precipitate sizes in the NG-50 sample are included. Three regions of different phases in the equilibrium phase diagram (fig. S4) are colored.

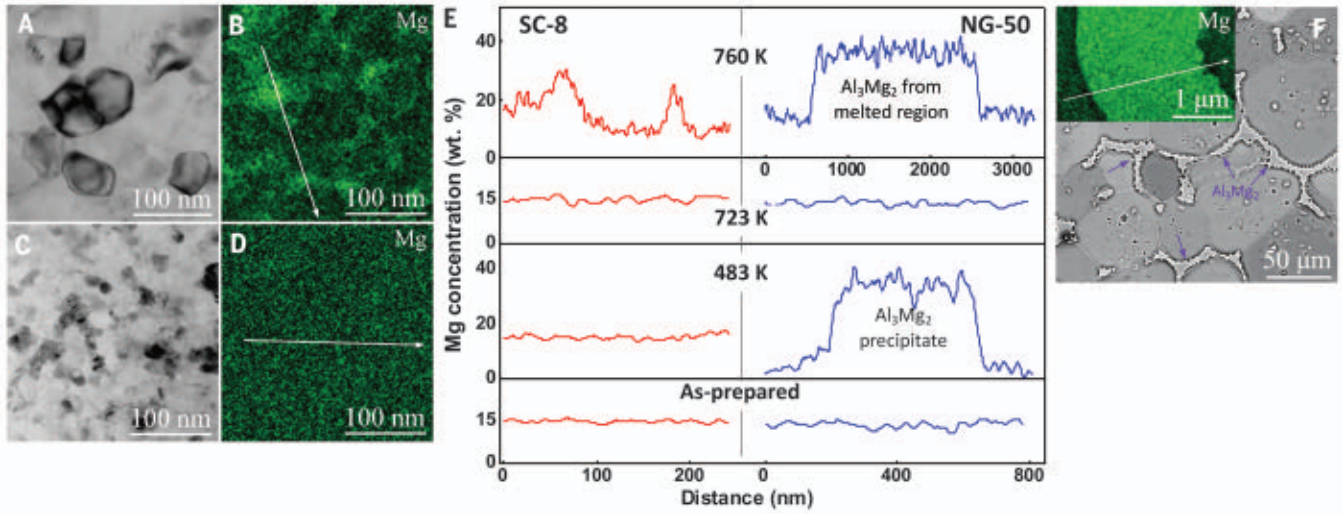


Fig. 4. Elemental distributions upon annealing. (**A** to **D**) Typical bright-field TEM images and corresponding EDS elemental mappings of Mg in the SC-8 samples annealed at 760 K [(A) and (B)] and 723 K [(C) and (D)] for 1 hour, respectively. (**E**) EDS line scans of Mg concentration in the as-prepared and the annealed samples at various temperatures (as indicated) for 1 hour. (**F**) SEM image of the annealed NG-50 sample and EDS elemental mapping (inset) of the Al_3Mg_2 phase solidified from the melted region (arrow).

atomic diffusion was markedly suppressed in the SC-8 sample with the stable Schwarz crystal structures. The melting processes of alloys are controlled by atomic diffusion and usually occur within a temperature range instead of at a specific temperature as in pure metals. The inhibited diffusion in the SC-8 sample is also verified by its melting behaviors.

As the coarse-grained Al-15%Mg alloy was heated above the solidus temperature, segregation of Mg atoms was induced by fast diffusion of both elements. The rapid redistribution of Mg atoms led to a split in melting temperature: Mg-rich regions with lowered melting temperatures and Mg-poor regions with higher ones. With increasing temperatures, the Mg-rich regions melted preferably and expanded by consuming the surrounding Mg-poor solids. This scenario is exactly what we observed in the NG-50 sample. We measured partial melting around GBs of Al grains as they were annealed >736 K. After quenching, the Al_3Mg_2 phase crystallized from the melted regions (Fig. 4).

However, when we annealed the SC-8 sample at 760 K for 1 hour, we observed an α-Al phase without detectable Al_3Mg_2. An increased fluctuation in Mg distribution (within 6.6 to 30.3% Mg) compared with that at lower annealing temperatures (723 K; Fig. 4) was induced, which is indicative of redistribution of Mg atoms. The highest Mg concentration was below the eutectic composition (35.8%) and was not high enough to induce melting. The grain sizes increased to ~45 nm after annealing, which was much smaller than that in the NG-50 sample. The lattice constant of α-Al grains dropped slightly with a depletion of Mg. We found partial melting only when the sample was annealed >805 K (figs. S10 and S11), 69 K higher than the solidus temperature and that of the NG-50 sample. The surprising elevation in melting temperature of the Al–Mg alloy in the SC-8 sample suggests that atomic diffusion was insufficient in the Schwarz crystal structures for inducing melting at the solidus temperature.

A Schwarz crystal is structurally characterized by minimal interfaces with zero mean curvature constrained by twin boundaries, which are extremely stable against thermal and mechanical loading. Molecular dynamic simulations have indicated that in Schwarz crystal structures (*17*), atoms at interfaces are constrained to vibrate about their rest positions to maintain the stable minimal interfaces at elevated temperatures even in vicinity of the melting point. The constraint is so effective that violent oscillation of atoms with large deviations from their rest positions is not allowed, limiting the local collective motion of atoms. In other words, the constraint greatly reduces the possibility for the interfacial atoms to escape from their rest positions. Therefore, the diffusion of atoms was inhibited in the entire temperature range studied because diffusion coefficients are proportional to the jumping probability (*2*).

With the extremely high density of GBs, the equilibrium vacancy concentration in the tiny α-Al grains would be so low that it would be difficult for a vacancy to be stabilized inside the grain. For substitutional diffusion processes, which are dominated by vacancy mechanisms (*1*, *2*), lattice diffusion coefficients are proportional to the probability of finding a vacancy adjacent to the diffusing atoms. Therefore, lattice diffusion of Al and Mg atoms in the extremely fine grains is also inhibited.

Our observations in this supersaturated Al–Mg alloy echo our earlier observation of inhibited coarsening of nanograins in pure Cu Schwarz crystal samples up to the melting point (*17*), which is a self-diffusion–controlled process. The diffusionless feature of the Schwarz crystal structure in metals is of great importance for understanding the fundamental diffusional processes in interfaces and the transport dynamics in solid states, especially at high temperatures. The Schwarz crystal seems to provide a robust barrier for arresting the diffusion of atoms in metals and substitutional alloys, boosting stability up to melting temperatures. This stability is much higher than that for conventional alloys. Developing advanced Al and other alloys using the Schwarz crystal structure should lead to materials with useful properties for high-temperature applications.

REFERENCES AND NOTES

1. H. Mehrer, *Diffusion in Solids: Fundamentals, Methods, Materials, Diffusion-Controlled Processes* (Springer, 2007), vol. 155.
2. D. A. Porter, K. E. Easterling, *Phase Transformations in Metals and Alloys* (CRC Press, 2009).
3. I. Polmear, D. St. John, J. F. Nie, M. Qian, *Light Alloys* (Butterworth-Heinemann, 2017).
4. M. Werinos *et al.*, *Acta Mater.* **118**, 296–305 (2016).
5. M. Song *et al.*, *Acta Mater.* **112**, 361–377 (2016).
6. F. P. Pickering, *Constitution and Properties of Steels*, vol. 7 of *Materials Science and Technology*, R. W. Cahn, P. Haasen, E. J. Kramer, Eds. (Wiley, 2005).
7. M. Meyers, K. Chawla, *Mechanical Behaviors of Materials* (Cambridge Univ. Press, 2009).
8. Q. Ding *et al.*, *Nature* **574**, 223–227 (2019).
9. K.-Y. Tsai, M.-H. Tsai, J.-W. Yeh, *Acta Mater.* **61**, 4887–4897 (2013).
10. S. V. Divinski, G. Reglitz, H. Rosner, Y. Estrin, G. Wilde, *Acta Mater.* **59**, 1974–1985 (2011).
11. J. Weissmüller, *Nanostruct. Mater.* **3**, 261–272 (1993).
12. R. Kirchheim, *Acta Mater.* **50**, 413–419 (2002).
13. C. C. Koch, R. O. Scattergood, K. A. Darling, J. E. Semones, *J. Mater. Sci.* **43**, 7264–7272 (2008).

14. T. Chookajorn, H. A. Murdoch, C. A. Schuh, *Science* **337**, 951–954 (2012).
15. D. Raabe *et al.*, *Curr. Opin. Solid State Mater. Sci.* **18**, 253–261 (2014).
16. R. Reed, *The Superalloys: Fundamentals and Applications* (Cambridge Univ. Press, 2006).
17. X. Y. Li, Z. H. Jin, X. Zhou, K. Lu, *Science* **370**, 831–836 (2020).
18. Y. Du *et al.*, *Mater. Sci. Eng.* **363**, 140–151 (2003).
19. A. R. Denton, N. W. Ashcroft, *Phys. Rev. A* **43**, 3161–3164 (1991).
20. X. Sauvage, N. Enikeev, R. Valiev, Y. Nasedkina, M. Murashkin, *Acta Mater.* **72**, 125–136 (2014).
21. C. Chieh, *Acta Crystallogr. A* **5**, 946–952 (1979).
22. K. Oh-ishi, Z. Horita, D. J. Smith, T. G. Langdon, *J. Mater. Res.* **16**, 583–589 (2011).
23. Y. Qi, R. K. Mishra, *Phys. Rev. B Condens. Matter Mater. Phys.* **75**, 224105 (2007).
24. E. Orowan, *Symposium on Internal Stresses in Metals and Alloys* (Institute of Metals, 1948), vol. 451.
25. M. Chen *et al.*, *Science* **300**, 1275–1277 (2003).
26. X. Zhou, X. Y. Li, K. Lu, *Science* **360**, 526–530 (2018).
27. H. Okamoto, T. Massalski, *Binary Alloy Phase Diagrams* (ASM International, 1990).
28. M. J. Starink, A. Zahra, "Precipitation kinetics of an Al-15% Mg alloy studied by microcalorimetry and TEM," in *Materials Science Forum* (Trans Tech Publications, 1996), vol. 217, pp. 795–800.
29. S. Nebti, D. Hamana, G. Cizeron, *Acta Metall. Mater.* **43**, 3583–3588 (1995).
30. W. Zhong, M. S. Hooshmand, M. Ghazisaeidi, W. Windl, J.-C. Zhao, *Acta Mater.* **189**, 214–231 (2020).

ACKNOWLEDGMENTS

We thank C. Y. Wu, B. Wang, and L. Fang for help with sample preparation and Z. H. Jin for discussion. **Funding:** This work was supported by the Ministry of Science & Technology of China (grant nos. 2017YFA0204401 and 2017YFA0700700) and the Chinese Academy of Sciences. **Author contributions:** X.Y.L. and K.L. developed the concept, designed the experiments, and supervised the projects. W.X. prepared the samples, characterized structures, and performed the measurements. W.X. and B.Z. performed the calculations. All authors discussed and analyzed the results. X.Y.L. and K.L. wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6555/683/suppl/DC1
Materials and Methods
Figs. S1 to S11
References (*31*, *32*)

13 February 2021; accepted 1 July 2021
10.1126/science.abh0700

NANOFLUIDICS

Modeling of emergent memory and voltage spiking in ionic transport through angstrom-scale slits

Paul Robin, Nikita Kavokine, Lydéric Bocquet*

Recent advances in nanofluidics have enabled the confinement of water down to a single molecular layer. Such monolayer electrolytes show promise in achieving bioinspired functionalities through molecular control of ion transport. However, the understanding of ion dynamics in these systems is still scarce. Here, we develop an analytical theory, backed up by molecular dynamics simulations, that predicts strongly nonlinear effects in ion transport across quasi–two-dimensional slits. We show that under an electric field, ions assemble into elongated clusters, whose slow dynamics result in hysteretic conduction. This phenomenon, known as the memristor effect, can be harnessed to build an elementary neuron. As a proof of concept, we carry out molecular simulations of two nanofluidic slits that reproduce the Hodgkin-Huxley model and observe spontaneous emission of voltage spikes characteristic of neuromorphic activity.

Neurotransmission relies on the harmonious transport of manifold ionic species across the cellular membrane (*1*, *2*). In particular, the nonlinear, history-dependent dynamics of biological ion channels (*3*) is key to many neuronal processes. Although considerable progress in the design of new nanofluidic devices has been achieved over the past decade (*4*–*10*), artificial systems still cannot compete with the ionic machinery of neurons. This points to the need for the invention and design of artificial iontronic devices with advanced functionalities (*4*). Most notably, an ion-based memristor—short for memory resistor, an electronic device with hysteretic conductance (*11*, *12*)—could serve as an elementary building block for ion-based neuromorphic systems. In the fabrication of nanofluidic systems, a milestone has very recently been reached, as planar confinement of an electrolyte down to a single molecular layer (see Fig. 1, A and B) was demonstrated and investigated experimentally (*13*–*15*). Extremely confined systems are promising from a functional point of view (*16*): Because they are most sensitive to the discrete nature of ions (*17*, *18*), they are prone to exhibiting exotic transport properties (*15*, *19*, *20*). Yet, modeling such subnanometric systems requires going beyond the traditional Poisson-Nernst-Planck (PNP) equations (*18*) because the reduced dimensionality drastically affects the nature of charge interactions and results in qualitative changes in the conduction dynamics (*20*). Here, we develop a theoretical and numerical framework to study such two-dimensional (2D) systems and show how they can be used to build new iontronic devices inspired by biological neurons.

In bulk electrolytes, ionic transport is essentially linear because the high dielectric constant of water reduces the strength of ionic interactions. Conversely, in 1D geometries (such as in carbon nanotubes), translational degrees of freedom are extremely constrained, limiting correlation times and potential memory effects. Here, we focus on an intermediate geometry: a monolayer of ions, molecularly confined in a subnanometric slit (see Fig. 1). In such strong planar confinement, ions experience not only a reduction in translational freedom but also stronger electrostatic interactions whose nature is intermediate between that of 3D and 2D systems, which we term "$2D^+$" interactions.

We first investigate the equilibrium properties of 2D electrolytes. Using all-atom molecular dynamics (MD) simulations, we consider aqueous solutions of various salts—such as NaCl, $CaCl_2$, and $CaSO_4$—confined in a narrow slit of tunable spacing h = 0.7, 1.0 or 1.4 nm (amounting to one, two, and three water layers, respectively) made of two sheets of graphene or hexagonal boron nitride. Details of the simulations are provided in the supplementary materials. Although these salts are known to be completely dissociated in bulk solutions, we find that in monolayer confinement, they may form tightly bound Bjerrum pairs (*21*) (see Fig. 1C)—and even triplets, in the case of $CaCl_2$. Even in two- or three-layer confinement, all salts except NaCl still associate into pairs. This pairing was recently predicted for electrolytes confined in 1D carbon nanotubes (*20*) and can similarly be explained in the $2D^+$ case in terms of the confinement-induced effective interaction between the ions. Because the water dielectric permittivity ϵ_w is much larger than that of the confining medium, the electric field lines created by an ion are forced to remain parallel to the channel walls over a typical length ξ. An exact computation, reported in the supplementary text, yields, in the case of a symmetric electrolyte, the corresponding pair-wise interaction potential (see fig. S3):

$$\beta V_{ij}(r) = -\frac{q_i q_j}{T^*}\log\frac{r}{r+\xi} \qquad (1)$$

where β is inverse temperature, V_{ij} is the interaction potential between ions i and j, r is the distance between the two ions, $q_{i,j} = \pm 1$ is the charge sign, and $1/T^*$ is a dimensionless coupling constant. We derive the precise

Laboratoire de Physique de l'École Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, F-75005 Paris, France.
***Corresponding author. Email: lyderic.bocquet@ens.fr**

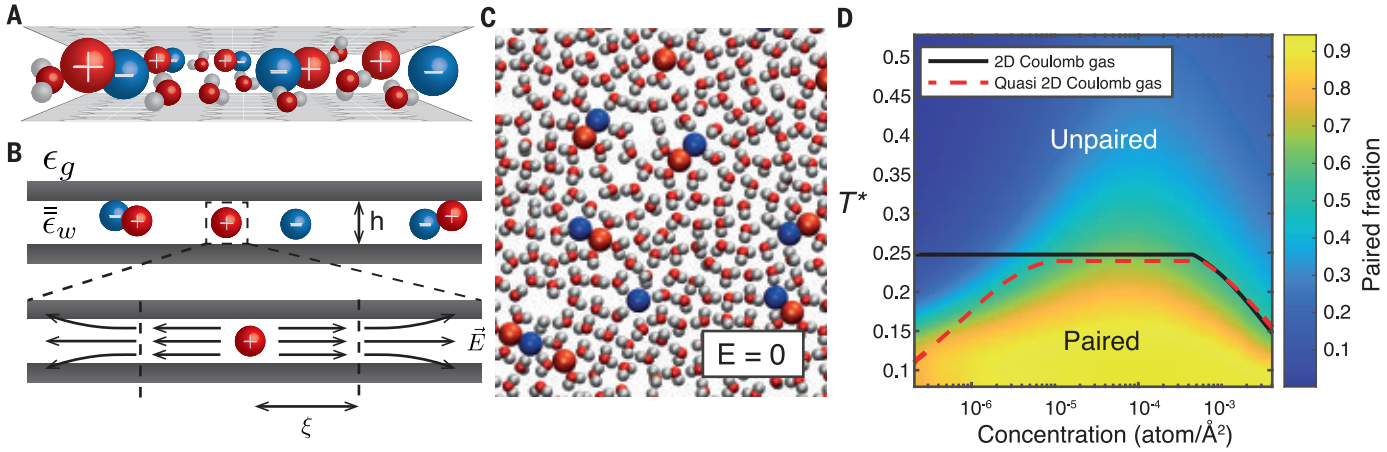


Fig. 1. Simulation and modelization of a 2D electrolyte. (**A**) Sketch of a 2D electrolyte. Red, sodium ions; blue, chloride ions; red and silver, water molecules. (**B**) Quasi-2D Coulomb gas model, with $2D^+$ Coulomb interactions given by Eq. 1. The electric field is confined in the channel over a length ξ owing to dielectric contrast between water and graphite. The vertical dashed lines indicate the shift from quasi-2D to 3D electrostatic interactions. ϵ_g, graphite dielectric permittivity; ϵ_w, water dielectric permittivity. (**C**) Typical configuration of a monolayer electrolyte as observed in all-atom MD simulations. Sodium ions (in red) and chloride ions (in blue) form tightly bound Bjerrum pairs at room temperature in slits of height h = 0.7 nm. (**D**) Phase diagram of a 2D electrolyte, obtained from Brownian dynamics simulations. At reduced temperature $T^* = 0.25$, the system undergoes a pairing phase transition, analogous to the KT transition. The solid black and dashed red curves correspond to the mean-field prediction for the transition line for an exactly 2D ($\xi = \infty$) or $2D^+$ (finite ξ) Coulomb gas, respectively (see supplementary text).

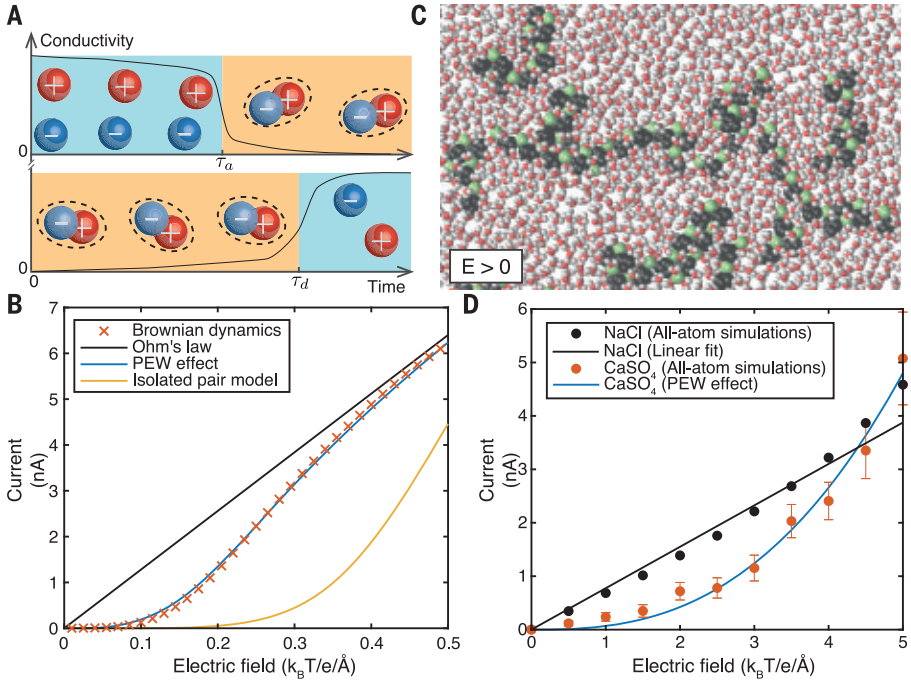


Fig. 2. Analytical theory of the PEW effect. (**A**) Schematic representation of Onsager's Wien effect for an isolated ion pair: Free ions have a lifetime of order τ_a, during which they contribute to conduction before assembling into pairs (top). Pairs typically break up after a time τ_d and do not take part in conduction (bottom). (**B**) *I-V* characteristic of a generic divalent salt. Taking into account the formation of polyelectrolytes substantially improves the agreement with simulations (blue line, see Eqs. 6 and 7) compared with the IP model (yellow line, see Eq. 4) for an ionic concentration $\rho = 10^{-3}$ atom/nm^2. (**C**) Under a constant external field, Bjerrum pairs assemble into complex structures: Bjerrum polyelectrolytes are shown here with $CaSO_4$. Calcium ions are in green, and sulfate ions are in black. (**D**) *I-V* characteristic of $CaSO_4$ in a 1-nm slit, as obtained from all-atom simulations, compared with the power law $I \propto E^a$ of the PEW effect theory, with the predicted exponent $a = a_{PEW} = 2.6$. NaCl does not form pairs in slits of that height and exhibits a linear response.

expressions of $V(r)$, T^*, and the dielectric length ξ in the supplementary text (see section 3.1), notably in the experimentally relevant case where the dielectric permittivity of confined water is anisotropic (*22*, *23*). In the case of divalent salt in a slit of height h = 0.7 nm at room temperature, we typically find $T^* = 0.11$ and $\xi = 14$ nm. The corresponding Bjerrum length ℓ_B, defined by $\beta V(\ell_B) = 1$, is 130 nm, which is much larger than the Bjerrum length in bulk water ($\ell_B = 0.7$nm). Therefore, the interaction potential in Eq. 1 is much stronger than its bulk counterpart $V_{bulk}(r) = e^2/4\pi\epsilon_0\epsilon_w r$; this qualitatively explains the confinement-induced ion pairing. The coupling constant $1/T^*$ is also proportional to Z^2/h, where Z is the valence of ions, explaining why monovalent ions only form pairs in the thinnest slits.

To identify the range of effective temperatures and ionic concentrations in which Bjerrum pairing occurs (Fig. 1D), we performed implicit solvent Brownian dynamics (BD) simulations of a symmetric electrolyte interacting with the derived potential in Eq. 1. At a short distance ($r \ll \xi$), our interaction potential is logarithmic and resembles a 2D Coulomb potential. The pairing transition in our monolayer electrolyte is therefore almost described by a 2D Coulomb gas model (*24*, *25*) and is analogous to the Kosterlitz-Thouless (KT) topological phase transition (*26*). However, the analogy is not perfect because the interaction potential recovers a bulk $1/r$ behavior at large distances $r \gg \xi$. To account for these observations, we use a mean-field approach inspired by Fuoss's theory of bulk

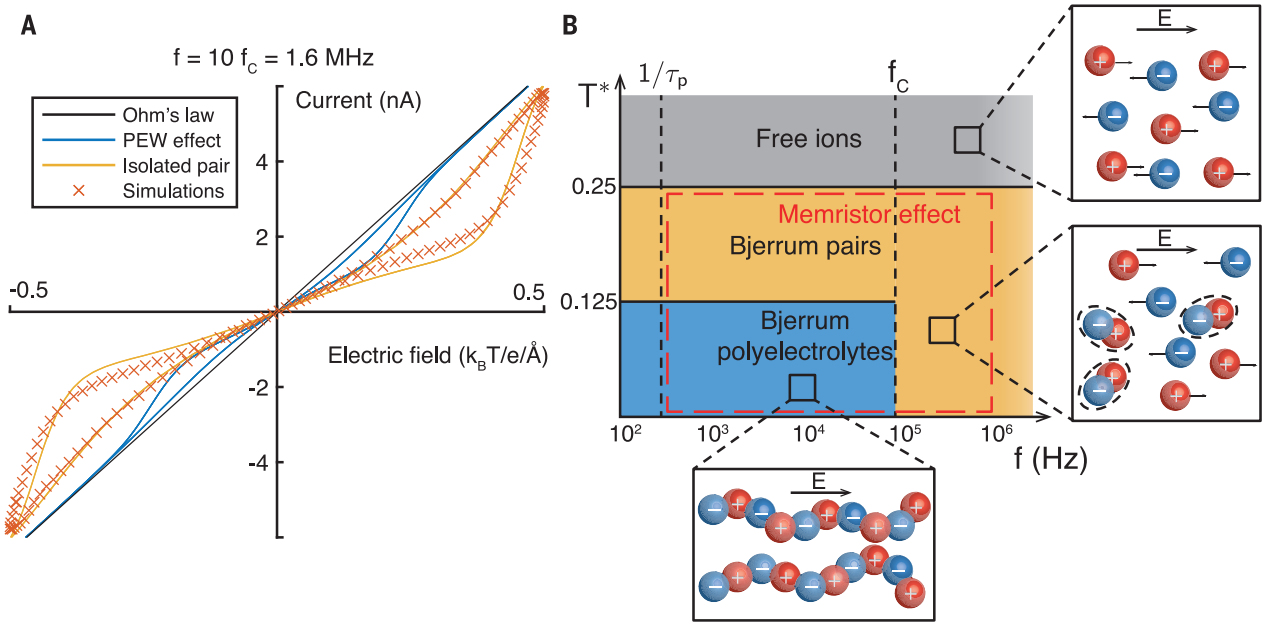


Fig. 3. Time-dependent regime and memristor effect. (**A**) "Pinched" *I-V* characteristic at high frequency (here, $f = 10 f_C = 1.6$ MHz and $T^* = 0.11$), typical of memristive devices. Here, the IP model (yellow line, Eqs. 2 and 4) provides a better prediction than the PEW effect model (blue line, Eqs. 2 and 6), showing that Bjerrum polyelectrolytes are unstable at high frequency. (**B**) Temperature-frequency diagram summarizing the existence domains of Bjerrum pairs and polyelectrolytes for a generic salt with $r_0 = 1$ Å, $D = 10^{-9}$ m^2/s, and ionic concentration $\rho \sim 10^{-3}$ atom/nm^2. Orders of magnitude for other values of parameters are obtained by noticing that the natural scale for frequencies is D/r_0^2.

electrolytes (*27*), in which ion pairs are incorporated as a separate species. We are able to determine analytically the pairing transition temperature, both in the ideal 2D and our 2D$^+$ setting (Fig. 1D), as detailed in the supplementary text, section 3. Our analytical computation reproduces quantitatively the BD simulations, and the ideal 2D description turns out to be valid for all but the lowest salt concentrations. The ion pairing quasi-KT phase transition appears as a specific feature of the monolayer electrolyte. In 1D confinement, there can be no phase transition (*20*), and in three dimensions, the ion-ion interactions are usually not strong enough for pairing to occur at room temperature (*24*). By contrast, our model predicts a transition temperature $T = 350$ K for divalent salts like $CaSO_4$ in slits with $h = 1.4$ nm or for $CaCl_2$ with $h = 0.7$ nm.

Such strong ionic correlations at equilibrium are at the source of highly nonlinear ion transport, illustrated by the current-voltage (*I-V*) characteristics obtained from MD simulations (Fig. 2, B and D). This nonlinearity is an indicator that conduction proceeds through the breaking of ion pairs. Consistently, a nonlinear response is observed for $CaSO_4$, which is paired at equilibrium, unlike NaCl, which does not form pairs in the considered conditions (Fig. 2D).

This type of ion transport is known as the second Wien effect, as famously pioneered by Onsager (*28*, *29*). We accordingly develop a theory for the Wien effect in the 2D$^+$ geometry. We consider a chemical equilibrium between pairs and "free" ions of the form $NaCl \leftrightarrows Na^+ + Cl^-$ (see Fig. 2A). Assuming that pairs dissociate with a time scale τ_d and free ions assemble into pairs with a time scale τ_a, we obtain an evolution equation for the fraction n_f of free ions not engaged in a pair:

$$\dot{n}_f = \frac{1 - n_f}{\tau_d} - \frac{n_f^2}{\tau_a} \tag{2}$$

Although computing the dependence of τ_d with parameters is usually a mathematical challenge, we were able to reduce it to a self-similar problem in the particular case of 2D$^+$ confined electrolytes (see supplementary text, section 3) yielding:

$$\tau_a = \frac{T^*}{4\pi D \rho} \tag{3}$$

$$\tau_d = \frac{r_0^2}{2D}\left(\frac{l_E}{r_0}\right)^{1/T^*} \tag{4}$$

where the length scale $l_E = k_B T/ZeE$ describes the strength of the external field with respect to thermal fluctuations, k_B is the Boltzmann constant, E is the electric field, D is the diffusion coefficient, ρ is the ionic concentration, and r_0 is the radius of considered ions. Because this model considers ion pairs as independent, noninteracting particles, we refer to it as the isolated pair (IP) model. When the steady state is reached, Eq. 2 can be solved, and the ionic current reads

$$I = N\frac{2ZeD}{k_B T L} E \frac{\tau_a}{2\tau_d}\left(\sqrt{1 + \frac{2\tau_d}{\tau_a}} - 1\right) \underset{E \to 0}{\propto} E^{a_{IP}(T^*)} \tag{5}$$

where N is the number of ions, L is the channel's length, and $a_{IP}(T^*) = 1 + 1/2T^*$ is the *I-V* curve exponent at low applied voltage. However, this prediction fails to reproduce simulation results, even at very low concentrations (see the solid yellow line in Fig. 2B). The all-atom MD simulations provide some hints to understanding this discrepancy. As shown in Fig. 2C, under an external electric field, ion pairs do not actually break but instead rearrange into gigantic clusters with chemistry-specific size and topology (see fig. S1 and supplementary text, section 2). These clusters, which we term "Bjerrum polyelectrolytes," lead to a radically new phenomenology.

The formation of macrostructures indicates that ion pairs cannot be treated independently, and thus the Wien effect dynamics is fundamentally different from Onsager's picture. Going beyond the IP model essentially amounts to taking into account the effect of Debye screening, characterized by the Debye length $\lambda_D = \sqrt{\frac{T^*}{\rho n_f}}$. Because under a (low) electric field E the density of the free ions is $n_f \sim \tau_d^{-1/2} \propto l_E^{-1/2T^*}$, $\lambda_D \propto l_E^{1/4T^*}$ sets the size of the screening cloud around a given ion. Yet, two ions are carried away from each other if they are separated by a distance exceeding $l_E/T^* \propto l_E$ in the direction of the field, say x. For sufficiently

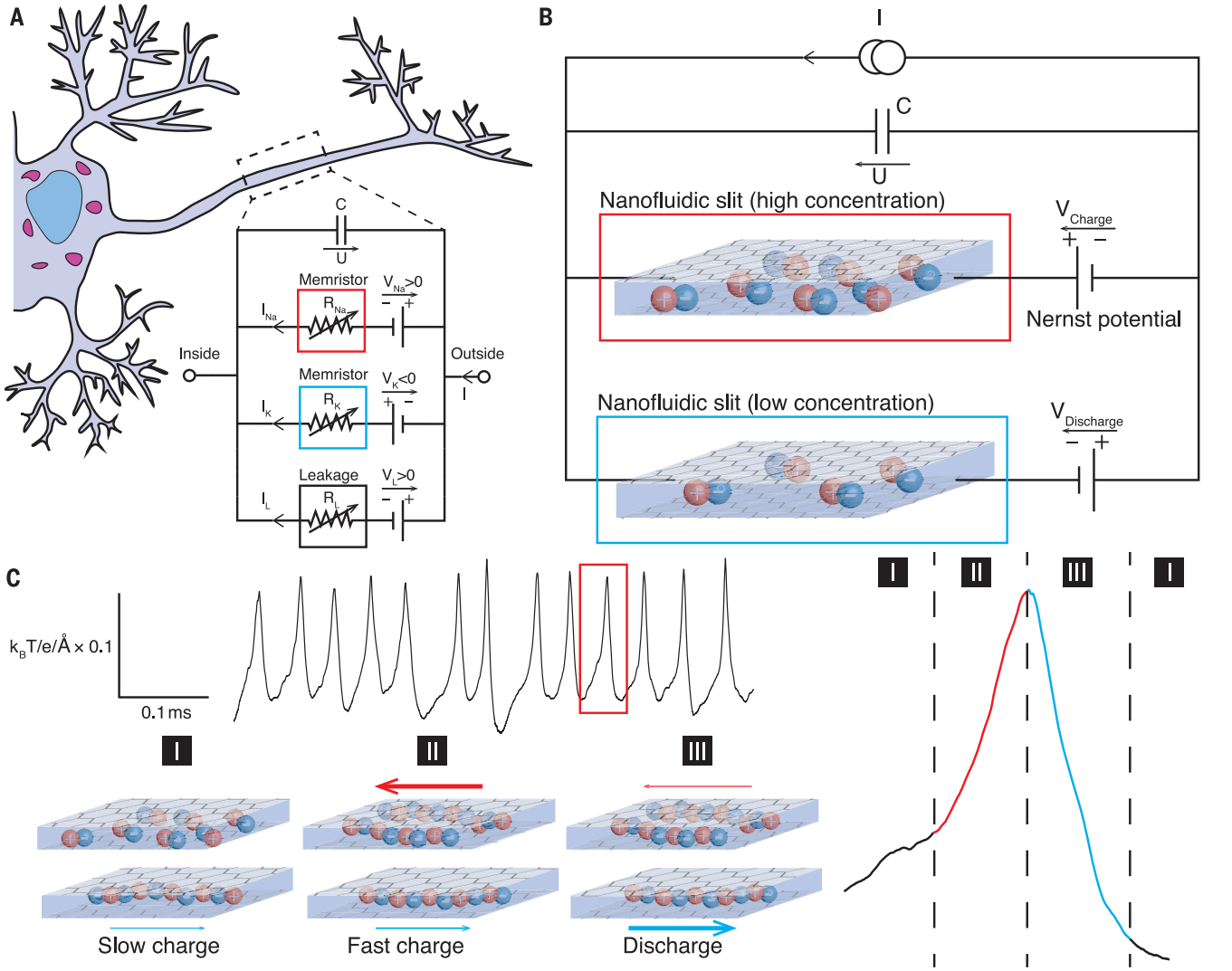


Fig. 4. Building an artificial Hodgkin-Huxley neuron from 2D ionic memristors. (**A**) Electronic representation of the Hodgkin-Huxley model, adapted from (*3*). I_{Na}, R_{Na}, and V_{Na}; I_K, R_K, and V_K; and I_L, R_L, and V_L are the current, resistance, and Nernst potential of sodium, potassium, and other species, respectively. (**B**) Prototype ionic machine implemented in BD simulations, exhibiting primitive neuronal behavior. Two slits with different ionic concentrations are simulated over long time scales (t ~ 1 ms). Each slit is connected to a pair of reservoirs, imposing a given Nernst potential on the slit, as in the original computation by Hodgkin and Huxley. (**C**) Spontaneous voltage spikes emitted by the prototype ionic machine and qualitative explanation for the observed spiking effect. The right panel shows different phases of a single spike, corresponding to the red rectangle on the left. Black line, quiescent neuron; red line, spike initiation; blue line, discharge.

low temperature (or sufficiently large field), $\lambda_D > l_E$ and the ionic atmosphere becomes anisotropic: It extends over l_E in the direction x and over λ_D in the perpendicular direction y. We show in the supplementary text (see section 3.6) that such an anisotropic screening cloud becomes unstable in the direction y below a critical temperature, explaining the formation of Bjerrum polyelectrolytes. In terms of the dynamics, this amounts to modifying the scaling exponent of the pair dissociation time according to (see supplementary text, section 3.6)

$$\tau_d = \frac{r_0^2}{2D}\left(\frac{l_E}{r_0}\right)^{1/2T^*} \quad (6)$$

As per Eq. 5, this results in a new exponent for the current-versus-voltage scaling:

$$I \underset{E\to 0}{\propto} E^{a_{PEW}(T^*)}, \quad a_{PEW}(T^*) = 1 + 1/4T^* \quad (7)$$

We term this new mechanism the polyelectrolytic Wien (PEW) effect, because conduction actually occurs inside the Bjerrum polyelectrolytes. In Fig. 2B, we show that the PEW prediction is in quantitative agreement with BD simulation results, in constrast to the IP model. We also compare it to results from all-atom simulations for $CaSO_4$ in Fig. 2D to demonstrate its robustness: The dynamics of the ionic assemblies is independent of the chemical nature of the salt and the confining medium or of simulation details (see fig. S2). Their formation is a direct consequence of 2D confinement.

The subtle nonlinear transport phenomena unveiled above set the stage for the memristor effect. The fraction of ions that is part of a Bjerrum polyelectrolyte structure is expected to behave as an internal variable, keeping memory of the voltage history. More formally, Eqs. 2 and 5 can be recast in a more transparent form, introducing the voltage U:

$$I = G(n_f) \times U \quad (8)$$

$$\frac{\partial}{\partial t} n_f = f(n_f, U) \quad (9)$$

The fraction of conducting ions, n_f, appears here as an internal state variable that depends

on the system's history. Furthermore, from the previous analysis, the relevant time scale governing the system's dynamics—τ_a—is much longer than usual molecular time scales, typically with $\tau_a \sim 1$ μs. This points to potential memory effects in the system. Altogether, Eqs. 8 and 9 formally define a voltage-controlled memristor, an electronic device whose resistance depends on its past (*11*, *12*).

A further proof and hallmark of the memristor behavior is shown in Fig. 3A, because the *I-V* characteristic under alternating voltage takes the form of a pinched loop for frequencies larger than a threshold f_p. The time scale $\tau_p = 1/f_p$ is found to decrease with ionic concentration ρ, and we interpret it as a formation time scale of Bjerrum polyelectrolyte. Qualitatively, because conduction occurs through the slow formation of macrostructures, the system cannot adapt to the instantaneous value of the electric field. Once the polyelectrolytes form, they contribute to conduction even if the field is turned off, before eventually dissolving. We compare the BD simulation results with our Wien effect theory to get insight into the underlying physics of the memristor behavior. We find that below a threshold frequency f_C (here $f_C = 160$ kHz), the simulated *I-V* curves are well reproduced by the PEW model, whereas for $f > f_C$, quantitative agreement is observed with the IP model (see Fig. 3B and fig. S4). This is consistent with the observation that at high frequencies, polyelectrolytes do not have time to form and conduction proceeds through the breaking of individual pairs. The threshold frequency $f_C = 160$ kHz is found to be independent of concentration and is therefore related to the dynamics of individual Bjerrum polyelectrolytes.

Although in our simulations the frequency of the alternating field is high owing to numerical constraints, our model predicts that it should also be relevant in a realistic experimental setting. We show in the supplementary text that a memristor effect could be observed at experimentally relevant frequencies $f \sim 100$ Hz and voltages $U \sim 0.1$ V (see fig. S5). The memory effect is driven by Bjerrum pairs and polyelectrolytes, which form in all salts regardless of their chemical nature. Therefore, the memristor effect, described by Eqs. 8 and 9, is a universal property of 2D electrolytes. We also find that optimal conditions correspond to a divalent salt with a slit height close to 1 nm, with other parameters such as the chemical structure of ions or of the confining material playing little role.

Memristors are the electric equivalents of voltage-gated ion channels. As such, they may serve as components of a primitive neuron, as first pointed out by Hodgkin and Huxley (*3*) (Fig. 4A). Hence, it is expected that assembling several nanofluidic memristors would allow neuromorphic behavior to be mimicked. To demonstrate this possibility, we perform parallel BD simulations of two monolayer electrolytes confined in two separate molecular channels (see Fig. 4B). For both electrolytes, the simulation yields the instantaneous relation between the applied voltage and resulting ionic current. A numeric circuitry is designed around the simulated molecular systems to reproduce the Hodgkin-Huxley model (see Fig. 4A). It includes a capacitor and opposite sign potentials applied on each channel that account for the Nernst potentials. The electronic circuit and the molecular channels are then simulated together, with the capacitor applying a voltage drop across the channels. Further details of the simulations are given in the supplementary text. We observe spontaneous voltage spikes at a frequency around 10 kHz (Fig. 4B). These spikes are the hallmark of neuromorphic behavior, obtained here from the sole properties of monolayer electrolytes.

The qualitative mechanism behind the generation of the spikes can be understood as follows: Because the two electrolytes are subject to (Nernst) potentials of opposite signs, they conduct current in opposite directions, but they are gated by the same voltage U. When U is small, only the discharging memristor is conducting, so that the capacitor receives a current $I_C < I$ and slowly charges. When U reaches a threshold value, the charging memristor starts conducting and the capacitor receives $I_C \gg I$, causing the voltage to spike up to $U \sim V_{charge}$. Then, the discharging memristor takes over: The capacitor receives a strongly negative current, and its voltage is lowered back to 0, at which point the process can start again (see Fig. 4C). This whole process is analogous to voltage spiking in biological neurons caused by the successive opening and closing of ion channels (sodium and potassium channels in the case of the Hodgkin-Huxley model of the giant squid axon). The observed working frequency of a few kilohertz is, considering the set of parameters used in simulations, just above the threshold frequency for the memristor effect, which is therefore crucial for observing the neuromorphic behavior.

The proper design of an experimentally accessible nanofluidic channel enables reproduction of the physical processes that occur in an elementary neuron, capable of emitting voltage spike trains. This result builds on the far-from-equilibrium transport properties of electrolytes in molecularly confined nanochannels, whose conductivity highlights a memristor effect. Our findings are supported by molecular simulations, combined with an extensive theoretical framework for the nonequilibrium transport, which generalizes Onsager's description of the Wien effect. These properties build on the $2D^+$ nature of such channels, where the magnitude of ionic correlations is intermediate between bulk and 1D systems. This theoretical prototyping is the first step toward an experimental demonstration of the ionic memristor, and exploration of memory phenomena in nanofluidic systems in general. The complex interplay between water, surfaces, and ions at the nanoscale gives rise to larger spatial structures—here, polyelectrolytes—which entail the emergence of slow dynamics and long memory times.

REFERENCES AND NOTES

1. B. Hille, *Biophys. J.* **22**, 283–294 (1978).
2. W. Gerstner, W. M. Kistler, *Spiking Neuron Models: Single Neurons, Populations, Plasticity* (Cambridge Univ. Press, 2002).
3. A. L. Hodgkin, A. F. Huxley, *J. Physiol.* **117**, 500–544 (1952).
4. L. Bocquet, *Nat. Mater.* **19**, 254–256 (2020).
5. S. Garaj *et al.*, *Nature* **467**, 190–193 (2010).
6. C. Y. Lee, W. Choi, J.-H. Han, M. S. Strano, *Science* **329**, 1320–1324 (2010).
7. K. Celebi *et al.*, *Science* **344**, 289–292 (2014).
8. J. Feng *et al.*, *Nature* **536**, 197–200 (2016).
9. E. Secchi *et al.*, *Nature* **537**, 210–213 (2016).
10. R. H. Tunuguntla *et al.*, *Science* **357**, 792–796 (2017).
11. L. Chua, *IEEE Trans. Circuit Theory* **18**, 507–519 (1971).
12. D. B. Strukov, G. S. Snider, D. R. Stewart, R. S. Williams, *Nature* **453**, 80–83 (2008).
13. B. Radha *et al.*, *Nature* **538**, 222–225 (2016).
14. A. Esfandiar *et al.*, *Science* **358**, 511–513 (2017).
15. T. Mouterde *et al.*, *Nature* **567**, 87–90 (2019).
16. S. Faucher *et al.*, *J. Phys. Chem. C* **123**, 21309–21326 (2019).
17. W. Sparreboom, A. van den Berg, J. C. T. Eijkel, *New J. Phys.* **12**, 015004 (2010).
18. N. Kavokine, R. R. Netz, L. Bocquet, *Annu. Rev. Fluid Mech.* **53**, 377–410 (2021).
19. A. Marcotte, T. Mouterde, A. Niguès, A. Siria, L. Bocquet, *Nat. Mater.* **19**, 1057–1061 (2020).
20. N. Kavokine, S. Marbach, A. Siria, L. Bocquet, *Nat. Nanotechnol.* **14**, 573–578 (2019).
21. N. Bjerrum, *Untersuchungen über Ionenassoziation* (Dankse Vid. Selskab, Mathematisk-fysiske meddelelser, København: A. F. Høst, 1926), pp. 1–48.
22. A. Schlaich, E. W. Knapp, R. R. Netz, *Phys. Rev. Lett.* **117**, 048001 (2016).
23. L. Fumagalli *et al.*, *Science* **360**, 1339–1342 (2018).
24. Y. Levin, *Rep. Prog. Phys.* **65**, 1577–1632 (2002).
25. P. Minnhagen, *Rev. Mod. Phys.* **59**, 1001–1066 (1987).
26. J. M. Kosterlitz, D. J. Thouless, *J. Phys. C Solid State Phys.* **6**, 1181–1203 (1973).
27. R. M. Fuoss, *J. Am. Chem. Soc.* **81**, 2659–2662 (1959).
28. L. Onsager, *J. Chem. Phys.* **2**, 599–615 (1934).
29. V. Kaiser, thesis, Ecole Normale Supérieure de Lyon (2014).
30. P. Robin, N. Kavokine, L. Bocquet, Molecular dynamics code for: Modeling of emergent memory and voltage spiking in ionic transport through angstrom-scale slits. Zenodo (2021); https://doi.org/10.5281/zenodo.4841109.

ACKNOWLEDGMENTS

We thank H. Yoshida for help and discussions on molecular dynamics. **Funding:** L.B. acknowledges funding from the EU H2020 Framework Programme/ERC Advanced Grant agreement number 785911-Shadoks and ANR project Neptune. This work has received the support of "Institut Pierre-Gilles de Gennes," program ANR-10-IDEX-0001-02 PSL and ANR-10-LABX-31. This work was granted access to the HPC resources of CINES under the allocation A0090710395 made by GENCI. **Author contributions:** L.B. conceived the project. P.R. carried out the theoretical analysis and the molecular dynamics simulations, with inputs from all authors. All authors discussed the results, co-wrote the article, and reviewed the final manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are in the main text or supplementary materials. The molecular dynamics codes that support the plots within this paper and other findings of this study are archived on Zenodo (*30*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6555/687/suppl/DC1
Supplementary Text
Figs. S1 to S5
Table S1
References (*31*–*47*)

19 November 2020; accepted 11 June 2021
10.1126/science.abf7923

COOLING TEXTILES

Hierarchical-morphology metafabric for scalable passive daytime radiative cooling

Shaoning Zeng[1]†, Sijie Pian[2]†, Minyu Su[1], Zhuning Wang[2], Maoqi Wu[1,3], Xinhang Liu[2], Mingyue Chen[3], Yuanzhuo Xiang[1], Jiawei Wu[1], Manni Zhang[1], Qingqing Cen[2], Yuwei Tang[2], Xianheng Zhou[2], Zhiheng Huang[1], Rui Wang[1], Alitenai Tunuhe[1], Xiyu Sun[3], Zhigang Xia[4], Mingwei Tian[5], Min Chen[6], Xiao Ma[7], Lvyun Yang[1], Jun Zhou[1], Huamin Zhou[8], Qing Yang[2], Xin Li[7], Yaoguang Ma[2]*, Guangming Tao[1]*

Incorporating passive radiative cooling structures into personal thermal management technologies could effectively defend humans against intensifying global climate change. We show that large-scale woven metafabrics can provide high emissivity (94.5%) in the atmospheric window and high reflectivity (92.4%) in the solar spectrum because of the hierarchical-morphology design of the randomly dispersed scatterers throughout the metafabric. Through scalable industrial textile manufacturing routes, our metafabrics exhibit desirable mechanical strength, waterproofness, and breathability for commercial clothing while maintaining efficient radiative cooling ability. Practical application tests demonstrated that a human body covered by our metafabric could be cooled ~4.8°C lower than one covered by commercial cotton fabric. The cost-effectiveness and high performance of our metafabrics present substantial advantages for intelligent garments, smart textiles, and passive radiative cooling applications.

As global warming intensifies, people exposed to outdoor environments are at increased risk of illnesses related to excessive heat stress (*1*, *2*). Personal thermal management (PTM) technology, which can control the microclimate of the human body, can facilitate thermal comfort with high efficiency (*3*). However, challenges related to energy consumption and bulky nature of current PTM designs have not yet been resolved (*4*–*7*). Radiative cooling, which radiates heat directly to outer space through the atmospheric transparent spectral window (ATSW; λ ~8 to 13 μm) (*8*, *9*), is a promising cooling method for overcoming these obstacles. Methods that harness nanophotonic structures such as multilayer photonic structures (*9*, *10*), metamaterials (*11*, *12*), and random media (*13*–*20*) have successfully yielded daytime radiative cooling devices and systems by simultaneously introducing solar irradiance rejection and enhancing the emitted thermal radiation. However, most radiative cooling structures (thin films, coatings, and paints, among others) still result in weak air-water permeability and inadequate wearability properties that limit these materials from being directly applied to PTM systems (*21*). Thus, as the dominant medium that protects the skin from the external environment, clothing may be an excellent candidate for implementation of daytime radiative cooling functionality.

Researchers have tested various types of mid-infrared (MIR)–transparent radiative cooling textiles (*22*–*26*). But, to maintain enough MIR transmittance, the thickness of the MIR-transparent structure is highly restricted to <~150 μm. Solar heating blockage and structural toughness become challenging at this thickness level. By contrast, the performance of MIR-emissive textiles does not depend on the spectrum of the underlying surface and thus relaxes the thickness restrictions. However, only a few studies on MIR-emissive radiative cooling textiles or fibers have been proposed for PTM (*27*, *28*).

We designed a multilayer metafabric knitted with composite microfibers that incorporates hierarchically designed random metamaterial structures to directly integrate radiative cooling technology for PTM applications. The hierarchical-morphology design mechanism directly provides an extended spectroscopic response that spans two orders of magnitude in wavelength (0.3 to 25 μm) and enables the metafabric to resonantly reject solar power and strongly emit in the MIR range (Fig. 1, A and B). The multilayer metafabric consists of a titanium oxide–polylactic acid (TiO_2-PLA) composite woven textile laminated with a thin polytetrafluoroethylene (PTFE) layer. The laminated top layer is designed as a 50-μm-thick PTFE clothing film that is commercially available and widely used in clothing industries. The porous composite contains nanobeads of 200 to 1000 nm in diameter and nanofibers with a length of several micrometers and a width <200 nm (fig. S1). The composite strongly reflects ultraviolet (UV) light from the incoming radiation before it reaches the metatextile layer. As scatterers with high refractive indices, TiO_2 nanoparticles with size distributions between 200 and 1600 nm can use the collective effect of multiple Mie resonances (*29*, *30*) and produce the scattering peaks required to cover the entire visible–to–near-infrared (VIS-NIR) band with high efficiency.

Selective emitters are desirable for radiative cooling applications when the working temperature is below the ambient temperature. For scenarios above ambient temperature, effective emission in the full MIR band can be useful. Therefore, PLA microfibers, which have C=O (1825 to 1725 cm^{-1}), $-CH_3$ (1454.49 cm^{-1}), –CH (1382 to 1300 cm^{-1}), C–O (1042 to 1267 cm^{-1}), and C–C (867.67 cm^{-1}) bonds (*31*), were selected to provide rich emittance in the MIR wave band (fig. S2). Also, PLA fibers exhibit excellent solar transmittance, high moisture absorption properties, and substantial biodegradability (*32*). The diameter of the fibers is designed to be ~30 μm for enhanced MIR absorption (emissivity) (fig. S3), enabling them to simultaneously exhibit multiple scattering for increased absorption (*33*) and a rugged textile surface for gradient refractive index antireflection (*14*, *34*).

We further optimized the radiative cooling performance using a numerical model based on Lorenz-Mie theory and Monte Carlo simulations (*29*, *35*, *36*). The optical properties described by the path length–dependent electromagnetic scattering process between incoming radiation and metafabric require that more structure parameters be rigorously determined to improve the radiative cooling capability of the metafabric. We selected nanoparticles with a log-normal distribution (μ = 6.2, σ^2 = 0.2, where μ and σ^2 are the mean and variance of the natural logarithm of the log-normal distribution) and peak diameters of 400 nm (fig. S4) because they induced the best solar rejection performance enhancement of the fabric (Fig. 1C). Such a wide distribution of nanoparticles, when combined with PTFE nanobeads, provides broad-spectrum scattering and reflectivity across the UV-VIS-NIR band (Fig. 1D). We numerically evaluated the radiative cooling performance of the metafabric by using a steady-state heat-transfer

[1]Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China. [2]State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, International Research Center for Advanced Photonics, Zhejiang University, Hangzhou 310058, China. [3]iFabric Lab, TD Energy, Inc., Weifang 261500, China. [4]State Key Laboratory of New Textile Materials and Advanced Processing Technologies, Wuhan Textile University, Wuhan 430200, China. [5]Research Center for Intelligent and Wearable Technology, College of Textiles and Clothing, State Key Laboratory of Bio-Fibers and Eco-Textiles, Intelligent Wearable Engineering Research Center of Qingdao, Qingdao University, Qingdao 266071, China. [6]Wuhan National Laboratory for Optoelectronics and School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China. [7]State Key Laboratory of Biobased Fiber Manufacturing Technology, China Textile Academy, Beijing 110105, China. [8]State Key Laboratory of Material Processing and Die and Mould Technology, School of Materials Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China.
*Corresponding author. Email: tao@hust.edu.cn (G.T.); mayaoguang@zju.edu.cn (Y.M.)
†These authors contributed equally to this work.

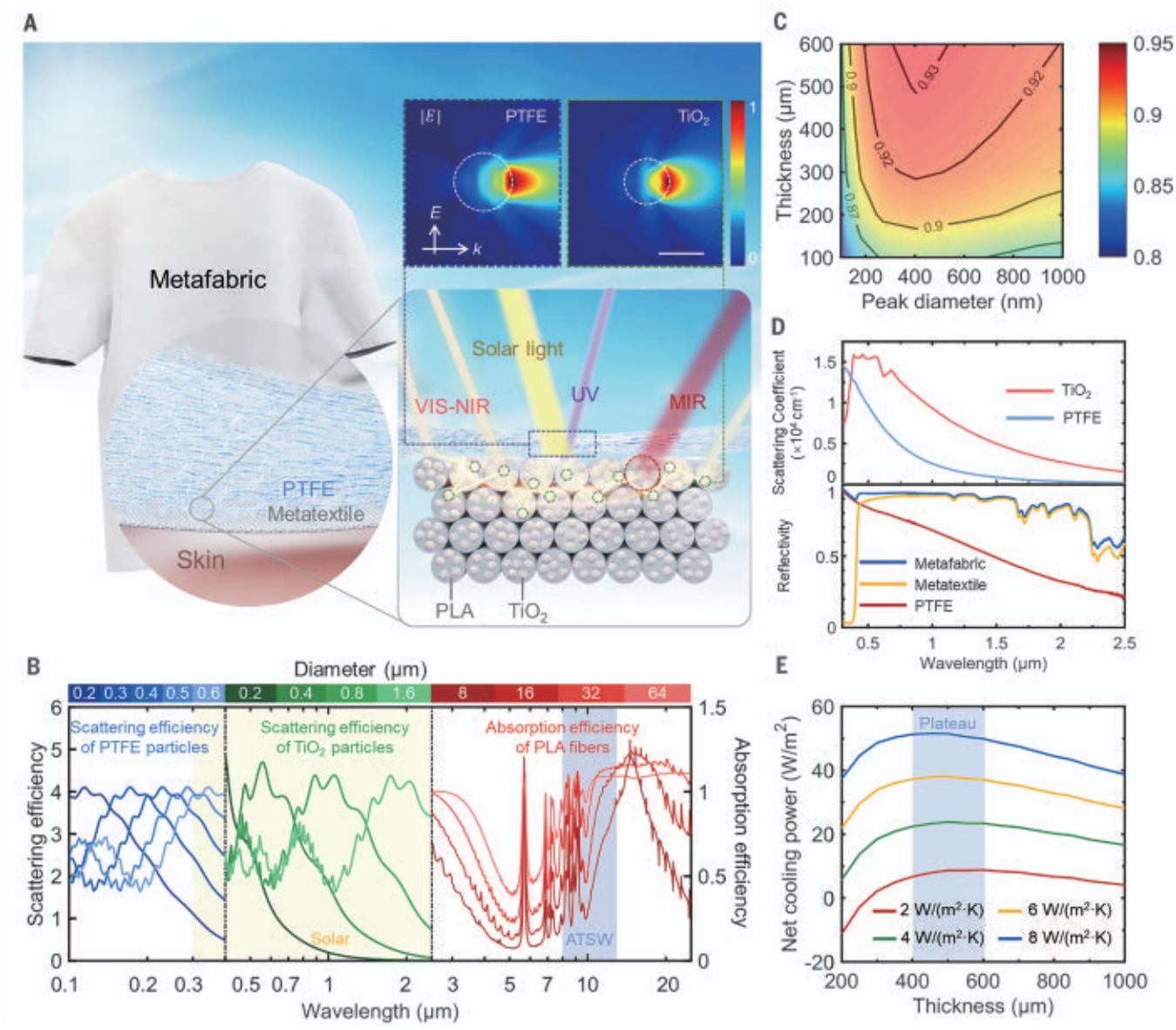


Fig. 1. Proposed structure and simulated properties of the metafabric. (**A**) Schematic of a metafabric for daytime radiative cooling. The blue, green, and red dashed boxes highlight the three-level hierarchical structure responding to the UV, VIS-NIR, and MIR bands, respectively. The insets show the calculated scattering fields of 300- and 550-nm light by a 500-nm PTFE particle and a 400-nm TiO_2 particle, respectively. Scale bar, 400 nm. *E*, electric field of the incident light; *k*, wave vector of the incident light. (**B**) Calculated scattering and absorption efficiencies for particles with different sizes encapsulated in the metafabric. PTFE particles, TiO_2 particles, and PLA fibers demonstrate strong scattering and absorption of UV light, VIS-NIR light, and MIR light, respectively. (**C**) Calculated solar reflectance (0.3 to 2.5 μm) of the metafabric with different thicknesses and TiO_2 particle sizes under the same doping concentration (15% volume fraction). For a consistent thickness, the log-normal distribution of TiO_2 particles with a peak diameter of 400 nm (μ = 6.2, σ^2 = 0.2) can provide the highest solar reflectance. (**D**) (Top) Scattering coefficients of PTFE and TiO_2 nanoparticles with optimized size distributions and volume fractions of 15%. (Bottom) Measured solar radiation band reflectivity curves of metafabric, metatextile, and PTFE clothing films. (**E**) Calculated net cooling power of the metafabric versus fabric thickness by using different heat convection coefficients with a skin temperature of 34°C and an ambient temperature of 22°C.

model of clothed human skin with a skin temperature of 34°C and an ambient temperature of 22°C (*36*). We observed a net cooling power plateau corresponding to the thickness range of 400 to 600 μm, with the most efficient cooling behavior occurring at ~500 μm (Fig. 1E).

Modern fabrication technologies have enabled substantial flexibility for the scalable manufacture of our metafabrics by realizing the hierarchical-morphology design at the subfiber level (Fig. 2A) (*36*). Our metafibers (Fig. 2B) exhibited superior tensile properties—with an elongation of 29.5% and a breakage strength of 1.886 cN per decitex (dtex) (Fig. 2C), which are flexible and strong enough to be stitched and embroidered via a commercial sewing machine (Fig. 2D and movie S1). This compatibility with commercial sewing techniques allows the cooling metafibers to be applied to any clothing type and textile pattern. Mechanical tests showed that our metafabric (Fig. 2E and movie S2) can endure a tensile force of ~482 N and an elongation of ~20%, comparable to the mechanical properties of other commercial fabrics (Fig. 2F). The hydrophobic top surface of the metafabric exhibits excellent waterproofness (Fig. 2G), whereas the metafabric still maintains a certain degree of air permeability. When sandwiched between air and water environments,

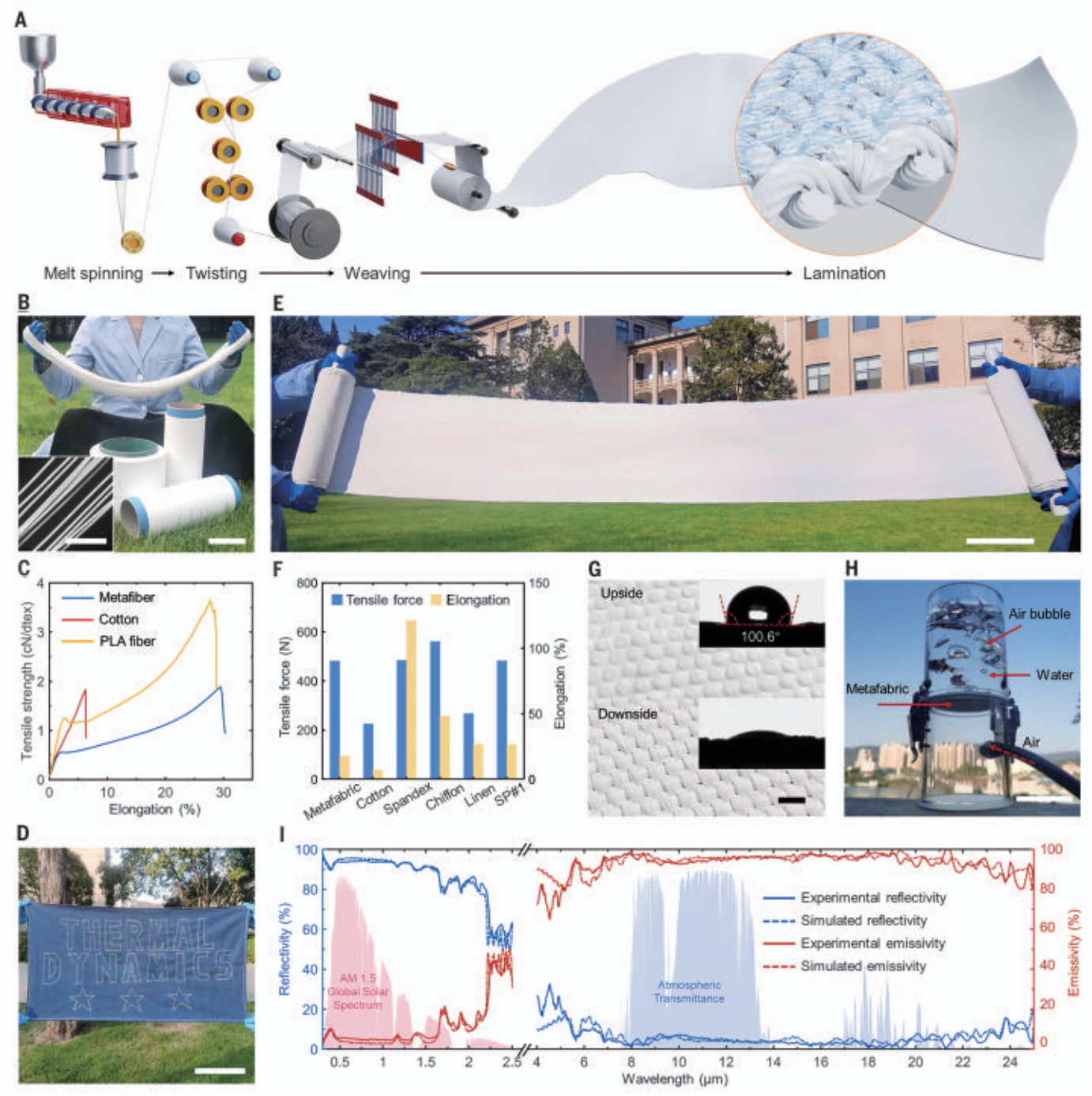


Fig. 2. Manufacture and characterization of the metafabric. (**A**) Fabrication process of the metafabric. (**B**) Photograph of the fabricated metafibers. Scale bar, 10 cm. The inset shows an optical micrograph of metafibers with diameters of ~30 μm. Scale bar, 500 μm. (**C**) Mechanical strength tests of the fiber strength versus elongation. (**D**) Photograph of metafibers embroidered with the letters "THERMAL DYNAMICS" on denim. Scale bar, 20 cm. (**E**) Photograph of the metafabric (0.3 m by 15 m). Scale bar, 20 cm. (**F**) Tensile strength test of the metafabric, cotton, spandex, chiffon, linen, and sun-protective clothing #1 (SP#1). (**G**) Morphologies of the upper and lower surfaces of the metafabric. The upper surface is hydrophobic, and the lower surface is hydrophilic, as shown in the inset images of contact angle tests. Scale bar, 1 mm. (**H**) Demonstration of the waterproofness and breathability of the metafabric. Scale bar, 5 cm. (**I**) Measured reflectivity and emissivity spectra of the metafabric (0.3 to 25 μm). AM, air mass.

our metafabric exhibits continuous bubble transmittance and tight waterproofness that facilitate conventional heat-exchange mechanisms for favorable wearability (Fig. 2H). The durability and washing resistance of our metafabric were also validated and suggest that our technology is readily compatible with commercial applications (fig. S5) (*36*). By benefitting from our hierarchical-morphology design, the fabric shows a broadband reflectivity of 92.4% in the solar radiation region (0.3 to 2.5 μm) and an average emissivity of 94.5% across the ATSW (Fig. 2I). The broad band of high emissivity between 4 and 25 μm can provide additional cooling power for the

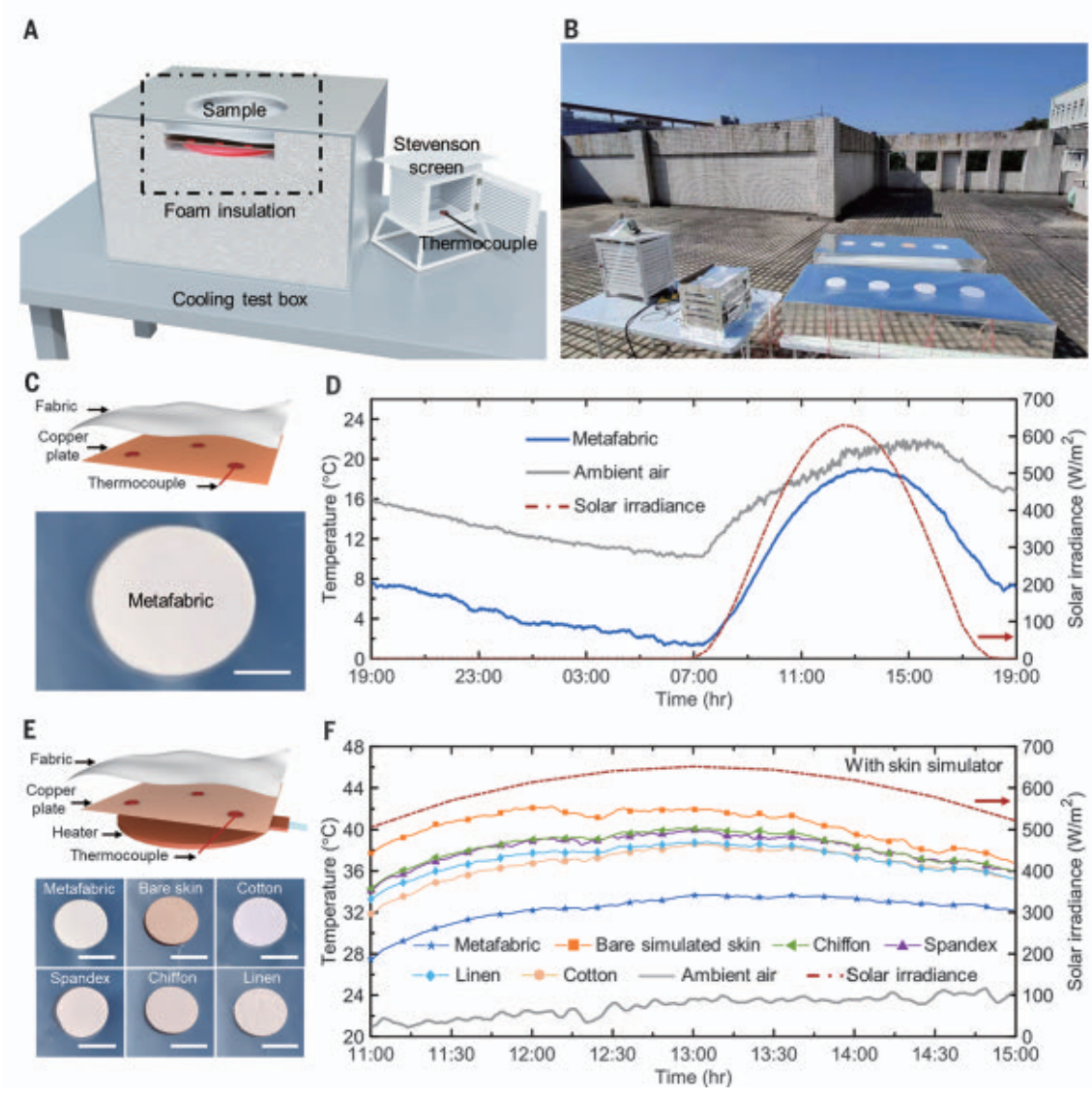


Fig. 3. Direct thermal analysis to determine the metafabric cooling performance. (**A** and **B**) Schematic (A) and photo (B) of the thermal measurement system used to characterize the radiative cooling performance. (**C**) Schematic and photo of the sample structure for the test. Scale bar, 5 cm. (**D**) Continuous temperature measurement of the subambient cooling performance test in Guangzhou, China (23°5′32″N, 113°23′45″E, 5 and 6 December 2020), over 24 hours. (**E**) Schematic of the device used for the skin simulator cooling test and photographs of the fabric samples used in the test. Scale bars, 10 cm. (**F**) Temperature difference of skin simulators under different fabric samples in the same location (28 November 2020).

working temperatures of our metafabric, which are higher than the ambient temperature when used in PTM scenarios (*8*).

We demonstrated the outdoor radiative cooling performance of our metafabric by direct thermal measurements under clear sky conditions in Guangzhou, China (Fig. 3, A and B) (*36*). The temperature of each fabric sample was monitored by three K-type thermocouples taped on the copper plate to ensure the accuracy and uniformity of thermal measurements (Fig. 3C). Throughout a continuous 24-hour measurement period, the metafabric consistently maintained a temperature below the ambient temperature (Fig. 3D). The minimum temperature difference was ~2°C, which occurred within half an hour of the peak solar irradiance. Furthermore, we tested the radiative cooling performance of our metafabric on a human skin simulator (Fig. 3E) (*36*). Under peak solar irradiance between 11:00 and 15:00 hours, the temperature of the metafabric was approximately 5.0°, 6.8°, 7.0°, 5.8°, and 10.2°C lower than that of the cotton, spandex, chiffon, linen, and bare skin simulators, respectively (Fig. 3F).

To validate the cooling performance of our metafabric with scalable, wearable features in pragmatic scenarios, we composed a homemade vest by sewing a commercial cotton fabric and a metafabric together. A volunteer wearing that vest reclined under direct sunlight for an hour while we monitored the thermal properties of the vest and the volunteer (Fig. 4A). A thermal camera (Fluke Ti400) showed a large temperature difference between the two sides of the vest (34.4° and 31.0°C). Thermocouples adhered under the vest also indicated a temperature difference of ~4.8°C between the parts of the body that were covered by different fabrics (Fig. 4B). A similar test for the radiative cooling vest was performed in Sipsongpanna, China, for repeatability. During the half-hour measurement process, the surface of each half of the vest gradually exhibited a notable temperature difference (Fig. 4C and movie S3). The two halves of the body showed a distinct temperature difference of >3°C directly after the vest was removed. In addition, we tested the cooling performance of the metafabric with three car models covered by a commercial vehicle cover (*36*), our metafabric, and nothing, respectively (Fig. 4A). During a 90-min experiment, the maximum interior temperature of the car model covered by the metafabric was ~30° and ~27°C cooler than that of the model with no cover and with a vehicle cover, respectively (Fig. 4D). These results demonstrate the potential for commercial applications in various complex scenarios, such as smart textiles, sunshade products, and logistics transportation, among others (fig. S6).

The metafabric exhibits efficient radiative cooling performance and provides necessary breathability and wearing comfort for PTM. When compared with passive cooling modalities such as films or paints, the woven structure enables the metafabric to easily accommodate

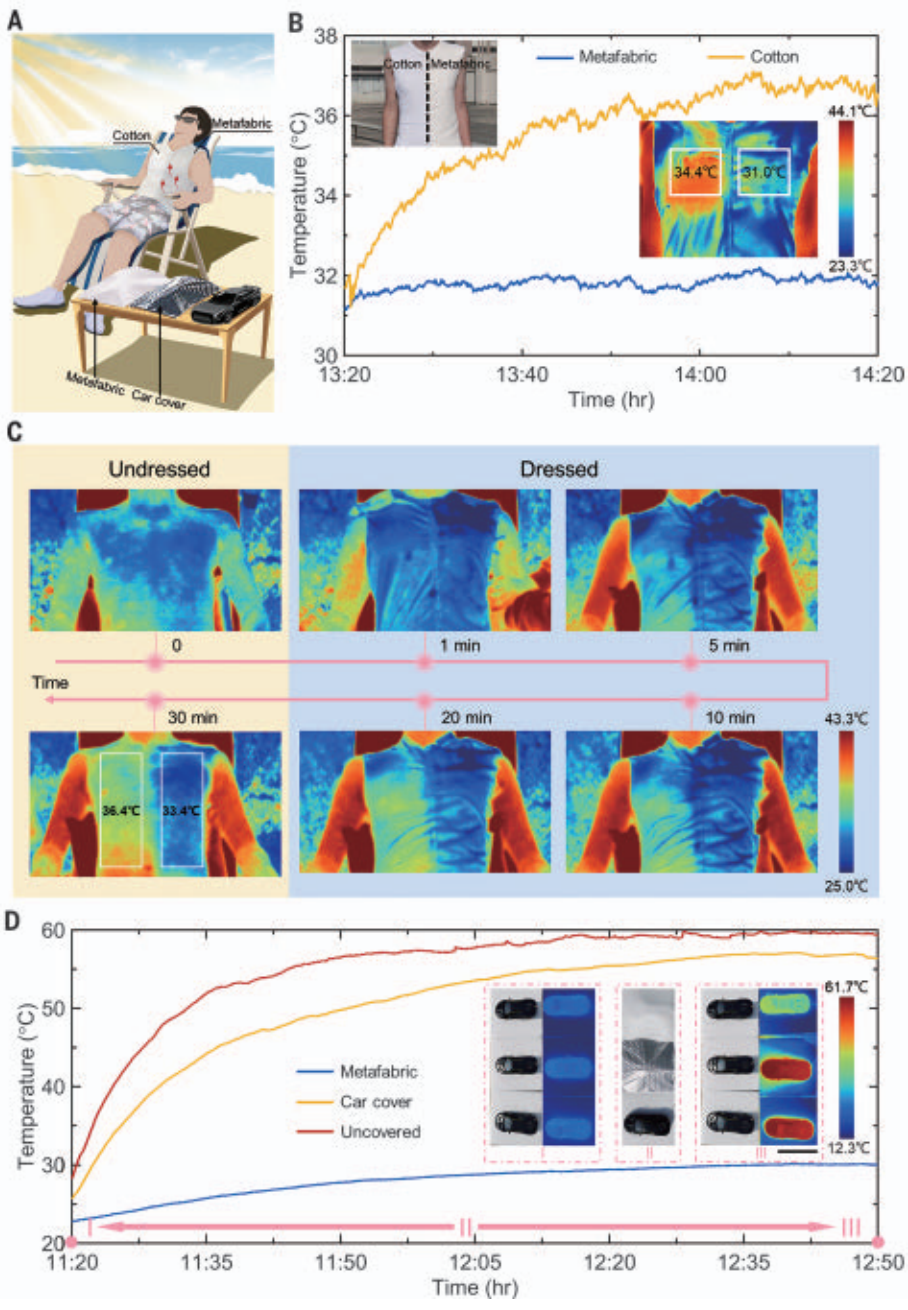


Fig. 4. Practical characterization of the metafabric with sun exposure. (**A**) Schematic of the metafabric cooling tests on the human body and model cars. (**B**) Temperature tracking for skin under different fabrics in direct sunlight in Guangzhou, China (23°5′32″N, 113°23′45″E, 7 December 2020). The insets show photographs and thermal images of the volunteer wearing a homemade vest. (**C**) Infrared images of the volunteer under direct sunlight in Sipsongpanna, China (22°10′22″N, 100°51′29″E, 13 December 2020). (**D**) Temperature test curves of three identical car models exposed to the sun. The inset shows photographs and IR images of the car models during the three time periods. "I" and "III" labels correspond to the initial and final measurements, respectively, and "II" shows a photograph of the entire test process. Scale bar, 10 cm.

complex deformations (bending, stretching, and twisting), which leads to versatile compatibility. Through embroidery, cutting, and sewing, the metafabric can be integrated into various products for different scenarios, such as clothing, tents, car covers, curtains, and awnings. Further optimization and exploration of the fiber's structural features are expected to improve the cooling efficiency by the combined effect of both radiative cooling and perspiration evaporation (*37*). Although we have not yet explored the color-compatible radiative cooling and cooling power modulation of the metafabric, on the basis of the idea of hierarchical-morphology design, applying a fourth-level dyeing layer and adopting asymmetrical photonics structures may address the above challenges.

REFERENCES AND NOTES

1. R. A. I. Lucas, Y. Epstein, T. Kjellstrom, *Extrem. Physiol. Med.* **3**, 14 (2014).
2. L. T. Biardeau, L. W. Davis, P. Gertler, C. Wolfram, *Nat. Sustain.* **3**, 25–28 (2020).
3. Y. Peng, Y. Cui, *Joule* **4**, 724–742 (2020).
4. M. Zhao *et al.*, *Int. J. Ind. Ergon.* **43**, 232–237 (2013).
5. T. Gao *et al.*, *ACS Nano* **11**, 11513–11520 (2017).
6. T. Zhang *et al.*, *Nano Energy* **41**, 35–42 (2017).
7. D. Zhao *et al.*, *Appl. Energy* **218**, 282–291 (2018).
8. D. Zhao *et al.*, *Appl. Phys. Rev.* **6**, 021306 (2019).
9. A. P. Raman, M. A. Anoma, L. Zhu, E. Rephaeli, S. Fan, *Nature* **515**, 540–544 (2014).
10. Z. Chen, L. Zhu, A. Raman, S. Fan, *Nat. Commun.* **7**, 13729 (2016).
11. M. M. Hossain, B. Jia, M. Gu, *Adv. Opt. Mater.* **3**, 1047–1051 (2015).
12. Y. Zhai *et al.*, *Science* **355**, 1062–1066 (2017).
13. Z. Huang, X. Ruan, *Int. J. Heat Mass Transf.* **104**, 890–896 (2017).
14. J. Mandal *et al.*, *Science* **362**, 315–319 (2018).
15. T. Li *et al.*, *Science* **364**, 760–763 (2019).
16. D. Li *et al.*, *Nat. Nanotechnol.* **16**, 153–158 (2021).
17. X. Li *et al.*, *Cell Rep. Phys. Sci.* **1**, 100221 (2020).
18. J. Mandal, Y. Yang, N. Yu, A. P. Raman, *Joule* **4**, 1350–1356 (2020).
19. X. Xue *et al.*, *Adv. Mater.* **32**, e1906751 (2020).
20. T. Wang *et al.*, *Nat. Commun.* **12**, 365 (2021).
21. W. Yan *et al.*, *Mater. Today* **35**, 168–194 (2020).
22. J. K. Tong *et al.*, *ACS Photonics* **2**, 769–778 (2015).
23. P.-C. Hsu *et al.*, *Science* **353**, 1019–1023 (2016).
24. L. Cai *et al.*, *Adv. Mater.* **30**, e1802152 (2018).
25. Y. Peng *et al.*, *Nat. Sustain.* **1**, 105–112 (2018).
26. L. Cai *et al.*, *Joule* **3**, 1478–1486 (2019).
27. X. A. Zhang *et al.*, *Science* **363**, 619–623 (2019).
28. N. N. Shi *et al.*, *Light Sci. Appl.* **7**, 37 (2018).
29. H. C. van de Hulst, *Light Scattering by Small Particles* (Courier, 1981).
30. P. Chýlek, J. Zhan, *J. Opt. Soc. Am. A* **6**, 1846–1851 (1989).
31. Ł. Łopusiewicz, F. Jędra, M. Mizielińska, *Polymers* **10**, 386 (2018).
32. N. F. Zaaba, M. Jaafar, *Polym. Eng. Sci.* **60**, 2061–2075 (2020).
33. R. Mupparapu, K. Vynck, T. Svensson, M. Burresi, D. S. Wiersma, *Opt. Express* **23**, A1472–A1484 (2015).
34. H. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 14657–14666 (2020).
35. L. Wang, S. L. Jacques, L. Zheng, *Comput. Methods Programs Biomed.* **47**, 131–146 (1995).
36. Materials and methods are available as supplementary materials.
37. R. K. Varshney, V. K. Kothari, S. Dhamija, *J. Textil. Inst.* **101**, 495–505 (2010).

ACKNOWLEDGMENTS

The authors acknowledge X. Zhang and J. Liu for encouragement, support, and vision. We thank X. Liu for helpful discussion; P. Wu for assistance with the melt-spinning experiment; C. Wang and K. Wang for assistance with textile weaving; and Y. Ma, S. Zhao, and W. Chen for performing the experimental measurements. **Funding:** This work was supported by the National Natural Science Foundation of China (grant 61875064 to G.T. and grant 61905213 to Y.M.). **Author contributions:** G.T. and Y.M. conceived of the idea, designed the experiments, and supervised the research project. S.Z. and S.P. conducted the experiments. S.Z. performed the manufacturing and characterization of the fibers and fabrics. S.P. performed the modeling and simulation. G.T., Y.M., S.Z., and S.P. analyzed the data and wrote the paper. All authors contributed to the experiments and discussion of the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Information requests should be directed to the corresponding authors.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6555/692/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 and S2
References (*38*–*45*)
Movies S1 to S6

16 March 2021; accepted 28 June 2021
Published online 8 July 2021
10.1126/science.abi5484

BIOMECHANICS

Acrobatic squirrels learn to leap and land on tree branches without falling

Nathaniel H. Hunt[1,2]*, Judy Jinn[3], Lucia F. Jacobs[3], Robert J. Full[2]

Arboreal animals often leap through complex canopies to travel and avoid predators. Their success at making split-second, potentially life-threatening decisions of biomechanical capability depends on their skillful use of acrobatic maneuvers and learning from past efforts. Here, we found that free-ranging fox squirrels (*Sciurus niger*) leaping across unfamiliar, simulated branches decided where to launch by balancing a trade-off between gap distance and branch-bending compliance. Squirrels quickly learned to modify impulse generation upon repeated leaps from unfamiliar, compliant beams. A repertoire of agile landing maneuvers enabled targeted leaping without falling. Unanticipated adaptive landing and leaping "parkour" behavior revealed an innovative solution for particularly challenging leaps. Squirrels deciding and learning how to launch and land demonstrates the synergistic roles of biomechanics and cognition in robust gap-crossing strategies.

Leaping across gaps in a disconnected canopy presents challenges (*1*–*4*) to arboreal animals (*5*), shaping interactions between learning and biomechanical capabilities (*6*, *7*). Animals traversing a compliant branch must decide on a launch point. As the distance to a landing target decreases, lessening required momentum, the compliance of the branch underfoot increases, which magnifies the momentum-generating impulse lost to branch bending during takeoff (*4*, *8*). The decision will depend on the interplay of an animal's biomechanical capabilities and perception of the physical environment [i.e., branch properties (*9*)]. Although the former may be innate, the latter is only partially observable (*10*) and is likely estimated through trial-and-error learning. We quantified how decision-making and learning capabilities complement biomechanical adaptations for targeted leaping in the canopy. Here, we report three experiments where free-ranging fox squirrels (*Sciurus niger*) leapt across gaps between simulated branches.

In the first experiment, branches were cantilevered, varying in compliance across their lengths (Fig. 1A). The landing perch was a narrow rod oriented perpendicular to the sagittal plane, which constrained successful leap trajectories to a small region.

[1]Department of Biomechanics, University of Nebraska, Omaha, Omaha, NE, USA. [2]Department of Integrative Biology, University of California at Berkeley, Berkeley, CA, USA. [3]Department of Psychology, University of California at Berkeley, Berkeley, CA, USA.
*Corresponding author. Email: nhunt@unomaha.edu

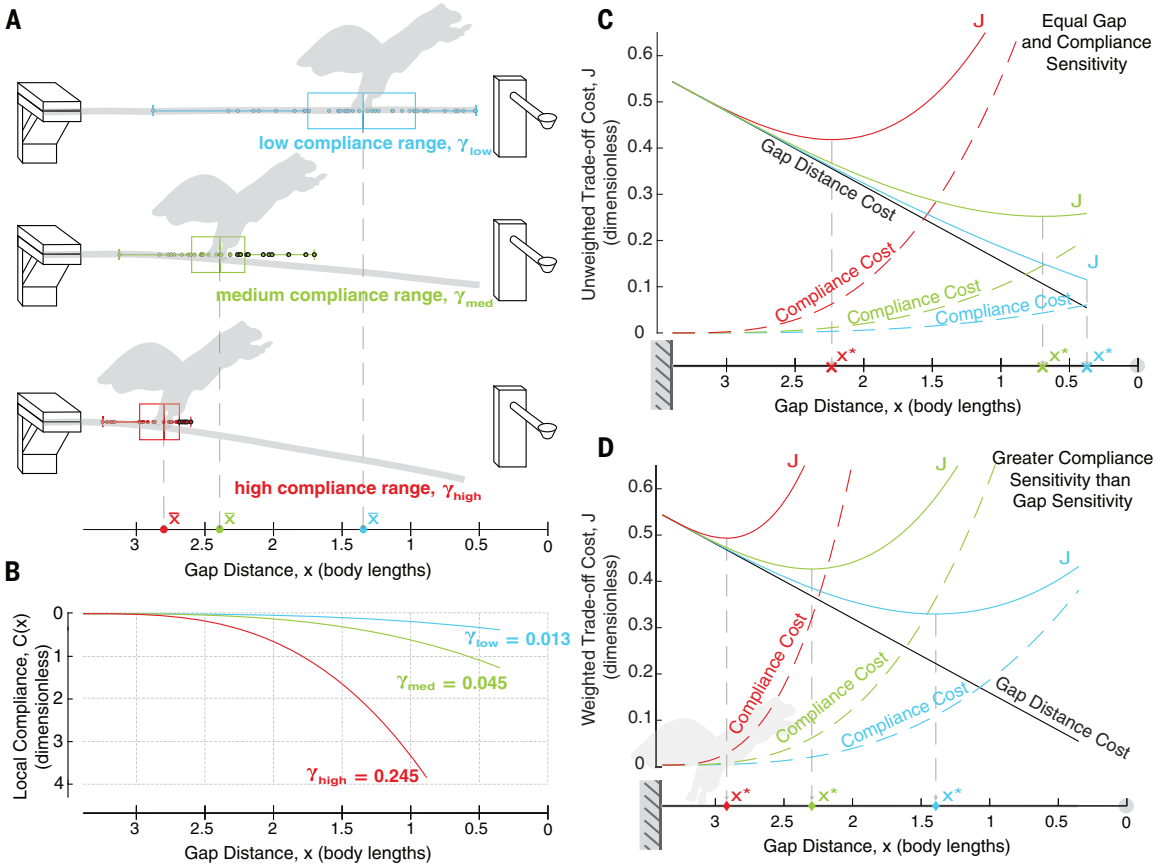


Fig. 1. Launch-point decisions from simulated branches with varying compliance. (**A**) Gap distance (x) launch-point decisions differed on simulated branches designed to represent three ranges of compliance: low (blue), medium (green), and high (red). (**B**) Local compliance increased cubically as gap distance decreased when traversing simulated branches from left to right. (**C** and **D**) Dimensionless costs modeled versus gap distance. Gap distance cost (solid black line) is identical for all branches. Compliance costs (colored dashed lines) increase as gap distance decreases. Solid colored lines represent the total trade-off cost, J, by summing gap cost and compliance cost. A trade-off model assuming equal weights for gap cost and compliance cost (C) and a trade-off model assuming independent weights (D) are shown.

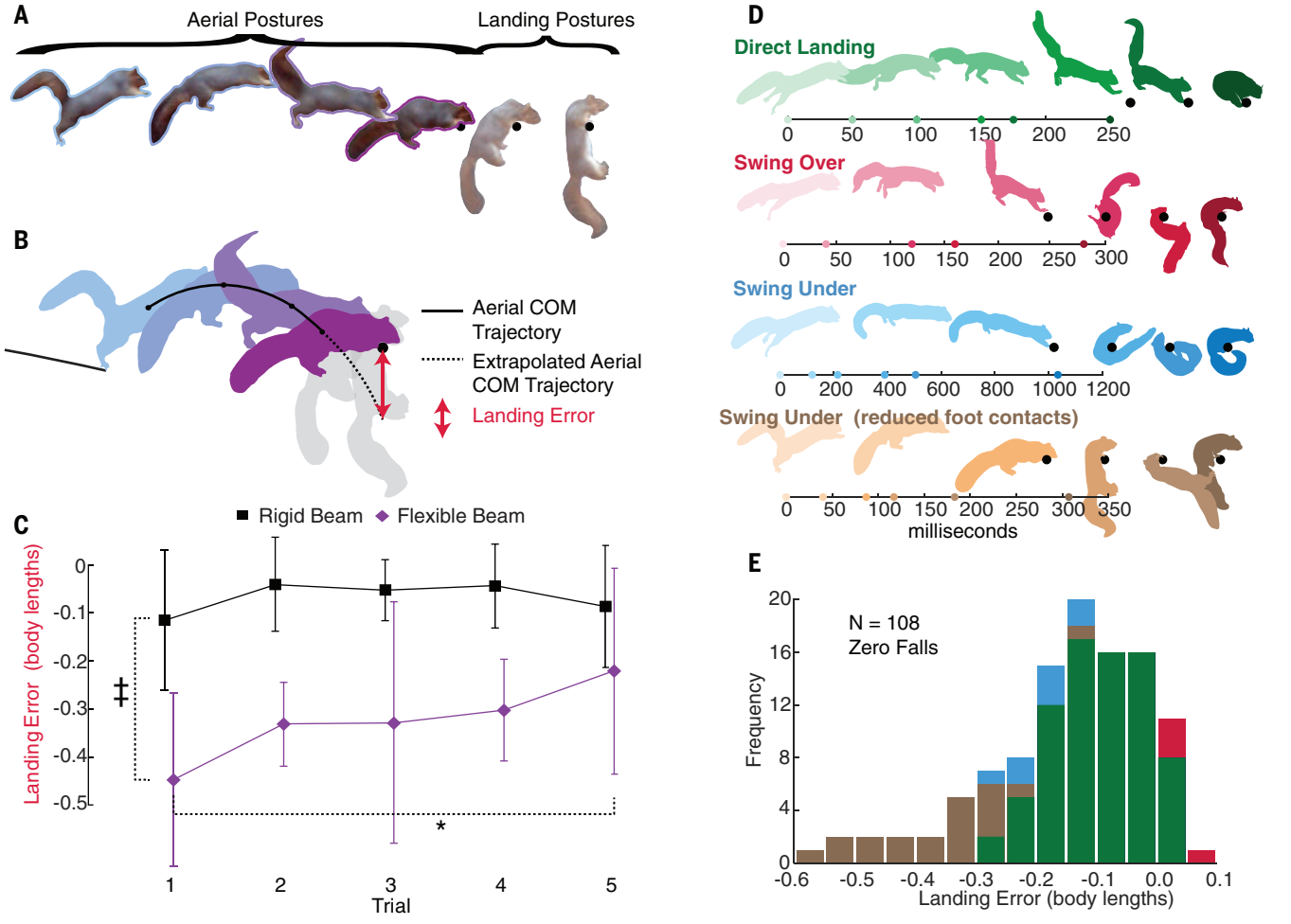


Fig. 2. Learning to leap from an unfamiliar compliant beam and land robustly. (**A**) Aerial and landing postures (fig. S2) (*12*). (**B**) Landing error was quantified as the height discrepancy between the landing perch and the extrapolated trajectory of the COM during the aerial phase (red arrow) (*12*). (**C**) Five individual squirrels performed sequential leaps from both the rigid and compliant beam (movie S2). Landing error (mean ± 95% CI) for the rigid beam (black) and compliant beam (purple) is shown for each trial. Negative landing error values indicate undershooting the perch. The asterisk and double dagger (‡) symbols indicate statistically significant differences. (**D**) Four landing maneuvers observed in the first two experiments. Color-coded markers on the time scale indicate the time of each body position. (**E**) Frequency distribution of landing maneuvers across a range of landing errors.

We hypothesized that squirrels leaping from a branch to a target perch in the canopy perceive and adapt their leaping behavior as a trade-off between branch compliance and leap distance. Launching closer to a branch attachment provides a stiffer (less compliant) substrate but increases leap distance. Launching farther from the attachment increases compliance—magnifying loss of impulse from substrate deformation (*4*)—but reduces leap distance. To test this hypothesis, we created three simulated branches with different ranges of local compliance represented by γ (Fig. 1, A and B) but of equal diameter to obscure visual cues of compliance.

Squirrel launch-point decisions showed a balanced trade-off between preferences for low compliance and short gap distance [likelihood ratio test, $\chi^2(1) = 27.69$, $P < 0.001$; Fig. 1A and movie S1]. On the least-compliant branch, launch points included gap distances less than a body length with no aerial phase. As squirrels traversed branches with higher compliance, they leapt earlier with gap distances greater than three body lengths (Fig. 1A).

To compare the relative sensitivity of squirrels' launch-point decisions to compliance and gap distance, we used selections made across the simulated branches to infer an optimality model of the launch-point decision (x^*) (*11*, *12*). The models represent the trade-off between gap distance and branch compliance with an objective (cost) function with weights, a_g and a_c, that quantified the response to gap distance and branch compliance, respectively. The unweighted model assumed equal sensitivity ($a_g = a_c$) and poorly predicted launch-point decisions [mean absolute error = 1.16 body lengths (BLs); Fig. 1C]. The weighted model, which allowed different sensitivities, accurately described the launch-point decisions across all three simulated branches (mean absolute error = 0.31 BLs; Fig. 1D). It indicated that launch-point decisions are more sensitive to compliance than to gap distance (a_c / a_g = 6.2). Although the trade-off objective function predicts an optimal launch point (fig. S3), squirrels leapt from a range of locations on each branch. The range of launch points increased with decreasing branch compliance range, indicating a wider range of sufficient (*13*) launch points (Fig. 1A). Constrained launch points on high-compliance branches may be a result of insufficient leaping momentum with a fixed-leg extension. Although not observed here, this biomechanical limitation may be mitigated through branch elastic energy storage and return (*4*, *14*, *15*).

Throughout the experiment, squirrels leapt gaps without falling. To quantify leaping performance, we defined landing error as the height discrepancy between the landing perch and a squirrel's extrapolated center of mass

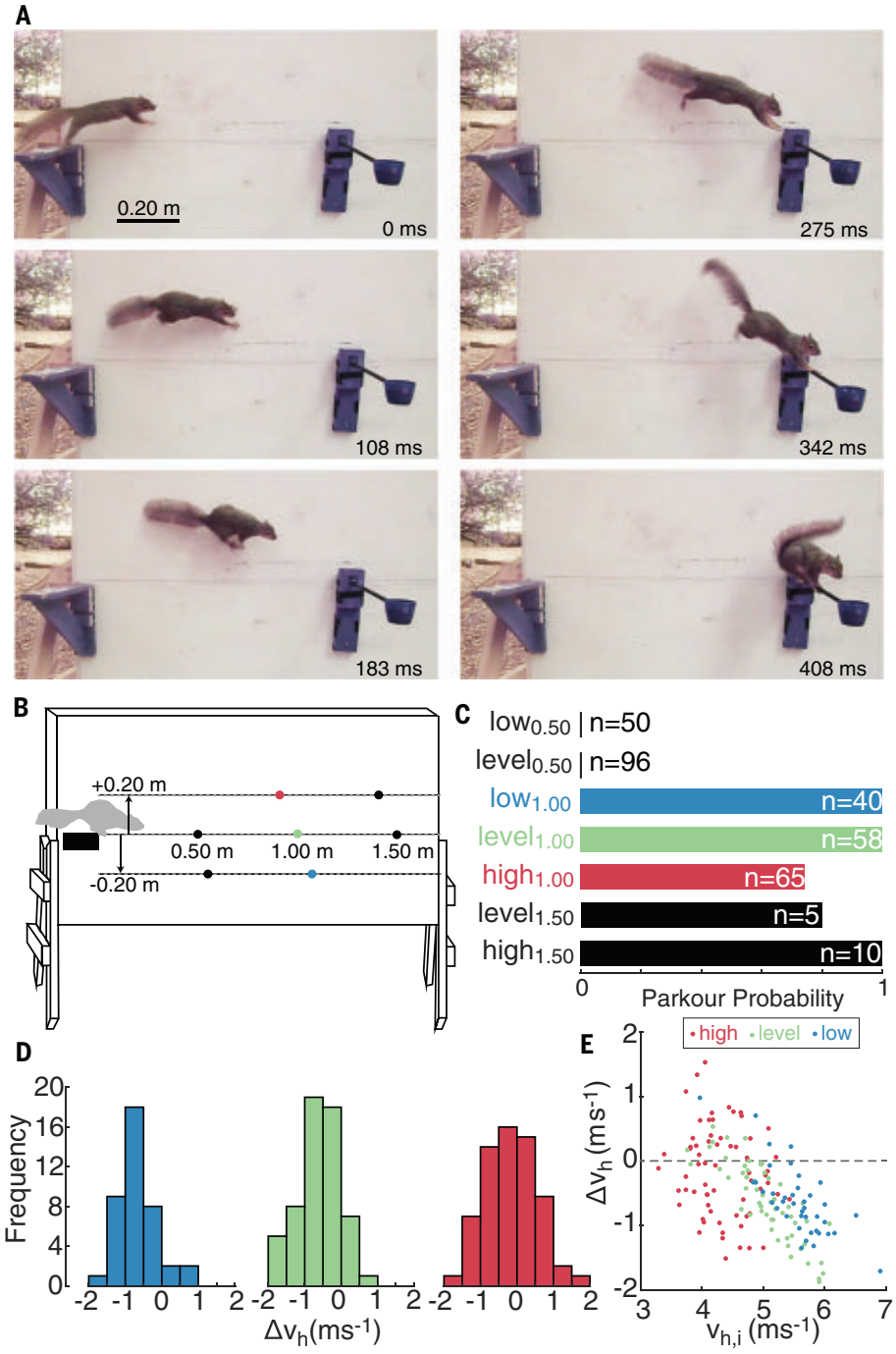


Fig. 3. Parkour leaping. (**A**) Parkour maneuver. (**B**) Diagram of the apparatus with vertical surface, launch platform, and landing perch positions. (**C**) Probabilities of using the parkour maneuver for different placements of the landing perch, where *n* equals the total number of trials. Probability estimates at the 1.50-m distance are uncertain because of the small sample size. (**D**) Change in horizontal velocity (Δv_h) for low, level, and high perch placements at the 1.00-m distance. (**E**) Change in horizontal velocity (Δv_h) versus the initial parkour contact phase landing velocity ($v_{h,i}$).

(COM) (*16*) using the ballistics of the aerial phase standardized by the squirrel's body length (Fig. 2B, fig. S2A, and movie S2). There was no significant correlation between launch-point decision and landing error [likelihood ratio test, $\chi^2(1) = 1.00$, $P = 0.32$]. Landing errors were small for low-compliance (−0.12 ± 0.08 BLs), medium-compliance (−0.12 ± 0.08 BLs), and high-compliance branches (−0.09 ± 0.08 BLs).

We hypothesized that constraining the squirrels to leap from a launch point with even greater compliance would result in greater initial landing errors but that they would demonstrate error-based learning to improve their leaping performance. Error-based motor learning—in contrast to other forms of motor learning like action selection, use-dependent, and skill learning (*17*, *18*)—is a recalibration of an existing motor control policy to compensate for differences between expected and experienced sensory information resulting from changing animal-environment dynamics (*19*), and it reoptimizes plans for future movements (*20*).

To increase the challenge of leaping in our second experiment, we introduced a thin steel beam with a compliance 2.9 times as great as that for any previous leaps (fig. S4) and a minimum gap distance 1.7 times as long as the previous average (2.2 BLs). A rigid beam was used as a control.

Five squirrels performed repeated leaps from both rigid and compliant beams. Leaping from the rigid beam led to relatively small landing errors, similar to those exhibited from the simulated branches (Fig. 2C). The compliant beam caused larger landing errors [two-way repeated measures analysis of variance (ANOVA), $F_{(1,4)} = 16.2$, $P = 0.016$] up to half a body length (Fig. 2C). Squirrels learned to improve their landings over five trials [$F_{(4,16)} = 4.60$, $P = 0.012$; movie S2]. Error reduction was achieved by changes in launch velocity [$F_{(4,16)} = 3.38$, $P = 0.035$] rather than launch angle [$F_{(4,16)} = 2.59$, $P = 0.076$], vertical launch COM position [$F_{(4,16)} = 0.75$, $P = 0.57$], or horizontal launch COM position [$F_{(4,6)} = 1.71$, $P = 0.20$].

Squirrels used swinging landing maneuvers that included caudal-led (swinging under; negative landing errors) and cranial-led (swinging over; positive landing errors) inversion of the COM about the landing perch (Fig. 2D and movie S3). Landing errors varied from −0.57 to 0.10 BLs (*12*), but no squirrels fell. The landing maneuver depended upon the landing error [Fig. 2E, mutual information measurement (*21*)].

In a third experiment, we further challenged the squirrels by manipulating gap distance and landing height. Unexpectedly, squirrels perceived the vertical surface of the apparatus as an additional affordance—an opportunity for action using its biomechanical capabilities (*9*, *22*)—and used an innovative strategy where they selectively established an additional control point midleap using a parkour leaping maneuver, a dynamic form of locomotion that extends mobility using additional forces. In this maneuver, they reoriented some or all of their legs by rolling toward the vertical surface, generating forces that could alter their trajectory (Fig. 3A and movie S4).

To determine how often squirrels use this parkour strategy, we varied the distance (0.50, 1.00, and 1.50 m) and the height of the landing perch relative to the launching platform (±0.20 m) at each perch distance along an

isocline of constant impulse (Fig. 3B). Squirrels consistently used the parkour maneuver for the medium and long leaps (ranging 3 to 5 BLs) but never for short leaps (1.5 BLs; Fig. 3C). There was a preference for using the parkour maneuver regardless of height. However, at 1.00-m distance, squirrels used the parkour maneuver less for the high perch position (74%) compared with the level and low positions (100%).

We hypothesized that squirrels use contact on a vertical surface as an additional control point to modify horizontal velocity before landing. To test this hypothesis, we quantified initial horizontal velocity before contact ($v_{h,i}$) and final horizontal velocity after contact ($v_{h,f}$) for the medium-distance leaps. Initial and final horizontal velocities ranged widely from ~3 to 7 ms^{-1} (Fig. 3E). Squirrels modified horizontal velocity (v_h) depending on the height of the landing perch [linear mixed-effects model controlling for individual, $t_{(161)} = 2.3$, $P = 0.02$]. When squirrels used the parkour strategy, they generally reduced horizontal velocity during the contact phase to reach lower perch heights ($\Delta v_h = -0.63 \pm 0.54$ ms^{-1}, mean ± SD), level perch heights ($\Delta v_h = -0.61 \pm 0.59$ ms^{-1}, mean ± SD), and high perch heights ($\Delta v_h = -0.17 \pm 0.69$ ms^{-1}, mean ± SD; Fig. 3D). In addition to an opportunity for deceleration before landing, squirrels used the parkour maneuver to make midleap velocity corrections that countered variations in initial horizontal velocity. Initial horizontal velocity predicted parkour horizontal velocity changes [linear mixed-effects model controlling for individual and perch height, $t_{(161)} = -11.0$, $P < 0.001$] {$\Delta v_h = [-0.86 \pm 0.15$, estimate ± confidence interval (CI)] $v_{h,i}$}.

Gap traversibility depends on the complement of environmental properties with an animal's locomotive capacities. The synergy between biomechanical energy management and learned information for launching and landing likely determines arboreal leaping and ultimately the path through the canopy. The role of fast and accurate leaping in driving the evolution of biomechanical capabilities, learning-based decision-making, and innovation promises to reveal the mechanisms and origins of arboreal agility (*7*, *23*–*25*).

REFERENCES AND NOTES

1. A. Schmidt, M. S. Fischer, *J. Exp. Biol.* **213**, 3615–3624 (2010).
2. M. Graham, J. J. Socha, *J. Exp. Zool.* **333**, 60–73 (2020).
3. Z. M. Jones, B. C. Jayne, *J. Exp. Biol.* **215**, 2096–2107 (2012).
4. C. A. Gilman, M. D. Bartlett, G. B. Gillis, D. J. Irschick, *J. Exp. Biol.* **215**, 220–226 (2012).
5. J. Fröbisch, R. R. Reisz, *Proc. R. Soc. B.* **276**, 3611–3618 (2009).
6. M. Cartmill, in *Primate Locomotion*, J. A. Jenkins, Ed. (Academic Press, 1974), pp. 45–83.
7. T. W. Fawcett *et al*., *Trends Cogn. Sci.* **18**, 153–161 (2014).
8. B. Demes, W. L. Jungers, T. S. Gross, J. G. Fleagle, *Am. J. Phys. Anthropol.* **96**, 419–429 (1995).
9. E. J. Gibson, A. D. Pick, *An Ecological Approach to Perceptual Learning and Development* (Oxford Univ. Press, 2000).
10. A. van Casteren *et al*., *PLOS ONE* **8**, e67877 (2013).
11. A. Y. Ng, S. Russell, in *Proceedings of the Seventeenth International Conference on Machine Learning* (2000), pp. 663–670.
12. See supplementary materials.
13. G. E. Loeb, *Biol. Cybern.* **106**, 757–765 (2012).
14. H. C. Astley, A. Haruta, T. J. Roberts, *J. Exp. Biol.* **218**, 3360–3363 (2015).
15. S. K. S. Thorpe, R. H. Crompton, R. M. N. Alexander, *Biol. Lett.* **3**, 253–256 (2007).
16. A. L. Hof, M. G. J. Gazendam, W. E. Sinke, *J. Biomech.* **38**, 1–8 (2005).
17. L. Shmuelof, J. W. Krakauer, *Neuron* **72**, 469–476 (2011).
18. J. Diedrichsen, O. White, D. Newman, N. Lally, *J. Neurosci.* **30**, 5159–5166 (2010).
19. J. Diedrichsen, R. Shadmehr, R. B. Ivry, *Trends Cogn. Sci.* **14**, 31–39 (2010).
20. J. Izawa, T. Rane, O. Donchin, R. Shadmehr, *J. Neurosci.* **28**, 2883–2891 (2008).
21. T. M. Cover, J. A. Thomas, *Entropy, Relative Entropy, and Mutual Information* (Wiley, 2005).
22. K. S. Jones, *Ecol. Psychol.* **15**, 107–114 (2003).
23. L. F. Jacobs, *Brain Behav. Evol.* **62**, 128–139 (2003).
24. T. W. Fawcett, S. Hamblin, L. A. Giraldeau, *Behav. Ecol.* **24**, 2–11 (2013).
25. N. T. Dunham *et al*., *J. Exp. Zool.* **331**, 103–119 (2019).
26. R. Full, N. Hunt, L. Jacobs, J. Jinn, Dryad (2021); https://doi.org/10.6078/D11Q5Q.

ACKNOWLEDGMENTS

We thank A. Robin, C. Lee, E. Frendberg-Mates, M. Lecoeuche, B. Weiss, K. Utsumi, R. Ta, J. Murata, E. Murphy, and N. Zangenah for assistance with data collection. **Funding:** We acknowledge funding from the National Science Foundation (NSF) Integrative Graduate Education and Research Traineeship Program (DGE-0903711) to R.J.F.; NSF CDI-Type-II: Collaborative Research: Cyber-Amplified Bioinspiration in Robotics (1028319) to R.J.F. and L.F.J.; Army Research Office (ARO MURI W911NF1810327) to R.J.F.; Chancellor's Fellowship, University of California, Berkeley, to N.H.H.; an NSF Graduate Fellowship to J.J.; Stochastic Labs (Minerva Foundation) Seed Funding Grant to L.F.J.; and National Institutes of Health (NIH) Research Enhancement (R15AG063106) and Centers of Biomedical Research Excellence (P20GM109090) to N.H.H. **Author contributions:** L.F.J. and N.H.H. initiated the project. N.H.H., J.J., and R.J.F. designed the experiments. N.H.H. and J.J. collected the data. N.H.H. analyzed the data. N.H.H. and R.J.F. wrote the paper with contributions from all of the authors. R.J.F. oversaw all aspects of the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data and code that support the findings of this study are available on Dryad (*26*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6555/697/suppl/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References (*27*, *28*)
MDAR Reproducibility Checklist
Movies S1 to S4

8 October 2020; accepted 27 May 2021
10.1126/science.abe5753

STRUCTURAL BIOLOGY

Maturation of the matrix and viral membrane of HIV-1

Kun Qu[1,2,3], Zunlong Ke[1]†, Vojtech Zila[4]†, Maria Anders-Össwein[4], Bärbel Glass[4], Frauke Mücksch[4]‡, Rainer Müller[5], Carsten Schultz[3,5,6], Barbara Müller[4], Hans-Georg Kräusslich[3,4,7], John A. G. Briggs[1,2,3]*

Gag, the primary structural protein of HIV-1, is recruited to the plasma membrane for virus assembly by its matrix (MA) domain. Gag is subsequently cleaved into its component domains, causing structural maturation to repurpose the virion for cell entry. We determined the structure and arrangement of MA within immature and mature HIV-1 through cryo–electron tomography. We found that MA rearranges between two different hexameric lattices upon maturation. In mature HIV-1, a lipid extends out of the membrane to bind with a pocket in MA. Our data suggest that proteolytic maturation of HIV-1 not only assembles the viral capsid surrounding the genome but also repurposes the membrane-bound MA lattice for an entry or postentry function and results in the partial removal of up to 2500 lipids from the viral membrane.

Assembly and budding of HIV-1 is initiated at the plasma membrane (PM) and driven primarily by a 55-kDa viral polyprotein named Gag. Gag consists of an N-terminal matrix (MA) domain, which is responsible for recruitment to the PM; the capsid (CA) domain, which induces Gag self-assembly through protein-protein interactions; the nucleocapsid (NC) domain, which recruits the viral RNA genome to the assembly site; and some small peptide domains (*1*, *2*). Protein-protein, protein-lipid, and protein-RNA interactions lead to clustering of Gag at the assembly site, membrane bending, and subsequent release of the membrane-enveloped, immature HIV-1 particle. Concomitant with or shortly after release, the viral protease (PR) cleaves Gag at multiple positions, leading to a substantial structural rearrangement to repurpose the virus particle

[1]Structural Studies Division, Medical Research Council Laboratory of Molecular Biology, CB2 0QH Cambridge, UK. [2]Structural and Computational Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany. [3]Molecular Medicine Partnership Unit, European Molecular Biology Laboratory and Universitätsklinikum Heidelberg, 69117 Heidelberg, Germany. [4]Department of Infectious Diseases, Virology, Universitätsklinikum Heidelberg, 69120 Heidelberg, Germany. [5]Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany. [6]Department of Chemical Physiology and Biochemistry, Oregon Health & Science University, Portland, OR 97239, USA. [7]German Center for Infection Research, Heidelberg, Germany.
***Corresponding author. Email: jbriggs@mrc-lmb.cam.ac.uk**
†These authors contributed equally to this work. ‡Present address: Laboratory of Retrovirology, The Rockefeller University, New York, NY 10065, USA.

for entry into a target cell (*3*, *4*). Maturation results in condensation of a ribonucleoprotein complex (RNP) from NC and RNA, which is then surrounded by a cone-shaped capsid made of CA, whereas mature MA is thought to remain associated with the viral membrane. Rearrangement of the HIV-1 envelope (Env) glycoproteins on the particle surface is presumed to render the virion fusogenic (*2*, *3*, *5*, *6*).

The structure of the heterologously expressed 17-kDa MA protein has been determined, revealing a small, folded domain composed of five α helices and one 3_{10} helix between helices 2 and 3 (*7*, *8*). MA crystallizes as a trimer (*8*) and multimerizes on artificial membrane monolayers into a hexameric lattice of trimers with holes at the sixfold symmetric positions in the lattice (*9*). It has been suggested that these holes provide binding sites for the C-terminal tail of HIV-1 Env and promote Env incorporation into assembling virions (*10*, *11*).

Membrane recruitment of Gag by MA is mediated by an N-terminal myristate moiety as well as a highly basic region (HBR, residues 17 to 31) (*12*, *13*). The N-terminal myristate is thought to be in equilibrium between sequestration in a pocket within the MA domain and an exposed conformation (*14*): This transition is known as the "myristoyl switch". Myristate exposure can be promoted by the trimerization of MA (*14*). Additionally, $PI(4,5)P_2$ (phosphatidylinositol 4,5-bisphosphate) in the PM is required for correct targeting of Gag (*15*, *16*) and for stably retaining Gag at the PM (*17*). Early NMR (nuclear magnetic resonance) studies suggested that insertion of the 2′ acyl chain of $PI(4,5)P_2$ in an extended lipid conformation into a pocket on the side of MA may promote exposure of the myristate and its insertion into the lipid bilayer (*18*). However, this mode of $PI(4,5)P_2$ binding, in which the 2′ acyl chain would have to be pulled out of the bilayer, is no longer widely considered relevant for PM binding. Instead, MA-membrane interactions are believed to involve the $PI(4,5)P_2$ headgroup and the membrane-facing HBR (*15*, *17*, *19*, *20*). MA interacts with nucleic acids in the cytosol, particularly tRNAs, through the HBR (*21*–*23*). These observations support a model in which an exchange of nucleic acid for the $PI(4,5)P_2$ headgroup promotes myristate exposure and stabilization of PM binding during virus assembly (*21*, *23*, *24*).

The known roles of MA are performed during virus assembly, and no function for MA during entry or postentry stages is currently established. Models for how MA functions during assembly are limited by a current lack of structural information on MA within virions. It is unclear what conformation and arrangement MA adopts in situ, and whether MA undergoes any structural changes in the process of virus maturation. In this study, we set out to determine the structure and arrangement of MA within immature and mature virus particles.

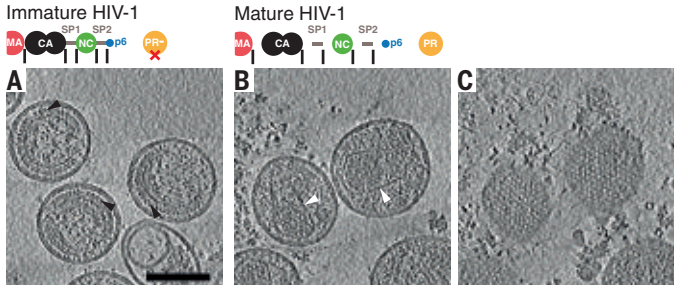


Fig. 1. Tomograms of immature and mature HIV-1 particles. (**A**) (Top) Schematic of Gag protein domain architecture with cleavage sites marked by vertical black lines. The $Asp^{25} \rightarrow Asn$ (D25N) mutation (red X) inactivates the viral PR. (Bottom) Computational slices through a representative tomogram of immature HIV-1 (cHIV PR-) particles. Density is black. The striated Gag layer (black arrowheads) is distinct in the immature particles. (**B**) Same as in (A) for cleaved, mature HIV-1 (cHIV) particles. Distinct conical CA cores are indicated by white arrowheads. (**C**) Slices grazing the inner surface of the membrane at the top of the particles in (B), revealing a regular MA lattice. Gold fiducials are removed from the images. Scale bar: 100 nm.

Purified mature particles were generated by transfection of a noninfectious HIV-1 proviral derivative that expresses all viral proteins except the virulence factor Nef (pcHIV plasmid that generates cHIV particles) (*25*). Immature cHIV particles were generated by transfection of pcHIV carrying an active-site mutation in PR (pcHIV PR-). Particles were imaged without chemical fixation by using cryo–electron tomography (cryo-ET). Visual inspection of tomograms revealed the expected features: a striated CA layer in immature virus particles and conical CA cores in mature virus particles (Fig. 1, A and B). In some regions underlying the lipid envelope of mature virions, we observed regular lattice-like features (Fig. 1C).

We subjected the region underlying the viral membrane in the immature cHIV particles to reference-free subtomogram averaging. This analysis revealed a hexameric lattice of MA trimers with hexamer-hexamer spacing of 9.8 nm, as well as large holes at the sixfold axes (Fig. 2A), reminiscent of the lattices observed in membrane-associated purified MA in vitro (*9*). To improve the spatial resolution of the immature MA structure, we analyzed a cryo-ET dataset of fixed, immature, complete $HIV\text{-}1_{NL4\text{-}3}$ particles (*26*). From this dataset, we obtained the same MA lattice structure at 7.2 Å resolution (Fig. 2, B to F, and fig. S1). The poorly ordered and sparse MA lattice covered only small patches of the inner membrane leaflet, whereas in other areas MA did not appear to form a lattice, suggesting heterogeneity in MA packing (Fig. 2G and fig. S2). Despite this heterogeneity, we could clearly resolve α helices (Fig. 2, C to F). The MA trimer structure, determined by crystallography [PDBID: 1HIW (*8*)], could be fitted into the density as a rigid body. The MA domain of Gag is therefore present as a trimer within immature HIV-1 particles (Fig. 2D and fig. S3A).

MA helices 1 and 2 lie approximately parallel to the membrane and form a surface rich in charged and hydrophobic residues that attach MA to the inner membrane leaflet (Fig. 2, C and F). MA trimers are packed together to form a lattice by a dimeric interface where the N-terminal residues and the N terminus of helix 1 interact with themselves and the 3_{10} helix (Fig. 2, D and E). We did not observe any density corresponding to lipids bound in the described $PI(4,5)P_2$ binding site on the side of MA, next to the HBR (*18*) (Fig. 2F). The absence of density is consistent with current models in which interactions between the headgroup of $PI(4,5)P_2$ and the membrane-proximal face of MA function as the major cellular determinant of efficient MA-PM targeting (*17*, *19*). The N-terminal region of MA (residues 1 to 11), which changes conformation during the myristoyl switch (*14*, *27*), is directly involved in trimer-trimer interactions (fig. S3B). At lower isosurface thresholds, a connection to the membrane appears in the vicinity of the N terminus of helix 1 (Fig. 2E). No notable density was observed in the myristoyl pocket (fig. S3C). The conformational equilibrium therefore shifts toward the exposed, membrane-inserted conformation in the immature virion, consistent with exposure of the myristate moiety upon MA oligomerization or membrane recruitment.

The C-terminal helix 5 of MA extends toward the center of the virus particle and is separated from the underlying CA layer by a disordered stretch containing the MA-CA cleavage site (Fig. 2C). Therefore, CA is not resolved when MA is aligned, and MA is not resolved when CA is aligned (fig. S4). The position of MA is thus constrained but not fixed relative to CA.

A number of studies have implied that holes in a hypothetical MA lattice may present a binding site for the cytoplasmic tail of Env, and mutations affecting MA trimerization have been shown to affect Env incorporation

(*9*, *10*, *28*). We found that MA does form a lattice containing holes, although we did not observe density for the Env tail in the holes, consistent with the low Env copy number per particle (*29*). A number of mutations in MA have been reported to result in defects in Env incorporation (*10*, *11*, *28*, *30*–*33*) (Fig. 2H, red, purple, and blue spheres). Many of the mutated residues (red and purple spheres) are not exposed toward the central hole; rather, they are located at sites close to the N-terminal residues of MA or the intra- and intertrimer interfaces, consistent with their phenotype in MA trimerization (Fig. 2H). These mutations are thus likely to modulate MA trimerization or lattice formation, and/or the myristoyl switch, but would not be presumed to directly control Env cytoplasmic tail binding within the holes. This is consistent with the model in which correct MA oligomerization is required for Env incorporation (*11*, *34*).

We next subjected the region underlying the membrane in the mature cHIV particles to reference-free subtomogram averaging. This analysis revealed large patches of a hexagonal lattice of MA trimers with hexamer-hexamer spacing of ~8.8 nm. These dimensions are similar to the 9-nm repeating MA lattice observed in anomalous, multicored, membranous particles (*35*). We resolved the structure of the lattice to a resolution of 9.5 Å (Fig. 3A). Like the immature lattice, the mature lattice is formed from MA trimers arranged in a hexagonal lattice, though it has a higher degree of regularity than the immature MA lattice (fig. S2). The arrangement of MA within the immature and mature MA lattices is, however, notably different (compare Figs. 2 and 3), as the MA lattice undergoes structural maturation to form a new, different hexagonal lattice in the mature virion.

We had previously determined the structure of the CA layer within cHIV derivatives in which either the PR cleavage site between MA and CA was blocked by mutation (HIV-1 MA-CA) or in which the cleavage site between CA and SP1 was blocked in addition (HIV-1 MA-SP1) (*36*–*38*). The CA domains and the CA lattices remain structurally immature in MA-CA and MA-SP1 particles. Upon analyzing the MA layer in these datasets, we found that the MA lattice in HIV-1, MA-CA, and MA-SP1 corresponds to that of mature particles, despite the absence of cleavage between MA and CA (Fig. 3B and fig. S5). Cleavages downstream of SP1 are therefore sufficient for MA lattice maturation, which can occur without CA lattice maturation. It is unlikely that proteolytic cleavage between CA and NC could directly trigger MA maturation through a structural signal passed through CA without inducing any detectable structural changes in CA. We therefore suggest that maturation of MA must be triggered in trans by another effector in the viroplasm or viral membrane. Further experiments will be required to elucidate the mechanism of MA maturation. Immature virions are fusion incompetent (*5*), but cleavages downstream of CA are sufficient to overcome this defect (*39*). These observations suggest that structural

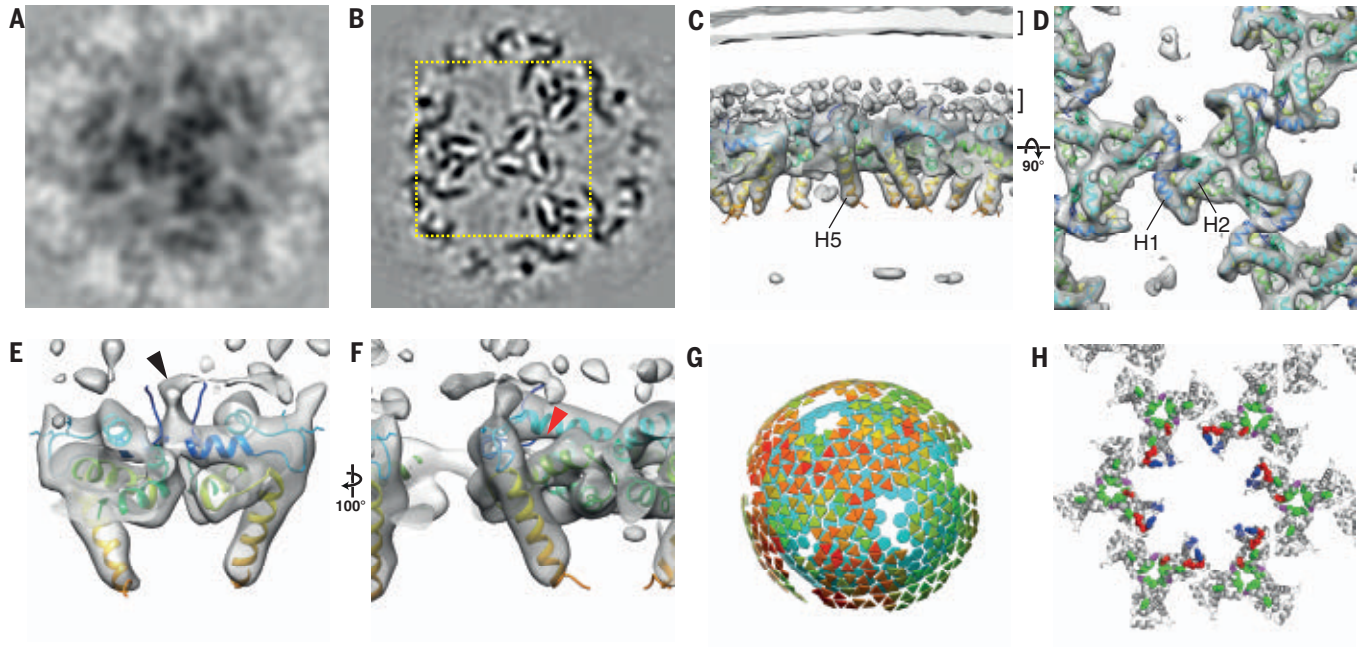


Fig. 2. Immature HIV-1 matrix structure. (**A** and **B**) Slices through reconstructions of the immature HIV-1 MA lattice in cHIV PR- (A) and HIV-1_{NL4-3} PR- (B). The region within the yellow box in (B) is shown at a higher magnification in (D). Density is black. (**C** and **D**) Isosurface views of a cryo-ET reconstruction (gray) of an immature HIV-1_{NL4-3} PR- MA lattice cut perpendicular to the membrane (C) or viewed from the top toward the virus center (D). In (C), the two layers of density corresponding to the lipid headgroup layers are indicated by brackets. The structure of monomeric MA previously determined by NMR [PDBID: 2H3Q (*18*), colored blue to red from the N- to C-terminus, respectively], was fitted as a rigid body into the density. Helices 1, 2, and 5 are marked. (**E**) Same as in (C), enlarged and cut to reveal density corresponding to the N-terminal residues (black arrowhead). (**F**) Same as in (E) but rotated. The red arrowhead indicates the unoccupied $PI(4,5)P_2$ binding pocket. (**G**) Lattice map derived from subtomogram averaging for the immature HIV-1_{NL4-3} PR- MA lattice. Positions and orientations of MA trimers are illustrated by triangles colored on a scale from red (lower cross-correlation to average structure) to green (higher cross correlation to average structure). The positions of the CA hexamers are indicated by cyan hexagons to illustrate the relationship between the MA and CA layers. (**H**) Immature MA lattice (gray ribbons) and residues, the mutations of which have been reported to modulate Env incorporation (colored spheres). Mutations at residues L12, L30, and L74 (red), E16 and E98 (blue), and T69 (purple) impair Env incorporation (*11*, *28*, *30*–*33*). Mutations in V34, F43, Q62, and S66 (green) can rescue Env incorporation defects (*11*, *28*, *31*, *43*), except for those caused by mutations at T69 (purple) (*28*). E16 and E98 (blue) face the hole in the MA lattice in the immature virion. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

maturation of MA, rather than CA, correlates with HIV-1 becoming fusogenic, potentially by allowing Env to redistribute on the virus surface (*6*).

We obtained a higher-resolution 7.0-Å structure of the mature MA lattice from MA-SP1 particles. As for the immature MA lattice, the crystallographic trimer could be fitted into the density as a rigid body (Fig. 3, B to E, and fig. S1). Comparison of the immature and mature MA structures revealed important differences (Fig. 4). In the immature lattice, the HBR faces the holes at the sixfold axes, which are therefore surrounded by a basic surface (Fig. 4B). By contrast, in the mature virus MA presents a largely neutral surface toward the holes (Fig. 4B). This change could influence structural changes in the cytoplasmic tails of Env trimers that may alter their distribution and fusogenicity. In the mature virus, basic residues in the HBR loop face acidic residues in the N terminus of helix 4 (E72) and the 3_{10} helix (E51) of the adjacent MA monomer, to form a dimeric interface that links trimers together (Fig. 3D, Fig. 4A, and fig. S3B). The PI(4,5)P_2 binding pocket on the side of MA, adjacent to the HBR (*18*), is positioned at the center of the dimer interface. Unlike the immature virus, wherein the PI(4,5)P_2 binding pocket is empty, it contains a density in the mature virus. Our structure does not resolve the headgroup or acyl chain of the bound lipid, but the density is consistent with the previously described extended-lipid conformation of bound PI(4,5)P_2 in which the 2′ acyl chain is removed from the bilayer, whereas the 1′ acyl chain extends upward toward the bilayer [PDBID: 2H3Q and 2H3V (*18*)] (Fig. 3, D and E, and fig. S6). We consider it reasonable to assign the observed density to PI(4,5)P_2. The N terminus of helix 1 prominently protrudes up into the PM, indicating that, in the mature MA lattice, the myristate moiety predominantly retains an exposed conformation, inserted into the membrane.

To biochemically determine a potential interaction between MA and PI(4,5)P_2 in the immature or mature lattice, we prepared immature and mature particles from cells incorporating a functionalized PI(4,5)P_2 derivative [f-PI(4,5)P_2] with a diazirine group in the 1′ acyl chain, which can be photo-crosslinked to nearby proteins (fig. S7A). We found that upon photoactivation, f-PI(4,5)P_2 was efficiently crosslinked to Gag within immature PR- particles, but not to MA within mature particles (fig. S7B). This observation confirms direct interaction of the MA domain of the Gag polyprotein with PI(4,5)P_2, reveals that proteolytic cleavage of Gag alters the relative positions of MA and PI(4,5)P_2, and provides direct biochemical evidence for a change in MA-lipid interactions upon HIV-1 maturation.

Taken together, our results (summarized in fig. S8) revealed that MA–like CA and NC–

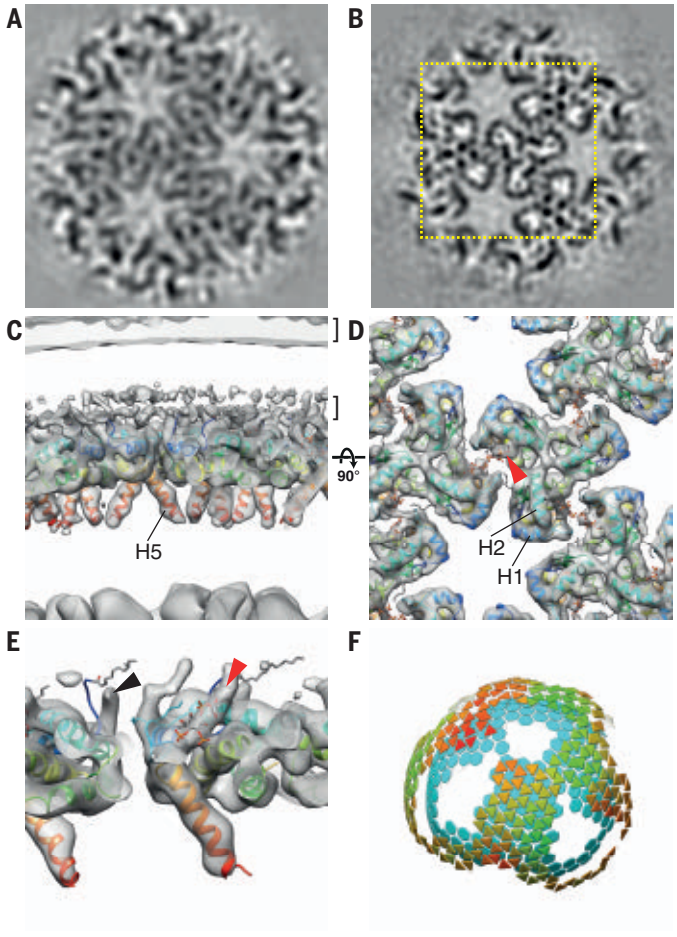


Fig. 3. Mature HIV-1 matrix structure. (**A** and **B**) Slices through reconstructions of the mature HIV-1 MA lattice in cHIV (A) and cHIV MA-SP1 (B). The region within the yellow box in (B) is shown at a higher magnification in (D). (**C** and **D**) An isosurface view of the cryo-ET reconstruction for mature cHIV MA-SP1 MA lattice fitted with the structure of monomeric MA, cut perpendicular to the membrane (C), or viewed from the top toward the virus center (D). In (C), the two layers of density corresponding to the lipid headgroup layers are indicated by brackets. The structure of the monomeric MA previously determined by NMR [PDBID: 2H3Q (*18*); colored blue to red from the N- to C-terminus, respectively], was fitted as a rigid body into the density. Helices 1, 2, and 5 are marked. Density is observed in the PI(4,5)P_2 binding site (red arrowhead), and the structure of PI(4,5)P_2 bound to MA as resolved by NMR [PDBID: 2H3V (*18*)], is shown as a stick model. (**E**) Same as in (C), enlarged and cut to reveal density corresponding to the N-terminal residues (black arrowhead), and PI(4,5)P_2 (red arrowhead). (**F**) Same as in Fig. 2G, lattice map for a mature cHIV MA-SP1 MA lattice and the underlying CA lattice.

undergoes substantial structural maturation to form very different lattices in immature and mature HIV-1 (movie S1). Immature virions contain MA trimers packed together through their N termini into a poorly ordered hexagonal lattice containing basic-charged holes and disordered regions, which could accommodate the cytoplasmic tails of Env. The N-terminal myristate is predominantly membrane-inserted in the immature and mature particle. The immature MA structure provides a framework for understanding the established roles of MA in virus assembly and Env incorporation. Mature virions, by contrast, contain MA trimers packed together by means of their HBRs, resulting in a well-ordered hexagonal lattice with neutral holes. The maturation of MA is reminiscent of that of CA, which matures between two different hexameric protein lattices with two different functional roles (*3*). Upon maturation, both the protein arrangement and the lipid binding properties of MA change. The lipid binding pocket on the side of MA is occupied in the mature structure, containing density consistent with the previously described extended-lipid conformation of PI(4,5)P_2 (fig. S6B), in which one acyl chain is removed from the membrane (*18*). The resolution of our density does not allow us to distinguish PI(4,5)P_2 from other lipids, but its position is such that any lipid positioned within this density is pulled outward from the lipid bilayer. The virion contains approximately three PI(4,5)P_2 molecules per Gag (*40*); we assume that the remaining PI(4,5)P_2 molecules may interact with the membrane-proximal surface of MA through their headgroups, as in the immature virus. Our data therefore imply that maturation of Gag results in the partial removal of up to ~2500 of the ~150,000 lipids in the inner leaflet of the viral membrane (*40*). The energy barrier to the removal of acyl chains from the bilayer may be overcome by the binding of PI(4,5)P_2 to MA and stabilization of the MA lattice.

In conclusion, our data suggest that the structural maturation of Gag not only condenses the RNP and achieves assembly of a viral capsid core surrounding the genome

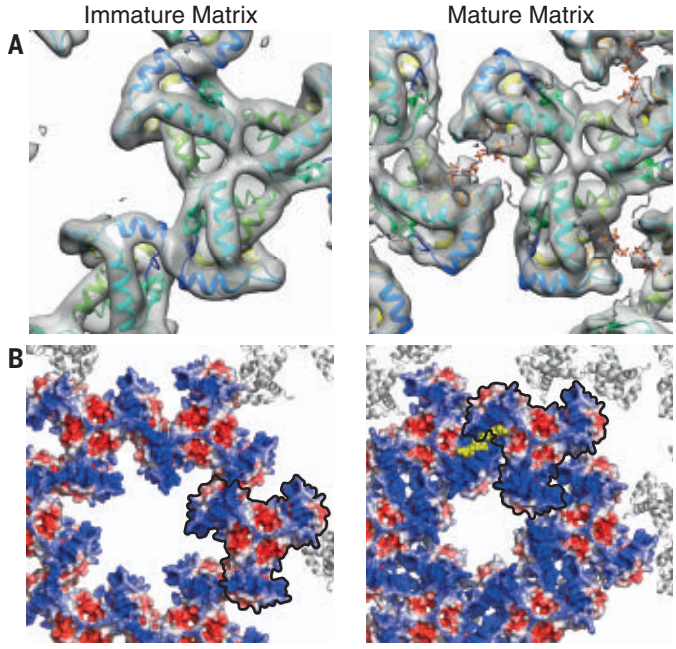


Fig. 4. Comparison of the immature and mature HIV-1 MA lattices. (**A**) Immature and mature MA lattices are shown with one trimer aligned. In the immature lattice, contact with neighboring trimers is mediated by N-terminal regions, and the $PI(4,5)P_2$ binding site is empty. In the mature lattice, contact with neighboring trimers is mediated by the region surrounding the occupied $PI(4,5)P_2$ binding site. (**B**) Electrostatic surface potential maps of the hexameric lattice of immature and mature MA trimers. The red [−5 *kT/e* (k, Boltzmann constant; T, temperature; e, electron charge)] and blue (+5 *kT/e*) colors represent negatively and positively charged electric potentials, respectively. $PI(4,5)P_2$ is shown in yellow. The negatively charged lipid headgroups are surrounded by positively charged residues. One MA trimer, aligned as in (A), is outlined in black. The surface potential of the MA facing the holes at the hexamer positions in the lattice changes from positive to neutral or negative during maturation. Additionally, the hole becomes smaller.

but also rearranges the membrane-bound MA layer and modulates the lipid bilayer itself. How these changes alter the properties of the virus remains to be elucidated, but we speculate that they affect not only MA and Env function but also the physical properties of the viral envelope, possibly correlating with the observed reduction in stiffness of the virus particle (*41*, *42*). The unexpected presence of a different, well-defined MA structure in mature virions suggests that MA performs additional roles in the mature virion or after entry into the target cell. We speculate that $PI(4,5)P_2$-stabilized MA lattices may remain associated with the membrane after virion fusion and have a postentry function, perhaps acting as signaling platforms to prepare the target cell for early HIV-1 replication.

REFERENCES AND NOTES

1. W. I. Sundquist, H.-G. Kräusslich, *Cold Spring Harb. Perspect. Med.* **2**, a006924–a006924 (2012).
2. E. O. Freed, *Nat. Rev. Microbiol.* **13**, 484–496 (2015).
3. S. Mattei, F. K. Schur, J. A. G. Briggs, *Curr. Opin. Virol.* **18**, 27–35 (2016).
4. O. Pornillos, B. K. Ganser-Pornillos, *Curr. Opin. Virol.* **36**, 47–55 (2019).
5. T. Murakami, S. Ablan, E. O. Freed, Y. Tanaka, *J. Virol.* **78**, 1026–1031 (2004).
6. J. Chojnacki *et al.*, *Science* **338**, 524–528 (2012).
7. M. A. Massiah *et al.*, *J. Mol. Biol.* **244**, 198–223 (1994).
8. C. P. Hill, D. Worthylake, D. P. Bancroft, A. M. Christensen, W. I. Sundquist, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 3099–3104 (1996).
9. A. Alfadhli, R. L. Barklis, E. Barklis, *Virology* **387**, 466–472 (2009).
10. P. R. Tedbury, E. O. Freed, *Trends Microbiol.* **22**, 372–378 (2014).
11. P. R. Tedbury, M. Novikova, A. Alfadhli, Y. Hikichi, I. Kagiampakis, *J. Virol.* **94**, e01526–e19 (2019).
12. M. Bryant, L. Ratner, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 523–527 (1990).
13. W. Zhou, L. J. Parent, J. W. Wills, M. D. Resh, *J. Virol.* **68**, 2556–2569 (1994).
14. C. Tang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 517–522 (2004).
15. A. Ono, S. D. Ablan, S. J. Lockett, K. Nagashima, E. O. Freed, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14889–14894 (2004).
16. V. Chukkapalli, I. B. Hogue, V. Boyko, W.-S. Hu, A. Ono, *J. Virol.* **82**, 2405–2417 (2008).
17. F. Mücksch, V. Laketa, B. Müller, C. Schultz, H.-G. Kräusslich, *eLife* **6**, e25287 (2017).
18. J. S. Saad *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11364–11369 (2006).
19. P. Y. Mercredi *et al.*, *J. Mol. Biol.* **428**, 1637–1655 (2016).
20. R. E. Murphy, J. S. Saad, *Viruses* **12**, 548 (2020).
21. V. Chukkapalli, S. J. Oh, A. Ono, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1600–1605 (2010).
22. A. Alfadhli *et al.*, *J. Mol. Biol.* **410**, 653–666 (2011).
23. C. R. Gaines *et al.*, *J. Mol. Biol.* **430**, 2113–2127 (2018).
24. A. Alfadhli, A. Still, E. Barklis, *J. Virol.* **83**, 12196–12203 (2009).
25. M. Lampe *et al.*, *Virology* **360**, 92–104 (2007).
26. F. K. M. Schur *et al.*, *Science* **353**, 506–508 (2016).
27. J. S. Saad *et al.*, *J. Mol. Biol.* **366**, 574–585 (2007).
28. P. R. Tedbury, S. D. Ablan, E. O. Freed, *PLOS Pathog.* **9**, e1003739 (2013).
29. E. Chertova *et al.*, *J. Virol.* **76**, 5315–5325 (2002).
30. E. O. Freed, M. A. Martin, *J. Virol.* **69**, 1984–1989 (1995).
31. E. O. Freed, M. A. Martin, *J. Virol.* **70**, 341–351 (1996).
32. A. Joshi, S. D. Ablan, F. Soheilian, K. Nagashima, E. O. Freed, *J. Virol.* **83**, 5375–5387 (2009).
33. L. Brandano, M. Stevenson, *J. Virol.* **86**, 2347–2359 (2012).
34. P. R. Tedbury, M. Novikova, S. D. Ablan, E. O. Freed, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E182–E190 (2016).
35. G. A. Frank *et al.*, *Nat. Commun.* **6**, 5854 (2015).
36. K. Wiegers *et al.*, *J. Virol.* **72**, 2846–2854 (1998).
37. A. de Marco *et al.*, *PLOS Pathog.* **6**, e1001215–e1001219 (2010).
38. S. Mattei *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E9401–E9410 (2018).
39. D. J. Wyma *et al.*, *J. Virol.* **78**, 3429–3435 (2004).
40. F. Mücksch *et al.*, *Sci. Rep.* **9**, 17661–13 (2019).
41. N. Kol *et al.*, *Biophys. J.* **92**, 1777–1783 (2007).
42. H.-B. Pang *et al.*, *Retrovirology* **10**, 4–11 (2013).
43. A. Ono, M. Huang, E. O. Freed, *J. Virol.* **71**, 4409–4418 (1997).

ACKNOWLEDGMENTS

We thank W. Hagen, D. Morado, A. Tan, F. Schur, and S. Mattei for technical assistance, and L. James and P. Chlanda for helpful discussions. This study made use of the electron microscopes at EMBL and the MRC-LMB EM Facility as well as high-performance computing resources set up and maintained by EMBL IT Services and LMB Scientific Computing. We thank the staff who maintain those resources. J.A.G.B. has a secondary affiliation as director at the Max Planck Institute of Biochemistry. **Funding:** Funding was provided to J.A.G.B. by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC-CoG-648432 MEMBRANEFUSION), the Medical Research Council (MC_UP_1201/16), and the European Molecular Biology Laboratory. Funding was also given to H.G.K. by the Deutsche Forschungsgemeinschaft (112927078-TRR 83 and 240245660-SFB1129 project 5), to B.M. by the Deutsche Forschungsgemeinschaft (MU885-1 and 240245660-SFB1129 project 6), and to C.S. by the NIH (R01GM127631) and the Deutsche Forschungsgemeinschaft (278001972-TRR186 project Z1). **Author contributions:** K.Q., B.M., H.G.K., and J.A.G.B. conceived the project. K.Q., Z.K., V.Z., B.M., H.G.K., and J.A.G.B. designed the experiments. V.Z., M.A.Ö., and B.G. purified virus particles. K.Q. and Z.K. performed electron microscopy and computational image processing. R.M. and C.S. provided the f-PI(4,5)P_2 derivative. F.M. and V.Z. respectively established and performed crosslinking experiments. K.Q. and J.A.G.B. wrote the original draft with assistance from B.M. and H.G.K.; the draft was edited and reviewed by all authors. K.Q. and V.Z. prepared figures. B.M., H.G.K., and J.A.G.B. acquired funding and supervised and managed the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All reasonable requests for materials will be fulfilled under a material transfer agreement with the providing institution. Representative tomograms are deposited in the Electron Microscopy Data Bank (EMDB) under accession codes EMD-4020, EMD-13079, EMD-13085, and EMD-13086; the immature and mature lattice structures are deposited under accession codes EMD-13087 and EMD-13088. The associated molecular models are deposited in the Protein Data Bank (PDB) under identifiers 7OVQ and 7OVR.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6555/700/suppl/DC1
Materials and Methods
Figs. S1 to S9
Table S1
References (*44*–*65*)
MDAR Reproducibility Checklist
Movie S1

7 September 2020; accepted 24 June 2021
10.1126/science.abe6821



Global Recruitment

Academic Talents for Zhongfa Aviation University

ABOUT US

Zhongfa Aviation University, funded by Zhejiang Province and the city of Hangzhou, is jointly established by Beihang University and Ecole Nationale de l'Aviation Civile (ENAC). The university is dedicated to multidisciplinary and innovation-oriented education as well as cutting-edge and core technology research. With the goal of becoming a world-class university, ZFAU will stay committed to establishing a new international university model that is modern and forward-thinking. The campus of ZFAU is situated adjacent to Liangzhu Ancient City, a UNESCO World Cultural Heritage site. The picturesque campus is expected to be completed by 2023 and will provide the most conducive environment for students and staff to live and work. In the present preparatory stage, the university operates from an interim campus in the northwest of Hangzhou.

RESEARCH AREAS

School of Civil Aviation
The School of Civil Aviation focuses on a broad range of research areas including but not limited to space-air-ground integrated network and aeronautical broadband communication, air traffic management and transportation planning, avionics and satellite navigation, general aviation and intelligent operations, and civil aviation finance and regulations.

School of Aerospace
The School of Aerospace focuses on a broad range of research areas including but not limited to trans-sonic aviation aerodynamics and thermodynamics, green aviation and alternative aviation fuels (AAFs), aerospace big data and its environmental and climatic implications, and unmanned/manned general aviation and unconventional aerospace engineering.

School of Informatics
The School of Informatics focuses on a broad range of research areas including but not limited to quantum precision measurement, high-performance quantum computation, artificial intelligence, integrated circuit materials and devices, electronic design automation of integrated circuits, aviation microelectronics, and control science and engineering.

School of Engineering
The School of Engineering focuses on a broad range of research areas including but not limited to high-temperature/ultra-high-temperature materials, high-temperature functional coating, high-temperature corrosion, advanced lightweight materials, advanced composite materials, new energy materials, optoelectronic materials, digital design and manufacturing integration technology, carbon fiber composites manufacturing process, and robot design and control technology.

School of Science
The School of Science focuses on a broad range of research areas including but not limited to new state of quantum matter, extreme condition physics, and medical physics.

ESSENTIAL QUALIFICATIONS

Applicants should fit one of the four following categories:

Type A (Chair Professor): world-leading scholars.

Type B (Associate Professor/Full Professor): tenured associate professors/full professors at top universities or research institutions, OR exceptional scholars with excellence in academic performance and global academic influence. Applicants should hold a doctoral degree from a top university or research institution.

Type C (Assistant Professor/Associate Professor): assistant professors/associate professors at top universities or research institutions, OR scholars with an outstanding track record of academic or technological performance in their fields of research. Applicants should hold a doctoral degree from a top university or research institution.

Type D (Postdoctoral Fellow): Ph.D. graduates from top universities with strong scientific research capabilities and recognised academic potential.







SALARY, BENEFITS, AND RESEARCH SUPPORT

Salary and benefits
Each successful applicant will receive an internationally competitive salary (e.g. up to US$77,000 for post-doctors, subsidies included), and a fringe benefit package that includes a rent-free apartment, access to high-quality nursery/primary education for children, etc.

Research support
Successful applicants will be provided with state-of-the-art research platforms, generous research funds, ample laboratory space and excellent office environment, etc (the establishment of tailored research labs will be strongly supported). Academic freedom and a multicultural work environment will be guaranteed.

APPLY

Please scan the QR code for more information

Scan for more info ▸

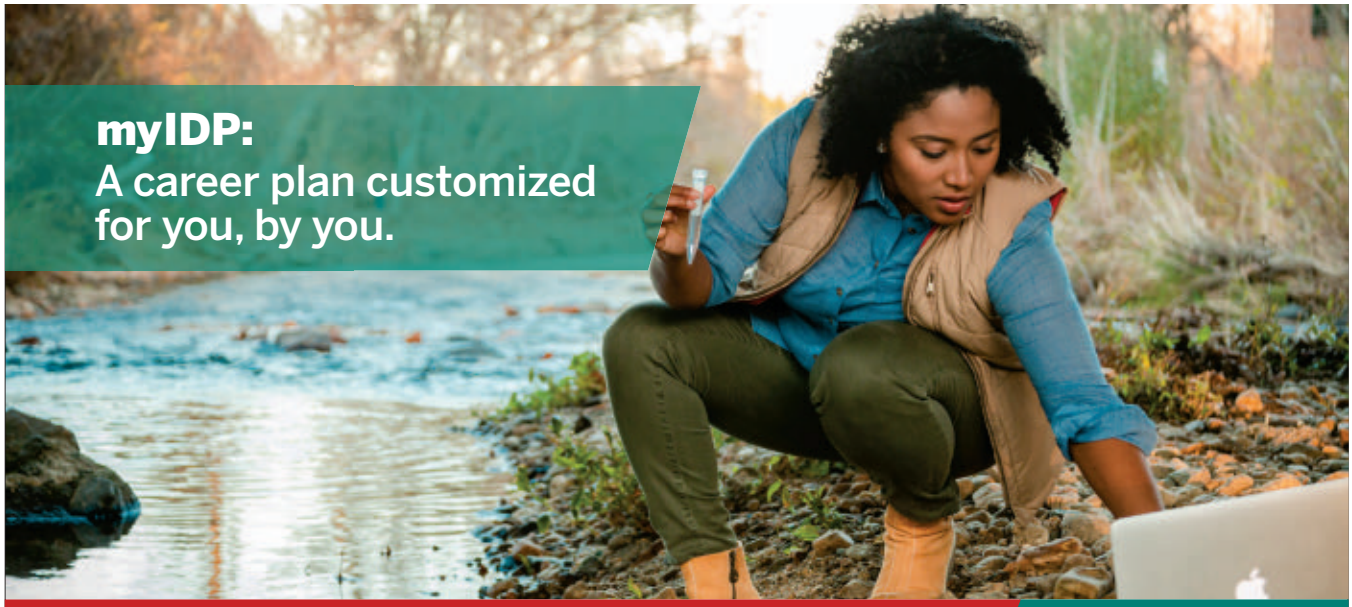
CONTACT US

Tel: +86 (0)571 89308159
Email: zhongfa@vip.sina.com
For more information, please follow our Wechat official account (Wechat ID.: ZFAU_HZ).

Scan to follow ▸

ScienceCareers online @sciencecareers.org
OPPORTUNITIES IN CHINA

ScienceCareers online @sciencecareers.org



For your career in science, there's only one Science

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.
- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track.
- Articles and resources to guide you through the process.
- Options to save materials online and print them for further review and discussion.
- Ability to select which portion of your IDP you wish to share with advisors, mentors, or others.
- A certificate of completion for users that finish myIDP.

myIDP

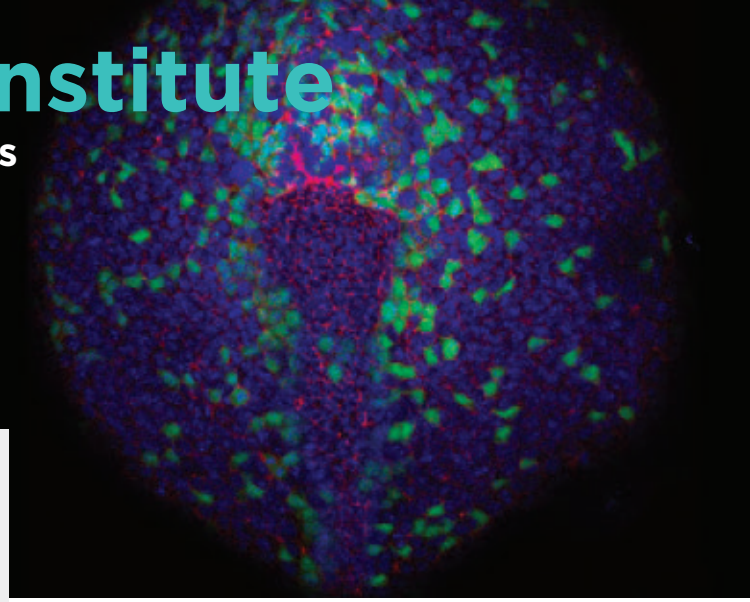
Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers AAAS In partnership with:

FASEB Federation of American Societies for Experimental Biology
UMASS University of Massachusetts Medical School
UCSF University of California San Francisco
UIC
BURROUGHS WELLCOME FUND

Sloan Kettering Institute

Celebrating 75 years of breakthroughs



Join our faculty and help set the pace of biomedical science.

JOIN US

Successful candidates will hold an appointment in one of SKI's research programs. Candidates may apply to up to two programs. MSK is an equal opportunity and affirmative action employer committed to diversity and inclusion in all aspects of recruiting and employment.

Cancer Biology & Genetics
Chair: Scott Lowe, PhD

Cell Biology
Chair (Interim): Xuejun Jiang, PhD

Chemical Biology
Chair: Derek Tan, PhD

Computational & Systems Biology
Chair: Dana Pe'er, PhD

Developmental Biology
Chair: Anna-Katerina Hadjantonakis, PhD

Immunology
Chair: Alexander Rudensky, PhD

Molecular Biology
Chair: John Petrini, PhD

Molecular Pharmacology
Chair: David Scheinberg, MD, PhD

Structural Biology
Chair: Christopher Lima, PhD



Memorial Sloan Kettering Cancer Center

AREAS OF BASIC AND TRANSLATIONAL RESEARCH

- Stem Cell Biology
- Developmental Oncology
- Regenerative Medicine
- Machine Learning
- 3D Single-Cell Analytics
- Biophysics & Imaging
- Organelle Biology
- Chromatin & Gene Regulation
- Genome Integrity and Functional Genomics
- Chemistry & Chemical Biology
- Metastasis & Tumor Microenvironment
- Immunity, Host-Microbial Interactions and Microbiomes
- Tumor Immunobiology
- Experimental Therapeutics, Imaging and Bioengineering

RESEARCH AND TRAINING

- 100 research laboratories housed in state-of-the-art buildings
- 26 Core facilities offering cutting-edge scientific services
- More than 900 pre- and postdoctoral trainees
- Appointments in the Gerstner Sloan Kettering Graduate School of Biomedical Sciences and the Weill Cornell Graduate School of Medical Sciences

Visit **www.ski.edu** to learn more.

MSK is an equal opportunity and affirmative action employer committed to diversity and inclusion in all aspects of recruiting and employment. All qualified individuals are encouraged to apply and will receive consideration without regard to race, color, gender, gender identity or expression, sexual orientation, national origin, age, religion, creed, disability, veteran status or any other factor which cannot lawfully be used as a basis for an employment decision.

Application deadline: **October 15, 2021** | Apply at: **https://facultysearch.ski.edu**

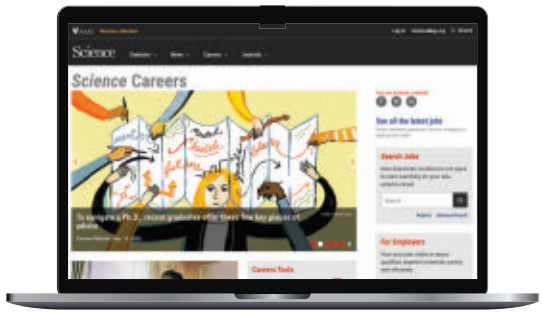
ScienceCareers online @sciencecareers.org



Science Careers helps you advance your career. Learn how !

- Register for a free online account on **ScienceCareers.org**.
- Search hundreds of job postings and find your perfect job.
- Sign up to receive e-mail alerts about job postings that match your criteria.
- Upload your resume into our database and connect with employers.
- Watch one of our many webinars on different career topics such as job searching, networking, and more.
- Download our career booklets, including Career Basics, Careers Beyond the Bench, and Developing Your Skills.
- Complete an interactive, personalized career plan at "my IDP."
- Visit our Employer Profiles to learn more about prospective employers.
- Read relevant career advice articles from our library of thousands.

Visit **ScienceCareers.org** today — all resources are free



ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

Shenzhen Institute of Advanced Technology
Chinese Academy of Sciences



Established in partnership between the Chinese Academy of Sciences and the Shenzhen Municipal Government, the Shenzhen Institute of Advanced Technology (SIAT) is a newly-created university with an objective to become the world's preeminent institute for emerging science and engineering programs. SIAT is equipped with state-of-art teaching and research facilities and is dedicated to cultivating international, visionary, and interdisciplinary talents while delivering research support to pursue innovation-driven development.

SIAT is located in Shenzhen, also known as the "Silicon Valley of China," a modern, clean, and green city, well-known for its stunning architecture, vibrant economy, and its status as a leading global technology hub. SIAT is seeking applications for faculty positions of all ranks in the following academic programs: Computer Science and Engineering, Bioinformatics, Robotics, Life Sciences, Material Science and Engineering, Biomedical Engineering, Pharmaceutical Sciences, Synthetic Biology, Neurosciences, etc. SIAT seeks individuals with a strong record of scholarship who possess the ability to develop and lead high-quality teaching and research programs. SIAT offers a comprehensive benefits package and is committed to faculty success throughout the academic career trajectory, providing support for ambitious and world-class research projects and innovative, interactive teaching methods.

Further information:
http://english.siat.cas.cn/DP/Va/

NIMS ICYS

ICYS Research Fellow at ICYS, NIMS, Japan

The International Center for Young Scientists (ICYS) of the National Institute for Materials Science (NIMS) invites applications for ICYS Research Fellow positions. ICYS will offer you the freedom to conduct independent and self-directed research in various areas of materials science with full access to NIMS advanced research facilities.

The common language at ICYS is English. Clerical and technical support in English will be given by the ICYS staff. An annual salary of approximately 5.35 million yen is guaranteed, which may be increased to a maximum of ~5.88 million yen depending on the performance of the Research Fellow*. In addition, a research grant of 2 million yen per year will be provided to each Research Fellow. The initial contract term is two years, which may be extended for another year depending on one's performance. Also, advantage is given when applying to NIMS permanent researcher position.

All applicants must have obtained a PhD degree within the last ten years. Applicants should submit an **application form** including a research proposal during the ICYS term, **CV, a list of DOI of journal publications, PDF files of three significant publications**, and **PhD Certificate** to the ICYS Recruitment Desk by **September 30, 2021 JST**. The format for the application documents can be downloaded from our website. The selection will be made on the basis of originality and quality of the research proposal as well as the research achievements. Please visit our website for more details.

* Approximately 23% of annual salary will be deducted as social insurance premium, residence tax and income tax.

ICYS Recruitment Desk
National Institute for Materials Science
www.nims.go.jp/icys/recruitment/

CASE WESTERN RESERVE UNIVERSITY EST. 1826

TENURE TRACK/TENURED FACULTY POSITION IN CANCER RESEARCH

Case Western Reserve University
School of Medicine
Case Comprehensive Cancer Center

The Case Comprehensive Cancer Center **(https://case.edu/cancer/)**, an NCI-designated Comprehensive Cancer Center at CWRU, with affiliates University Hospitals Cleveland Medical Center and Cleveland Clinic, invites applications for tenure track/tenured faculty positions at the level of Assistant, Associate, or Full Professor. Applicants are being sought that have a PhD and basic/translational cancer research expertise. Applicants for Assistant Professor must have a track record of outstanding cancer research and publications and potential for extramural funding. Candidates applying for consideration at a senior rank must possess national (Associate Professor) or international (Professor) reputations in collaborative cancer research, a distinguished record of publication and funding, a commitment to mentoring, teaching, and leadership, and they must fulfill other qualifications necessary for a tenured appointment at CWRU.

Successful candidates will be expected to interact closely with members of established basic and translational cancer center research programs. Primary appointment will be in the CWRU School of Medicine department of expertise or in the Cancer Center.

Applicants are also asked to submit a statement explaining how their research, teaching, and/or service have contributed to diversity, equity and inclusion within their scholarly field(s) and/or how their individual and/or collaborative efforts have promoted structural justice inside and outside institutions of higher learning. This statement should also reflect on the ways in which the candidate's continued efforts will foster a culture of diversity, pluralism, and individual difference at Case Western Reserve University into the future.

Please send curriculum vitae and a cover letter outlining your research interests, and your efforts in diversity, equity and inclusion to: Ruth Keri, PhD Chair, Case Comprehensive Cancer Center faculty search committee, c/o **cancersearch@case.edu**. After initial review, you will be asked for a list of three or more references. Please include "Cancer Faculty Search" in the subject line.

In employment, as in education, Case Western Reserve University is committed to Equal Opportunity and Diversity. Women, veterans, members of underrepresented minority groups, and individuals with disabilities are encouraged to apply.

Case Western Reserve University provides reasonable accommodations to applicants with disabilities. Applicants requiring a reasonable accommodation for any part of the application and hiring process should contact the Office of Inclusion, Diversity and Equal Opportunity at 216-368-8877 to request a reasonable accommodation. Determinations as to granting reasonable accommodations for any applicant will be made on a case-by-case basis.

ScienceCareers online @sciencecareers.org

By **Martha Nelson**

Over the baby bump

"Well, at least *you* have permanent job security," my aunt casually remarked last summer, referring to my role in the federal government's COVID-19 response. I shuddered. Little did she know I had just learned the National Institutes of Health (NIH) would not be renewing my annual contract. In my 12 years at NIH, job security had never been a priority; instead, I prized independence. But now, new motherhood followed by the pandemic seemed to be sinking my career, and I had to find a new path.

I had waited to have a baby until my career was well established, thinking that would help buffer me from some of the challenges I knew scientist mothers face, but I still struggled. When my son was born in 2018, I had no paid parental leave (U.S. government workers were finally granted paid parental leave in October 2020), and my husband got just 2 weeks from his small company. I could telework from home for the baby's first months, and I appreciated the flexibility at first. But it ended up being the worst of all worlds. I had neither full-time leave to adjust to motherhood nor a presence in the office to sway decisions pertaining to my work.

When I returned to the office, my supervisor repeatedly warned that I was on thin ice. I acknowledged that adjusting to being a working mom was a steep learning curve, but I was exhausted trying to manage all my obligations and didn't know what I could do differently. When it came time to renew my annual contract, my supervisor said he was inclined not to sign it. I cried and convinced him to renew me for one more year. But 3 months later, the pandemic hit and even more of my time was devoted to caring for my son. The next time around, my supervisor made good on his threat.

I was gutted. My virtual going away party felt more like a funeral. But then my son burst into the living room, triumphantly clutching his ukulele and filling the space with his light. I realized I was prepared to do anything to safeguard my family—including switching to a more secure, family-friendly career path.

I began to speak with colleagues in science policy and communication about alternatives. But when I gathered a group of trusted women in my field for a videochat, they insisted I could stay in research. All I needed was a crash course in survival as a woman and mother in science.

For months we met weekly to postmortem the nonscientific reasons my job fell apart and give me confidence to find a new one. I was surprised to learn that the setbacks I had faced were not unusual for women scientists. My colleagues also pointed out how my own gender biases were crippling my job search. My husband and I were determined to both be highly involved parents. Still, I was rejecting positions with longer hours or commutes, fearing they might interfere with my self-imposed expectations for my role as a mother rather than giving myself credit for securing my family financially while my husband launched a new company.



"I was in danger of repeating a past mistake of quitting prematurely."

I also realized I was in danger of repeating a past mistake of quitting prematurely instead of giving myself time to adapt. Twenty years ago, I was a promising high school distance runner, with a boyish frame that helped me capture title after title. But during puberty I temporarily got fleshier and slower. Dieting and intense workouts only accelerated burnout until I gave up and quit, feeling broken. Years later, I dusted myself off and began trail running in the mountains. When I finally joined the college track team, I raced faster than ever and proved I did not need to be a flat-bodied pixie, I just needed time and space for my changing body to adjust. I have not stopped racing since.

I've learned, twice now, to keep lacing up. The pandemic was my tipping point as a working mom, but I survived by reaching out to savvier women who gave me perspective. I've now secured a better position at another branch of NIH that maintains my research independence while offering greater job security. This fall I'll be heading to my new office just as my toddler begins his first day of preschool. Other kids will surely push him down and snatch his toys. But I'll let him know the same thing happens to his mom, and we dust off, redirect, and embark on new adventures together. ■

Martha Nelson will be starting as a staff scientist at the National Center for Biotechnology Information's intramural research program this fall.

ILLUSTRATION: ROBERT NEUBECKER



BII | Science | Prize for Innovation

Apply for the new BII & Science Prize for Innovation today

Behind every life-changing solution is an entrepreneurial scientist–a creative mind who proved an idea in the lab and dared to carry it out in the world.

To encourage more scientists to translate their research, BioInnovation Institute (BII) & Science present a new annual award.

Our three winners will have their essays published in Science magazine and will be invited into BII's entrepreneurial ecosystem. In addition, the Grand Prize winner will receive a prize of USD 25,000 and each runner-up will receive USD 10,000 at a grand award show celebration in Copenhagen, Denmark.

The call for applications has just opened.
Apply before November 1, 2021.
www.bii.dk/scienceprize

Presented by BII & Science

BII BioInnovation Institute Science

Apply before November 1, 2021
www.bii.dk/scienceprize



Everyone Remembers Their First GRC!
October-November 2021

> *"I remember my first GRC experience because the talks and discussions provided a great learning opportunity for a young scientist. The free time and meals allowed for one-on-one meetings and the chance to form professional connections with peers and leaders in the field."*
>
> DR. ALICIA BYRNE, Past Chair of the Human Genetics and Genomics GRS

Alcohol-Induced End Organ Diseases + GRS
OCTOBER 24 – 29, 2021
FOUR POINTS SHERATON / HOLIDAY INN EXPRESS, VENTURA, CA, US
CHAIRS: Gyongyi Szabo and Bernd Schnabl
VICE CHAIRS: Elizabeth Kovacs and Gavin Arteel

Biotherapeutics and Vaccines Development + GRS
NOVEMBER 14 – 19, 2021
VENTURA BEACH MARRIOTT, VENTURA, CA, US
CHAIRS: Michael Tarlov and Linda Narhi
VICE CHAIRS: Peter Tessier and Jessica Molek

Cannabinoid Function in the CNS + GRS
NOVEMBER 7 – 12, 2021
FOUR POINTS SHERATON / HOLIDAY INN EXPRESS, VENTURA, CA, US
CHAIRS: Matthew Hill and Sachin Patel
VICE CHAIRS: Miriam Melis and Mario Stelt

Carbon Capture, Utilization and Storage + GRS
OCTOBER 24 – 29, 2021
JORDAN HOTEL AT SUNDAY RIVER, NEWRY, ME, US
CHAIR: Roger Aines
VICE CHAIR: Andre Bardow

Cardiac Arrhythmia Mechanisms + GRS
OCTOBER 17 – 22, 2021
FOUR POINTS SHERATON / HOLIDAY INN EXPRESS, VENTURA, CA, US
CHAIRS: David Christini and Ursula Ravens
VICE CHAIRS: Crystal Ripplinger and Edward Vigmond

Cell Biology of Metals + GRS
OCTOBER 17 – 22, 2021
MOUNT SNOW, WEST DOVER, VT, US
CHAIRS: Caryn Outten and Mitchell Knutson
VICE CHAIRS: Angela Wilks and Kerry Kornfeld

Central Nervous System Injury and Repair + GRS
OCTOBER 17 – 22, 2021
JORDAN HOTEL AT SUNDAY RIVER, NEWRY, ME, US
CHAIRS: Yimin Zou and Simone Di Giovanni
VICE CHAIRS: Elizabeth Bradbury and Yishi Jin

Chronobiology + GRS
OCTOBER 10 – 15, 2021
MOUNT SNOW, WEST DOVER, VT, US
CHAIR: Samer Hattar
VICE CHAIR: Charlotte Helfrich-Forster

Computer Aided Drug Design + GRS
NOVEMBER 7 – 12, 2021
MOUNT SNOW, WEST DOVER, VT, US
CHAIR: Georgia McGaughey
VICE CHAIR: Matthias Rarey

Lung Development, Injury and Repair + GRS
NOVEMBER 7 – 12, 2021
WATERVILLE VALLEY, WATERVILLE VALLEY, NH, US
CHAIRS: Anne-Karina Perl and Daniel Tschumperlin
VICE CHAIRS: Rory Morty and Xin

Mammalian DNA Repair + GRS
OCTOBER 31 – NOVEMBER 5, 2021
FOUR POINTS SHERATON / HOLIDAY INN EXPRESS, VENTURA, CA, US
CHAIR: Lee Zou
VICE CHAIR: Patricia Opresko

Medicinal Chemistry + GRS
OCTOBER 24 – 29, 2021
MOUNT SNOW, WEST DOVER, VT, US
CHAIR: Janeta Popovici-Muller
VICE CHAIR: Cynthia Parrish

Phagocytes + GRS
OCTOBER 24 – 29, 2021
WATERVILLE VALLEY, WATERVILLE VALLEY, NH, US
CHAIR: Michelle Lennartz
VICE CHAIR: Sergio Catz

Physical Science of Cancer + GRS
NOVEMBER 14 – 19, 2021
FOUR POINTS SHERATON / HOLIDAY INN EXPRESS, VENTURA, CA, US
CHAIRS: Josef Käs and Sharon Gerecht
VICE CHAIRS: David Odde and Jennifer Munson

Stochastic Physics in Biology + GRS
OCTOBER 10 – 15, 2021
FOUR POINTS SHERATON / HOLIDAY INN EXPRESS, VENTURA, CA, US
CHAIR: Yuhai Tu
VICE CHAIR: Jie Xiao

Stress Proteins in Growth, Development and Disease + GRS
NOVEMBER 7 – 12, 2021
JORDAN HOTEL AT SUNDAY RIVER, NEWRY, ME, US
CHAIR: Brian Freeman
VICE CHAIR: Harm Kampinga

Three Dimensional Electron Microscopy + GRS
OCTOBER 31 – NOVEMBER 5, 2021
WATERVILLE VALLEY, WATERVILLE VALLEY, NH, US
CHAIR: Sharon Wolf
VICE CHAIR: Daniela Nicastro

Apply now at www.grc.org!